Research Article

Statistics in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Modeling the random effects covariance matrix for longitudinal data with covariates measurement error

Md Erfanul Hoque[a], Mahmoud Torabi[a, b]*

Longitudinal data occur frequently in practice such as medical studies and life sciences. Generalized linear mixed models (GLMMs) are commonly used to analyze such data. It is typically assumed that the random effects covariance matrix is constant among subjects in these models. In many situations, however, the correlation structure may differ among subjects and ignoring this heterogeneity can lead to biases in model parameters estimate. Recently, Lee *et al.* developed a heterogeneous random effects covariance matrix for GLMMs for error-free covariates. Covariates measured with error also happen frequently in the longitudinal data set-up (e.g., blood pressure, cholesterol level). Ignoring this issue in the data may produce bias in model parameters estimate and lead to wrong conclusions. In this paper, we propose an approach to properly model the random effects covariance matrix based on covariates in the class of GLMMs where we also have covariates measured with error. The resulting parameters from the decomposition of random effects covariance matrix have a sensible interpretation and can be easily modeled without the concern of positive definiteness of the resulting estimator. Performance of the proposed approach is evaluated through simulation studies which show that the proposed method performs very well in terms of bias, mean squared error, and coverage rate. An application of the proposed method is also provided using a longitudinal data from Manitoba Follow-up study.

Keywords: Cholesky decomposition, Longitudinal data, Measurement error, Monte Carlo Expectation-maximization algorithm, Random effects

1. Introduction

In many medical studies, longitudinal data or repeated measurement data occur frequently where often changes in a particular characteristic in the participating individuals are investigated by observing repeatedly over time. The generalized linear mixed models (GLMMs) are commonly used to analyze such data which enable us to account for between and within individuals heterogeneity [1–3]. In these models, it is assumed that the random effects covariance matrix is constant across the subject and also the high dimensionality and positive definite constraints make the structure of random effects covariance matrix

[a]*Department of Statistics, University of Manitoba, MB, R3T 2N2, Canada*
[b]*Departments of Community Health Sciences and Statistics, University of Manitoba, MB, R3E 0W3, Canada*
[*]*Correspondence to: Mahmoud Torabi, Department of Community Health Sciences, University of Manitoba, MB, R3E 0W3, Canada.*
E-mail: Mahmoud.Torabi@umanitoba.ca

restricted. However, the covariance matrix may vary by measured covariates in many situations, and biased estimates of the fixed effects may result by ignoring this heterogeneity [4].

There are few literatures to discuss the issue of accounting the heterogeneity in covariance matrix. Chiu *et al.* [5] modeled the covariance matrix using a log matrix parametrization in marginal models and estimates were obtained using estimating equations. Using the modified Cholesky decomposition, Pourahmadi [6, 7] developed covariance matrix which depends on subject-specific covariates. Following the idea of modified Cholesky decomposition, Pourahmadi and Daniels [8] developed a class of dynamic conditionally mixed models by allowing to vary the marginal covariance matrix across subjects. However, they considered the random effects covariance matrix to be fixed across individuals. Daniels and Pourahmadi [9] proposed Bayesian priors for the parameters associated with modified Cholesky decomposition in linear mixed model, and Daniels and Zhao [10] introduced an approach to model the entire random effects covariance matrix for linear mixed model where all the parameters can be modeled. Recently, Lee *et al.* [11] introduced a heterogeneous random effects covariance matrix for the GLMMs by using modified Cholesky decomposition for error-free covariates [6, 7].

It is important to note that in the models above it is assumed that the covariates are perfectly measured for the validity of inferential methods. Covariates measured with error also happen frequently in practice. Extensive works have been done on covariates measurement error [12–21]. The GLMMs framework are adopted to make the inference procedure in most of these cases. In practice, longitudinal data are prone to be not perfect and seriously biased results can be led by ignoring this issue. Covariate measurement error is a common typical feature of longitudinal study [22]. Recently, the attention has been increased to address the effects of covariate measurement error in the analysis of longitudinal data [22]. In 2011, Yi *et al.* [23] presented a fairly general framework to make inference for longitudinal data with covariates measurement error and missing responses simultaneously by adopting the framework of GLMMs. They employed the expectation-maximization (EM) algorithm to conduct inference for parameters of interest [24]. However, in their work the random effects covariance matrix was left unspecified.

In this paper, our goal is to properly model the random effects covariance matrix under the framework of GLMMs with covariates measurement errors. For this purpose, we extend the model introduced by Lee *et al.* [11] for the random effects covariance matrix for the GLMMs to the case when the covariates are also subject to measurement error. An important feature of using random effects covariance matrix by defining covariates and Cholesky decomposition is that each subject has a specific covariance structure and also our proposed model accounts for the covariates measured with error.

The rest of the paper is organized as follows. In Section 2, notation and specification of the model are given. The proposed approach to model the random effects covariance matrix for the GLMMs with covariates measurement error is given in Section 3. In Section 4, the general framework is provided to make the inference procedure to estimate model parameters. Performance of the proposed method is evaluated through simulation studies and also by analyzing a motivating data set arising from Manitoba Follow-up study (MFUS) in Sections 5 and 6, respectively. Finally, concluding remarks are given in Section 7. Details of the proposed approach in the case of binary outcome are illustrated in Appendix.

2. Notation and model specification

The model of interest defines repeated measurement on each of m individuals with responses that follow a generalized linear model with random intercept for each individual, with time-specific covariates that are subject to error and time-specific error-free covariates. Suppose that Y_{ij} be the response variable at time point j for subject i. Let $\mathbf{X}_{ij}$ be the vector of error-prone covariates, and $\mathbf{Z}_{ij}$ be the vector of error-free covariates, $i = 1, 2, \ldots, m$, and $j = 1, 2, \ldots, n_i$. Further, let $\mathbf{W}_{ij}$ be an observed version of $\mathbf{X}_{ij}$. Denote the response vector for the ith subject by $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^T$ and also denote $\mathbf{X}_i = (\mathbf{X}_{i1}^T, \ldots, \mathbf{X}_{in_i}^T)^T$, $\mathbf{Z}_i = (\mathbf{Z}_{i1}^T, \ldots, \mathbf{Z}_{in_i}^T)^T$, and $\mathbf{W}_i = (\mathbf{W}_{i1}^T, \ldots, \mathbf{W}_{in_i}^T)^T$.

2.1. *Response process*

We assume that Y_{ij} belongs to the exponential family conditionally on the random effects $\mathbf{u}_i$:

$$f(y_{ij}|\mathbf{x}_{ij},\mathbf{u}_i;\eta_{ij},\zeta) = \exp\{(y_{ij}\eta_{ij} - b(\eta_{ij}))/a(\zeta) + c(y_{ij},\zeta)\}, \tag{1}$$

where $a(\cdot)$, $b(\cdot)$, and $c(\cdot,\cdot)$ are known functions and η_{ij}, canonical parameters, can be further modeled to accommodate within-subject variability. The dispersion parameter ζ is assumed to be known/estimated (e.g., $\zeta = 1$ for binary response). To emphasize on the estimation of parameters of interest we treat ζ as known here. We model a transformation of the conditional mean, $\mu_{ij} = E(Y_{ij}|\mathbf{x}_{ij},\mathbf{u_i})$, as a linear model of the both fixed and random effects:

$$g(\mu_{ij}) = \eta_{ij} = \beta_0 + \mathbf{X}_{ij}^T\boldsymbol{\beta}_x + \mathbf{Z}_{ij}^T\boldsymbol{\beta}_z + u_{ij}, \tag{2}$$

where $g(\cdot)$ is the link function and $\boldsymbol{\beta} = (\beta_0,\boldsymbol{\beta}_x^T,\boldsymbol{\beta}_z^T)^T$ is the vector of fixed parameters. We suppose that the random effects $\mathbf{u}_i = (u_{i1},\ldots,u_{in_i})^T$ are independent and also independent of the explanatory variables. Here, $\mathbf{u}_i \sim N(\mathbf{0},\boldsymbol{\Sigma}_i)$, where the random effects covariance matrix $\boldsymbol{\Sigma}_i$ is indexed by subject i. The vector $\mathbf{u}_i$ represents the vector of random effect values for subject i. We also assume that $f(\mathbf{y}_i|\mathbf{u}_i,\mathbf{x}_i,\mathbf{w}_i,\mathbf{z}_i) = f(\mathbf{y}_i|\mathbf{u}_i,\mathbf{x}_i,\mathbf{z}_i)$ which is connected to the usual non-differential error mechanism [22], but is different because of the dependence on the random effects.

2.2. *Measurement error process*

To feature the measurement error process we employ a multiple regression model [22] as follows:

$$\mathbf{W}_{ij} = \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_x\mathbf{X}_{ij} + \boldsymbol{\Gamma}_z\mathbf{Z}_{ij} + \mathbf{e}_{ij}\ , \tag{3}$$

where error terms $\mathbf{e}_{ij}$'s are assumed to be independent among different j and also independent of $\mathbf{X}_{ij}$, $\mathbf{Z}_{ij}$, Y_{ij} and random effects $\mathbf{u_i}$, and follow a distribution, $f(\mathbf{e}_{ij},\sigma^2)$, where σ^2 is the associated parameter. It is often assumed that $\mathbf{e}_{ij}$ has zero mean. Note that the assumption of independence of e_{ij} over replicates is valid in the case of large number of subjects (m) or the long-term average if the replicates within a subject are taken far enough apart in time [22]; we will further discuss this assumption in the data application (Section 6). Let $\boldsymbol{\Gamma}_0 = (\gamma_{01},\ldots,\gamma_{0p})^T$ be the vector of intercept coefficients, $\boldsymbol{\Gamma}_x = (\Gamma_{x1},\ldots,\Gamma_{xp})^T$ and $\boldsymbol{\Gamma}_z = (\Gamma_{z1},\ldots,\Gamma_{zp})^T$ denote the vector of regression coefficients of $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$, respectively. Also let $\boldsymbol{\gamma} = (\boldsymbol{\Gamma}_0^T,\boldsymbol{\Gamma}_x^T,\boldsymbol{\Gamma}_z^T)^T$ be the vector of all the regression coefficients. By setting $\boldsymbol{\Gamma}_0 = \mathbf{0}, \boldsymbol{\Gamma}_z = \mathbf{0}\ \&\ \boldsymbol{\Gamma}_x = \mathbf{1}_p$, where $\mathbf{1}_p$ is the vector of ones with dimension p, the model above (3) can be then written as a classical additive error model [22]: $\mathbf{W}_{ij} = \mathbf{X}_{ij} + \mathbf{e}_{ij}$.

3. A model for the random effects covariance matrix

The main contribution of this paper is to use a specified model for the random effects covariance matrix in the class of GLMMs with covariates measurement error where heterogeneity of the random effects covariance matrix has been ignored. For this purpose, we employ the structure of the random effects covariance matrix given by Lee *et al.* [11] and propose a heterogeneous random effects covariance matrix for GLMMs which depends on subject-specific covariates. To that end, we decompose the random effects covariance matrix based on the modified Cholesky decomposition [6, 7] which results in a set of dependence parameters, generalized autoregressive parameters (GARPs), and a set of variance parameters called innovation variances (IVs). The basic idea of this proposed structure is that the covariance matrix $\boldsymbol{\Sigma}_i$ of the random effects vector $\mathbf{u}_i$, can be diagonalized by a lower triangular matrix which is constructed from the regression coefficients when u_{ij} is regressed on its predecessors $u_{i1}, u_{i2},\ldots,u_{ij-1}$.

More specifically, it can be written as:

$$u_{i1} = \epsilon_{i1} \tag{4}$$

$$u_{ij} = \sum_{t=1}^{j-1} \phi_{i,jt} u_{it} + \epsilon_{ij}, \quad \text{for} \quad j = 2, 3, \dots, n_i, \tag{5}$$

where $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2}, \dots, \epsilon_{in_i})^T \sim N(\mathbf{0}, \mathbf{D}_i)$ with $\mathbf{D}_i = \text{diag}(\sigma_{i1}^2, \sigma_{i2}^2, \dots, \sigma_{in_i}^2)$. Then for the $j = 2, 3, \dots, n_i$ we can write (4) and (5) in matrix form as follows:

$$\mathbf{T}_i \mathbf{u}_i = \boldsymbol{\epsilon}_i, \tag{6}$$

where $\mathbf{T}_i$ is a unit lower triangular matrix having ones on its diagonal and $-\phi_{i,jt}$ in the (j,t)th element for $2 \leq j \leq n_i$ and $t = 1, 2, \dots, j-1$. From (6) we can write $\mathbf{T}_i \boldsymbol{\Sigma}_i \mathbf{T}_i^T = \text{var}(\boldsymbol{\epsilon}_i) = \mathbf{D}_i$. The GARPs are denoted by ϕ, and σ_{ij}^2 represents the IVs. Time-and/or subject-specific covariate vectors can be used to model the GARPs and IVs by setting

$$\phi_{i,jt} = \mathbf{k}_{i,jt}^T \boldsymbol{\delta}, \quad \log\left(\sigma_{ij}^2\right) = \mathbf{h}_{i,j}^T \boldsymbol{\lambda}, \tag{7}$$

where $\boldsymbol{\delta}$ and $\boldsymbol{\lambda}$ are $a \times 1$ and $b \times 1$ vectors of unknown dependence and variance parameters, respectively. The design vectors $\mathbf{k}_{i,jt}$ and $\mathbf{h}_{i,j}$ are covariates to model the GARP/IV parameters as a function of the subject-specific covariates [7, 8, 10, 11]. The parametrization of GARPs or IVs has various advantages. Firstly, we can model the random effects covariance matrix in terms of covariates because of unconstrained characteristics of GARPs and IVs. Secondly, as in (7) there is a linear combination of covariates which result that the parameters have a reasonable interpretation and easy to model and the positive definiteness of $\boldsymbol{\Sigma}_i$ is also guaranteed because of the positive σ_{ij}^2 [10, 11]. We can also have specific $\boldsymbol{\Sigma}_i$ for each subject i. The covariance matrix also covers e.g. auto-regressive model with order 1, AR (1), as a special case.

4. General inference method

To derive the likelihood function for the GLMM with covariates measurement error, let us define the parameters as $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \gamma^T, \sigma^2, \boldsymbol{\delta}^T, \boldsymbol{\lambda}^T)^T$. The complete data likelihood function of $\mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i$ for subject i can be written as:

$$L_i(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i) = f(\mathbf{y}_i | \mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta}) f(\mathbf{x}_i | \mathbf{w}_i, \mathbf{z}_i; \gamma, \sigma^2) f(\mathbf{u}_i; \boldsymbol{\delta}, \boldsymbol{\lambda}),$$

where $f(\mathbf{y}_i | \mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta})$ belongs to the exponential family (1) given the random effects. We assume $f(\mathbf{x}_i | \mathbf{w}_i, \mathbf{z}_i; \gamma, \sigma^2)$ has a multivariate Normal density with mean vector $\boldsymbol{\mu}_f$ and covariance matrix $\boldsymbol{\sigma}_f^2$, which is independent of $\mathbf{u}_i$, and the details of $\boldsymbol{\mu}_f$ and $\boldsymbol{\sigma}_f^2$ are given in following Section (Section 4.1) e.g. in the case of scalar x_{ij}. Also, $\mathbf{u}_i \sim f(\mathbf{u}_i | \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\delta}, \boldsymbol{\lambda}) = f(\mathbf{u}_i; \boldsymbol{\delta}, \boldsymbol{\lambda})$ which $f(\mathbf{u}_i; \boldsymbol{\delta}, \boldsymbol{\lambda})$ has a multivariate Normal density with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_i$ and this can be simplified based on the proposed structure of the random effects covariance matrix [11] as follows:

$$f(\mathbf{u}_i; \boldsymbol{\delta}, \boldsymbol{\lambda}) = (2\pi)^{-n_i/2} \left[\prod_{j=1}^{n_i} \left(\sigma_{ij}^2\right)^{-1/2} \right] \exp\left(-\frac{1}{2} \sum_{j=1}^{n_i} \frac{\epsilon_{ij}^2}{\sigma_{ij}^2} \right) \quad \text{with} \quad \epsilon_{i1} = u_{i1}.$$

We can then write the complete data log-likelihood as:

$$l_c(\boldsymbol{\theta}) = \sum_{i=1}^{m} \left\{ \log f(\mathbf{y}_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta}) + \log f(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\gamma}, \sigma^2) + \log f(\mathbf{u}_i; \boldsymbol{\delta}, \boldsymbol{\lambda}) \right\}. \tag{8}$$

The EM algorithm is employed to evaluate the log-likelihood function above due to its intractable form [24]. The E-step can be written as (at iteration l):

$$\begin{aligned} Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(l)}) &= \mathrm{E}\left[l_c(\boldsymbol{\theta}) | \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)} \right] \\ &= \sum_{i=1}^{m} \int \int \left\{ \log f(\mathbf{y}_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta}) + \log f(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\gamma}, \sigma^2) + \log f(\mathbf{u}_i; \boldsymbol{\delta}, \boldsymbol{\lambda}) \right\} \\ &\quad f\left(\mathbf{x}_i, \mathbf{u}_i | \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)} \right) \, d\mathbf{x}_i d\mathbf{u}_i, \end{aligned} \tag{9}$$

where $f\left(\mathbf{x}_i, \mathbf{u}_i | \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}\right)$ is the conditional density of random effects $(\mathbf{x}_i, \mathbf{u}_i)$ given the observed data $(\mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i)$ at the initial value $\boldsymbol{\theta}^{(l)}$. As in the $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l)})$ function above, the multiple integrals are not in closed form for computation, it is not often possible to evaluate this expectation directly. Hence, Monte Carlo EM (MCEM) algorithm can be employed for this purpose [2]. To this end, we need to generate a sample from $f(\mathbf{x}_i, \mathbf{u}_i|\mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)})$ for each i. Gibbs sampling or Metropolis-Hasting algorithm can be used to accomplish this [25–27]. Essentially, we can iteratively sample from $f(\mathbf{x}_i|\mathbf{y}_i, \mathbf{u}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)})$ and $f(\mathbf{u}_i|\mathbf{y}_i, \mathbf{x}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)})$. These conditional distributions have the form respectively as follows:

$$\begin{aligned} f(\mathbf{x}_i|\mathbf{y}_i, \mathbf{u}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}) &\propto f(\mathbf{y}_i|\mathbf{x}_i, \mathbf{u}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}) f(\mathbf{x}_i|\mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}) \\ f(\mathbf{u}_i|\mathbf{y}_i, \mathbf{x}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}) &\propto f(\mathbf{y}_i|\mathbf{x}_i, \mathbf{u}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)}) f(\mathbf{u}_i; \boldsymbol{\theta}^{(l)}) \end{aligned}$$

Suppose that, we take a pseudo-random sample of size M, $\left\{ (\mathbf{x}_i^{(1)}, \mathbf{u}_i^{(1)}), (\mathbf{x}_i^{(2)}, \mathbf{u}_i^{(2)}), \ldots, (\mathbf{x}_i^{(M)}, \mathbf{u}_i^{(M)}) \right\}$ for individual i, from the joint distribution $f(\mathbf{x}_i, \mathbf{u}_i|\mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\theta}^{(l)})$ via the Metropolis-Hasting algorithm. Then, the E-step at the $(l+1)$th EM iteration can be written as:

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(l)}) &\approx \sum_{i=1}^{m} \left\{ \frac{1}{M} \sum_{k=1}^{M} l_c\left(\boldsymbol{\theta}^{(l)}; \mathbf{y}_i, \mathbf{x}_i^{(k)}, \mathbf{u}_i^{(k)} \right) \right\} \\ &= \sum_{i=1}^{m} \sum_{k=1}^{M} \frac{1}{M} \log f\left(\mathbf{y}_i, |\mathbf{x}_i^{(k)}, \mathbf{z}_i, \mathbf{u}_i^{(k)}; \boldsymbol{\beta} \right) + \sum_{i=1}^{m} \sum_{k=1}^{M} \frac{1}{M} \log f\left(\mathbf{x}_i^{(k)} | \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\gamma}, \sigma^2 \right) \\ &+ \sum_{i=1}^{m} \sum_{k=1}^{M} \frac{1}{M} \log f\left(\mathbf{u}_i^{(k)}; \boldsymbol{\delta}, \boldsymbol{\lambda} \right). \end{aligned} \tag{10}$$

In the M-step of MCEM algorithm, an optimization procedure can be employed to maximize $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(l)})$ with respect to $\boldsymbol{\theta}$ to produce an updated estimate $\boldsymbol{\theta}^{(l+1)}$. These E and M steps will continue until convergence and then the current values of $\boldsymbol{\theta}$ will be declared as MLEs of $\boldsymbol{\theta}$, namely $\hat{\boldsymbol{\theta}}$.

The standard errors of the MLEs cannot be automatically obtained from the EM algorithm. To obtain the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$, we can use

$$\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\theta}}) \approx \left[\sum_{i=1}^{m} \sum_{k=1}^{M} \frac{1}{M} \mathbf{S}_{ik}(\hat{\boldsymbol{\theta}}) \mathbf{S}_{ik}^T(\hat{\boldsymbol{\theta}}) \right]^{-1}, \tag{11}$$

where, $\mathbf{S}_{ik}(\hat{\boldsymbol{\theta}}) = \left. \dfrac{\partial l_c\left(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{x}_i^{(k)}, \mathbf{u}_i^{(k)} \right)}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ [28]. Then by taking the square root of the diagonal elements of $\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\theta}})$, the standard errors of the MLEs can be obtained. One can also use an optimization procedure to get the standard error. For example, **optim** function in *R-program* provides the hessian matrix which can be used to calculate the standard errors [29]. As an illustration, we provide a detailed calculation of the above proposed approach in the case of binary outcome in Appendix. However, similar framework can be extended to other exponential family distributions.

5. Simulation study

We conduct a simulation study to evaluate the performance of our proposed approach. In particular, we make a comparison of the efficiency of our proposed method to the method where covariates measurement error are ignored (Naive 1). Moreover, we also compare the efficiency of the proposed method to the method for GLMMs with covariates measurement error where constant random effects covariance matrix is considered across the subjects (Naive 2). We use the following model to simulate data:

$$\text{logit}\{P_{ij}\} = \beta_0 + \beta_x X_{ij} + \beta_z Z_{ij} + \beta_{z^*} Z_i^* + u_{ij}, \tag{12}$$

$i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n_i$, with $P_{ij} = P(Y_{ij} = 1 | x_{ij}, \mathbf{z}_{ij}^+)$, with $\mathbf{Z}_{ij}^+ = (Z_{ij}, Z_i^*)$ and Z_i^* equals to 0 or 1 with an equal sample size per group. The error-free covariate Z_{ij} is generated as the random variables $v_i + \varphi_{ij}$, where v_i's and φ_{ij}'s are independently identically distributed following $N(0, 0.5^2)$, and they are independent. The Z_{ij} and Z_i^* are then treated fixed during the simulation study.

We consider classical additive model for the structural measurement error for which we generate surrogate variable W_{ij} as $W_{ij} = X_{ij} + e_{ij}$, where e_{ij}'s are independently identically distributed following $N(0, \sigma^2)$, where σ^2 indicates the measurement error variation in covariate X_{ij}. The true covariate X_{ij} is generated from model $X_{ij} = \mu_x + a_i + \xi_{ij}$, where $\mu_x = 1$, a_i's and ξ_{ij}'s are independently identically distributed following $N(0, 1)$. We consider random effects $\mathbf{u}_i = (u_{i1}, \ldots, u_{in_i}) \sim N(\mathbf{0}, \mathbf{\Sigma}_i)$. To simplify variance components of $\mathbf{\Sigma}_i$, the parameters of random effects covariance matrix are defined as:

$$\phi_{i,jt} = \delta_0 I(|j-t| = 1) + \delta_1 I(|j-t| = 1) Z_i^* \quad \text{and} \quad \log(\sigma_{ij}^2) = \lambda_0 + \lambda_1 Z_i^*, \tag{13}$$

Here, we consider $m = 350$ and two cases of number of follow-ups ($n_i = 5, 10$). The initial values are $\boldsymbol{\beta} = (\beta_0, \beta_x, \beta_z, \beta_{z^*}) = (1, 2, 2, 2)$. We set three different values $(0, 0.4, 0.8)$ for σ to evaluate the impact of varying degree of measurement error on estimation. The initial values of the parameters of the random effects covariance matrix are $\boldsymbol{\delta} = (\delta_0, \delta_1) = (0.5, 0.3)$ and $\boldsymbol{\lambda} = (\lambda_0, \lambda_1) = (0.1, 0.2)$. We generate $B = 200$ independent data sets. We then fit three methods: the first method is the proposed approach; the second method (Naive 2) is the model which considers homogeneous covariance matrix, that is, $u_{ij} = u_{i0} \sim N(0, \sigma_u^2)$, where σ_u^2 is the average of the σ_{ij}^2 defined in the equation (13) and covariate X_{ij} with measurement error; and the third method (Naive 1) is the model which ignores covariate measurement error and assumes homogeneous covariance matrix.

To check the performance of our proposed approach in different scenarios, we also consider two additional cases: 1) misspecification of the measurement error distribution; and 2) different random effects covariance matrix structures rather than (13). We describe these two cases accordingly.

Case 1: We generate $B = 200$ datasets from the model (12) with a sample size of $m = 350$ and $n_i = 5$ using a misspecified measurement error distribution. In particular, we generate e_{ij} from the skew-normal distribution with location parameter $\lambda = 0$, scale parameter $\delta = 10$, and skewness parameter $\alpha = 15$, that is, $e_{ij} \sim SN(\lambda = 0, \delta = 10, \alpha = 15)$, where SN stands for skewed Normal. We then fit the three methods (Proposed, Naive 1, Naive 2) considering the same structure of random effects covariance matrix defined in (13). We assume true model parameters are $\boldsymbol{\beta} = (\beta_0, \beta_x, \beta_z, \beta_{z^*}) = (1, 1, 2, 2)$, $\boldsymbol{\delta} = (\delta_0, \delta_1) = (0.5, 0.3)$, and $\boldsymbol{\lambda} = (\lambda_0, \lambda_1) = (0.1, 0.2)$.

Case 2: We also generate $B = 200$ datasets from the model (12) with a sample size of $m = 350$ and $n_i = 5$

with various structured covariance matrix $\mathbf{\Sigma}_i$ for the proposed approach. In this case, we only focus on the severe measurement error variation [$\sigma = 0.8$]; note that this is the worst scenario (severe measurement error and small number of follow-ups). We consider three different random effects covariance matrix models to fit the proposed approach. Model 1 is the heterogeneous special Autoregressive 2 (AR2) covariance matrix depending on Z_i^* : $\phi_{i,jt} = \delta_0 I(|j-t|=1) + \delta_1 I(|j-t|=2) + \delta_2 I(|j-t|=2)Z_i^*$ and $\log(\sigma_{ij}^2) = \lambda_0 + \lambda_1 Z_i^*$. We assume true values for the parameters δ and λ are $(\delta_0, \delta_1, \delta_2) = (0.5, 0.3, 0.2)$ and $(\lambda_0, \lambda_1) = (0.1, 0.2)$. Model 2 is the heterogeneous non-stationary covariance matrix depending on Z_i^*. The parameters of $\mathbf{\Sigma}_i$ are specified as: $\phi_{i,jt} = \delta_0 + \delta_1|j-t|Z_i^* + \delta_2|j-t|^2 Z_i^*$ and $\log(\sigma_{ij}^2) = \lambda_0 + \lambda_1 Z_i^*$. We assume true values for the parameters δ and λ are $(\delta_0, \delta_1, \delta_2) = (0.5, 0.3, 0.2)$ and $(\lambda_0, \lambda_1) = (0.1, 0.2)$. In Model 3, we consider heterogeneous non-stationary structure for $\mathbf{\Sigma}_i$ which is given by $\phi_{i,jt} = \delta_0 + \delta_1|j-t|Z_i^* + \delta_2|j-t|^2 Z_i^*$ and $\log(\sigma_{ij}^2) = \lambda_0 + \lambda_1 \text{time}_j + \lambda_2 \text{time}_j^2$, where $j = 1, \ldots, 5$ and $\text{time}_j = (j-1)/10$. We assume true values for the parameters δ and λ are $(\delta_0, \delta_1, \delta_2) = (0.5, 0.3, 0.2)$ and $(\lambda_0, \lambda_1, \lambda_2) = (0.05, 0.1, 0.2)$.

5.1. Simulation results

In the following, we report the simulation results of empirical bias (Bias), root mean squared error (RMSE), and coverage rate (CR) for the 95% confidence intervals (CI) of the model parameters estimate where e.g. for the β_0 (fixed intercept) we have: $\text{Bias}_{\beta_0} = \frac{1}{B}\sum_{b=1}^{B} \hat{\beta}_0^{(b)} - \beta_0$, where $\hat{\beta}_0^{(b)}$ is the estimated value of β_0 in the simulation run b and β_0 is the true value of this parameter. Also, $\text{RMSE}_{\beta_0} = \sqrt{\text{Bias}_{\beta_0}^2 + \text{Var}_{\beta_0}}$ and 95% $\text{CI}_{\beta_0}^{(b)} : \hat{\beta}_0^{(b)} \pm 1.96\sqrt{\text{Var}(\hat{\beta}_0^{(b)})}$, where the Var_{β_0} are the average of model-based variances $\text{Var}(\hat{\beta}_0)$ over B simulation runs, and CR is the proportion of times (out of $B = 200$) that the true parameter falls in the corresponding 95% CI.

The following Tables 1 and 2 represent the results of the three methods (Naive 1, Naive 2, and Proposed) for the two cases $n_i = 5$ and $n_i = 10$ under different measurement error variations (no error [$\sigma = 0.0$], moderate error [$\sigma = 0.4$] and severe error [$\sigma = 0.8$]) under the random effects covariance matrix structure (13). Table 1 shows the results of the fixed effects and random effects parameters for the three methods in the case of the absence of measurement error, moderate error, and severe error when number of follow-ups is 5 ($n_i = 5$). It is evident from the results that the proposed approach works well under this situation in terms of bias, RMSE, and CR for 95% CI. In the absence of measurement errors, for the cases of Naive 1 and Naive 2 methods, we can see considerable biases in fixed effects estimates (0.0203 for β_x, 0.0153 for β_z, 0.0061 for β_{z^*}, in the case of Naive 1), (0.0011 for β_x, 0.0025 for β_z, 0.0011 for β_{z^*}, in the case of Naive 2) whereas in the proposed approach, the estimates show fairly small biases (-0.0001 for β_x, 0.0002 for β_z, 0.0003 for β_{z^*}). The proposed approach also shows the good CR for 95% CI. The RMSEs are also smaller for the proposed approach compared to the other two methods. We can see considerable biases in the fixed effects parameters for mis-specification of the distribution of random effects. In random effects covariance matrix parameters, the GARPs (δ_0, δ_1) and the IVs (λ_0, λ_1) parameters have small biases.

From the results of moderate measurement error variation [$\sigma = 0.4$], it is obvious here that ignoring the measurement error in data results in considerable biases. The biases tend to increase with the increase of magnitude of measurement error. The results indicate the higher biases for Naive 1 and Naive 2, while the proposed approach shows considerably smaller biases in the fixed effects parameters. Moreover, the proposed approach indicates smaller RMSEs as well as good CR for 95% CI in the estimates of fixed effects compared to the other two methods.

In case of severe measurement error variation [$\sigma = 0.8$], it is clear from the results that the performance of Naive 1 is noticeably affected with the increase of magnitude of measurement error. Based on the results, it can be seen that there is considerably large finite-sample biases in the fixed effects estimates (0.2463 for β_0, -0.7067 for β_x, -0.3753 for β_z, -0.3181 for β_{z^*}) and very low CRs for the 95% CI for the Naive 1 approach. The biases become smaller for the Naive 2 method (0.0018 for β_0, -0.0010 for β_x, -0.0003 for β_z, -0.0015 for β_{z^*}). As expected, the proposed approach performs very well with respect to biases, RMSEs as well as the CRs. In particular, the biases (-0.0005 for β_0, 0.0002 for β_x, 0.0007 for β_z, 0.0009 for β_{z^*}) are fairly small with good CRs. The RMSEs obtained from the Naive 1 and Naive 2 are much bigger than those obtained

Degree of error	Parameter	Naive 1 Bias	Naive 1 RMSE	Naive 1 CR	Naive 2 Bias	Naive 2 RMSE	Naive 2 CR	Proposed Bias	Proposed RMSE	Proposed CR
No error										
($\sigma = 0$)	β_0	0.0082	0.1440	0.93	0.0013	0.0297	0.93	-0.0006	0.0076	0.94
	β_x	0.0203	0.1151	0.93	0.0011	0.0237	0.94	-0.0001	0.0060	0.95
	β_z	0.0153	0.1838	0.95	0.0025	0.0355	0.95	0.0002	0.0100	0.96
	β_{z^*}	0.0061	0.2029	0.95	0.0011	0.0432	0.93	0.0003	0.0132	0.96
	σ_u^2	0.0032	0.1338	0.94	0.0019	0.0282	0.95	-	-	-
	δ_0	-	-	-	-	-	-	0.1464	0.6609	0.93
	δ_1	-	-	-	-	-	-	-0.0385	0.1123	0.86
	λ_0	-	-	-	-	-	-	-0.0053	0.0397	0.87
	λ_1	-	-	-	-	-	-	-0.0188	0.0777	0.85
Moderate error										
($\sigma = 0.4$)	β_0	0.0796	0.1583	0.92	0.0047	0.0296	0.93	0.0005	0.0071	0.95
	β_x	-0.2487	0.2663	0.27	-0.0005	0.0244	0.95	0.0001	0.0059	0.94
	β_z	-0.1570	0.2344	0.85	0.0004	0.0340	0.95	0.0012	0.0101	0.95
	β_{z^*}	-0.1163	0.2261	0.90	-0.0044	0.0452	0.94	-0.0007	0.0117	0.96
	σ_u^2	0.0032	0.1338	0.94	0.0026	0.0309	0.95	-	-	-
	δ_0	-	-	-	-	-	-	0.0920	0.5574	0.95
	δ_1	-	-	-	-	-	-	-0.0297	0.1060	0.86
	λ_0	-	-	-	-	-	-	-0.0060	0.0347	0.87
	λ_1	-	-	-	-	-	-	-0.0137	0.0726	0.87
Severe error										
($\sigma = 0.8$)	β_0	0.2453	0.2756	0.49	0.0018	0.0246	0.94	-0.0005	0.0075	0.96
	β_x	-0.7067	0.7101	0.00	-0.0010	0.0203	0.92	0.0002	0.0060	0.97
	β_z	-0.3753	0.4063	0.33	-0.0003	0.0305	0.93	0.0007	0.0098	0.95
	β_{z^*}	-0.3181	0.3637	0.57	-0.0015	0.0376	0.93	0.0009	0.0120	0.95
	σ_u^2	0.0032	0.1338	0.94	-0.0005	0.0246	0.95	-	-	-
	δ_0	-	-	-	-	-	-	0.1488	0.6743	0.92
	δ_1	-	-	-	-	-	-	-0.0422	0.1161	0.83
	λ_0	-	-	-	-	-	-	-0.0070	0.0360	0.87
	λ_1	-	-	-	-	-	-	-0.0203	0.0772	0.86

Table 1. Bias, RMSE, and CR for 95% CI of the parameters estimate in case of $n_i = 5$ for the three methods (Naive 1, Naive 2, Proposed) under the random effects covariance matrix structure (13)

from the the proposed approach. Also, the RMSEs for the Naive 2 are much smaller than the corresponding values of the Naive 1.

Table 2 represents the results of the estimates of the coefficients of fixed and random effects for the three approaches with 10 number of follow-ups ($n_i = 10$). In terms of bias, RMSE, and CR of 95% CI, it is apparent from the results that the proposed approach works perfectly well compared to the other two approaches for fixed effects parameters estimate in the case of no measurement error. The parameters of the random effects covariance matrix (GARPs, IVs) also show smaller biases, compared to the 5 number of follow-ups ($n_i = 5$), and good CRs. As expected, the RMSEs of the model parameters estimate for the proposed approach are smaller than the corresponding values of the both Naive 1 and Naive 2, and also the Naive 2 has smaller RMSEs compared to the Naive 1.

It is observed under moderate measurement error and 10 number of follow-ups that the proposed approach provides big improvement on biases for fixed effects estimate (-0.0003 for β_0, -0.0001 for β_x, -0.0004 for β_z, 0.0007 for β_{z^*}) than Naive 1 (0.1045 for β_0, -0.2404 for β_x, -0.1236 for β_z, -0.1326 for β_{z^*}) and moderate improvement than Naive 2 (-0.0013 for β_0, -0.0001 for β_x, 0.0005 for β_z, 0.0024 for β_{z^*}). These improvements are also true for the RMSEs. In this case, we can also

Degree of error	Parameter	Naive 1 Bias	Naive 1 RMSE	Naive 1 CR	Naive 2 Bias	Naive 2 RMSE	Naive 2 CR	Proposed Bias	Proposed RMSE	Proposed CR
No error										
($\sigma = 0$)	β_0	0.0011	0.0922	0.95	0.0010	0.0107	0.93	0.0003	0.0048	0.96
	β_x	0.0065	0.0774	0.98	-0.0010	0.0090	0.95	0.0000	0.0041	0.96
	β_z	0.0084	0.1143	0.96	-0.0006	0.0134	0.95	0.0001	0.0055	0.95
	β_{z^*}	0.0095	0.1508	0.96	-0.0007	0.0174	0.94	-0.0003	0.0083	0.95
	σ_u^2	-0.0106	0.1441	0.97	-0.0007	0.0151	0.91	-	-	-
	δ_0	-	-	-	-	-	-	-0.0255	0.1388	0.94
	δ_1	-	-	-	-	-	-	-0.0146	0.0798	0.93
	λ_0	-	-	-	-	-	-	-0.0045	0.0307	0.89
	λ_1	-	-	-	-	-	-	-0.0099	0.0520	0.92
Moderate error										
($\sigma = 0.4$)	β_0	0.1045	0.1364	0.80	-0.0013	0.0122	0.93	-0.0003	0.0048	0.95
	β_x	-0.2404	0.2487	0.03	-0.0001	0.0088	0.94	-0.0001	0.0042	0.94
	β_z	-0.1236	0.1639	0.79	0.0005	0.0137	0.95	-0.0004	0.0064	0.95
	β_{z^*}	-0.1326	0.1959	0.86	0.0024	0.0207	0.94	0.0007	0.0077	0.94
	σ_u^2	-0.0106	0.1441	0.97	-0.0003	0.0188	0.94	-	-	-
	δ_0	-	-	-	-	-	-	-0.0189	0.1184	0.93
	δ_1	-	-	-	-	-	-	-0.0142	0.0730	0.93
	λ_0	-	-	-	-	-	-	0.0002	0.0270	0.92
	λ_1	-	-	-	-	-	-	-0.0069	0.0534	0.91
Severe error										
($\sigma = 0.8$)	β_0	0.2772	0.2884	0.05	0.0004	0.0088	0.94	0.0002	0.0046	0.94
	β_x	-0.7046	0.7061	0.00	-0.0004	0.0086	0.92	0.0002	0.0039	0.96
	β_z	-0.3239	0.3380	0.10	0.0008	0.0154	0.92	-0.0001	0.0057	0.95
	β_{z^*}	-0.3290	0.3541	0.26	0.0003	0.0191	0.93	-0.0002	0.0076	0.94
	σ_u^2	-0.0106	0.1441	0.97	-0.0016	0.0160	0.94	-	-	-
	δ_0	-	-	-	-	-	-	-0.0343	0.1479	0.91
	δ_1	-	-	-	-	-	-	-0.0191	0.0768	0.92
	λ_0	-	-	-	-	-	-	-0.0030	0.0289	0.91
	λ_1	-	-	-	-	-	-	-0.0102	0.0507	0.93

Table 2. Bias, RMSE, and CR for 95% CI of the parameters estimate in case of $n_i = 10$ for the three methods (Naive 1, Naive 2, Proposed) under the random effects covariance matrix structure (13)

see the smaller amount of biases (with good CRs) of random effects covariance parameters in comparison with the 5 number of follow-ups $(n_i = 5)$.

Performance of the three methods in the context of the severe measurement error indicates that the proposed approach performs well with small biases and good CRs for the fixed effects parameters and random effects covariance parameters. Also, the RMSEs show the consistency of the performance as other scenarios. On the other hand, the Naive 1 performs poorly with relatively large amount of biases and low CRs particularly for the parameter of measurement error variable (-0.7046 for β_x with 0% CR). Performance of the proposed approach and Naive 2 are similar in terms of biases but the proposed approach has much smaller RMSEs. The estimates of GARPs and IVs represent fairly small biases (with good CRs) and smaller RMSEs compared to the corresponding values for the 5 number of follow-ups $(n_i = 5)$.

We also evaluate performance of the proposed approach under model mis-specification in measurement error distribution (Case 1). It is clear from Table 3 that there are smaller biases and RMSEs for the fixed parameters and random effects parameters of the proposed approach compared to the Naive methods.

Table 4 represents the results of the fixed and random effects parameters for the proposed approach under different

	Naive 1			Naive 2			Proposed		
Parameter	Bias	RMSE	CR	Bias	RMSE	CR	Bias	RMSE	CR
β_0	0.3117	0.3458	0.46	-0.0022	0.0789	0.92	0.0019	0.0235	0.92
β_x	-0.9575	0.9575	0.00	0.0008	0.0256	0.93	-0.0005	0.0079	0.95
β_z	-0.3504	0.3730	0.18	-0.0037	0.0657	0.94	0.0037	0.0241	0.94
β_{z^*}	-0.3624	0.3902	0.29	0.0060	0.1037	0.95	0.0005	0.0331	0.94
σ_u^2	0.0039	0.1338	0.94	-0.0045	0.0321	0.92	-	-	-
δ_0	-	-	-	-	-	-	0.2314	0.8136	0.89
δ_1	-	-	-	-	-	-	-0.0423	0.1218	0.83
λ_0	-	-	-	-	-	-	-0.0034	0.0515	0.86
λ_1	-	-	-	-	-	-	-0.0146	0.0950	0.89

Table 3. Bias, RMSE, and CR for 95% CI of the parameters estimate in case of $n_i = 5$ for the three methods (Naive 1, Naive 2, Proposed) under the random effects covariance matrix structure (13) and mis-specified error distribution

structured covariance matrix in the case of severe error $[\sigma = 0.8]$ when number of follow-ups is 5 $(n_i = 5)$ (Case 2). It is evident from the results that the proposed approach performs consistently better than the other two Naive methods (from Table 1) in terms of biases and RMSE for the all three models. However, in the case of Model 1 (heterogeneous AR2) the performance of the proposed approach is similar to the Naive 2 method (from Table 1) in terms of biases and CRs but the proposed approach shows much smaller RMSEs.

	Model 1			Model 2			Model 3		
Parameter	Bias	RMSE	CR	Bias	RMSE	CR	Bias	RMSE	CR
β_0	0.0006	0.0068	0.97	-0.0001	0.0090	0.96	0.0002	0.0090	0.94
β_x	0.0020	0.0062	0.95	-0.0000	0.0056	0.95	0.0009	0.0057	0.96
β_z	0.0026	0.0088	0.93	0.0009	0.0083	0.95	0.0000	0.0108	0.95
β_{z^*}	0.0030	0.0121	0.96	0.0006	0.0134	0.94	0.0009	0.0132	0.95
δ_0	0.0092	0.2037	0.97	0.0139	0.1677	0.93	0.0300	0.1802	0.93
δ_1	0.0138	0.2036	0.98	0.0220	0.1202	0.92	0.0296	0.1096	0.92
δ_2	0.0109	0.0487	0.93	0.0194	0.0885	0.91	0.0379	0.1018	0.91
λ_0	-0.0029	0.0197	0.95	0.0429	0.1329	0.93	0.0991	0.2770	0.96
λ_1	0.0059	0.0514	0.93	0.0989	0.3766	0.96	0.0896	0.2505	0.95
λ_2	-	-	-	-	-	-	0.0803	0.2273	0.95

Table 4. Bias, RMSE, and CR for 95% CI of the parameters estimate in case of $n_i = 5$ for the proposed method under different heterogeneous structures of random effects covariance matrix and severe error $[\sigma = 0.8]$

Overall, based on the simulation results, it is evident that the larger biases can occur in the fixed effects parameters by ignoring the measurement error in covariate and also not specifying the distribution of random effects correctly. The simulation results also demonstrate that the proposed approach performs very well with small biases and RMSEs as well as good CRs for 95% CI. Moreover, as expected, the RMSEs for the all methods tend to decrease when the number of follow-ups (n_i) increases.

6. Application: Manitoba Follow-up study

The Manitoba Follow-up study (MFUS) is the longest running study of cardiovascular disease and ageing in Canada. It is believed that the MFUS is the only cohort study in the world which is financed by the members who are being studied. The

MFUS cohort consists of 3983 men who were recruits in the Royal Canadian Air Force during the early years of World War II and was established at the University of Manitoba on July 1, 1948 [30]. The mean age of the men in the cohort was around 31 years, with about 90% between age 20 and 39 years. It was declared that all men were free of clinical evidence of ischaemic heart disease (IHD). The baseline measurement of systolic and diastolic blood pressure and body mass index (mean $\pm$ standard deviation) were found 121 ± 10 mmHg, 76 ± 8 mmHg and 23.8 ± 2.7kg/m^2, respectively [31]. At present, the MFUS continues with its 68th year of uninterrupted study.

For the purpose of our study, a sub-sample of the MFUS data has been used. For the selection of sample from MFUS participants, approximately 500 men, $1/8$th of the cohort was chosen. In particular, from the registry file (one record per MFUS man) a random number in the interval $[0, 1]$ was selected using the *ranuni* function in SAS software [32], and the man retained in our file if the random number was less than 0.125. The selected people were merged by ID number with the blood pressure and body weight file, and all measurements recorded between July 1, 1948 and July 1, 2008 were kept. Finally, the complete data set contains 373 members with 10 first follow-up observations for each member which has been used in the analysis.

An important risk factor for cardiovascular disease is high blood pressure which is one of the major cause of mortality. However, environmental and genetic factors and their interactions may be the cause for complex disorder disease like high blood pressure [33]. In this work, it is of interest how hypertension which is based on blood pressure is associated with corresponding risk factors and how individual measurements vary within individuals. For the purpose of analysis, we divide individuals into two categories: having hypertension and not having hypertension. An individual is said to have hypertension if his/her systolic blood pressure is greater than 140 mmHg or his/her diastolic blood pressure is greater than 90 mmHg [34]. The main covariates of interest include age, body mass index (BMI), and IHD status [Yes, No (ref)]. An individual is said to have IHD if he develops IHD in any of his 10 first follow-ups. These covariates are taken into consideration as they are found to have significant impact on the occurrence of hypertension in many studies [35–39]. In particular, there have been many literatures which show the association between IHD and blood pressure [40, 41]. For this study, July 1, 1948 or the closest date from July 1, 1948 is considered as the baseline. Based on the baseline information, it is observed that among the individuals, the mean age is 30.6 years with standard deviation 5.24 and the minimum and maximum age are 20.2 and 52.6, respectively. The mean BMI is 23.87 kg/m^2 with standard deviation 2.66 kg/m^2. Note that at the baseline all individuals are IHD free. There are 9.1% individuals who had IHD till 10th follow-up, and 25.64% of individuals developed IHD till July 1, 2008.

In many longitudinal studies, the reported BMI of a person for a certain age is actually the long term average values of BMI for that person in that year. BMI is the the ratio of weight (kg) and height (m^2) and weight has a daily as well as seasonal variations. That is why the true and observed BMI are different. Moreover, since only the overall/baseline weight and height are considered to calculate BMI, there is always an overestimation/underestimation issue of true BMI. Many literatures have shown that the BMI is subject to measurement error [42–45].

In MFUS, the all BMIs are reported by the physicians, however, the BMIs are based on the follow-up weights and heights from the baseline. It means that the physicians in the follow-up times only ask for the weights and then report the BMI using the weights with the heights at the baseline. Hence, it can be considered that there exists a variability of any BMI measurement taken at a specific assessment time for an individual. In particular, it can be said that the observed BMI may overestimate the true BMI.

Let the response variable Y_{ij} be the binary response, taking 1 if the subject i has hypertension at assessment j, and 0 otherwise. We can consider the model as follows:

$$\text{logit}\{P_{ij}\} = \beta_0 + \beta_1 \text{Age}_{ij} + \beta_2 \text{BMI}_i + \beta_3 \text{IHD}_i + u_{ij}, \tag{14}$$

where $i = 1, 2, \ldots, 373$, $j = 1, 2, \ldots, 10$, with $P_{ij} = P(Y_{ij} = 1|\text{Age}_{ij}, \text{BMI}_i, \text{IHD}_i, u_{ij})$ and each individual has 10 visits. It is assumed that $\mathbf{u}_i = (u_{i1}, \ldots, u_{ij})$ follows a multivariate Normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}_i$.

Here BMI_i, representing the true body mass index over time for subject i, which cannot be observed in practice and is treated as the error-contaminated covariate. To feature the measurement error variation we employ the following classical structural measurement error model $\text{BMI}_{ij} = \text{BMI}_i + e_{ij}$, where BMI_{ij} is the measurement taken for subject i at assessment time point j, $\text{BMI}_i \sim N(\mu_x, \sigma_x^2)$, and e_{ij}'s are assumed to follow independent Normal distribution with mean 0 and variance σ^2. It is known that the errors over replicates can be considered conditionally independent given the long-term average if the replicates within an individual are taken far enough apart in time [22]. In MFUS, during first 15 years, routine medical examination including BMI, blood pressure were requested from study members at 5-year intervals, then 3-year intervals, every year, twice a year and now three times each year [31]. This actually demonstrates the BMI measurements within an individual were taken in long time distance and suggested that the errors can be considered independent over these replicates.

6.1. *Analysis of impact of covariates on hypertension*

We analyze the data with the three approaches: the proposed approach which allows the random effects covariance matrix to vary by IHD for each subject; Naive 2 approach where constant random effects covariance matrix with AR(1) structure is considered across the subjects; and Naive 1 approach where measurement error in covariate is also ignored. The results for the all three approaches are reported in Table 5. In particular, the associated model parameters estimates, their standard errors, and corresponding 95% confidence intervals for the all three approaches are provided. The GARP and IV parameters are obtained by specifying $k_{i,jt}$ and $h_{i,j}$ as follows:

$$\mathbf{k}_{i,j,j-1} = (1, \text{IHD}_i) \quad \text{and} \quad \mathbf{h}_{i,j} = (1, \text{IHD}_i),$$

and estimated value of $\mathbf{\Sigma}_i$ is calculated using $\mathbf{\Sigma}_i = \mathbf{T}_i^{-1}\mathbf{D}_i(\mathbf{T}_i^T)^{-1}$, where $\mathbf{D}_i = \text{diag}(\sigma_{i1}^2, \sigma_{i2}^2, \ldots, \sigma_{in_i}^2)$, $\log\left(\sigma_{ij}^2\right) = \mathbf{h}_{i,j}^T\boldsymbol{\lambda}$, and $\mathbf{T}_i$ is a unit lower triangular matrix having ones on its diagonal and $-\phi_{i,jt}$ ($\phi_{i,jt} = \mathbf{k}_{i,jt}^T\boldsymbol{\delta}$) in the (j,t)th element for $2 \le j \le 10$. Here we specify the following structure for the parameters of random effects covariance matrix:

$$\phi_{i,jt} = \delta_0 I(|j-t|=1) + \delta_1 I(|j-t|=1)\text{IHD}_i \quad \text{and} \quad \log(\sigma_{ij}^2) = \lambda_0 + \lambda_1\text{IHD}_i. \tag{15}$$

Moreover, the adjusted odds ratio (OR) is used to compare the odd of occurrence of hypertension with covariates. In logistic regression, estimate of OR e.g. for covariate $\text{Age}_{ij}(i = 1, \ldots, m;\ j = 1, \ldots, n_i)$ can be obtained as $\widehat{OR} = \exp(\hat{\beta}_1)$. The $100(1-\alpha)\%$ confidence interval for OR is

$$\widehat{OR} \pm z_{\alpha/2}\sqrt{\text{var}(\widehat{OR})},$$

where $\text{var}(\widehat{OR}) = \left(\widehat{OR}\right)^2\text{var}(\hat{\beta}_1) = \exp(2\hat{\beta}_1)\text{var}(\hat{\beta}_1)$ using the delta method. The estimated OR and its standard error with 95% confidence interval for the covariates under the all three approaches are given in Table 6.

6.1.1. Estimation of parameters From Table 5, in case of Naive 1, it is observed that BMI, age, and IHD are positively associated with the occurrence of hypertension. The 95% confidence intervals of these fixed effects estimate indicate the significant effect of these covariates on hypertension. For Naive 2, BMI is positively associated with the development of hypertension and this effect is found to be significant. Age and IHD also have significant positive effects on the hypertension. Here, the reliability ratio, $\dfrac{\sigma_x^2}{\sigma_x^2 + \sigma^2}$, is 0.83 which indicates the amount of error associated with the covariate BMI. Hence, it is clear that there is 17% error associated with the covariate BMI.

The estimate of GARP and IV parameters indicates that the covariance matrix varies according to the IHD group. This result demonstrates that the random effects covariance matrix differs by measured covariates and neglecting this heterogeneity can cause the biased estimate of model parameters [4]. In the estimate of IV, the coefficient of IHD was found significant which indicates that the estimated IV was higher for those individuals

	Naive 1				Naive 2				Proposed			
Parameter	Est	SE	LB	UB	Est	SE	LB	UB	Est	SE	LB	UB
β_0	-13.72	0.580	-14.86	-12.58	-13.16	0.573	-14.28	-12.04	-12.09	0.588	-13.24	-10.94
β_{BMI}	0.278	0.021	0.237	0.319	0.269	0.020	0.230	0.308	0.208	0.021	0.167	0.249
β_{Age}	0.089	0.006	0.077	0.101	0.083	0.006	0.071	0.095	0.089	0.005	0.079	0.099
β_{IHD}	1.227	0.162	0.909	1.545	1.033	0.161	0.717	1.349	0.561	0.180	0.208	0.914
σ_u^2	1.982	0.046	1.892	2.072	1.959	0.045	1.870	2.047	-	-	-	-
μ_x	-	-	-	-	23.98	0.042	23.90	24.06	24.03	0.042	23.95	24.12
σ_x^2	-	-	-	-	6.612	0.153	6.312	6.912	6.533	0.151	6.237	6.829
σ^2	-	-	-	-	1.380	0.032	1.317	1.443	1.380	0.032	1.317	1.443
δ_0	-	-	-	-	-	-	-	-	0.146	0.005	0.136	0.156
δ_1	-	-	-	-	-	-	-	-	-0.111	0.001	-0.113	-0.109
λ_0	-	-	-	-	-	-	-	-	1.561	0.012	1.537	1.585
λ_1	-	-	-	-	-	-	-	-	1.973	0.012	1.949	1.997

Table 5. Estimate (Est), standard error (SE), and lower bound (LB) and upper bound (UB) of 95% CI of model parameters estimate for the Naive 1, Naive 2, and Proposed approaches

who have IHD than the individuals without IHD. By following equation (15), the estimated values for $\hat{\mathbf{D}}$ for each group of IHD are $\log \hat{\mathbf{D}}_{(IHD=0)} = \text{diag}\,(1.56, 1.56, 1.56, 1.56, 1.56, 1.56, 1.56, 1.56, 1.56, 1.56)$ and $\log \hat{\mathbf{D}}_{(IHD=1)} = \text{diag}\,(3.53, 3.53, 3.53, 3.53, 3.53, 3.53, 3.53, 3.53, 3.53, 3.53)$. This demonstrates that the estimated IV parameters vary across the status of IHD. The coefficient of IHD in the estimates of GARP is also found significant and this also indicates the substantial variation of GARPs across the status of IHD. As $\hat{\delta}_1$ and $\hat{\lambda}_1$ are statistically significant from zero, the proposed approach also works better than assuming the same correlation structural AR(1) for the all subjects (Naive 2). By following equation (15), the estimated values of $\hat{\mathbf{T}}$ for each group of IHD are given by

$$\hat{\mathbf{T}}_{(IHD=0)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.15 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.15 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.15 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.15 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.15 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.15 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.15 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.15 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.15 & 1 \end{bmatrix},$$

$$\hat{\mathbf{T}}_{(IHD=1)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.04 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.04 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.04 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.04 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.04 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.04 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.04 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.04 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.04 & 1 \end{bmatrix}.$$

The significant estimate of coefficient of BMI indicates that the estimated conditional probability of hypertension given the random effects increases with the increase of individual's BMI. Also, age was found significant and the conditional probability of hypertension increases as age increases. The reliability ratio for the proposed approach is also 0.82 which indicates the amount of 18% error involvement with the covariate BMI.

The fixed effects estimates for the all three approaches reveal the same nature of covariates effects on hypertension status, but there exists a variation in the magnitude of the fixed effects for the naive methods compared to the proposed approach, especially in case of IHD covariate. The naive approaches suggest a significant positive IHD effect of individuals who have IHD compared to the individuals without IHD on hypertension status that is nearly more than two times higher than the corresponding value in the proposed approach. This might be due to ignoring the heterogeneity or temporal dependence on random effects covariance matrix in the naive methods. The estimates of BMI effect are also larger in naive methods than the proposed approach. We also observe that the random effects variance estimate (and its standard error) for the both naive methods are close to each other, and the variance of measurement error (σ^2) and corresponding mean and variance of BMI (μ_x, σ_x^2) are also similar for the both Naive 2 and proposed approaches.

We also calculate the Akaike information criterion (AIC) as a goodness-of-fit criterion to compare the three models. The AIC values indicated that the proposed approach provided a better fit than the Naive models (3780.3, 6388.8, and 4627.2 for the proposed approach, Naive 2, and Naive 1, respectively).

6.1.2. Estimation of odds ratio To examine the association of a covariate with the occurrence of hypertension controlling other covariates in the model, one may use adjusted OR. The adjusted OR and its standard error with 95% confidence interval for the covariates under the three approaches are given in Table 6.

	Naive 1			**Naive 2**			**Proposed**		
Quantity	BMI	Age	IHD	BMI	Age	IHD	BMI	Age	IHD
OR	1.320	1.093	3.411	1.309	1.087	2.809	1.231	1.093	1.752
SE	0.028	0.007	0.553	0.026	0.007	0.452	0.026	0.005	0.315
95% LB	1.266	1.080	2.328	1.257	1.074	1.923	1.181	1.082	1.134
95% UB	1.375	1.106	4.494	1.360	1.099	3.696	1.282	1.104	2.371

Table 6. Odds ratio (OR), standard error (SE), lower bound (LB) and upper bound (UB) of 95% CI of model parameters estimate for the Naive 1, Naive 2, and Proposed approaches

From Table 6, in case of Naive 1, it is found that the OR for BMI is 1.320 which implies that with one unit increase in BMI, the odds of developing hypertension is expected to increase 32%. It is interesting to observe that the development of hypertension increases with the increase of age as well (OR=1.09). In case of IHD, it is observed that an individual with IHD is 241% more likely to have hypertension compared to an individual without IHD. For the Naive 2, it is observed that the estimated OR for BMI is 1.309 with p-value 0.00. It implies that the odds of having hypertension is significantly increased by 31% with one unit increase of BMI. The OR of age also reveals the same nature which means with one unit increase in age, the odds of developing hypertension is increased by 9%. For the IHD, the OR is 2.809 which means that the individuals with IHD are 181% more likely to develop hypertension than the individuals without IHD. Same nature can be revealed for the BMI, age, and IHD for the proposed approach. The OR for the BMI indicates the significant increase of odds (1.231) of developing hypertension with the increase of one unit in BMI. The odds of having hypertension is expected to increase 9% with one unit increase in age. For the IHD, the OR is 1.75 which means that the individuals with the IHD are 75% more likely to develop hypertension than the individuals without the IHD. We also observe that the all three covariates are statistically significant from zero.

7. Concluding remarks

In this paper, our aim was to properly model the random effects covariance matrix under the GLMMs with covariates measurement error. For this purpose, we extended the model introduced by Lee *et al.* [11] to model the random effects covariance matrix for the GLMMs to the case when the covariates are subject to measurement error using modified Cholesky decomposition. This covariance matrix was decomposed to the GARPs and IVs parameters and this structure is able to accommodate the heterogeneous covariance matrix which depends on subject-specific covariates. In this paper, we analytically derived the necessary formulae for the case of binary outcome, however, one can follow the same steps to derive for other exponential family distributions. From the simulation studies, we have demonstrated that the proposed approach performs very well in terms of bias, RMSE as well as coverage rate of the model parameters estimate. The simulation studies also indicated that the larger biases can occur in the fixed effects parameters by ignoring the measurement error in covariates and also not specifying the distribution of random effects correctly. The proposed approach for modelling random effects covariance matrix is also computationally attractive and provides parameters which have sensible interpretation for modelling trajectories over time. Especially, the dependence and variability of the random effects can be characterized by the covariance parameters. To incorporate the heterogeneity in the random effects, we worked with the random intercept model to explicitly show performance of our proposed approach, however, one can also consider the random effects in terms of design matrix.

In the longitudinal data analysis with covariates measurement error, if the main interest is the covariance structure or subject-specific prediction, then proper care needs to be taken in modeling the covariance structure as well as the measurement error. However, taking proper care is important even if these are not the main or direct interests. For instance, if the random effects covariance matrix is not modeled correctly when it is function of subject-specific covariates, then the inference will be incorrect which may lead to wrong conclusions.

It should be noted that in measurement error problem, model identifiability is an important issue. In case of measurement error, additional data source such as a validation sub-sample or replications is needed to perform a measurement error analysis [22]. In longitudinal studies, repeated measurements are collected for error-prone variables, and for identifying model parameters these measurements can be used as replicates. In particular, the parameters of an error model are often identifiable if the number of repeated assessment of measurement error covariates is larger than the number of parameters in the error model [23]. If the parameters are not identifiable, then in numerical iterative procedures, fast divergence can occur. For example, if there is a non-identifiability problem then the EM algorithm would diverge quickly [46]. Our numerical experience, however, did not indicate that there is an issue with non-identifiability for the models considered (equations (1) to (3)) in this work (simulation study as well as the real data application).

In the past twenty years, significant contributions have been made in the area of longitudinal data with covariates measurement error. However, there are still a lot of interesting and important problems related to this work need to be explored as future works. For example, our plan is to study our random effects subject-specific variance-covariance matrix for the marginalized random effects models as well. It would be also interesting to model the random effects covariance matrix in case of longitudinal data with response measurement error. The proposed model is based on measurement error in the mean model, however, one can extend it to study the measurement error in the variance model as well. We also assumed that the design vectors $\mathbf{k}$ and $\mathbf{h}$ needed to model the GARP/IV parameters are subsets of error-free covariates, however, one can extend our proposed approach to also consider the covariates with measurement error to build the model to estimate the GARP/IV parameters. We have planned to study these approaches in our future studies.

In conclusion, in the presence of covariates measurement error in longitudinal data, incorrectly modeling the random effects covariance matrix can have significant effects on inference in model parameters. Hence, it is important to properly model the random effects covariance matrix in the presence of covariates measurement error.

Appendix A

A.1. An illustration

This appendix contains an illustration of the general inference procedure of our proposed method in the case of binary outcome. In particular, let us consider the longitudinal binary data and assume the following logistic mixed model:

$$\text{logit}\big\{P(Y_{ij} = 1 \mid x_{ij}, z_{ij}, u_{ij})\big\} = \beta_0 + \beta_x x_{ij} + \beta_z z_{ij} + u_{ij},$$

and the density function for y_{ij} is given by $f(y_{ij}|\mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta}) = P_{ij}(u_{ij})^{y_{ij}}[1 - P_{ij}(u_{ij})]^{1-y_{ij}}$, $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n_i$, with

$$P_{ij}(u_{ij}) = P(Y_{ij} = 1|x_{ij}, z_{ij}, u_{ij}) = \frac{\exp(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij})}{1 + \exp(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij})}, \quad \text{where} \quad \mathbf{x}_{ij} = (1, x_{ij}, z_{ij})^T, \boldsymbol{\beta}^T = (\beta_0, \beta_x, \beta_z),$$

and $1 - P_{ij}(u_{ij}) = \dfrac{1}{1 + \exp(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij})}$. We can then write the following density function:

$$\begin{aligned} f(\mathbf{y}_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta}) &= \frac{\exp\{\sum_{j=1}^{n_i} y_{ij}\big(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij}\big)\}}{\prod_{j=1}^{n_i}\{1 + \exp\big(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij}\big)\}} \\ &= \exp\left[\sum_{j=1}^{n_i} y_{ij}\big(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij}\big) - \left(\sum_{j=1}^{n_i} \log\left\{1 + \exp\big(\mathbf{x}_{ij}^T\boldsymbol{\beta} + u_{ij}\big)\right\}\right)\right]. \end{aligned} \tag{16}$$

Also, $f(\mathbf{u}_i)$ has a multivariate Normal density with mean vector **0** and covariance matrix $\boldsymbol{\Sigma}_i$ which can be written as:

$$f(\mathbf{u}_i; \boldsymbol{\delta}, \boldsymbol{\lambda}) = (2\pi)^{-n_i/2}\left[\prod_{j=1}^{n_i} \big(\sigma_{ij}^2\big)^{-1/2}\right]\exp\left(-\frac{1}{2}\sum_{j=1}^{n_i}\frac{\epsilon_{ij}^2}{\sigma_{ij}^2}\right) \text{ with } \quad \epsilon_{i1} = u_{i1}. \tag{17}$$

Moreover, let us consider classical additive structural measurement error model as $w_{ij} = x_{ij} + e_{ij}$, where $e_{ij} \sim N(0, \sigma^2)$ and x_{ij} be the covariate subject to error and has Normal distribution $N(\mu_x, \sigma_x^2)$. To avoid identifiability issue, one should assume whether the parameter σ_x^2 or σ^2 is known. We can write $w_{ij}|x_{ij} \sim N(x_{ij}, \sigma^2)$. Then the conditional distribution of $x_{ij}|w_{ij}$ can be written as follows:

$$x_{ij}\Big|w_{ij}; \mu_x, \sigma_x^2, \sigma^2 \sim N(\mu_f, \sigma_f^2), \quad \text{where} \quad \mu_f = \frac{\mu_x\sigma^2 + w_{ij}\sigma_x^2}{\sigma^2 + \sigma_x^2} \text{ and } \sigma_f^2 = \frac{\sigma_x^2\sigma^2}{\sigma^2 + \sigma_x^2}. \tag{18}$$

Therefore, we can write the complete data log-likelihood following (8) as:

$$
\begin{aligned}
l_c(\boldsymbol{\theta}) &= \sum_{i=1}^{m} \log L_i(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i) \\
&= \sum_{i=1}^{m} \left[\sum_{j=1}^{n_i} y_{ij}\big(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_{ij}\big) - \left(\sum_{j=1}^{n_i} \log \left\{1 + \exp\big(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_{ij}\big)\right\} \right) \right] \\
&+ \sum_{i=1}^{m} \left[-\frac{n_i}{2} \log\left(2\pi \frac{\sigma_x^2 \sigma^2}{\sigma^2 + \sigma_x^2}\right) - \frac{1}{2} \sum_{j=1}^{n_i} \frac{\left(x_{ij} - \frac{\mu_x \sigma^2 + w_{ij} \sigma_x^2}{\sigma^2 + \sigma_x^2}\right)^2}{\frac{\sigma_x^2 \sigma^2}{\sigma^2 + \sigma_x^2}} \right] \\
&+ \sum_{i=1}^{m} \left[-\frac{n_i}{2} \log(2\pi) - \sum_{j=1}^{n_i} \frac{1}{2} \log \sigma_{ij}^2 - \frac{1}{2} \sum_{j=1}^{n_i} \frac{\epsilon_{ij}^2}{\sigma_{ij}^2} \right],
\end{aligned}
\tag{19}
$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T = (\mu_x, \sigma_x^2)^T, \sigma^2, \boldsymbol{\delta}^T, \boldsymbol{\lambda}^T\right)^T$ is the associated parameters to develop the EM algorithm. We can then write the observed data likelihood as follows:

$$
L(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i) = \int\int f(\mathbf{y}_i | \mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta}) f(\mathbf{x}_i | \mathbf{w}_i; \boldsymbol{\gamma}, \sigma^2) f(\mathbf{u}_i; \boldsymbol{\delta}, \boldsymbol{\lambda}) d\mathbf{x}_i d\mathbf{u}_i \tag{20}
$$

This likelihood function is not in closed form, so we rely on MCEM algorithm to evaluate this. The ML estimators of $\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \sigma^2$ can be obtained by solving the following estimating equations:

$$
\sum_{i=1}^{m} \mathrm{E}\left\{ \frac{\partial \log f\big(\mathbf{y}_i \mid \mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta}\big)}{\partial \boldsymbol{\beta}} \middle| \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i \right\} = 0,
$$

$$
\sum_{i=1}^{m} \mathrm{E}\left\{ \frac{\partial \log f\big(\mathbf{x}_i; \boldsymbol{\gamma}, \sigma^2\big)}{\partial \boldsymbol{\gamma}} \middle| \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i \right\} = 0,
$$

$$
\sum_{i=1}^{m} \mathrm{E}\left\{ \frac{\partial \log f\big(\mathbf{x}_i; \boldsymbol{\gamma}, \sigma^2\big)}{\partial \sigma^2} \middle| \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i \right\} = 0,
$$

$$
\sum_{i=1}^{m} \mathrm{E}\left\{ \frac{\partial \log f\big(\mathbf{u}_i; \boldsymbol{\delta}, \boldsymbol{\lambda}\big)}{\partial \boldsymbol{\delta}} \middle| \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i \right\} = 0,
$$

$$
\sum_{i=1}^{m} \mathrm{E}\left\{ \frac{\partial \log f\big(\mathbf{u}_i; \boldsymbol{\delta}, \boldsymbol{\lambda}\big)}{\partial \boldsymbol{\lambda}} \middle| \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i \right\} = 0,
$$

where the conditional expectations are with respect to the conditional distribution of random effects $\big(\mathbf{x}_i, \mathbf{u}_i\big)$ given the observed data $\big(\mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i\big)$. Here for the parameters $\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \sigma^2$, the above score functions for individual i can be expressed as:

$$
\frac{\partial \log L(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{x}_i, \mathbf{u}_i)}{\partial \boldsymbol{\beta}} = \frac{1}{L(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{w}_i, \mathbf{z}_i)} \int\int \frac{\partial f(\mathbf{y}_i | \mathbf{x}_i, \mathbf{z}_i, \mathbf{u}_i; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} f(\mathbf{x}_i | \mathbf{w}_i, \mathbf{z}_i; \boldsymbol{\gamma}, \sigma^2) f(\mathbf{u}_i; \boldsymbol{\delta}, \boldsymbol{\lambda})\, d\mathbf{x}_i d\mathbf{u}_i, \tag{21}
$$

where,

$$\frac{\partial f(\mathbf{y}_i|\mathbf{x}_i,\mathbf{z}_i,\mathbf{u}_i;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \exp\left[\sum_{j=1}^{n_i} y_{ij}\left(\mathbf{x}_{ij}^T\boldsymbol{\beta}+u_{ij}\right) - \left(\sum_{j=1}^{n_i}\log\left\{1+\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta}+u_{ij}\right)\right\}\right)\right]$$
$$\left[\sum_{j=1}^{n_i} y_{ij}\mathbf{x}_{ij} - \frac{1}{1+\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta}+u_{ij}\right)}\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta}+u_{ij}\right)\mathbf{x}_{ij}\right]$$
$$= \exp\left[\sum_{j=1}^{n_i} y_{ij}\left(\mathbf{x}_{ij}^T\boldsymbol{\beta}+u_{ij}\right) - \left(\sum_{j=1}^{n_i}\log\left\{1+\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta}+u_{ij}\right)\right\}\right)\right]\left[\sum_{j=1}^{n_i}\left(y_{ij}-P_{ij}(u_{ij})\right)\mathbf{x}_{ij}\right].$$

Also,

$$\frac{\partial \log L(\boldsymbol{\theta};\mathbf{y}_i,\mathbf{x}_i,\mathbf{u}_i)}{\partial\sigma^2} = \frac{1}{L(\boldsymbol{\theta};\mathbf{y}_i,\mathbf{w}_i,\mathbf{z}_i)}\int\int f(\mathbf{y}_i|\mathbf{x}_i,\mathbf{z}_i,\mathbf{u}_i;\boldsymbol{\beta})\frac{\partial f(\mathbf{x}_i|\mathbf{w}_i;\boldsymbol{\gamma},\sigma^2)}{\partial\sigma^2}f(\mathbf{u}_i;\boldsymbol{\delta},\boldsymbol{\lambda})\,d\mathbf{x}_i d\mathbf{u}_i, \quad (22)$$

where $\frac{\partial f(\mathbf{x}_i|\mathbf{w}_i;\boldsymbol{\gamma},\sigma^2)}{\partial\sigma^2} = \frac{\partial}{\partial\sigma^2}\left[\exp\{\log f(\mathbf{x}_i|\mathbf{w}_i;\boldsymbol{\gamma},\sigma^2)\}\right] = f(\mathbf{x}_i|\mathbf{w}_i;\boldsymbol{\gamma},\sigma^2)\frac{\partial}{\partial\sigma^2}\{\log f(\mathbf{x}_i|\mathbf{w}_i;\boldsymbol{\gamma},\sigma^2)\}$,
and

$$\frac{\partial \log L(\boldsymbol{\theta};\mathbf{y}_i,\mathbf{x}_i,\mathbf{u}_i)}{\partial\boldsymbol{\gamma}} = \frac{1}{L(\theta;\mathbf{y}_i,\mathbf{w}_i,\mathbf{z}_i)}\int\int f(\mathbf{y}_i|\mathbf{x}_i,\mathbf{z}_i,\mathbf{u}_i;\boldsymbol{\beta})\frac{\partial f(\mathbf{x}_i|\mathbf{w}_i;\boldsymbol{\gamma},\sigma^2)}{\partial\boldsymbol{\gamma}}f(\mathbf{u}_i;\boldsymbol{\delta},\boldsymbol{\lambda})\,d\mathbf{x}_i d\mathbf{u}_i, \quad (23)$$

where $\frac{\partial f(\mathbf{x}_i|\mathbf{w}_i;\boldsymbol{\gamma},\sigma^2)}{\partial\boldsymbol{\gamma}} = \frac{\partial}{\partial\boldsymbol{\gamma}}\left[\exp\{\log f(\mathbf{x}_i|\mathbf{w}_i;\boldsymbol{\gamma},\sigma^2)\}\right] = f(\mathbf{x}_i|\mathbf{w}_i;\boldsymbol{\gamma},\sigma^2)\frac{\partial}{\partial\boldsymbol{\gamma}}\{\log f(\mathbf{x}_i|\mathbf{w}_i;\boldsymbol{\gamma},\sigma^2)\}$.
Furthermore,

$$\frac{\partial \log L(\boldsymbol{\theta};\mathbf{y}_i,\mathbf{x}_i,\mathbf{u}_i)}{\partial\boldsymbol{\delta}} = \frac{1}{L(\boldsymbol{\theta};\mathbf{y}_i,\mathbf{w}_i,\mathbf{z}_i)}\int\int f(\mathbf{y}_i|\mathbf{x}_i,\mathbf{z}_i,\mathbf{u}_i;\boldsymbol{\beta})f(\mathbf{x}_i|\mathbf{w}_i,\mathbf{z}_i;\boldsymbol{\gamma},\sigma^2)\frac{\partial f(\mathbf{u}_i;\boldsymbol{\delta},\boldsymbol{\lambda})}{\partial\boldsymbol{\delta}}\,d\mathbf{x}_i d\mathbf{u}_i, \quad (24)$$

where $f(\mathbf{u}_i;\boldsymbol{\delta},\boldsymbol{\lambda})$ is defined in (17) and $\phi_{i,jt} = \mathbf{k}_{i,jt}^T\boldsymbol{\delta}$, $log\left(\sigma_{ij}^2\right) = \mathbf{h}_{i,j}^T\boldsymbol{\lambda}$, $u_{ij} = \sum_{t=1}^{j-1}\phi_{i,jt}u_{it}+\epsilon_{ij}$, for $j = 2,3,\ldots,n_i$. We can then write

$$\frac{\partial f(\mathbf{u}_i;\boldsymbol{\delta},\boldsymbol{\lambda})}{\partial\boldsymbol{\delta}} = \frac{\partial}{\partial\boldsymbol{\delta}}\left[\exp\{\log f(\mathbf{u}_i;\boldsymbol{\delta},\boldsymbol{\lambda})\}\right] = f(\mathbf{u}_i;\boldsymbol{\delta},\boldsymbol{\lambda})\frac{\partial}{\partial\boldsymbol{\delta}}\{\log f(\mathbf{u}_i;\boldsymbol{\delta},\boldsymbol{\lambda})\} = -f(\mathbf{u}_i;\boldsymbol{\delta},\boldsymbol{\lambda})\sum_{j=1}^{n_i}\frac{\epsilon_{ij}}{\sigma_{ij}^2}\frac{\partial\epsilon_{ij}}{\partial\boldsymbol{\delta}}$$

$$\text{with}\quad \frac{\partial\epsilon_{i1}}{\partial\boldsymbol{\delta}} = 0 \quad\text{as}\quad \epsilon_{i1} = u_{i1} \quad\text{and}\quad \frac{\partial\epsilon_{ij}}{\partial\boldsymbol{\delta}} = -\sum_{t=1}^{j-1}u_{it}k_{i,jt}.$$

Similarly, we can write

$$\frac{\partial \log L(\boldsymbol{\theta};\mathbf{y}_i,\mathbf{x}_i,\mathbf{u}_i)}{\partial\boldsymbol{\lambda}} = \frac{1}{L(\boldsymbol{\theta};\mathbf{y}_i,\mathbf{w}_i,\mathbf{z}_i)}\int\int f(\mathbf{y}_i|\mathbf{x}_i,\mathbf{z}_i,\mathbf{u}_i;\boldsymbol{\beta})f(\mathbf{x}_i|\mathbf{w}_i,\mathbf{z}_i;\boldsymbol{\gamma},\sigma^2)\frac{\partial f(\mathbf{u}_i;\boldsymbol{\delta},\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\,d\mathbf{x}_i d\mathbf{u}_i, \quad (25)$$

where $\frac{\partial f(\mathbf{u}_i;\boldsymbol{\delta},\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}} = \frac{\partial}{\partial\boldsymbol{\lambda}}\left[\exp\{\log f(\mathbf{u}_i;\boldsymbol{\delta},\boldsymbol{\lambda})\}\right]$. After some algebra, we can write $\frac{\partial f(\mathbf{u}_i;\boldsymbol{\delta},\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}} = f(\mathbf{u}_i;\boldsymbol{\delta},\boldsymbol{\lambda})\sum_{j=1}^{n_i}\left(\frac{\epsilon_{ij}^2}{\sigma_{ij}^2}-1\right)h_{i,j}$. As we do not have closed forms of the above equations, we use the MCEM to get the approximation of the integrals in (21)-(25). We then follow the equation (9) to get the E-step of the MCEM. To evaluate the E-steps, we generate samples from these conditional distributions using Metropolis-Hasting algorithm. In M-steps, we used **optim** function in *R-program* (R 3.1.1) using quasi-Newton based method to get the updated estimates, and we continued these steps until convergence [29].

Supplementary materials

The supplementary materials contain R codes and corresponding “readme” files for the simulation and real data application conducted in this paper.

Acknowledgements

We would like to thank two referees for constructive comments and suggestions, which led to an improved version of the manuscript. This research is a part of MSc thesis work of Erfanul Hoque. Mahmoud Torabi gratefully acknowledges the research support of the Natural Sciences and Engineering Research Council of Canada (NSERC). Erfanul Hoque’s research was supported by the University of Manitoba Graduate Fellowship (UMGF) and Manitoba Graduate Scholarship (MGS). We also thank Dr. Robert B. Tate for providing the MFUS dataset.

References

[1] Breslow NE, Clayton DG, Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, 1993; **88**(421):9–25, doi:10.1080/01621459.1993.10594284.

[2] McCulloch CE, Searle SR, *Generalized, Linear, and Mixed Models*, John Wiley & Sons, 2001.

[3] Diggle P, *Analysis of Longitudinal Data*, Oxford University Press, 2002.

[4] Heagerty PJ, Kurland BF, Misspecified maximum likelihood estimates and generalised linear mixed models, *Biometrika*, 2001; **88**:973–985, doi:10.1093/biomet/88.4.973.

[5] Chiu TYM, Leonard T, Tsui KW, The matrix-logarithmic covariance model, *Journal of the American Statistical Association*, 1996; **91**(433):198–210, doi:10.2307/2291396.

[6] Pourahmadi M, Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation, *Biometrika*, 1999; **86**(3):677–690, doi:10.1093/biomet/86.3.677.

[7] Pourahmadi M, Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix, *Biometrika*, 2000; **87**(2):425–435, doi:10.1093/biomet/87.2.425.

[8] Pourahmadi M, Daniels MJ, Dynamic conditionally linear mixed models for longitudinal data, *Biometrics*, 2002; **58**(1):225–231, doi:10.1111/j.0006-341X.2002.00225.x.

[9] Daniels MJ, Pourahmadi M, Bayesian analysis of covariance matrices and dynamic models for longitudinal data, *Biometrika*, 2002; **89**(3):553–566.

[10] Daniels MJ, Zhao YD, Modelling the random effects covariance matrix in longitudinal data, *Statistics in Medicine*, 2003; **22**(10):1631–1647, doi:10.1002/sim.1470.

[11] Lee K, Lee J, Hagan J, Yoo JK, Modeling the random effects covariance matrix for generalized linear mixed models, *Computational Statistics & Data Analysis*, 2012; **56**(6):1545–1551, doi:10.1016/j.csda.2011.09.011.

[12] Cook JR, Stefanski LA, Simulation-extrapolation estimation in parametric measurement error models, *Journal of the American Statistical Association*, 1994; **89**(428):1314–1328.

[13] Carroll RJ, Ruppert D, Stefanski LA, *Measurement Error in Nonlinear Models*, London: Chapman & Hal l/CRC press, 1995.

[14] Lin XH, Breslow NE, Bias correction in generalized linear mixed models with multiple components of dispersion, *Journal of the American Statistical Association*, 1996; **91**(435):1007–1016, doi:10.1080/01621459.1996.10476971.

[15] Wang N, Lin X, Gutierrez RG, Carroll RJ, Bias analysis and SIMEX approach in generalized linear mixed measurement error models, *Journal of the American Statistical Association*, 1998; **93**(441):249–261, doi:10.2307/2669621.

[16] Fuller WA, *Measurement Error Models*, John Wiley & Sons, 2009.

[17] Torabi M, Datta GS, Rao JNK, Empirical Bayes estimation of small area means under a nested error linear regression model with measurement errors in the covariates, *Scandinavian Journal of Statistics*, 2009; **36**(2):355–369, doi: 10.1111/j.1467-9469.2008.00623.x.

[18] Datta GS, Rao JNK, Torabi M, Pseudo-empirical Bayes estimation of small area means under a nested error linear regression model with functional measurement errors, *Journal of Statistical Planning and Inference*, 2010; **140**(11):2952–2962, doi:10.1016/j.jspi.2010.03.046.

[19] Torabi M, Small area estimation using survey weights under a nested error linear regression model with structural measurement error, *Journal of Multivariate Analysis*, 2012; **109**:52–60, doi:10.1016/j.jmva.2012.02.015.

[20] Torabi M, Likelihood inference in generalized linear mixed measurement error models, *Computational Statistics & Data Analysis*, 2013; **57**(1):549–557, doi:10.1016/j.csda.2012.07.018.

[21] Torkashvand E, Jozani JM, Torabi M, Pseudo-empirical Bayes estimation of small area means based on James-Stein estimation in linear regression models with functional measurement error, *Canadian Journal of Statistics*, 2015; **43**(2):265–287, doi:10.1002/cjs.11245.

[22] Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM, *Measurement Error in Nonlinear Models: A Modern Perspective*, Chapman & Hall /CRC press, 2006.

[23] Yi GY, Liu W, Wu L, Simultaneous inference and bias analysis for longitudinal data with covariate measurement error and missing responses, *Biometrics*, 2011; **67**(1):67–75.

[24] Meng XL, van Dyk D, Fast EM-type implementations for mixed effects models, *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 1998; **60**:559–578, doi:10.1111/1467-9868.00140.

[25] Levine RRA, Casella G, Implementations of the Monte Carlo EM algorithm, *Journal of Computational and Graphical Statistics*, 2001; **10**(3):422–439, doi:10.1198/106186001317115045.

[26] Fort G, Moulines E, Convergence of the Monte Carlo expectation maximization for curved exponential families, *Annals of Statistics*, 2003; **31**(4):1220–1259, doi:10.1214/aos/1059655912.

[27] Caffo BS, Jank W, Jones GL, Ascent-based Monte Carlo expectation-maximization, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2005; **67**(2):235–251, doi:10.1111/j.1467-9868.2005.00499.x.

[28] McLachlan G, Krishnan T, *The EM Algorithm and Extensions*, John Wiley & Sons, 2007.

[29] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016.

[30] Mathewson FA, Brereton CC, Keltie WA, Paul GI, The University of Manitoba Follow-up study: A prospective investigation of cardiovascular disease, general description-mortality and incidence of coronary heart disease, *Canadian Medical Association journal*, 1965; **92**:947–953.

[31] Tate RB, Cuddy TE, Mathewson FAL, Cohort profile: The Manitoba Follow-up study (MFUS), *International Journal of Epidemiology*, 2015; **44**(5):1528–1536, doi:10.1093/ije/dyu141.

[32] SAS Institute Inc, *SAS/STAT Software, Version 9.1*, Cary, NC, 2003.

[33] Kraft P, Bauman L, Yuan JY, Horvath S, Multivariate variance-components analysis of longitudinal blood pressure measurements from the Framingham Heart Study, *BMCGenet*, 2003; **4 Suppl 1**:S55.

[34] Stockwell DH, Madhavan S, Cohen H, Gibson G, Alderman MH, The determinants of hypertension awareness, treatment, and control in an insured population, *American Journal of Public Health*, 1994; **84**(11):1768–1774, doi: 10.2105/AJPH.84.11.1768.

[35] Stamler J, Epidemiologic findings on body mass and blood pressure in adults, *Annals of Epidemiology*, 1991; **1**(4):347–362, doi:10.1016/1047-2797(91)90045-E.

[36] Kaufman JS, Asuzu MC, Mufunda J, Forrester T, Wilks R, Luke A, Long AE, Cooper RS, Relationship between blood pressure and body mass index in lean populations, *Hypertension*, 1997; **30**(6):1511–1516, doi:10.1161/01.HYP.30.6.1511.

[37] Humayun A, Shah AS, Alam S, Hussein H, Relationship of body mass index and dyslipidemia in different age groups of male and female population of Peshawar, *Journal of Ayub Medical College, Abbottabad : JAMC*, 2009; **21**:141–144.

[38] Zhang YX, Wang SR, Comparison of blood pressure levels among children and adolescents with different body mass index and waist circumference: study in a large sample in Shandong, China, *European Journal of Nutrition*, 2014; **53**(2):627–634, doi:10.1007/s00394-013-0571-1.

[39] Hoque ME, Khokan MR, Bari W, Impact of stature on non-communicable diseases: evidence based on Bangladesh Demographic and Health Survey, 2011 data, *BMC public health*, 2014; **14**:1007, doi:10.1186/1471-2458-14-1007.

[40] Rabkin SW, Mathewson AL, Tate RB, Predicting risk of ischemic heart disease and cerebrovascular disease from systolic and diastolic blood pressures, *Annals of Internal Medicine*, 1978; **88**(3):342–345.

[41] Tate RB, Manfreda J, Cuddy TE, The effect of age on risk factors for ischemic heart disease: the Manitoba Follow-Up study, 1948-1993, *Annals of Epidemiology*, 1998; **8**(7):415–421.

[42] Prentice RL, Measurement error and results from analytic epidemiology: dietary fat and breast cancer., *Journal of the National Cancer Institute*, 1996; **88**(23):1738–1747, doi:10.1093/jnci/88.23.1738.

[43] Rothman KJ, BMI-related errors in the measurement of obesity, *International Journal of Obesity*, 2008; **32 Suppl 3**:S56–S59, doi:10.1038/ijo.2008.87.

[44] O'Neill D, Sweetman O, The consequences of measurement error when estimating the impact of obesity on income, *IZA Journal of Labor Economics*, 2013; **2**(1):3, doi:10.1186/2193-8997-2-3.

[45] Abarin T, Li H, Wang L, Briollais L, On method of moments estimation in linear mixed effects models with measurement error on covariates and response with application to a longitudinal study of gene-environment interaction, *Statistics in Biosciences*, 2014; **6**(1):1–18, doi:10.1007/s12561-012-9074-5.

[46] Stubbendick AL, Ibrahim JG, Maximum likelihood methods for nonignorable missing responses and covariates in random effects models, *Biometrics*, 2003; **59**(4):1140–1150, doi:10.1111/j.0006-341X.2003.00131.x.