Published in final edited form as:

Stat Med. 2018 January 15; 37(1): 119–136. doi:10.1002/sim.7509.

Generalized linear mixed model for binary outcomes when covariates are subject to measurement errors and detection limits

Xianhong Xie^{a,*†}, Xiaonan Xue^a, and Howard D. Strickler^a

^aDepartment of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, U.S.A

Abstract

Longitudinal measurement of biomarkers is important in determining risk factors for binary endpoints such as infection or disease. However, biomarkers are subject to measurement error and some are also subject to left-censoring due to a lower limit of detection. Statistical methods to address these issues are few. We herein propose a generalized linear mixed model and estimate the model parameters using the Monte Carlo Newton-Raphson (MCNR) method. Inferences regarding the parameters are made by applying Louis's method and the delta method. Simulation studies were conducted to compare the proposed MCNR method with existing methods including the maximum likelihood (ML) method and the ad hoc approach of replacing the left-censored values with half of the detection limit (HDL). The results showed that the performance of the MCNR method is superior to ML and HDL with respect to the empirical standard error, as well as the coverage probability for the 95% confidence interval. The HDL method uses an incorrect imputation method and the computation is constrained by the number of quadrature points; while the ML method also suffers from the constrain for the number of quadrature points, the MCNR method does not have this limitation and approximates the likelihood function better than the other methods. The improvement of the MCNR method is further illustrated with real-world data from a longitudinal study of local cervicovaginal HIV viral load and its effects on oncogenic HPV detection in HIV-positive women.

Keywords

longitudinal data; generalized linear mixed model; measurement error; detection limit; monte carlo newton-raphson

*Correspondence to: Xianhong Xie, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, 1300 Morris Park Ave, Belfer Bldg. #1312D, Bronx, NY 10461, U.S.A.

†xianhong.xie@einstein.yu.edu

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site.

1. Introduction

Biomarkers play a critical role in clinical and epidemiologic research, as well as in clinical practice. From a research perspective, biomarkers can provide important information regarding the molecular pathway(s) that may underlie specific disease states. In clinical practice, biomarkers may be useful for risk stratification (e.g., to determine the need for screening or treatment) or may themselves be used as targets for the development of new therapeutic strategies. For example, in HIV-infected patients CD4+ T-cell count and HIV viral load are measured periodically to monitor patient health and determine whether and when it is necessary to change the anti-retroviral therapy (ART) regimen. These longitudinal data are also useful scientifically in studying the impact of host immune status on risk of various outcomes, such as the impact of HIV-related immunosuppression on the prevalence of human papillomavirus (HPV), the viral cause of cervical cancer, which we use as a real-world example in this paper. Additional examples include the role of inflammatory factors in the development of cardiovascular disease, as well as the effects of sex hormones on risk of several cancers. In short, prospective and longitudinally measured biomarkers and their relation with serially assessed disease outcomes (e.g., tests for HPV) are important and common areas of investigation.

A common limitation to biomarkers, however, is that they often have lower limits of detection (LLD), i.e., they are left-censored below which the exact values are unknown. Moreover, all biomarkers are subject to measurement error. HIV viral load subjects to measurement error and lower limits of detection, and CD4+ count subjects to measurement error [1].

The methods for studying longitudinally measured biomarker levels and disease outcomes without measurement error or detection limit have been well developed, such as generalized estimating equation (GEE) method [2] and generalized linear mixed model (GLMM) [3–5]. When only the measurement errors but not the detection limits are present with the longitudinal outcomes, the following methods have been proposed: regression calibration [6], simulation extrapolation (SIMEX) [7], corrected scores, maximum likelihood, and Bayesian method. The description of these methods can be found in Carroll *et al.* [8]. Recently, a method called data cloning has also been applied to the generalized linear mixed measurement error models [9]. When only the detection limits but not the measurement errors are present with the longitudinal outcomes, Giovanini [10] studied the maximum likelihood and regression calibration methods. However, there is a paucity of research regarding statistical methods for the common situation of concurrent measurement error and detection limits.

To address this issue, a frequently employed ad-hoc method is to impute the left-censored values as half of the detection limit (HDL), or some other arbitrary fraction of the detection limit. While these imputation methods are convenient they can lead to biased results [11]. A more rigorous method is the maximum likelihood (ML) method in which the likelihood function is integrated out using the Gauss-Hermite quadrature, and optimized with common algorithms such as Newton-Raphson method or other similar approaches. However, problems arise with the ML method when there are multiple covariates subject to

measurement errors, because there are more random effects to integrate out besides the between-subject random variations for the outcomes. Furthermore, the presence of detection limits exacerbates the problem with the need for additional integrations for those biomarkers below detection limits. The problem with the ML method is that the number of quadrature points needed increases exponentially with the dimension of the random effects, the computational burden severely restricts the number of quadrature points. The approximation to the likelihood by quadrature ML is not expected to be good when the dimension of the random effects is high. The HDL method suffers the same problem as the ML method and also the approximation due to the imputation of the detection limits. To our knowledge, the ML method in this setting has not been evaluated.

An alternative to the above methods is the Monte Carlo Newton-Raphson (MCNR) method, which has been shown to give accurate estimates in the setting of generalized linear mixed models without measurement error or detection limits [12]. We propose a MCNR method for the more general setting and compare with the HDL and the ML methods.

In the following section, we describe the above methods for analyzing longitudinal biomarker data, while accounting measurement error (but assuming no left censoring), and then develop methods to concurrently address both left-censoring and measurement error. In the Results section, we show simulation studies comparing and contrasting the three methods. These same three methods are then applied to a real-world dataset from a longitudinal study of local cervicovaginal HIV viral load and its effects on oncogenic HPV detection in HIV-positive women. The derivations for the proposed method are given in the Appendix.

2. Methods

2.1. Generalized linear mixed model that addresses measurement errors but assumes no left-censoring

2.1.1. Model—Let $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^T$ denote the observed binary outcomes for the i -th subject, $1 \leq i \leq m$, where m is the total number of subjects, n_i is the number of observations for subject i . $\mathbf{X}_i (n_i \times q)$ is the observed biomarkers measured with errors, where q is the number of biomarkers. $\mathbf{D}_i^y (n_i \times p_1)$ and $\mathbf{D}_i^x (n_i \times p_2)$ are the covariates without measurement error for the binary outcomes and for the biomarkers respectively, which include the intercept and other covariates. Conditional on an unobservable random effect vector $\mathbf{b}_i$ for the subject i , $\mathbf{b}_i = (b_{i1}, b_{i2}, \dots, b_{i,q+1})^T$, the components of data $\mathbf{y}_i$ follow the Bernoulli distribution with mean vector $\mathbf{p}_i$ which is linked to the linear predictors through the following equation:

$$g(\mathbf{p}_i) = \mathbf{D}_i^y \boldsymbol{\beta}_y + \mathbf{1}_i \mathbf{b}_{i,-1}^T \boldsymbol{\beta}_b + \mathbf{1}_i b_{i1}, \quad (1)$$

where g is the logit function: $g(p_{ij}) = \log(p_{ij}/(1-p_{ij}))$, $\boldsymbol{\beta}_y$ is the effects of the covariates on the binary outcomes, $\mathbf{1}_i$ is the column vector of all 1's with the length n_i , $\mathbf{b}_{i,-1} = (b_{i2}, \dots, b_{i,q+1})^T$, and $\boldsymbol{\beta}_b$ is the effects of the biomarkers on the outcomes. For the biomarkers, we assume the following linear mixed-effects model:

$$\mathbf{X}_i = \mathbf{D}_i^x \boldsymbol{\beta}_x + \mathbf{1}_i \mathbf{b}_{i,-1}^T + \boldsymbol{\varepsilon}_i, \quad (2)$$

where $\boldsymbol{\beta}_x$ ($p_2 \times q$) is the effects of the covariates on the biomarkers arranged in a matrix format, and $\boldsymbol{\varepsilon}_i$ is the measurement error matrix. We also assume the random effects and the measurement errors have normal distributions with the following variance structures:

$$\mathbf{b}_i = \begin{pmatrix} b_{i1} \\ \mathbf{b}_{i,-1} \end{pmatrix} \sim \mathbf{N}(\mathbf{0}, \begin{pmatrix} \sigma_{b_{i1}}^2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi} \end{pmatrix}), \quad (3)$$

$$\text{vec}(\boldsymbol{\varepsilon}_i) \sim \mathbf{N}(\mathbf{0}, \sum \otimes \mathbf{I}_{n_i}), \quad (4)$$

where $\text{vec}(\cdot)$ is the stacking of the columns of a matrix below one another, $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ are general positive definite matrices of dimension q , $\mathbf{I}_{n_i}$ is the identity matrix of dimension n_i , $\otimes$ is the Kronecker product. We assume the random effects are independent from the measurement errors for each subject i and the data between the subjects are independent. Note the random intercept b_{i1} on $\mathbf{y}_i$ is assumed to be independent from the random effects $\mathbf{b}_{i,-1}$ on $\mathbf{X}_i$. The b_{i1} is the subject-specific random effect on the binary response, representing individual's own level of risk of event and $\mathbf{b}_{i,-1}$ is the subject-specific random effect on the biomarkers, representing individual's own level of biomarker value. The independence between b_{i1} and $\mathbf{b}_{i,-1}$ is confirmed in our application data using model selection methods Akaike information criterion (AIC) and Bayesian information criterion (BIC). The independence is also assumed in Wang *et al.* [7], page 250 of the paper. A relatively straightforward model, namely, the random intercept model, is used for this initial presentation and application of our new methods to real data, in order to best explain and establish this approach; given that with multiple biomarkers subject to measurement errors and some also subject to the left-censoring, fitting this model is complex. More sophisticated and complex random effects in the biomarkers will be considered in follow-up studies.

In the above formulation, define $\mathbf{X}_i^* = \mathbf{D}_i^x \boldsymbol{\beta}_x + \mathbf{1}_i \mathbf{b}_{i,-1}^T$, the true covariate for $\mathbf{X}_i$ without measurement error. Another way of expressing equations (1) and (2) is

$$\begin{aligned} g(\mathbf{p}_i) &= (\mathbf{D}_i^y \boldsymbol{\beta}_y - \mathbf{D}_i^x \boldsymbol{\beta}_x \boldsymbol{\beta}_b) + \mathbf{X}_i^* \boldsymbol{\beta}_b + \mathbf{1}_i b_{i1}, \\ \mathbf{X}_i &= \mathbf{X}_i^* + \boldsymbol{\varepsilon}_i, \end{aligned}$$

which has the same form as the generalized linear mixed measurement error model in Wang *et al.* [7]. Under our model assumptions, there is a one to one correspondence between the model parameters in our model (1)–(2) and the model parameters in Wang *et al.* [7]. So the $\boldsymbol{\beta}$'s are identifiable. This is easy to see when $\mathbf{D}_i^y = \mathbf{D}_i^x$, and it becomes slightly more complex when these two matrices are not the same. A matrix with more columns containing both the

covariates in $\mathbf{D}_i^y$ and $\mathbf{D}_i^x$ needs to be constructed and is used to replace these two matrices in the next to last equation, and 0's need to be filled to β_y and $\beta_x\beta_b$ in the above mentioned equation for those columns in the augmented matrix but not in $\mathbf{D}_i^y$ and $\mathbf{D}_i^x$ respectively. Also with the model assumptions, the effect $\mathbf{b}_{i,-1}$ and measurement error $\mathbf{e}_i$ are identifiable in model (2) alone because of repeated observations of biomarkers, and $\mathbf{b}_{i,-1}$ is shared by the models for both the binary outcomes and the biomarkers, b_{11} is identifiable too. The parameters related to variance components in our model are identifiable.

2.1.2. Estimation—When the biomarker data is uncensored, the likelihood has the form

$$L(\theta; \mathbf{y}, \mathbf{X}) = \prod_{i=1}^m \int_{R^{q+1}} f(\mathbf{y}_i | \mathbf{b}_i; \beta_y, \beta_b) h(\mathbf{X}_i | \mathbf{b}_i; \beta_x, \sum) \varphi_1(b_{11}; \sigma_{b_{11}}^2) \varphi_q(\mathbf{b}_{i,-1}; \Psi) d\mathbf{b}_i, \quad (5)$$

where $\theta = (\beta_y, \beta_b, \beta_x, \sigma_{b_{11}}^2, \Psi, \sum)$ is the parameters in the model, $\mathbf{y}$ and $\mathbf{X}$ are the collections of observed outcomes and biomarkers for all the subjects, respectively, f is the density of the binary outcomes $\mathbf{y}_i$ given the random effects $\mathbf{b}_i$,

$f(\mathbf{y}_i | \mathbf{b}_i; \beta_y, \beta_b) = \prod_{j=1}^{n_i} p_{ij}^{y_{ij}} (1-p_{ij})^{1-y_{ij}}$, p_{ij} is the j -th component of $\mathbf{p}_i$, h is the multivariate normal density with the mean vector $\mu_{\mathbf{x}_i} = \text{vec}(\mathbf{D}_i^x \beta_x + \mathbf{1}_i \mathbf{b}_{i,-1}^T)$ and the variance $\Sigma \otimes \mathbf{I}_{n_p} \varphi_1$ is the univariate normal density with the mean 0 and variance $\sigma_{b_{11}}^2$, φ_q is the q -dimensional normal density with the mean vector $\mathbf{0}$ and variance Ψ . To calculate the likelihood, one can use the numeric integration method: e.g., Gauss-Hermite quadrature method. It uses a fixed set of K ordinates (quadrature points) and weights to approximate the integrals in equation (5). The number of quadrature points needed increases exponentially with the dimension of the integration. Alternatively, the adaptive Gauss-Hermite quadrature method, which uses a varying set of ordinates and weights for each integral, can be applied. For those integrands that are easier to approximate, fewer quadrature points are used; on the other hand, for those integrands that are harder to approximate, more quadrature points are used. The method starts with the minimal number of quadrature points and the number of quadrature points is increased until the difference between two successive log-likelihoods is small. SAS software provides the facilities for both of these quadrature methods. Specifically, the SAS procedure NLMIXED can be used for the general purpose integration in the likelihood optimization framework. Once the likelihood is calculated, the score equations are solved, and the parameters are found. The variance is estimated by the inverse of the observed information matrix.

2.2. Generalized linear mixed model that addresses both measurement errors and left-censoring

2.2.1. Model—When the biomarker data are left-censored, we observe (Q_{ijk}, C_{ijk}) instead of X_{ijk} , where

$$\begin{aligned} X_{ijk} &< Q_{ijk} \text{ if } C_{ijk}=1, \\ X_{ijk} &= Q_{ijk} \text{ if } C_{ijk}=0, \\ i &= 1, \dots, m, \quad j = 1, \dots, n_i, \quad k = 1, \dots, q, \end{aligned} \quad (6)$$

k indexes the biomarkers. To obtain the likelihood, the normal density $h(\mathbf{X}_i | \mathbf{b}_i; \boldsymbol{\beta}_x, \boldsymbol{\Sigma})$ in the equation (5) is replaced by

$$I(\mathbf{Q}_i, \mathbf{C}_i | \mathbf{b}_i; \boldsymbol{\beta}_x, \boldsymbol{\Sigma}) = \prod_{j=1}^{n_i} \varphi_{n_{ij}^o}(\mathbf{X}_{ij}^o | \mathbf{b}_i; \boldsymbol{\mu}_{\mathbf{x}_{ij}}^o, \sum_{\mathbf{x}_{ij}}^o) \int_{R(\mathbf{Q}_{ij}, \mathbf{C}_{ij})} \varphi_{n_{ij}^c}(\mathbf{X}_{ij}^c | \mathbf{b}_i, \mathbf{X}_{ij}^o; \boldsymbol{\mu}_{\mathbf{x}_{ij}}^{c|o}, \sum_{\mathbf{x}_{ij}}^{c|o}) d\mathbf{X}_{ij}^c,$$

where $\mathbf{Q}_i$ and $\mathbf{C}_i$ are the collections of Q_{ijk} and C_{ijk} over j and k , n_{ij}^o and n_{ij}^c are the number of biomarkers among $\{(Q_{ijk}, C_{ijk}), 1 \leq k \leq q\}$ that are observed and censored, $\mathbf{X}_{ij}^o$ and $\mathbf{X}_{ij}^c$ are the corresponding observed and censored biomarkers, φ 's are the multivariate normal

densities with the given mean and covariance, $\boldsymbol{\mu}_{\mathbf{x}_{ij}}^o$ is the submatrix of

$\boldsymbol{\mu}_{\mathbf{x}_{ij}} = (\mathbf{D}_{ij}^x \boldsymbol{\beta}_x + \mathbf{b}_{i,-1})^T$ corresponding to the conditional mean for $\mathbf{X}_{ij}^o$ given $\mathbf{b}_i$, where $\mathbf{D}_{ij}^x$

is the j -th row of $\mathbf{D}_i^x$, $\sum_{\mathbf{x}_{ij}}^o$ is the submatrix of $\boldsymbol{\Sigma}$ corresponding to the conditional variance

for $\mathbf{X}_{ij}^o$ given $\mathbf{b}_i$, $\boldsymbol{\mu}_{\mathbf{x}_{ij}}^{c|o}$ and $\sum_{\mathbf{x}_{ij}}^{c|o}$ are the conditional mean and variance for $\mathbf{X}_{ij}^c$ given $\mathbf{X}_{ij}^o$ (and $\mathbf{b}_i$) from the joint normal distribution with the mean $\boldsymbol{\mu}_{\mathbf{x}_{ij}}$ and the variance $\boldsymbol{\Sigma}$. The integration region $R(\mathbf{Q}_{ij}, \mathbf{C}_{ij})$ is $(-\infty, Q_{ijk_1}) \times \dots \times (-\infty, Q_{ijk_{n_{ij}^c}})$ for the left-censored biomarkers $\mathbf{X}_{ij}^c$.

2.2.2. Existing methods—The ML method can be applied in this setting. From the above description, we see that another layer of integration is needed to calculate the cumulative distribution functions (CDFs) for the left-censored data. The computational burden is more challenging compared to the case of no left-censoring. For datasets involving at most two left-censored biomarkers, the integrations for the CDFs can be achieved in SAS by calling the univariate and the bivariate normal cumulative distribution functions PROBUNORM and PROBBNRM respectively. For datasets involving more than two left-censored biomarkers, iterated integral or some other methods has to be used to calculate the CDFs. Although the program code is easy to write, as mentioned, the accuracy of the ML method in this setting with left-censored biomarkers has not been assessed using formal simulation studies.

The common ad-hoc method to address the left-censored data uses the detection limit itself, half the detection limit, or detection limit / $\sqrt{2}$ to impute these data. However, each of these approaches has been shown by other researchers to introduce bias [11]. The methods tend to underestimate the variance and have biases on the parameters related to the biomarkers.

2.2.3. Monte Carlo Newton-Raphson method—An alternative to the maximum likelihood method is the Monte Carlo Newton-Raphson method. Assume the complete data

is $(\mathbf{y}, \mathbf{X}, \mathbf{b})$, which denotes the collection of all $\{(\mathbf{y}_i, \mathbf{X}_i, \mathbf{b}_i), 1 \leq i \leq m\}$, the likelihood on the complete data has the form

$$L_c(\theta; \mathbf{y}, \mathbf{X}, \mathbf{b}) = \prod_{i=1}^m f(\mathbf{y}_i | \mathbf{b}_i; \boldsymbol{\beta}_y, \boldsymbol{\beta}_b) h(\mathbf{X}_i, \mathbf{b}_i; \boldsymbol{\beta}_x, \boldsymbol{\Sigma}) \varphi_1(b_{i1}; \sigma_{b11}^2) \varphi_q(\mathbf{b}_{i,-1}; \boldsymbol{\Psi}). \quad (7)$$

Denote $\boldsymbol{\beta} = (\boldsymbol{\beta}_y^T, \boldsymbol{\beta}_b^T, \text{vec}(\boldsymbol{\beta}_x)^T)^T$, and let $\mathbf{Q}$ and $\mathbf{C}$ be the collections of all $\mathbf{Q}_i$'s and $\mathbf{C}_i$'s, the estimate of $\boldsymbol{\beta}$ is given by the following iterative equation at step $s + 1$,

$$\boldsymbol{\beta}^{(s+1)} = \boldsymbol{\beta}^{(s)} - E \left[\frac{\partial^2 \log(L_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} | \mathbf{y}, \mathbf{Q}, \mathbf{C} \right]^{-1} E \left[\frac{\partial \log(L_c)}{\partial \boldsymbol{\beta}} | \mathbf{y}, \mathbf{Q}, \mathbf{C} \right], \quad (8)$$

where the expectations are evaluated at the parameters at step s . The details of the first and second order derivatives on the complete data log-likelihood are given in the Appendix. The estimates of σ_{b11}^2 , $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ are given by

$$\sigma_{b11}^{2(s+1)} = \frac{1}{m} \sum_{i=1}^m E \left[b_{i1}^2 | \mathbf{y}_i, \mathbf{Q}_i, \mathbf{C}_i \right], \quad (9)$$

$$\boldsymbol{\Psi}^{(s+1)} = \frac{1}{m} \sum_{i=1}^m E \left[\mathbf{b}_{i,-1} \mathbf{b}_{i,-1}^T | \mathbf{y}_i, \mathbf{Q}_i, \mathbf{C}_i \right], \quad (10)$$

$$\boldsymbol{\Sigma}^{(s+1)} = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} E \left[\mathbf{r}_{ij} \mathbf{r}_{ij}^T | \mathbf{y}_i, \mathbf{Q}_i, \mathbf{C}_i \right], \quad (11)$$

where $n = \sum_{i=1}^m n_i$, the total number of binary responses, $\mathbf{r}_{ij} = (\mathbf{X}_{ij} - \mathbf{D}_{ij}^x \boldsymbol{\beta}_x - \mathbf{b}_{i,-1})^T$. As in the equation (8), the above expectations in the equations (9) to (11) are also evaluated using the estimates from the previous step.

To get the conditional distributions of the censored $\mathbf{X}_i$'s and $\mathbf{b}_i$'s given the observed data $(\mathbf{y}, \mathbf{Q}, \mathbf{C})$, we note that the joint density of $\mathbf{y}_i$, $\mathbf{X}_i$ and $\mathbf{b}_i$ is as follows:

$$C \cdot \exp \left(\sum_{j=1}^{n_i} [y_{ij} \eta_{ij} - \log(1 + \exp(\eta_{ij}))] - \frac{1}{2} \log(\sigma_{b11}^2) - \frac{1}{2} \log(\det(\Psi)) - \frac{b_{i1}^2}{2\sigma_{b11}^2} - \frac{1}{2} \mathbf{b}_{i,-1}^T \Psi^{-1} \mathbf{b}_{i,-1} - \frac{n_i}{2} \log(\det(\Sigma)) - \frac{1}{2} \sum_{j=1}^{n_i} (\mathbf{X}_{ij}^T - \mathbf{b}_{i,-1} - \tilde{\mathbf{D}}_{ij}^{\mathbf{x}} \text{vec}(\boldsymbol{\beta}_{\mathbf{x}}))^T \Sigma^{-1} (\mathbf{X}_{ij}^T - \mathbf{b}_{i,-1} - \tilde{\mathbf{D}}_{ij}^{\mathbf{x}} \text{vec}(\boldsymbol{\beta}_{\mathbf{x}})) \right),$$

where $\eta_{ij} = \mathbf{D}_{ij}^y \boldsymbol{\beta}_y + (1 \ \boldsymbol{\beta}_b^T) \mathbf{b}_i$, $\mathbf{D}_{ij}^y$ is the j -th row of the matrix $\mathbf{D}_i^y$, $\tilde{\mathbf{D}}_{ij}^{\mathbf{x}} = \mathbf{I}_q \otimes \mathbf{D}_{ij}^{\mathbf{x}}$, $\mathbf{C}$ is some constant not dependent on $\mathbf{y}_i$, $\mathbf{X}_i$ and $\mathbf{b}_i$. Given the data, the $\mathbf{X}_{ijk}$'s (j denotes the replicate and k denotes the biomarker) satisfy the constraints that $\mathbf{X}_{ijk} < \mathbf{Q}_{ijk}$ for those left-censored biomarkers. So the conditional distribution of the censored $\mathbf{X}_i$ given $\mathbf{b}_i$ and $(\mathbf{y}, \mathbf{Q}, \mathbf{C})$ is the truncated multivariate normal distribution. Note from the above expression that given $\mathbf{b}_i$ and $(\mathbf{y}, \mathbf{Q}, \mathbf{C})$, $\mathbf{X}_{ij}$'s are independent for different j 's. The censored $\mathbf{X}_{ij}$ conditional on $\mathbf{b}_i$ and $(\mathbf{y}, \mathbf{Q}, \mathbf{C})$ has the truncated multivariate normal distribution with the mean

parameter $\boldsymbol{\mu}_{\mathbf{x}_{ij}}^{c|o}$ and the variance parameter $\sum_{\mathbf{x}_{ij}}^{c|o}$, where the mean and variance parameters are from the conditional distribution of multivariate normal distribution

$$N(\mathbf{b}_{i,-1} + \tilde{\mathbf{D}}_{ij}^{\mathbf{x}} \text{vec}(\boldsymbol{\beta}_{\mathbf{x}}), \Sigma), \boldsymbol{\mu}_{\mathbf{x}_{ij}}^{c|o} = \boldsymbol{\mu}_{\mathbf{x}_{ij}}^c + \sum_{\mathbf{x}_{ij}}^{co} \sum_{\mathbf{x}_{ij}}^o \Sigma_{\mathbf{x}_{ij}}^{-1} (\mathbf{X}_{ij}^o - \boldsymbol{\mu}_{\mathbf{x}_{ij}}^o), \sum_{\mathbf{x}_{ij}}^{c|o} = \sum_{\mathbf{x}_{ij}}^c - \sum_{\mathbf{x}_{ij}}^{co} \sum_{\mathbf{x}_{ij}}^o \Sigma_{\mathbf{x}_{ij}}^{-1} \sum_{\mathbf{x}_{ij}}^{co T}.$$

In the expressions, $\mathbf{X}_{ij}^o$, $\boldsymbol{\mu}_{\mathbf{x}_{ij}}^o$ and $\sum_{\mathbf{x}_{ij}}^o$ are defined in Section 2.2.1, $\boldsymbol{\mu}_{\mathbf{x}_{ij}}^c$ is the subvector of $\boldsymbol{\mu}_{\mathbf{x}_{ij}} = \mathbf{b}_{i,-1} + \tilde{\mathbf{D}}_{ij}^{\mathbf{x}} \text{vec}(\boldsymbol{\beta}_{\mathbf{x}})$ corresponding to the conditional mean of censored biomarkers

given $\mathbf{b}_i$, $\sum_{\mathbf{x}_{ij}}^c$ and $\sum_{\mathbf{x}_{ij}}^{co}$ are the submatrices of Σ corresponding to the conditional variance of censored biomarkers given $\mathbf{b}_i$, and the conditional covariance between censored and observed biomarkers given $\mathbf{b}_i$ respectively. The conditional distribution of $\mathbf{b}_i$ given $\mathbf{X}_i$ and $(\mathbf{y}, \mathbf{Q}, \mathbf{C})$ is proportional to the above joint density and does not have a familiar form in $\mathbf{b}_i$, we used Metropolis-Hastings algorithm for the sampling.

The detailed Monte Carlo Newton-Raphson algorithm is as follows:

1. **Initialization of the parameters:** A two-step approach is used. First, the biomarker data is fitted with the approach in [13] by ignoring the binary responses $\mathbf{y}_i$'s and considering only the models for $\mathbf{Q}_i$'s and $\mathbf{C}_i$'s (Equations (2)–(4) and (6)); the estimates for the initial $\boldsymbol{\beta}_{\mathbf{x}}$, Ψ and Σ are obtained along with the conditional expectations of $\mathbf{b}_{i,-1}$'s given $\mathbf{Q}_i$'s and $\mathbf{C}_i$'s. Second, these conditional expectations are treated as “covariates”; a generalized linear mixed model with the random intercept is fitted to the outcomes $\mathbf{y}_i$'s versus the other covariates and the conditional expectations of $\mathbf{b}_{i,-1}$'s to get the initial $\boldsymbol{\beta}_{\mathbf{y}}$, $\boldsymbol{\beta}_{\mathbf{b}}$, and σ_{b11}^2 .
2. **Random sampling of $\mathbf{X}_i$'s and $\mathbf{b}_i$'s:**
 - a. **Initialization of $\mathbf{X}_i$'s and $\mathbf{b}_i$'s** – At each iteration, the initial values for $\mathbf{X}_i$'s are set to their expected values from the previous iteration (for the

first iteration, the expected values are from Step 1). For the initial values of $\mathbf{b}_i$'s, we treat the log-likelihood on the complete data as a function of $\mathbf{b}_i$ for each i , with $\mathbf{X}_i$'s replaced by their initial values, then we find the modes of the log-likelihoods with the Newton-Raphson algorithm, and calculate the Hessians at the modes. The initial $\mathbf{b}_i$'s are drawn randomly from the normal distributions with the means equal to the modes and the variances equal to the inverses of the negative Hessians.

- b. *Metropolized Gibbs sampling* – To continue the sampling of the $\mathbf{X}_i$'s and $\mathbf{b}_i$'s, we iterate between the Gibbs sampling and the Metropolis-Hastings algorithm. At each iteration, the $\mathbf{X}_i$'s are drawn once from the Gibbs sampler of the censored $\mathbf{X}_i$'s conditional on the $\mathbf{b}_i$'s and $(\mathbf{y}, \mathbf{Q}, \mathbf{C})$ (truncated multivariate normal distributions). Each $\mathbf{X}_{ij}$, $1 \leq j \leq n_i$, is sampled independently. For the sampling of $\mathbf{b}_i$'s, a single iteration of the Metropolis-Hastings algorithm is used. More specifically, the target distribution is the conditional distribution of $\mathbf{b}_i$ given the current $\mathbf{X}_i$ and $(\mathbf{y}, \mathbf{Q}, \mathbf{C})$, the candidate distribution (proposal distribution) is the normal distribution with the mean equal to the mode and the variance equal to the inverse of the curvature of the log density of $\mathbf{b}_i$ given $\mathbf{X}_i$ and $(\mathbf{y}, \mathbf{Q}, \mathbf{C})$ at the mode, where the $\mathbf{X}_i$ is evaluated at the initial value at step 2(a). The usual Metropolis-Hastings acceptance ratio is calculated and the sample of $\mathbf{b}_i$ is accepted or rejected accordingly. This iterative sampling of $\mathbf{X}_i$ and $\mathbf{b}_i$ is repeated many times to get the simulation samples.

3. Evaluation of the expectations: All the conditional expectations in the equations (8)–(11) are evaluated with the above Monte Carlo samples after the burn-in period.
4. Updating of the parameters: The parameters are updated using the equations (8)–(11).
5. Assessment of convergence: The estimated values of the parameters at the current iteration are compared to those from the previous iteration in terms of the relative difference. The convergence criterion is defined as

$\max\{|\tilde{\theta}_l^{(s+1)} - \tilde{\theta}_l^{(s)}| / (|\tilde{\theta}_l^{(s)}| + \delta) : 1 \leq l \leq L\}$, where $\tilde{\theta}_l^{(s)}$'s are the transformed parameters at step s with l indexes the parameters, L is the total number of parameters, δ is a small constant to prevent the denominators become too close to 0. The parameter vector $\boldsymbol{\beta}$ is untransformed, while σ_{b11}^2 and the variance matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ are transformed with the square root and the Cholesky decompositions respectively. If the convergence criterion is less than the tolerance level, the algorithm is declared converged and the iterations stop; otherwise, the absolute differences in the changes of the estimates of $\boldsymbol{\beta}_y$ and $\boldsymbol{\beta}_b$ between two adjacent iterations are compared to the Monte Carlo standard errors of these parameters [14, 15], the Monte Carlo sample size is doubled if the ratios

of the absolute differences to the standard errors are less than some constant [16]. The iteration returns to step 2.

At the last iteration, the cross-products of the log-likelihood gradients are calculated, and the information matrix is calculated with the Louis's method [17]. The information matrix on the transformed parameters (with further transformations on the variance parameters by taking the logarithms on $\sigma_{b_{11}}$ and the diagonals of the Cholesky decompositions of Ψ and Σ) are found by using the matrix derivative. The standard errors of the original parameters are obtained with the delta method, and the confidence intervals are calculated with the normal approximations.

The sampling of $\mathbf{X}_i$'s and $\mathbf{b}_i$'s is based on a general Gibbs sampling approach: the usual Gibbs sampling uses the conditional distributions of $\mathbf{X}_i$'s and $\mathbf{b}_i$'s as the candidate distributions and has an acceptance ratio of 1 for $\mathbf{X}_i$'s and $\mathbf{b}_i$'s; Approximations to the conditional distributions of $\mathbf{b}_i$'s are used as the candidate distributions for $\mathbf{b}_i$'s because the conditional distributions of $\mathbf{b}_i$'s are difficult to sample. The Markov chain converges to the conditional distribution of censored $\mathbf{X}_i$ and $\mathbf{b}_i$ given the observed data ($\mathbf{y}$, $\mathbf{Q}$, $\mathbf{C}$) as the simulation sample size gets large ([18]).

The proposed MCNR algorithm is a general version of the MCNR method, as described by McCulloch in [12]. The explanation of the convergence of the MCNR algorithm in the McCulloch paper applies here. Briefly, for sufficiently large simulation sample size, and under suitable regularity conditions, MCNR inherits the properties of the exact version of Newton-Raphson algorithm. If the surface to be maximized (i.e., the likelihood function) is concave or locally concave in a neighborhood of the maximum with the initial value close to the maximum, the proposed Monte Carlo Newton-Raphson algorithm is guaranteed to converge.

The computer codes in R and Fortran for implementing the above method are available online as supplemental materials and also available at https://sites.google.com/site/samxiepage/GLMM_merr_left_censoring_pkg_v1.zip?attredirects=0&d=1.

3. Simulation Studies

To compare the HDL, ML, and MCNR methods, six sets of simulations under the model (Equations (1)–(4) and (6)) were conducted assuming parameter values similar to the data from the example below. We varied the extent of left-censored data, and sample size. Each simulation included one covariate and three biomarkers. The sequential replicates within a subject: 1, 2, ..., n_i was the covariate. The covariate matrices $\mathbf{D}_i^x$ and $\mathbf{D}_i^y$ were the same which included the intercept and the sequential replicates. Among the biomarkers, the first two had measurement errors and were left-censored; the third one had measurement error but was not left-censored. The model parameters were set to the following values:

$$\begin{aligned}
 (\beta_y^T; \beta_b^T) &= (-3.4, 0.1; 4.1, -4.8, -0.7), \\
 \beta_x &= \begin{pmatrix} 1.2 & 3.1 & 2.00 \\ 0.1 & -0.1 & -0.01 \end{pmatrix}, \\
 \sigma_{b_{11}}^2 &= 4.3, \\
 \Psi &= \begin{pmatrix} 1.2 & 1.0 & -0.3 \\ 1.0 & 0.9 & -0.3 \\ -0.3 & -0.3 & 0.3 \end{pmatrix}, \\
 \Sigma &= \begin{pmatrix} 2.9 & 0.7 & -0.2 \\ 0.7 & 1.0 & -0.1 \\ -0.2 & -0.1 & 0.1 \end{pmatrix}.
 \end{aligned}$$

Specifically, we used three sets of left-censored proportions for the first and second biomarkers: (0.20, 0.10), (0.40, 0.30) and (0.60, 0.50). For the number of subjects, we considered two scenarios: 300 subjects each with four repeated measures; and 100 subjects each with four repeated measures. These values were chosen so that the simulated data would be similar to the real-world HIV study described below. To set the threshold values for each left-censored biomarker, we used the Monte Carlo method to simulate 10000 sets for each biomarker under the model (2–4). The thresholds were the values under which the expected censoring proportions were the target values. Five hundred data sets were generated under each set of assumed conditions and the HDL, ML, MCNR methods were applied to each simulated data set. The HDL method and the ML method were computed with the PROC NLMIXED in SAS, the Gaussian quadrature method was used and the number of points used in each dimension of the random effects $\mathbf{b}_i$ was 3, and the optimization method used was the double-dogleg method. All the other parameters in the SAS procedure were left as default. For the MCNR approach, the size of burn-in was 500, δ was set to 0.001, and the tolerance for convergence was 0.01.

The methods were compared with respect to the mean bias, empirical standard error (standard deviation of the estimates over the simulations), model based standard error (square root of the mean model based variances), and mean coverage probability of the 95% confidence intervals. For some of the simulations, the methods generated information matrices that were not positive definite, especially when the censoring proportions were high or the sample size was small. These simulations were excluded in calculating the biases, SEs, and the coverage probabilities. In [13], extensive comparisons between these methods on β_x and the variances Ψ and Σ were evaluated (note that the MCNR method reduces to the Monte Carlo Expectation Maximization method when we ignore the responses $\mathbf{y}_i$'s). We omitted the comparisons of these parameters in this article and refer the readers to [13]. Although there are only two biomarkers subject to detection limits, there are four random effects in $\mathbf{b}_i$ that need to be integrated out, which made the Gauss-Hermite quadrature ML method unsuitable for determining the MLE and drawing inference in this setting. The HDL method is inherently biased. Therefore, only the MCNR method is expected to perform well.

The results for $m = 300$ subjects are given in Table I. At the low censoring proportions: (0.20, 0.10), the bias of HDL ranged from -0.002 for $\beta_{y,2}$ (the slope for the replicate) to

−0.58 for σ_{b11}^2 , and the magnitude of relative biases ranged from 1% for β_{b3} (the slope for the biomarker 2) to 13% for σ_{b11}^2 . For the ML method, the magnitude of relative biases on the parameters β_{b2} and β_{y2} improved about four folds and one fold to 2% and 1% respectively compared to HDL, slightly improved on the parameters β_{y1} , became slightly worse on β_{b4} and σ_{b11}^2 , and two times worse on β_{b3} ; for the MCNR method, the magnitude of relative bias on β_{b2} improved one fold compared to the HDL method. The MCNR had the smallest bias on β_{y1} among all three methods (approximately 3% of the true value). It had higher biases on the other parameters in Table I than the other two methods for the low censoring proportions. The high bias in σ_{b11}^2 for the MCNR method was due to the fact that this parameter is influenced more by the Monte Carlo error than the fixed effects parameters β_{y1} and β_{y2} in the Table. It is harder to reach the optimum for the variance parameter. For the standard errors, the empirical and the model based SEs were generally close. The ratio of the model based SE to the empirical SE ranged from 0.61 to 1.47, 0.60 to 1.73, and 0.91 to 1.94 for HDL, ML, and MCNR respectively. The MCNR method gave smaller standard errors on the parameters β_{b2} , β_{b3} than either the ML or HDL method. The SEs on β_{y1} , β_{y2} , and σ_{b11}^2 were slightly larger for MCNR; the empirical SE on β_{b4} was the smallest for MCNR among all three methods compared, the model based SE on the parameter was larger for MCNR than the other two methods. As for the coverage probabilities of the 95% CIs, the MCNR gave the best coverage among these three methods on all the parameters except the parameter β_{y2} , on which all three methods had similar performance. The number of non positive definite matrices was 86, 64, and 39 for HDL, ML, and MCNR respectively.

At the medium censoring proportions (0.40, 0.30), the biases on β_{b2} and β_{b3} for all three methods increased compared to those at the low censoring proportions; the bias on β_{b4} increased for HDL and ML, but decreased for MCNR; the biases on the other parameters remained about the same for HDL and ML, decreased for MCNR. The standard errors on all the parameters for HDL and ML did not change too much; the SEs on β_{y1} , β_{y2} , σ_{b11}^2 for MCNR remained about the same, and the SEs on β_{b2} , β_{b3} , β_{b4} increased. The coverage probabilities on all the parameters remained largely unchanged for all three methods; MCNR still had the best coverage probability on the parameters except β_{y2} as in the low censoring proportion case. As the censoring proportions became even higher, the biases on β_{b2} , β_{b3} , β_{b4} and σ_{b11}^2 were getting larger for ML; for MCNR, only the biases on β_{b2} , β_{b3} were getting larger compared to the lower censoring proportions. The model based SEs on the parameters β_{b2} , β_{b3} , β_{b4} and σ_{b11}^2 increased for ML and MCNR compared to the medium censoring proportions; and the SEs on β_{b2} , β_{b3} increased for HDL. The coverage probabilities on β_{b2} , β_{b3} , β_{b4} for HDL became smaller compared to the medium censoring proportions; the coverage probabilities on β_{b2} , β_{b3} for ML slightly increased, and the coverage probabilities on β_{b2} , β_{b3} , β_{b4} for MCNR remained about the same.

The biases at the smaller sample size $m = 100$ (Table II) were not necessarily larger than the biases at $m = 300$, as expected, while the standard error for all the methods increased as expected because of the reduced number of observations: the model based SEs on the parameters β_{b2} , β_{b3} , β_{b4} increased more for MCNR compared to HDL and ML at the low and high censoring proportions in terms of relative increase; for the medium censoring proportions, the MCNR and ML methods had about the same relative increase on the model based SEs of these parameters, while the HDL method had the largest relative increase. The

coverage probabilities of all three methods decreased, while the MCNR still had the closest coverages to the nominal 95% level on the parameters $\beta_{b,2}$, $\beta_{b,3}$, $\beta_{b,4}$.

Although the model based standard errors for MCNR on the parameters $\beta_{b,2}$, $\beta_{b,3}$ were smaller than those of ML and HDL (Tables I and II), the median standard errors of the MCNR method were actually larger than that of ML (Figures 1 and 2). The discrepancy is due to the fact that the standard errors of the ML method had more variations than those of the MCNR method, and the model based standard error takes both the mean and SD of the standard errors into account. Based on Figures 1 and 2, the ML method consistently gives smaller estimates of the standard errors than the MCNR method, which causes the ML method to have lower coverage probabilities. While some standard errors in Tables I and II were very large because of the outliers in the standard error estimates, Figures 1 and 2 showed that a large portion of standard error estimates were not large.

The model based SE's are quite different from the empirical SE's in the simulation studies for some of the parameters, because the likelihood for the generalized linear measurement error problem with left-censoring is complex and the SE estimates for those parameters are highly variable for all three statistical approaches we assessed, especially for the parameters $\beta_{b,2}$ and $\beta_{b,3}$. We further looked into the simulation results, when the standard deviation of the SE estimates over the simulations is close to or above the mean of the SE estimates over the simulations, the model based SE becomes quite different from the empirical SE. Although the non-converged simulations were excluded for each method, there were still simulations with outlying estimates of SEs, and the model based SEs (square root of the mean square of the SEs over the simulations) are sensitive to the outlying SE estimates. The empirical SEs are sensitive to the outliers in the parameter estimates too, but to a less extent. All three methods suffer from this problem, and there is no clear winner for the model based SE in terms of its relationship to the empirical SE. The parameters did not appear to be significantly different from 0 in the simulation studies could be due to the parameter settings. None of the methods perfectly solves the problem at hand; the MCNR gives more conservative standard error estimates (Figure 1, medians of SEs) and a coverage probability close to the nominal level at low to medium censoring proportions (Table I). There is a trade-off between getting smaller standard error estimates and the desired nominal coverage probabilities.

Overall, the performance of MCNR was superior to the HDL and ML approaches in terms of the empirical SE and coverage probability on the parameters for the biomarkers with respect to the binary outcome. The comparative bias varied according to the scenario, with no one method showing superiority.

The computation time for ML and MCNR in simulations was 16 minutes and 25 minutes per simulation respectively at $m = 300$ and censoring proportions (0.40, 0.30). For $m = 100$ and censoring proportions (0.40, 0.30), the computation time for ML and MCNR in simulations was 8 minutes and 15 minutes per simulation. The computation time increases as censoring proportions get higher. This was run on a Windows XP PC. The %non-convergence for ML and MCNR in simulations at $m = 300$ was 12.8% and 7.8% respectively for censoring proportions = (0.20, 0.10), 20% and 29.2% for censoring proportions = (0.40, 0.30), 34.8%

and 58% for censoring proportions = (0.60, 0.50); the %non-convergence for ML and MCNR in simulations at $m = 100$ was 25.4% and 34.8% for censoring proportions = (0.20, 0.10), 33.2% and 50.6% for censoring proportions = (0.40, 0.30), 49.8% and 70% for censoring proportions = (0.60, 0.50).

4. Application to Real-World Data

The data that we used for our real-world example to illustrate the methods is from the Women's Interagency HIV Study (WIHS); a multi-center longitudinal study of HIV infection in US women. Briefly, a random sample of 248 HIV seropositive women was selected from the WIHS to study the effect of cervicovaginal HIV viral loads on the detection of oncogenic HPV, while adjusting the plasma HIV viral loads, the CD4+ T-cell count in whole blood, and other covariates [19]. There are over 100 HPV types, most of which cause no symptoms and resolve spontaneously, while 13 HPV types are oncogenic, i.e., have potential to cause cervical cancer. The cervical HIV RNA is a measure of the HIV copy number in the cervix of a woman, based on swab samples taken from the cervix, whereas the plasma HIV RNA measures HIV copy number in the blood. Each variable was assessed at semi-annual clinical visits, with the number of visits ranging from one to four (1.5 years of follow-up). The total number of person-visits was 1664. A significant proportion of HIV RNA levels were left censored: 57% of cervical HIV RNA and 26% of plasma HIV RNA values, with the corresponding LLDs of 50 and 80 copies/mL, respectively. We applied the HDL, ML, and MCNR methods to the $\log_{10}$ transformed HIV viral load data and the scaled square root transformed CD4+ counts (square root of CD4 divided by 10). The covariates included age in decades, race, smoking, number of male sexual partner in the past 6 months, and recent use of the highly active antiretroviral therapy (HAART) as an indicator variable. The results are shown in Table III.

From the table, we find that age had an inverse association with having oncogenic HPV infection, whereas current smoking, the number of recent sex partners, and current use of HAART each had a positive association with the oncogenic HPV infection, though none reached statistical significance based on MCNR. Being a black HIV positive woman in US had a significantly higher risk of contracting oncogenic HPV infections compared to a white HIV positive woman in the US across all three methods. For the coefficients of primary interest, the two HIV viral loads and the CD4+ cell count, both HDL and ML methods provided higher effect estimates and highly significant p-values, compared with MCNR. Based on the simulations, HDL and ML methods tend to have lower coverage probabilities than the nominal level; the results from the MCNR are in keeping with expectations based on the simulations. It shows that the odds ratio for $\log_{10}$ of cervical HIV viral load was 5.70 ($=\exp(1.74)$) with the 95% confidence interval (0.98, 33.25), while adjusting the plasma HIV viral load, the CD4+ count and other covariates.

For a more complete parameter estimates, including the effects of the covariates on the cervical HIV viral load, plasma HIV viral load and CD4+ count, the scale of between-subject random variation of these three biomarkers, and the measurement error parameters of these biomarkers, please see the Supplemental Tables S1-A and S1-B. Recently using HAART was significantly associated with a decrease in both cervical and plasma HIV viral

loads and an increase in CD4+ count; compared with White HIV+ those of Hispanic descent or other races had lower CD4+ counts.

5. Discussion

In this paper we proposed an MCNR method for the generalized linear mixed model with multiple repeatedly measured covariates subject to measurement errors and left-censoring. The performance of this method was shown in our simulations to be superior to the HDL and ML approaches in terms of the empirical SE and coverage probability on the parameters for the biomarkers with respect to the binary outcome. Whether at high censoring proportions or smaller sample size, the MCNR still performed well, while the HDL and ML performed poorly in terms of the coverage probability. In this paper, data from an HIV study was used to illustrate the proposed method but the approach naturally extends to other fields in which repeated measurements with errors and left-censoring of biomarkers are both present in the binary response setting. Furthermore, the MCNR method extends to measurement errors and right-censoring of biomarkers as well.

Both Gauss-Hermite quadrature ML and MCNR methods try to optimize the observed data likelihood, but in different ways. The Gauss-Hermite quadrature ML directly approximates the likelihood - product of integrals on the random effect $\mathbf{b}_i$ - with quadratures, the number of quadrature points needed for each integral on $\mathbf{b}_i$ is M^{q+1} , where M is the number of points along each axis of random effects $\mathbf{b}_i$, and $q + 1$ is the number of random effects in $\mathbf{b}_i$. When q is large, in our simulation, $q = 3$, increase of M by 1 significantly increases the computational burden and time for the ML method by Gauss-Hermite quadrature. Because of the practical computational constraint and the limited number of quadrature points, the ML method smoothed the likelihood function in the approximation, causing the method to have lower standard error estimates, hence lower coverage probabilities. In contrast, the MCNR method optimizes the Monte Carlo approximation of the expected complete data log-likelihood given the observed data, with expectation taken at the current parameter estimate, the increase of Monte Carlo sample size comes relatively cheaper than the increase of quadrature points, the MCNR method does a better job in approximating this high dimensional integration and optimization problem. And MCNR gives more conservative standard error estimates than the ML method and has higher coverage probabilities. The HDL method suffers the same problem as the ML method with the additional simplification of the likelihood by imputing the left-censored data with half of the detection limits, and it induces biases in the estimates as well. The HDL method was not using the correct model because of the imputation, and the computational constraint due to limited quadrature points adds another layer of error. The ML method was using the correct model, but suffers from the computational constraint due to limited quadrature points.

The problem at hand is very computationally challenging. For our application data, it required approximately 20 minutes using a PC to run the MCNR method. With more biomarkers or more random effects, the time would be longer. So we used a straightforward model and covariance structure. As more powerful computers become available, the computational cost problem can be alleviated to some extent.

In the simulation studies, all three methods HDL, ML, MCNR had simulations with non positive definite information matrices. This is a problem with the complex model and data structure. To overcome the problem, the bootstrapping method can be used to estimate the variances in the case of non positive definite information matrices. In addition, the variance estimate based on the gradient [1] can also be used with the MCNR method.

We have assumed that the drop-out process does not provide information on the likelihood of the observed data. For a short study period, such as the four semi-annual visits considered in our study, this is a reasonable assumption. The drop-out process has been studied by Albert and Follmann [20], Shen and Weissfeld [21], and Thiebaut *et al.* [22], although in simpler settings. Incorporating the drop-out process would make the likelihood more complex and the computation more challenging; in addition, more complex random effects model on the biomarkers may be considered. Both are potential areas of future research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was partially supported by NCI grants 5R01CA085178, 1R21CA139388 and a grant from Lupus Foundation of America.

References

1. Wu L. A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. *Journal of the American Statistical Association*. 2002; 97:955–964.
2. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73:13–22. DOI: 10.1093/biomet/73.1.13
3. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*. 1993; 88:9–25.
4. Stroup, WW. *Generalized Linear Mixed Models*. CRC Press; Boca Raton, FL: 2012.
5. Jiang, J. *Linear and Generalized Linear Mixed Models and Their Applications*. Springer; New York: 2007.
6. Follmann DA, Hunsberger SA, Albert PS. Repeated probit regression when covariates are measured with error. *Biometrics*. 1999; 55:403–409. [PubMed: 11318193]
7. 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; 93:249–261. DOI: 10.2307/2669621
8. Carroll, RJ., Ruppert, D., Stefanski, LA., Crainiceanu, CM. *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman and HALL/CRC; Boca Raton: 2006.
9. Torabi M. Likelihood inference in generalized linear mixed measurement error models. *Computational Statistics and Data Analysis*. 2013; 57:549–557.
10. Giovanini, J. Unpublished doctoral dissertation. Oregon State University; Corvallis: 2008. Generalized linear mixed models with censored covariates.
11. Hughes JP. Mixed effects models with censored data with application to HIV RNA levels. *Biometrics*. 1999; 55:625–629. DOI: 10.1111/j.0006-341X.1999.00625.x [PubMed: 11318225]
12. McCulloch CE. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*. 1997; 92:162–170.

13. Xie X, Xue X, Gange SJ, Strickler HD, Kim MY. WIHS HPV Study Group. Estimation and inference on correlations between biomarkers with repeated measures and left-censoring due to minimum detection levels. *Statistics in Medicine*. 2012; 31:2275–2289. DOI: 10.1002/sim.5371 [PubMed: 22714546]
14. Geyer CJ. Practical Markov Chain Monte Carlo. *Statistical Science*. 1992; 7:473–511.
15. Kuk AY, Cheng YW. The Monte Carlo Newton-Raphson algorithm. *Journal of Statistical Computation and Simulation*. 1997; 59:233–250.
16. Booth JG, Hobert JP. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society Series B*. 1999; 61:265–285. DOI: 10.1111/1467-9868.00176
17. Louis TA. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society Series B*. 1982; 44:226–233.
18. Gelman, A., Carlin, JB., Stern, HS., Rubin, DB. *Bayesian Data Analysis*. 1. Chapman and Hall/CRC; Boca Raton, FL: 1995.
19. Gharvey J, Kovacs A, Burk RD, Stewart ML, Minkoff H, Xie X, D'souza G, Xue X, Heather WD, Levine AM, Einstein MH, Colie C, Anastos K, Eltoum IE, Herold BC, Palefsky JM, Strickler HD. Genital tract HIV RNA levels and their associations with human papillomavirus infection and risk of cervical precancer. *Journal of Acquired Immune Deficiency Syndromes*. 2014; 66:316–323. [PubMed: 24694931]
20. Albert PS, Follmann DA. A random effects transition model for longitudinal binary data with informative missingness. *Statistica Neerlandica*. 2003; 57:100–111.
21. Shen C, Weissfeld LA. A copula model for repeated measurements with non-ignorable non-monotone missing outcome. *Statistics in Medicine*. 2006; 25:2427–2440. DOI: 10.1002/sim.2355 [PubMed: 16143999]
22. Thiebaut R, Jacqmin-Gadda H, Babiker A, Commenges D. The CASCADE Collaboration. Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of CD4+ cell count and HIV RNA viral load in response to treatment of HIV infection. *Statistics in Medicine*. 2005; 24:65–82. DOI: 10.1002/sim.1923 [PubMed: 15523706]
23. Magnus, JR., Neudecker, H. *Matrix Differential Calculus with Application in Statistics and Econometrics*. 3. John Wiley and Sons; Chichester: 2007.

Appendix: Derivation of the gradient and information matrix for the MCNR method

Denote $\eta_i = \mathbf{D}_i^y \beta_y + \mathbf{1}_i (\mathbf{1} \beta_b^T) \mathbf{b}_i$ and consider the complete data log-likelihood contribution from subject i ,

$$l_i = \sum_{j=1}^{n_i} [y_{ij} \eta_{ij} - \log(1 + \exp(\eta_{ij}))] - \frac{1}{2} \log(\sigma_{b11}^2) - \frac{1}{2} \log(\det(\Psi)) - \frac{b_{i1}^2}{2\sigma_{b11}^2} - \frac{1}{2} \mathbf{b}_{i,-1}^T \Psi^{-1} \mathbf{b}_{i,-1} - \frac{n_i}{2} \log(\det(\Sigma)) - \frac{1}{2} \sum_{j=1}^{n_i} (\mathbf{X}_{ij}^T - \mathbf{b}_{i,-1} - \tilde{\mathbf{D}}_{ij}^x \text{vec}(\beta_x))^T \Sigma^{-1} (\mathbf{X}_{ij}^T - \mathbf{b}_{i,-1} - \tilde{\mathbf{D}}_{ij}^x \text{vec}(\beta_x)) + C,$$

where $\eta_{ij} = \mathbf{D}_{ij}^y \beta_y + (\mathbf{1} \beta_b^T) \mathbf{b}_i$, $\mathbf{D}_{ij}^y$ is the j -th row of the matrix $\mathbf{D}_i^y$, $\tilde{\mathbf{D}}_{ij}^x = \mathbf{I}_q \otimes \mathbf{D}_{ij}^x$, C is some constant not dependent on any of the parameters. Let $p_{ij} = \exp(\eta_{ij}) / (1 + \exp(\eta_{ij}))$, $\mathbf{r}_{ij} = \mathbf{X}_{ij}^T - \mathbf{b}_{i,-1} - \tilde{\mathbf{D}}_{ij}^x \text{vec}(\beta_x)$. Denote $\text{vech}(\mathbf{A})$ the function of extracting the unique parameters of a symmetric matrix $\mathbf{A}$ into a vector by eliminating all the supradiagonal elements of $\mathbf{A}$ from $\text{vec}(\mathbf{A})$, $\mathbf{W}_b = \Psi^{-1}$, $\mathbf{W}_e = \Sigma^{-1}$, $\mathbf{D}_q$ the duplication matrix of dimension

q^2 by $q^*(q+1)/2$ which satisfies the equation $\mathbf{D}_q \text{vech}(\mathbf{A}) = \text{vec}(\mathbf{A})$ for any q by q symmetric matrix $\mathbf{A}$, and $\mathbf{K}_q$ the communication matrix of dimension q^2 by q^2 which satisfies the equation $\mathbf{T}_q \text{vec}(\mathbf{B}) = \text{vec}(\mathbf{B}^T)$ for any q by q general matrix $\mathbf{B}$ [23]. Omitting the lower and upper bounds of the summations over j and k in the following expressions, we obtain the gradients of log-likelihood ll_i with respect to the parameters:

$$\begin{aligned} \mathbf{S}_{\beta_y}^i &= \left[\sum_j (y_{ij} - p_{ij}) \mathbf{D}_{ij}^y \right]^T, \\ \mathbf{S}_{\beta_b}^i &= \sum_j (y_{ij} - p_{ij}) \mathbf{b}_{i,-1}, \\ \mathbf{S}_{\text{vec}(\beta_x)}^i &= \left[\sum_j \mathbf{r}_{ij}^T \sum^{-1} \tilde{\mathbf{D}}_{ij}^x \right]^T, \\ S_{\sigma_{b11}^2}^i &= -\frac{1}{2\sigma_{b11}^2} \left(1 - \frac{b_{i1}^2}{\sigma_{b11}^2} \right), \\ \mathbf{S}_{\text{vech}(\Psi)}^i &= \left[-\frac{1}{2} (\text{vec}(\mathbf{W}_b)^T - (\mathbf{b}_{i,-1}^T \otimes \mathbf{b}_{i,-1}^T) (\mathbf{W}_b \otimes \mathbf{W}_b)) \mathbf{D}_q \right]^T, \\ \mathbf{S}_{\text{vech}(\Sigma)}^i &= \left[-\frac{1}{2} (n_i \text{vec}(\mathbf{W}_e)^T - (\sum_j \mathbf{r}_{ij}^T \otimes \mathbf{r}_{ij}^T) (\mathbf{W}_e \otimes \mathbf{W}_e)) \mathbf{D}_q \right]^T. \end{aligned}$$

For the information matrices (negative Hessians) of ll_i with respect to the parameters:

$$\begin{aligned} \mathbf{I}_{\beta_y}^i &= \sum_j p_{ij} (1 - p_{ij}) [\mathbf{D}_{ij}^y]^T \mathbf{D}_{ij}^y, \\ \mathbf{I}_{\beta_y, \beta_b}^i &= \sum_j [\mathbf{D}_{ij}^y]^T p_{ij} (1 - p_{ij}) \mathbf{b}_{i,-1}^T, \\ \mathbf{I}_{\beta_b}^i &= \sum_j p_{ij} (1 - p_{ij}) \mathbf{b}_{i,-1} \mathbf{b}_{i,-1}^T, \\ \mathbf{I}_{\text{vec}(\beta_x)}^i &= \sum_j (\tilde{\mathbf{D}}_{ij}^x)^T \sum^{-1} \tilde{\mathbf{D}}_{ij}^x, \\ I_{\sigma_{b11}^2}^i &= \frac{1}{2\sigma_{b11}^4} \left(\frac{2b_{i1}^2}{\sigma_{b11}^2} - 1 \right), \\ \mathbf{I}_{\text{vech}(\Psi)}^i &= -\frac{1}{2} \mathbf{D}_q^T [\mathbf{T}_q (\mathbf{W}_b \otimes \mathbf{W}_b) - ((\mathbf{W}_b \mathbf{b}_{i,-1} \mathbf{b}_{i,-1}^T \mathbf{W}_b) \otimes \mathbf{W}_b + \mathbf{W}_b \otimes (\mathbf{W}_b \mathbf{b}_{i,-1} \mathbf{b}_{i,-1}^T \mathbf{W}_b)) \mathbf{T}_q] \mathbf{D}_q, \\ \mathbf{I}_{\text{vec}(\beta_x), \text{vech}(\Sigma)}^i &= \frac{1}{2} \left[\sum_j ((\mathbf{W}_e \tilde{\mathbf{D}}_{ij}^x) \otimes (\mathbf{W}_e \mathbf{r}_{ij}) + (\mathbf{W}_e \mathbf{r}_{ij}) \otimes (\mathbf{W}_e \tilde{\mathbf{D}}_{ij}^x)) \right]^T \mathbf{D}_q, \\ \mathbf{I}_{\text{vech}(\Sigma)}^i &= -\frac{1}{2} \mathbf{D}_q^T [n_i \mathbf{T}_q (\mathbf{W}_e \otimes \mathbf{W}_e) - \sum_j ((\mathbf{W}_e \mathbf{r}_{ij} \mathbf{r}_{ij}^T \mathbf{W}_e) \otimes \mathbf{W}_e + \mathbf{W}_e \otimes (\mathbf{W}_e \mathbf{r}_{ij} \mathbf{r}_{ij}^T \mathbf{W}_e)) \mathbf{T}_q] \mathbf{D}_q. \end{aligned}$$

The expressions for $\log(L_c)/\beta$ and ${}^2\log(L_c)/\beta\beta^T$ in Equation (8) by the notations above are as follows:

$$\frac{\partial \log(L_c)}{\partial \beta} = \sum_{i=1}^m (\mathbf{S}_{\beta_y}^i{}^T, \mathbf{S}_{\beta_b}^i{}^T, \mathbf{S}_{\text{vec}(\beta_x)}^i{}^T)^T,$$

$$\frac{\partial^2 \log(L_c)}{\partial \beta \partial \beta^T} = - \sum_{i=1}^m \begin{pmatrix} \mathbf{I}_{\beta_y}^i & \mathbf{I}_{\beta_y, \beta_b}^i & \mathbf{0} \\ \mathbf{I}_{\beta_y, \beta_b}^i{}^T & \mathbf{I}_{\beta_b}^i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{\text{vec}(\beta_x)}^i \end{pmatrix}.$$

For the cross-products of the gradients of l_i with respect to the parameters:

$$\mathbf{S}_{\beta_y}^i \mathbf{S}_{\beta_y}^i{}^T = \sum_{j,k} (y_{ij} - p_{ij})(y_{ik} - p_{ik}) [\mathbf{D}_{ij}^y]^T \mathbf{D}_{ik}^y,$$

$$\mathbf{S}_{\beta_y}^i \mathbf{S}_{\beta_b}^i{}^T = \sum_{j,k} [\mathbf{D}_{ij}^y]^T (y_{ij} - p_{ij})(y_{ik} - p_{ik}) \mathbf{b}_{i,-1}^T,$$

$$\mathbf{S}_{\beta_b}^i \mathbf{S}_{\beta_b}^i{}^T = \sum_{j,k} (y_{ij} - p_{ij})(y_{ik} - p_{ik}) \mathbf{b}_{i,-1} \mathbf{b}_{i,-1}^T,$$

$$\mathbf{S}_{\beta_y}^i \mathbf{S}_{\text{vec}(\beta_x)}^i{}^T = \sum_{j,k} [\mathbf{D}_{ij}^y]^T (y_{ij} - p_{ij}) \mathbf{r}_{ik}^T \sum^{-1} \tilde{\mathbf{D}}_{ik}^x,$$

$$\mathbf{S}_{\beta_b}^i \mathbf{S}_{\text{vec}(\beta_x)}^i{}^T = \sum_{j,k} (y_{ij} - p_{ij}) \mathbf{b}_{i,-1} \mathbf{r}_{ik}^T \sum^{-1} \tilde{\mathbf{D}}_{ik}^x,$$

$$\mathbf{S}_{\text{vec}(\beta_x)}^i \mathbf{S}_{\text{vec}(\beta_x)}^i{}^T = \sum_{j,k} [\tilde{\mathbf{D}}_{ij}^x]^T \sum^{-1} \mathbf{r}_{ij} \mathbf{r}_{ik}^T \sum^{-1} \tilde{\mathbf{D}}_{ik}^x,$$

$$\mathbf{S}_{\beta_y}^i \mathbf{S}_{\sigma_{b11}^2}^i = - \frac{1}{2\sigma_{b11}^2} \sum_j (y_{ij} - p_{ij}) \left(1 - \frac{b_{i1}^2}{\sigma_{b11}^2}\right) [\mathbf{D}_{ij}^y]^T,$$

$$\mathbf{S}_{\beta_b}^i \mathbf{S}_{\sigma_{b11}^2}^i = - \frac{1}{2\sigma_{b11}^2} \sum_j (y_{ij} - p_{ij}) \left(1 - \frac{b_{i1}^2}{\sigma_{b11}^2}\right) \mathbf{b}_{i,-1},$$

$$\mathbf{S}_{\text{vec}(\boldsymbol{\beta}_{\mathbf{x}})}^i S_{\sigma_{b11}^2}^i = -\frac{1}{2\sigma_{b11}^2} \sum_j [\tilde{\mathbf{D}}_{ij}^{\mathbf{x}}]^T \boldsymbol{\Sigma}^{-1} \left(1 - \frac{b_{i1}^2}{\sigma_{b11}^2}\right) \mathbf{r}_{ij},$$

$$S_{\sigma_{b11}^2}^i S_{\sigma_{b11}^2}^i = \frac{1}{4\sigma_{b11}^4} \left(1 - \frac{b_{i1}^2}{\sigma_{b11}^2}\right)^2,$$

$$\mathbf{S}_{\boldsymbol{\beta}_{\mathbf{y}}}^i \mathbf{S}_{\text{vech}(\boldsymbol{\Psi})}^i{}^T = -\frac{1}{2} \sum_j [\mathbf{D}_{ij}^{\mathbf{y}}]^T (y_{ij} - p_{ij}) (\text{vec}(\mathbf{W}_{\mathbf{b}})^T - (\mathbf{b}_{i,-1}^T \otimes \mathbf{b}_{i,-1}^T) (\mathbf{W}_{\mathbf{b}} \otimes \mathbf{W}_{\mathbf{b}})) \mathbf{D}_q,$$

$$\mathbf{S}_{\boldsymbol{\beta}_{\mathbf{b}}}^i \mathbf{S}_{\text{vech}(\boldsymbol{\Psi})}^i{}^T = -\frac{1}{2} \sum_j (y_{ij} - p_{ij}) \mathbf{b}_{i,-1} (\text{vec}(\mathbf{W}_{\mathbf{b}})^T - (\mathbf{b}_{i,-1}^T \otimes \mathbf{b}_{i,-1}^T) (\mathbf{W}_{\mathbf{b}} \otimes \mathbf{W}_{\mathbf{b}})) \mathbf{D}_q,$$

$$\mathbf{S}_{\text{vec}(\boldsymbol{\beta}_{\mathbf{x}})}^i \mathbf{S}_{\text{vech}(\boldsymbol{\Psi})}^i{}^T = -\frac{1}{2} \sum_j [\tilde{\mathbf{D}}_{ij}^{\mathbf{x}}]^T \boldsymbol{\Sigma}^{-1} \mathbf{r}_{ij} (\text{vec}(\mathbf{W}_{\mathbf{b}})^T - (\mathbf{b}_{i,-1}^T \otimes \mathbf{b}_{i,-1}^T) (\mathbf{W}_{\mathbf{b}} \otimes \mathbf{W}_{\mathbf{b}})) \mathbf{D}_q,$$

$$S_{\sigma_{b11}^2}^i S_{\text{vech}(\boldsymbol{\Psi})}^i{}^T = \frac{1}{4\sigma_{b11}^2} \left(1 - \frac{b_{i1}^2}{\sigma_{b11}^2}\right) (\text{vec}(\mathbf{W}_{\mathbf{b}})^T - (\mathbf{b}_{i,-1}^T \otimes \mathbf{b}_{i,-1}^T) (\mathbf{W}_{\mathbf{b}} \otimes \mathbf{W}_{\mathbf{b}})) \mathbf{D}_q,$$

$$\begin{aligned} & \mathbf{S}_{\text{vech}(\boldsymbol{\Psi})}^i \mathbf{S}_{\text{vech}(\boldsymbol{\Psi})}^i{}^T \\ &= \frac{1}{4} \mathbf{D}_q^T (\text{vec}(\mathbf{W}_{\mathbf{b}}) \text{vec}(\mathbf{W}_{\mathbf{b}})^T \\ & \quad - \text{vec}(\mathbf{W}_{\mathbf{b}}) \text{vec}(\mathbf{b}_{i,-1} \mathbf{b}_{i,-1}^T)^T (\mathbf{W}_{\mathbf{b}} \otimes \mathbf{W}_{\mathbf{b}}) \\ & \quad - (\mathbf{W}_{\mathbf{b}} \otimes \mathbf{W}_{\mathbf{b}}) \text{vec}(\mathbf{b}_{i,-1} \mathbf{b}_{i,-1}^T) \text{vec}(\mathbf{W}_{\mathbf{b}})^T \\ & \quad + (\mathbf{W}_{\mathbf{b}} \otimes \mathbf{W}_{\mathbf{b}}) \text{vec}(\mathbf{b}_{i,-1} \mathbf{b}_{i,-1}^T) \text{vec}(\mathbf{b}_{i,-1} \mathbf{b}_{i,-1}^T)^T (\mathbf{W}_{\mathbf{b}} \otimes \mathbf{W}_{\mathbf{b}})) \mathbf{D}_q, \end{aligned}$$

$$\mathbf{S}_{\boldsymbol{\beta}_{\mathbf{y}}}^i \mathbf{S}_{\text{vech}(\boldsymbol{\Sigma})}^i{}^T = -\frac{1}{2} \sum_{j,k} [\mathbf{D}_{ij}^{\mathbf{y}}]^T (y_{ij} - p_{ij}) (\text{vec}(\mathbf{W}_{\mathbf{e}})^T - (\mathbf{r}_{ik}^T \otimes \mathbf{r}_{ik}^T) (\mathbf{W}_{\mathbf{e}} \otimes \mathbf{W}_{\mathbf{e}})) \mathbf{D}_q,$$

$$\mathbf{S}_{\boldsymbol{\beta}_{\mathbf{b}}}^i \mathbf{S}_{\text{vech}(\boldsymbol{\Sigma})}^i{}^T = -\frac{1}{2} \sum_{j,k} (y_{ij} - p_{ij}) \mathbf{b}_{i,-1} (\text{vec}(\mathbf{W}_{\mathbf{e}})^T - (\mathbf{r}_{ik}^T \otimes \mathbf{r}_{ik}^T) (\mathbf{W}_{\mathbf{e}} \otimes \mathbf{W}_{\mathbf{e}})) \mathbf{D}_q,$$

$$\mathbf{S}_{\text{vec}(\boldsymbol{\beta}_x)}^i \mathbf{S}_{\text{vech}(\sum)}^i{}^T = -\frac{1}{2} \sum_{j,k} [\tilde{\mathbf{D}}_{ij}^x]^T \sum^{-1} \mathbf{r}_{ij} (\text{vec}(\mathbf{W}_e)^T - (\mathbf{r}_{ik}^T \otimes \mathbf{r}_{ik}^T)(\mathbf{W}_e \otimes \mathbf{W}_e)) \mathbf{D}_q,$$

$$S_{\sigma_{b11}^2}^i \mathbf{S}_{\text{vech}(\sum)}^i{}^T = \frac{1}{4\sigma_{b11}^2} \left(1 - \frac{b_{i1}^2}{\sigma_{b11}^2}\right) (n_i \text{vec}(\mathbf{W}_e)^T - (\sum_j \mathbf{r}_{ij}^T \otimes \mathbf{r}_{ij}^T)(\mathbf{W}_e \otimes \mathbf{W}_e)) \mathbf{D}_q,$$

$$\mathbf{S}_{\text{vech}(\Psi)}^i \mathbf{S}_{\text{vech}(\sum)}^i{}^T = \frac{1}{4} \mathbf{D}_q^T (\text{vec}(\mathbf{W}_b) - (\mathbf{W}_b \otimes \mathbf{W}_b)(\mathbf{b}_{i,-1} \otimes \mathbf{b}_{i,-1})) (n_i \text{vec}(\mathbf{W}_e)^T - (\sum_j \mathbf{r}_{ij}^T \otimes \mathbf{r}_{ij}^T)(\mathbf{W}_e \otimes \mathbf{W}_e)) \mathbf{D}_q,$$

$$\begin{aligned} & \mathbf{S}_{\text{vech}(\sum)}^i \mathbf{S}_{\text{vech}(\sum)}^i{}^T \\ &= \frac{1}{4} \mathbf{D}_q^T (n_i^2 \text{vec}(\mathbf{W}_e) \text{vec}(\mathbf{W}_e)^T \\ & \quad - n_i \text{vec}(\mathbf{W}_e) (\sum_j \mathbf{r}_{ij}^T \otimes \mathbf{r}_{ij}^T)(\mathbf{W}_e \otimes \mathbf{W}_e) \\ & \quad - n_i (\mathbf{W}_e \otimes \mathbf{W}_e) (\sum_j \mathbf{r}_{ij} \otimes \mathbf{r}_{ij}) \text{vec}(\mathbf{W}_e)^T \\ & \quad + (\mathbf{W}_e \otimes \mathbf{W}_e) (\sum_j \mathbf{r}_{ij}^T \otimes \mathbf{r}_{ij}^T) (\sum_j \mathbf{r}_{ij} \otimes \mathbf{r}_{ij}) (\mathbf{W}_e \otimes \mathbf{W}_e)) \mathbf{D}_q. \end{aligned}$$

Let $\boldsymbol{\theta}_u = (\boldsymbol{\beta}_y^T, \boldsymbol{\beta}_b^T, \text{vec}(\boldsymbol{\beta}_x)^T, \sigma_{b11}^2, \text{vech}(\Psi)^T, \text{vech}(\sum)^T)^T$, which is the unique parameter vector of the model, the gradient $\mathbf{S}_{\boldsymbol{\theta}_u}^i$ on $\boldsymbol{\theta}_u$ for the i -th subject, the information matrix $\mathbf{I}_{\boldsymbol{\theta}_u}^i$, and the cross products $\mathbf{S}_{\boldsymbol{\theta}_u}^i [\mathbf{S}_{\boldsymbol{\theta}_u}^i]^T$ can be calculated from the above equations. The observed information matrix on $\boldsymbol{\theta}_u$ is calculated as

$$\mathbf{I}_{\boldsymbol{\theta}_u}^{\text{obs}} = \sum_{i=1}^m \left(E[\mathbf{I}_{\boldsymbol{\theta}_u}^i | \mathbf{y}_i, \mathbf{Q}_i, \mathbf{C}_i] - E[\mathbf{S}_{\boldsymbol{\theta}_u}^i (\mathbf{S}_{\boldsymbol{\theta}_u}^i)^T | \mathbf{y}_i, \mathbf{Q}_i, \mathbf{C}_i] + E[\mathbf{S}_{\boldsymbol{\theta}_u}^i | \mathbf{y}_i, \mathbf{Q}_i, \mathbf{C}_i] E[\mathbf{S}_{\boldsymbol{\theta}_u}^i | \mathbf{y}_i, \mathbf{Q}_i, \mathbf{C}_i]^T \right).$$

The expectations in the above equation are calculated by the Monte Carlo method.

Figure 1.

Box-plot comparisons of standard errors in simulation studies between 3 methods: HDL, ML, and MCNR (number of subjects = 300). Outliers were not shown. The low, medium, and high censoring correspond to the censoring proportions (0.20, 0.10), (0.40, 0.30) and (0.60, 0.50) for biomarkers 1 and 2, respectively.

Figure 2.

Box-plot comparisons of standard errors in simulation studies between 3 methods: HDL, ML, and MCNR (number of subjects = 100). Outliers were not shown. The low, medium, and high censoring correspond to the censoring proportions (0.20, 0.10), (0.40, 0.30) and (0.60, 0.50) for biomarkers 1 and 2, respectively.

Table 1
Comparison of three methods: HDL, ML, MCNR in simulation studies (number of subjects = 300).

Censoring Proportions*	Performance Measure [†]	Method	$\beta_{j,1} (-3.4)^{\ddagger}$	$\beta_{j,2} (0.1)$	$\beta_{j,2} (4.1)$	$\beta_{j,3} (-4.8)$	$\beta_{j,4} (-0.7)$	$\sigma_{\theta 11}^2 (4.3)$
(0.20, 0.10)	Bias	HDL	0.23	-0.002	0.49	-0.03	-0.08	-0.58
		ML	0.22	-0.001	-0.09	0.08	-0.10	-0.60
		MCNR	-0.12	0.004	-0.24	0.54	0.15	0.77
	EmpSE	HDL	0.38	0.096	7.95	8.08	1.19	0.84
		ML	0.37	0.094	6.49	7.38	1.24	0.86
		MCNR	0.44	0.099	2.53	2.81	0.78	1.43
	ModSE	HDL	0.36	0.095	11.72	11.52	0.73	0.74
		ML	0.36	0.095	11.26	12.23	0.75	0.74
		MCNR	0.40	0.096	4.90	5.14	1.08	2.14
	CovProb	HDL	0.845	0.947	0.850	0.809	0.795	0.879
		ML	0.862	0.950	0.768	0.775	0.782	0.874
		MCNR	0.933	0.941	0.913	0.905	0.928	0.933
(0.40, 0.30)	Bias	HDL	0.23	-0.002	1.87	-1.15	-0.16	-0.59
		ML	0.22	-0.002	0.23	-0.30	-0.25	-0.63
		MCNR	-0.10	0.000	1.09	-0.71	0.02	0.54
	EmpSE	HDL	0.38	0.096	7.63	7.42	1.25	0.85
		ML	0.37	0.095	6.37	7.29	1.39	0.85
		MCNR	0.43	0.102	4.57	4.82	1.06	1.40
	ModSE	HDL	0.37	0.095	12.28	11.67	1.21	0.77
		ML	0.36	0.094	9.48	10.48	0.82	0.74
		MCNR	0.41	0.096	7.88	8.32	1.72	2.23
	CovProb	HDL	0.851	0.945	0.872	0.829	0.816	0.882
		ML	0.848	0.945	0.793	0.795	0.778	0.870
		MCNR	0.946	0.932	0.952	0.949	0.958	0.955

Censoring Proportions [*]	Performance Measure [†]	Method	$\beta_{\gamma,1} (-3.4)^{\ddagger}$	$\beta_{\gamma,2} (0.1)$	$\beta_{\gamma,3} (4.1)$	$\beta_{\gamma,4} (-4.8)$	$\beta_{\gamma,5} (-0.7)$	$\sigma_{\gamma,1}^2 (4.3)$
(0.60, 0.50)	Bias	HDL	0.21	-0.003	1.41	-0.15	0.08	-0.62
		ML	0.20	-0.006	1.32	-1.58	-0.31	-0.65
		MCNR	-0.10	-0.001	1.86	-1.07	0.05	0.52
	EmpSE	HDL	0.40	0.095	8.15	7.77	1.24	0.83
		ML	0.37	0.092	9.15	10.51	1.58	0.92
		MCNR	0.43	0.104	6.14	6.01	1.24	1.39
	ModSE	HDL	0.37	0.095	12.01	10.79	0.76	0.74
		ML	0.37	0.095	24.85	28.41	0.88	0.77
		MCNR	0.41	0.096	14.06	13.11	2.36	3.09
	CovProb	HDL	0.845	0.948	0.822	0.746	0.791	0.885
		ML	0.880	0.957	0.801	0.807	0.745	0.880
		MCNR	0.943	0.924	0.962	0.957	0.948	0.957

^{*}The censoring proportions (left-censoring) are for biomarkers 1 and 2 respectively.

[†]Abbreviations: EmpSE = Empirical SE, ModSE = Model based SE, CovProb = Coverage probability.

[‡]The numbers in the parentheses in the header row are the true parameter values.

Note: The number of simulations with non positive definite information matrices was 86, 64, 39 for HDL, ML, MCNR at low censoring proportions; 103, 100, 146 at medium censoring proportions; and 74, 174, 290 at high censoring proportions.

Table II

Comparison of three methods: HDL, ML, MCNR in simulation studies (number of subjects = 100).

Censoring Proportions	Performance Measure	Method	$\beta_{j,1} (-3.4)$	$\beta_{j,2} (0.1)$	$\beta_{j,3} (4.1)$	$\beta_{j,4} (-4.8)$	$\beta_{j,4} (-0.7)$	$\sigma_{\theta,11}^2 (4.3)$
(0.20, 0.10)	Bias	HDL	-0.08	0.001	3.08	-2.68	-0.48	-0.03
		ML	-0.10	0.001	1.30	-1.33	-0.18	0.06
		MCNR	-0.18	0.008	0.41	-0.16	-0.18	0.59
	EmpSE	HDL	1.18	0.171	14.38	15.04	3.20	5.86
		ML	1.31	0.170	11.29	13.45	2.75	6.91
		MCNR	0.73	0.180	7.53	8.32	3.04	2.41
	ModSE	HDL	3.20	0.168	32.54	31.61	1.92	28.32
		ML	4.51	0.167	35.94	39.56	1.59	41.82
		MCNR	0.72	0.169	21.82	22.93	7.95	4.01
	CovProb	HDL	0.943	0.964	0.802	0.778	0.696	0.951
		ML	0.941	0.962	0.756	0.753	0.651	0.949
		MCNR	0.960	0.948	0.896	0.911	0.972	0.969
(0.40, 0.30)	Bias	HDL	-0.20	0.001	5.81	-4.92	-0.13	0.82
		ML	-0.08	0.006	1.37	-1.39	-0.05	-0.27
		MCNR	-0.20	0.018	0.83	-0.40	-0.10	0.49
	EmpSE	HDL	1.83	0.175	20.71	20.60	3.32	11.57
		ML	0.98	0.176	13.76	16.29	3.47	3.69
		MCNR	0.76	0.178	10.24	10.18	2.20	2.41
	ModSE	HDL	25.82	0.169	78.48	76.98	7.39	255.63
		ML	4.04	0.168	25.92	29.46	2.23	33.46
		MCNR	0.81	0.169	22.19	22.32	4.29	5.05
	CovProb	HDL	0.935	0.954	0.824	0.786	0.707	0.938
		ML	0.937	0.958	0.766	0.784	0.701	0.952
		MCNR	0.947	0.955	0.899	0.919	0.968	0.972

Censoring Proportions	Performance Measure	Method	$\beta_{y,1}$ (−3.4)	$\beta_{j,2}$ (0.1)	$\beta_{b,2}$ (4.1)	$\beta_{b,3}$ (−4.8)	$\beta_{b,4}$ (−0.7)	σ_{b1}^2 (4.3)
(0.60, 0.50)	Bias	HDL	−0.03	0.003	3.08	−1.65	−0.02	−0.24
		ML	−0.10	0.001	1.53	−1.85	−0.16	−0.24
		MCNR	−0.18	0.019	−0.30	1.05	0.24	0.58
	EmpSE	HDL	0.77	0.180	18.41	16.78	2.80	2.64
		ML	1.01	0.177	18.21	19.74	3.18	3.90
		MCNR	0.79	0.176	5.69	5.34	1.46	2.68
	ModSE	HDL	0.77	0.169	24.44	20.29	1.66	2.19
		ML	6.74	0.169	32.87	35.93	1.55	55.77
		MCNR	0.73	0.169	30.48	28.79	8.48	5.28
	CovProb	HDL	0.937	0.943	0.808	0.763	0.691	0.937
		ML	0.936	0.952	0.725	0.733	0.673	0.924
		MCNR	0.933	0.947	0.900	0.913	0.973	0.980

Note: The number of simulations with non positive definite information matrices was 112, 127, 174 for HDL, ML, MCNR at low censoring proportions; 131, 166, 253 at medium censoring proportions; and 183, 249, 350 at high censoring proportions.

Table III

Effect estimates and standard errors for oncogenic HPV detection on the application data with HDL, ML, and MCNR methods.

Parameter	HDL				ML				MCNR			
	Estimate	SE	Z	P-value	Estimate	SE	Z	P-value	Estimate	SE	Z	P-value
Intercept:onc	-3.95	1.57			-3.89	1.60			-3.92	1.70		
Age in years/10:onc	-0.32	0.30	-1.04	0.2989	-0.31	0.31	-0.99	0.3235	-0.25	0.34	-0.75	0.4558
Race:onc												
White	ref				ref				ref			
Black	1.67	0.67	2.51	0.0121	1.49	0.70	2.14	0.0320	1.55	0.76	2.03	0.0421
Hispanic & others	0.20	0.70	0.29	0.7729	0.06	0.74	0.08	0.9360	0.37	0.83	0.44	0.6579
Smoking:onc												
Never	ref				ref				ref			
Former	-0.09	0.63	-0.14	0.8875	-0.07	0.66	-0.10	0.9209	-0.02	0.71	-0.03	0.9722
Current	1.10	0.58	1.92	0.0549	1.10	0.60	1.84	0.0658	1.00	0.67	1.50	0.1329
No. of male sexual partner past 6 mo.:onc												
0	ref				ref				ref			
1	0.43	0.46	0.93	0.3526	0.43	0.46	0.94	0.3460	0.36	0.45	0.80	0.4229
2	1.90	0.74	2.56	0.0103	1.87	0.72	2.59	0.0097	1.32	0.77	1.72	0.0856
3+	1.37	0.90	1.52	0.1278	1.40	0.90	1.55	0.1220	1.36	1.09	1.25	0.2100
HAART use past 6mo.:onc												
No	ref				ref				ref			
Yes	0.46	0.40	1.13	0.2569	0.47	0.40	1.15	0.2483	0.37	0.41	0.89	0.3758
Cervical swab HIV VL	5.79	1.52	3.81	0.0001	3.05	0.96	3.18	0.0015	1.74	0.90	1.94	0.0530
Plasma HIV VL	-4.75	1.31	-3.64	0.0003	-3.87	1.22	-3.16	0.0016	-2.00	1.04	-1.92	0.0548
Sqrt(CD4+ count)/10	-2.64	0.61	-4.35	<.0001	-2.70	0.65	-4.16	<.0001	-1.48	0.61	-2.42	0.0156
$\sigma_{\theta 1}^2$	3.09	1.01	NA	NA	3.13	1.01	NA	NA	4.88	1.73	NA	NA

Abbreviations: onc = Oncogenic HPV, VL = viral load, ref = reference level.