Title	A ROBUST CALIBRATION-ASSISTED METHOD FOR LINEAR MIXED EFFECTS MODEL UNDER CLUSTER-SPECIFIC NONIGNORABLE MISSINGNESS
Manuscript ID	SS-2016-0317.R1
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0317
Complete List of Authors	Myunghee Paik Yongchan Kwon Jae Kwang Kim and Hongsoo Kim
Corresponding Author	Myunghee Paik
E-mail	myungheechopaik@gmail.com

Statistica Sinica
doi:http://dx.doi.org/10.5705/ss.20??.???

A ROBUST CALIBRATION-ASSISTED METHOD FOR LINEAR MIXED EFFECTS MODEL UNDER CLUSTER-SPECIFIC NONIGNORABLE MISSINGNESS

Yongchan Kwon[1], Jae Kwang Kim[2], Myunghee Cho Paik[1]
and Hongsoo Kim[1]

[1]*Seoul National University and* [2]*Iowa State University*

Abstract: We propose a method for linear mixed effects models when the covariates are completely observed but the outcome of interest is subject to missing under cluster-specific nonignorable (CSNI) missingness. Our strategy is to replace missing quantities in the full-data objective function with unbiased predictors derived from inverse probability weighting and calibration technique. The proposed approach can be applied to estimating equations or likelihood functions with modified E-step, and does not require numerical integration as previous methods. Unlike usual inverse probability weighting, the proposed method does not require correct specification of the response model as long as the CSNI assumption is correct, and renders inference under CSNI without a full distributional assumption. Consistency and asymptotic normality are shown with a consistent variance estimator. Simulation results and a real data example are presented.

Key words and phrases, Calibration method, Cluster-specific nonignorable missingness, Inverse probability weighting, Nonignorable missingness.

1. Introduction

Missing data occur for various reasons and become frequent problems in surveys, clustered or longitudinal data. We consider the case in regression setting with clustered data when the outcome variable is subject to missing but the covariates are completely observed. Rubin (1976) in his seminal paper used the term missing at random if the response or observation indicator for the outcome is independent of the outcome given the covariates. When the data are missing at random, inverse probability weighting and imputation approaches, aside from likelihood approach, have been developed to handle missing values (Robins et al., 1995; Paik, 1997). The validity of these approaches depend on correct specification of the response and the imputation models, respectively. Many authors have investigated doubly robust methods which utilize both auxiliary models, but require correct specification of either model for the validity of the method while achieving semiparametric efficiency when both are correct (Robins et al., 1994; Bang and Robins, 2005; Kang and Schafer, 2007; Han, 2014). In the case of nonignorable missingness, the probability of response depends on unobserved data, and the analysis becomes challenging. The methods handling nonignorable missingness require both auxiliary models to be correctly specified. Many authors attacked the nonignorability problem using likelihood approach (Follmann and Wu, 1995; Ibrahim et al., 2001; Gao, 2004; Zhang and Paik, 2009), imputation approach (Paik, 1997; Yang et al., 2013), and inverse probability weighting approach (Rotnitzky and Robins, 1995; Shao and Wang, 2016). Nonignorability often causes nonidentifiability which should be carefully addressed in developing methods (Wang et al., 2014; Molenberghs et al., 2008).

In cluster data analysis, missing data should be handled while taking the correlation within cluster. Furthermore, the response indicators may be correlated within cluster. A popular way to model clustered data is mixed effects model where random effects are shared among the outcomes within the cluster to induce correlation. The random effects are not directly observable, which opens a possibility that data can be nonignorably missing when the response indicator depends on the random effect. It is plausible that unmeasured common factor that explains the outcome also explains the response indicators. When the response indicator depends on the random or cluster effects, but is independent of outcome given covariates and cluster effects, Yuan and Little (2007) called cluster-specific nonignorable (CSNI) missingness. The CSNI mechanism is a subclass of nonignorable missingness, but due to the conditional independence, is less serious than the case where the response indicators depend on the unobserved outcomes that are planned to be measured. Yuan and Little (2007) considered a special case of CSNI where the response indicator depends on cluster-specific covariates. A few methods have been proposed in the context of survey sampling under CSNI in the presence of covariates that vary within cluster (Skinner and D'Arrigo, 2011; Kim et al., 2016).

In the mixed effects model setting under CSNI missingness, likelihood approach has been proposed by Ibrahim et al. (2001) and Gao (2004) using Monte Carlo expectation-maximization (EM) algorithm and Laplace approximation method, respectively. Both methods provide good parameter estimation with a full distributional assumption but computations are extensive. Recently, Shao and Zhang (2015) proposed a clever solution to estimate the regression parameter under CSNI without any auxiliary model assumptions by transforming the model so that random effects are eliminated. This method works for general structure of random effects and simplifies computation dramatically, but due to elimination, the variance component cannot be estimated.

In this paper we propose methods for linear mixed effects models under CSNI missingness without correctly specifying the response model. Our strategy is to replace missing quantities in the full-data objective function with their unbiased predictors derived using inverse probability weighting and calibration technique. We apply the proposed approach both to estimating equations and likelihood functions with modified E-step. While previous methods require a full distributional assumption, the proposed method can use assumptions on the first two moments. The proposed method is robust in a sense that the validity of the method relies on the CSNI aspect of the response model not on the correct specification of the functional form. While the proposed estimator does not require numerical integration, it provides a consistent estimator for the variance component. Consistency and asymptotic normality of the proposed estimator are shown along with a consistent variance estimator. The rest of this paper is organized as follows. In Section 2, we present basic notations and the existing methods. Section 3, we introduce the proposed method and present asymptotic properties. In Section 4, finite sample properties are examined via simulation studies. Section 5 illustrates a real data application.

2. Basic setup

Let y_{ij} be an outcome of interest, x_{ij} be a row vector of covariate for the jth unit $(j = 1, \ldots, n_i)$ in the ith cluster $(i = 1, \ldots, K)$. First consider the linear mixed effect models,

$$y_{ij} = x_{ij}\beta + a_i + e_{ij} \tag{2.1}$$

where β is an unknown regression parameter, random effects a_i's are distributed with mean zero and variance D, and error e_{ij}'s are conditionally independent given a_i and x_{ij} with $E(e_{ij} \mid x_{ij}, a_i) = 0$ and $\text{Var}(e_{ij} \mid x_{ij}, a_i) = \sigma^2$. The main goal is to estimate parameters $\theta = (\beta^T, \sigma^2, D)^T$. Suppose that all fixed covariates x_{ij}'s are completely observed but the outcomes y_{ij}, $j = 1, \cdots, n_i$ are subject to missing. Let δ_{ij} be the response indicator whose value is one if the outcome y_{ij} is observed, zero, otherwise. Assume that

$$P(\delta_{ij} = 1 \mid x_{ij}, a_i, y_{ij}) = P(\delta_{ij} = 1 \mid x_{ij}, a_i). \tag{2.2}$$

The mechanism given in (2.2) is called cluster-specific nonignorable (CSNI) by Yuan and Little (2007). The CSNI missingness states that the outcome y_{ij} is independent of response indicator δ_{ij} given x_{ij} and a_i. Yuan and Little

(2007) considered the special case $x_{ij} = x_i$. We use the following working model,

$$P(\delta_{ij} = 1 \mid x_{ij}, a_i) \equiv \pi(x_{ij}, \alpha_i; \gamma) = \frac{\exp(\alpha_i + x_{ij}\gamma)}{1 + \exp(\alpha_i + x_{ij}\gamma)}, \tag{2.3}$$

where $\alpha_i = \gamma_0 a_i$, and (γ_0, γ) are unknown parameters. We call it working model since the validity of the method does not depend on the functional form of π, but depends only on the CSNI assumption itself. We require $\pi(x_{ij}, \alpha_i; \gamma) > 0$ and $\sum_{j=1}^{n_i} \delta_{ij} > 0$ and fix $\pi(x_{ij}, \alpha_i; \gamma) = 1$ if $\sum_{j=1}^{n_i} \delta_{ij} = n_i$. We postulate the same cluster-specific factor is responsible for within-cluster correlation in (2.1) and (2.3). This type of models have been developed as a shared parameter model (Follmann and Wu, 1995) or a shared random effects model (Gao, 2004).

While imputation and inverse probability weighting approach are popular under missing-at-random mechanism due to their own merit, most existing works under nonignorable missingness utilize likelihood method. Assuming both the linear mixed effects model (2.1) and the response model (2.2), a marginal likelihood function has a form,

$$\prod_{i=1}^{K} \int \int \prod_{j=1}^{n_i} f(y_{ij} \mid x_{ij}, a_i) g(\delta_i \mid x_i, \alpha_i; \gamma) \phi(a_i) dy_{i,mis} da_i, \tag{2.4}$$

where $f(\cdot \mid \cdot)$ denotes the conditional density of y_{ij} given x_{ij} and a_i, $\delta_i = (\delta_{i1}, \delta_{i2}, \ldots, \delta_{in_i})^T$, $x_i = (x_{i1}, x_{i2}, \ldots, x_{in_i})^T$, $y_{i,mis}$ denotes missing parts of $y_i = (y_{i1}, \ldots, y_{in_i})$, $\phi(a_i)$ is a density of a_i, and $g(\delta_i \mid x_i, \alpha_i; \gamma) = \prod_{j=1}^{n_i} \pi(x_{ij}, \alpha_i; \gamma)^{\delta_{ij}} \{1 - \pi(x_{ij}, \alpha_i; \gamma)\}^{(1-\delta_{ij})}$. Ibrahim et al. (2001) proposed a Monte Carlo expectation-maximization (EM) algorithm to estimate the unknown parameters. Maximizing the marginal likelihood function (2.4) using the EM algorithm requires calculating the conditional expectation given observed data,

$$E(a_i \mid x_i, y_{i,obs}, \delta_i) = \frac{\int \int \prod_{j=1}^{n_i} a_i f(y_{ij} \mid x_{ij}, a_i) g(\delta_i \mid x_i, \alpha_i; \gamma) \phi(a_i) da_i dy_{i,mis}}{\int \int \prod_{j=1}^{n_i} f(y_{ij} \mid x_{ij}, a_i) g(\delta_i \mid x_i, \alpha_i; \gamma) \phi(a_i) da_i dy_{i,mis}}, \tag{2.5}$$

where $y_{i,obs}$ denotes observed parts of y_i, and

$$\begin{aligned} &E(y_{ij} | x_i, y_{i,obs}, \delta_i) \\ &= \delta_{ij} y_{ij} + (1 - \delta_{ij}) \frac{\int \int \prod_{j=1}^{n_i} y_{ij} f(y_{ij} \mid x_{ij}, a_i) g(\delta_i \mid x_i, \alpha_i; \gamma) \phi(a_i) da_i dy_{i,mis}}{\int \int \prod_{j=1}^{n_i} f(y_{ij} \mid x_{ij}, a_i) g(\delta_i \mid x_i, \alpha_i; \gamma) \phi(a_i) da_i dy_{i,mis}}. \end{aligned} \tag{2.6}$$

Evaluating (2.5) and (2.6) is computationally demanding. Implementing Monte Carlo version of the EM algorithm is also computationally extensive because the Gibbs sampling from the above prediction model involves multiple Monte Carlo integrations. Another approach is to approximate the marginal likelihood (2.4) using the Laplace approximation. As Gao (2004) pointed out, accuracy of the Laplace approximation is questionable, which could cause lack of convergence in practice.

3. Proposed method

The proposed approach starts from identifying functions with missing data in the objective function when data are fully observed. The next step is to derive unbiased predictors of the functions with missing data using inverse probability and calibration technique, and replace them in the full-data estimating function. Our approach can be applied to estimating equations or likelihood functions. We first examine the calibration method in estimating the marginal mean and the required assumptions needed for the validity of the method.

3.1 Calibration method

When the goal is to estimate the marginal mean, say, μ, $(\sum_{i=1}^K n_i)^{-1} \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \mu) = 0$ provides a consistent estimate. When some values are missing, Kim et al. (2016) proposed $(\sum_{i=1}^K n_i)^{-1} \sum_{i=1}^K \sum_{j=1}^{n_i} (\delta_{ij}/\hat{\pi}_{ij}) y_{ij}$ where $\hat{\pi}_{ij}$ satisfies

$$E[\sum_{i=1}^{K} \{\sum_{j=1}^{n_i} y_{ij} - \sum_{j=1}^{n_i} (\delta_{ij}/\hat{\pi}_{ij}) y_{ij}\}] = 0. \tag{3.1}$$

This approach can be viewed as replacing quantities with missing data, $\sum_{j=1}^{n_i} y_{ij}$, with unbiased predictor, $\sum_{j=1}^{n_i} (\delta_{ij}/\hat{\pi}_{ij}) y_{ij}$. When π does not depend on random effects, the usual inverse probability weighting method estimates π from maximum likelihood. On the surface, the difference between the calibration method and the inverse probability weighting method in the case of non-clustered data seems trivial since they only differ in how to estimate auxiliary model π. However, an important difference lies in the model assumption. To proceed we inspect the calibration condition of Kim et al. (2016) as follows:

$$E\{\sum_{i=1}^{K}\sum_{j=1}^{n_i} y_{ij}-\sum_{i=1}^{K}\sum_{j=1}^{n_i}(\delta_{ij}/\pi_{ij})y_{ij}\}=E[\sum_{i=1}^{K}\sum_{j=1}^{n_i}\{1-(\delta_{ij}/\pi_{ij})\}(x_{ij}\beta+a_i+e_{ij})]=0. \tag{3.2}$$

Due to CSNI, $E[\{1-(\delta_{ij}/\pi_{ij})\}e_{ij}]$ is zero. For $E[\sum_{i=1}^{K}\sum_{j=1}^{n_i}\{1-(\delta_{ij}/\pi_{ij})\}(x_{ij}\beta+a_i)]$ to be zero, Kim et al. (2016) forced the following two constraints,

$$\begin{aligned}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{\delta_{ij}}{\pi(x_{ij},\alpha_i;\gamma)}x_{ij} &= \sum_{i=1}^{K}\sum_{j=1}^{n_i}x_{ij}\\ \sum_{j=1}^{n_i}\frac{\delta_{ij}}{\pi(x_{ij},\alpha_i;\gamma)} &= \sum_{j=1}^{n_i}1 \qquad \forall i\in\{1,2,\cdots,K\}.\end{aligned} \tag{3.3}$$

The validity of the inverse probability weighting relies mainly on

$$E\{\sum_{i=1}^{K}\sum_{j=1}^{n_i}(\delta_{ij}/\tilde{\pi}_{ij})(y_{ij}-\mu)\}=0,$$

where $\tilde{\pi}_{ij}$ is evaluated at the maximum likelihood estimator (MLE). Therefore correct model specification of π is required for the valid inference. As for the calibration method, the validity mainly depends on (3.1). The main requirement for (3.2) is the CSNI assumption (3.3). To wit, it does not require correct specification of the functional form of the response model as long as the data are CSNI. In this sense, (2.3) is only a working model. If the goal is to estimate the marginal mean μ, the imputation or outcome model is required to be partially correct in that only the part regarding the variables, x_{ij} needs to be correct. For example, if the true model for outcome y_{ij} is $x_{ij}\beta+g(z_{ij})+a_i+e_{ij}$, where $E\{g(z)\}=0$, and δ_{ij} is independent of z_{ij} given a_i and x_{ij}, $E[\{1-(\delta_{ij}/\pi_{ij})\}g(z_{ij})]$ is zero. Outcome model can be misspecified regarding z_{ij} or even z_{ij} could be omitted to estimate μ consistently. In the regression setting, the conditional model for y should be correctly specified to estimate the conditional mean even when data are fully observed. As for the response model, we show in the next Section that the correct specification of π is not required, but only CSNI assumption is.

Under the working logistic model, (2.3), the two calibration conditions reduce to

$$\psi(\gamma)=\sum_{i=1}^{K}\sum_{j=1}^{n_i}\psi_{ij}(\gamma)=\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left\{\frac{\delta_{ij}}{\hat{\pi}_{ij}(\gamma)}-1\right\}x_{ij}=0, \tag{3.4}$$

where $\hat{\pi}_{ij}(\gamma) = \pi\{x_{ij}, \hat{\alpha}_i(\gamma); \gamma\}$ and $\exp\{\hat{\alpha}_i(\gamma)\} = \sum_{j=1}^{n_i} \delta_{ij} \exp(-x_{ij}\gamma)/(n_i - \sum_{j=1}^{n_i} \delta_{ij})$. In the next two sections, we derive calibration-assisted objective function based on (3.4).

3.2 Calibration-assisted Estimating Equation

Under model (2.1), $Var(y_i \mid x_i) \equiv V_i = \sigma^2 I_i + DJ_i$, where I_i and J_i are $n_i \times n_i$ identity matrix and matrix of 1, respectively, and $V_i^{-1} = aI_i + b_i J_i$, where $a = \sigma^{-2}$ and $b_i = -D\sigma^{-2}(\sigma^2 + n_i D)^{-1}$. The weighted sum of squares has a form

$$\sum_{i=1}^{K}(y_i - x_i\beta)^T V_i^{-1}(y_i - x_i\beta) = \sum_{i=1}^{K}\{a(y_i - x_i\beta)^T(y_i - x_i\beta) + b_i(y_i - x_i\beta)^T J_i(y_i - x_i\beta)\}.$$

When data are fully observed, consistent estimators for β, σ^2 and D can be obtained based on the following moment-based estimating equation without any distributional assumptions:

$$S(\theta) = \sum_{i=1}^{K} \begin{bmatrix} S_{1i}^T(\theta) \\ S_{2i}(\theta) \\ S_{3i}(\theta) \end{bmatrix}, \tag{3.5}$$

where

$$S_{1i}(\theta) = \sum_{j=1}^{n_i}(x_{ij} - \tau_i \bar{x}_i)(y_{ij} - x_{ij}\beta)$$

$$S_{2i}(\theta) = \sum_{j=1}^{n_i}\left\{(y_{ij} - x_{ij}\beta)^2 - \tau_i(\bar{y}_i - \bar{x}_i\beta)^2 - \sigma^2\right\}$$

$$S_{3i}(\theta) = \sum_{j=1}^{n_i}\{\tau_i(\bar{y}_i - \bar{x}_i\beta)^2 - D\}$$

and

$$\tau_i = \frac{n_i D}{\sigma^2 + n_i D}.$$

Since the expectation of the equation (3.5) equals zero, a solution to the equation $S(\theta) = 0$ is consistent under certain regularity conditions. When there are missing data and data are missing at random, naively modified estimating equation using observed records alone gives a consistent estimate and is the restricted maximum likelihood estimator (REML). However, under CNSI, the estimating function does not have mean zero and the REML is biased. For example, the expectation of the estimating function for β,

$$\begin{aligned} & E\left[\sum_{j=1}^{n_i} \delta_{ij}(x_{ij} - \tau_i \bar{x}_i)\{E(y_{ij} \mid x_{ij}, a_i, \delta_{ij}) - x_{ij}\beta\}\right] \\ = \; & E\{\sum_{j=1}^{n_i} \delta_{ij}(x_{ij} - \tau_i \bar{x}_i) a_i\} \neq 0. \end{aligned}$$

The last inequality is because δ_{ij} depends on a_i given x_{ij}.

Our strategy is to find estimating function $U(\eta)$ that satisfies

$$E\{S(\theta) - U(\eta)\} = 0 \tag{3.6}$$

under constraints (3.3), where $\eta = (\theta^T, \gamma^T)^T$. The estimating function (3.5) has components including missing data, $\sum_{j=1}^{n_i} x_{ij}(y_{ij} - x_{ij}\beta)$, $\bar{x}_i \sum_{j=1}^{n_i}(y_{ik} - x_{ik}\beta)$, $\sum_{j=1}^{n_i}(y_{ij} - x_{ij}\beta)^2$, and $\left\{\sum_{j=1}^{n_i}(y_{ij} - x_{ij}\beta)\right\}^2$. Under constraints (3.3), we can verify that

$$E\left\{\bar{x}_i \sum_{j=1}^{n_i}(y_{ij} - x_{ij}\beta)\right\} = E\left\{\bar{x}_i \sum_{j=1}^{n_i} \frac{\delta_{ij}}{\hat{\pi}_{ij}(\gamma)}(y_{ij} - x_{ij}\beta)\right\} \tag{3.7}$$

and

$$E\left\{\sum_{j=1}^{n_i}(y_{ij} - x_{ij}\beta)^2\right\} = E\left\{\sum_{j=1}^{n_i} \frac{\delta_{ij}}{\hat{\pi}_{ij}(\gamma)}(y_{ij} - x_{ij}\beta)^2\right\}. \tag{3.8}$$

For $\sum_{j=1}^{n_i} x_{ij}(y_{ij} - x_{ij}\beta)$ and $\left\{\sum_{j=1}^{n_i}(y_{ij} - x_{ij}\beta)\right\}^2$, similar identities do not hold. Instead, we have the following lemma. Sketch of a proof is given in Supplementary material.

Lemma 1.

$$E\left\{\sum_{j=1}^{n_i} x_{ij}(y_{ij} - x_{ij}\beta)\right\} = E\left\{\sum_{j=1}^{n_i} \frac{\delta_{ij}}{\hat{\pi}_{ij}}(x_{ij} - \tilde{x}_i)(y_{ij} - x_{ij}\beta)\right\}, \tag{3.9}$$

where $\tilde{x}_i = {n_i}^{-1} \sum_{j=1}^{n_i} x_{ij} \{\delta_{ij}/\hat{\pi}_{ij}(\gamma) - 1\}$, *and*

$$E\left[\left\{\sum_{j=1}^{n_i}(y_{ij} - x_{ij}\beta)\right\}^2\right] = E\left[\left\{\sum_{j=1}^{n_i} \frac{\delta_{ij}}{\hat{\pi}_{ij}(\gamma)}(y_{ij} - x_{ij}\beta)\right\}^2 - C_i(\eta)\right], \tag{3.10}$$

where $C_i(\eta) = \sum_{j=1}^{n_i} \left\{\delta_{ij}/\hat{\pi}_{ij}^2(\gamma) - 1\right\} \sigma^2$.

Using (3.7), (3.8), (3.9), and (3.10), we can construct calibration-assisted estimating equation $U(\eta)$ that satisfies (3.6) as follows:

$$U(\eta)=\sum_{i=1}^{K}\sum_{j=1}^{n_i}\{U_{1ij}(\eta),U_{2ij}(\eta),U_{3ij}(\eta)\}^T, \tag{3.11}$$

where

$$\begin{aligned}
U_{1ij}(\eta) &= \frac{\delta_{ij}}{\hat{\pi}_{ij}(\gamma)}(x_{ij}-\tilde{x}_i)(y_{ij}-x_{ij}\beta)-\bar{x}_i\tau_i\frac{\delta_{ij}}{\hat{\pi}_{ij}(\gamma)}(y_{ij}-x_{ij}\beta)\\
U_{2ij}(\eta) &= \frac{\delta_{ij}}{\hat{\pi}_{ij}(\gamma)}(y_{ij}-x_{ij}\beta)^2-\tau_i n_i^{-2}\xi_i(\eta)-\sigma^2\\
U_{3ij}(\eta) &= \tau_i n_i^{-2}\xi_i(\eta)-D,
\end{aligned}$$

with

$$\xi_i(\eta)=\left\{\sum_{j=1}^{n_i}\frac{\delta_{ij}}{\hat{\pi}_{ij}(\gamma)}(y_{ij}-x_{ij}\beta)\right\}^2-C_i(\eta),$$

and $\tilde{x}_i$ and $C_i(\eta)$ are defined in (3.9) and (3.10), respectively.

Let $\Psi(\eta)=\sum_{i=1}^{K}\Psi_i(\eta)=\sum_{i=1}^{K}\{U_i(\eta),\psi_i(\gamma)\}^T$ where $U_i(\eta)=\sum_{j=1}^{n_i}U_{ij}(\eta)$ and $\psi_i(\gamma)=\sum_{j=1}^{n_i}\psi_{ij}(\gamma)$. Let $\hat{\eta}$ be the solution of $\Psi(\eta)=0$. Computation can be carried out by using Newton-Raphson algorithm from the calibration-assisted estimating equation. The method can be applied when covariates are either continuous, categorical or mixed of the two kinds. The proposed method does not require numerical integration as in some of previous methods. Furthermore, it does not require π to be correctly specified but only require CSNI to hold. Consistency and asymptotic normality of the calibrated parameter estimator $\hat{\eta}$ can be obtained mainly due to the equation (3.6). We define η^* which satisfies $E\{\Psi(\eta^*)\}=0$. Under CSNI and (2.1), $\eta^*=(\theta_0^T,\gamma^{*T})^T$, where θ_0 is the true parameter, and γ^* satisfies $E\{\psi(\gamma)\}=0$. Then by Taylor's expansion we have

$$K^{\frac{1}{2}}(\hat{\eta}-\eta^*)=K^{-\frac{1}{2}}\sum_{i=1}^{K}i(\eta^*)^{-1}\Psi_i(\eta^*)+o_p(1),$$

where $N=\sum_{i=1}^{K}n_i$ and

$$i(\eta)=E\left\{-\frac{1}{K}\frac{\partial\Psi(\eta)}{\partial\eta}\right\}.$$

Under regularity conditions, $\Psi_i(\eta^*) = \{U_{1i}(\eta^*), U_{2i}(\eta^*), U_{3i}(\eta^*), \psi_i(\gamma^*)\}^T$'s are independently distributed as normal with mean zero. This leads us that $K^{1/2}(\hat\eta - \eta^*)$ is asymptotically normal with mean zero and variance, $V_1 \equiv K^{-1}\sum_{i=1}^K E[\{i(\eta^*)^{-1}\Psi_i(\eta^*)\}^{\otimes 2}]$, which can be consistently estimated by $K^{-1}\sum_{i=1}^K \{\hat{i}(\hat\eta)^{-1}\Psi_i(\hat\eta)\}^{\otimes 2}$, where $\hat{i}(\eta) = -K^{-1}\sum_{i=1}^K \partial\Psi_i(\eta)/\partial\eta$. Finally, we have the following theorem.

Theorem 1. *Suppose that $\hat\eta$ is the solution of $\Psi(\eta) = 0$. Assume that $\{n_1, \cdots, n_K\}$ satisfies*

$$\frac{K^{-1}\sum_{i=1}^K n_i^2}{(K^{-1}\sum_{i=1}^K n_i)^2} = O(1) \tag{3.12}$$

and

$$\frac{\sum_{i=1}^K n_i^{2+\delta}}{(\sum_{i=1}^K n_i^2)^{(2+\delta)/2}} = o(1) \tag{3.13}$$

for some $\delta > 0$, as $K \to \infty$. Under some regularity conditions, $K^{1/2}(\hat\eta - \eta^*)$ is asymptotically normally distributed with mean zero and variance V_1 as $K \to \infty$, which can be consistently estimated by the sandwich variance $K^{-1}\sum_{i=1}^K \{\hat{i}(\hat\eta)^{-1}\Psi_i(\hat\eta)\}^{\otimes 2}$, where $B^{\otimes 2} = BB^T$.

Condition (3.12) roughly states that $\max_{1\le i\le K} n_i = O(K^{-1/2}N)$, where $N = \sum_{i=1}^K n_i$. Condition (3.13) is essentially a Liapounov condition for the central limit theorem. It means that no single n_i dominates the others in the asymptotic sense. It is noteworthy that the result holds when we do not assume a full distributional assumption of e_{ij}.

3.3 Likelihood method with EM algorithm

Now we consider the case of full distributional assumption with likelihood given by (2.4) when $f(\cdot \mid \cdot)$ and $\phi(\cdot)$ are normal. When there are no missing data in y, the EM algorithm treats (y, a) as full data, (y) as observed data, and (a) as missing data. The M-step is to solve $W(\theta) = 0$, where

$$W(\theta) = \begin{bmatrix} \sum_{i=1}^K \{\sum_{j=1}^{n_i} x_{ij}(y_{ij} - x_{ij}\beta) - \sum_{j=1}^{n_i} x_{ij}E(a_i \mid x_i, y_i)\}^T \\ \sum_{i=1}^K \sum_{j=1}^{n_i} E\{(y_{ij} - x_{ij}\beta - a_i)^2 - \sigma^2 \mid x_i, y_i\} \\ \sum_{i=1}^K E(a_i^2 - D \mid x_i, y_i) \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^K W_{1i}(\eta) \\ \sum_{i=1}^K W_{2i}(\eta) \\ \sum_{i=1}^K W_{3i}(\eta) \end{bmatrix}. \tag{3.14}$$

When $n_i = n$ for all i, W is equivalent to S. When n_i varies across clusters, (W_{2i}, W_{3i}) are different from (S_{2i}, S_{3i}). When data are missing, $E(a_i \mid x_i, y_i)$ and $E(a_i^2 \mid x_i, y_i)$ contain missing data and cannot be evaluated. When data are missing at random, the E-step is to evaluate $E(a_i \mid x_i, y_{i,obs})$, which has a closed form, but when data are CSNI missing, $E(a_i \mid x_i, y_{i,obs}, \delta_i)$ and $E(y_{ij} \mid x_i, y_{i,obs}, \delta_i)$ should be evaluated according to (2.5) and (2.6). Instead of evaluating these quantities, our strategy is to replace $E(a_i \mid x_i, y_i)$ and $E(a_i^2 \mid x_i, y_i)$ with their unbiased predictors. That is to modify the E-step, by imputing the unbiased predictors of $E(a_i \mid x_i, y_i)$ and $E(a_i^2 \mid x_i, y_i)$ instead of evaluating $E(a_i^p \mid x_i, y_{i,obs}, \delta_i)$ and $E(y_{ij}^p \mid x_i, y_{i,obs}, \delta_i)$, $p = 1, 2$. This avoids numerical integration whose lack of accuracy sometimes leads to computational instability. Note that $E(a_i \mid x_i, y_i) = D\mathbf{1}_i^T V_i^{-1}(y_i - \mu_i)$, and let

$$\tilde{E}(a_i|x_i, y_i, \delta_i) = D\mathbf{1}_i^T V_i^{-1}\Delta_i(y_i - x_i\beta) = \tau_i n_i^{-1}\sum_{k=1}^{n_i}\frac{\delta_{ik}}{\hat{\pi}_{ik}(\gamma)}(y_{ik} - x_{ik}\beta),$$

where Δ_i is a diagonal matrix with the j^{th} element $\delta_{ij}/\hat{\pi}_{ij}(\gamma)$. Using (3.7) through (3.10), we can find

$$E\{\sum_{j=1}^{n_i} x_{ij}E(a_i \mid x_i, y_i)\} = E\{\sum_{j=1}^{n_i} x_{ij}\tilde{E}(a_i \mid x_i, y_{i,obs}, \delta_i)\},$$

and an unbiased predictor of $E(a_i^2 \mid x_i, y_i)$. After replacing them in (3.14), the resulting M-step with the modified E-step provides the following equations:

$$Q(\eta) = \sum_{i=1}^{K}\{Q_i(\eta)\}^T = \sum_{i=1}^{K}\{Q_{1i}(\eta), Q_{2i}(\eta), Q_{3i}(\eta)\}^T,$$

where

$$Q_{1i}(\eta) = \sum_{j=1}^{n_i}(x_{ij} - \tilde{x}_i)\frac{\delta_{ij}}{\hat{\pi}_{ij}(\gamma)}(y_{ij} - x_{ij}\beta) - \sum_{j=1}^{n_i} x_{ij}\tilde{E}(a_i \mid x_i, y_i, \delta_i),$$

$$Q_{2i}(\eta) = \sum_{j=1}^{n_i} \frac{\delta_{ij}}{\hat{\pi}_{ij}(\gamma)}(y_{ij} - x_{ij}\beta)^2 - (2\tau_i - \tau_i^2)n_i^{-1}\xi_i(\eta) - n_i\sigma^2(1 - n_i^{-1}\tau_i),$$

$$Q_{3i}(\eta) = \tau_i \sum_{j=1}^{n_i} \left\{\tau_i n_i^{-2}\xi_i(\eta) - D\right\},$$

$$\xi_i(\eta) = \left\{\sum_{j=1}^{n_i} \frac{\delta_{ij}}{\hat{\pi}_{ij}(\gamma)}(y_{ij} - x_{ij}\beta)\right\}^2 - C_i(\eta),$$

and $\tilde{x}_i$, $\xi(\eta)$ and $C_i(\eta)$ are defined in Section 3.2. Let $\Xi(\eta) = \sum_{i=1}^K \Xi_i(\eta) = \sum_{i=1}^K \{Q_i(\eta), \psi_i(\gamma)\}^T$, $\tilde{\eta}$ be the solution of $\Xi(\eta) = 0$, $C_K \equiv K^{-1}\sum_{i=1}^K E[\{i^*(\tilde{\eta})^{-1} \Xi_i(\tilde{\eta})\}^{\otimes 2}]$, and $i^*(\eta) = E\{-K^{-1}\partial\Xi(\eta)/\partial\eta\}$. The asymptotic distribution of $K^{1/2}(\tilde{\eta} - \eta^*)$ can be obtained similarly to Theorem 1.

Theorem 2. *Suppose that $\tilde{\eta}$ is the solution of $\Xi(\eta) = 0$. Under the conditions in Theorem 1, $K^{1/2}(\tilde{\eta} - \eta^*)$ is asymptotically normally distributed with mean zero and variance V_2 as $K \to \infty$, which can be consistently estimated by the sandwich variance formula $K^{-1}\sum_{i=1}^K \{\hat{i}^*(\tilde{\eta})^{-1}\Xi_i(\tilde{\eta})\}^{\otimes 2}$, where $\hat{i}^*(\eta) = -K^{-1}\sum_{i=1}^K \partial\Xi_i(\eta)/\partial\eta$.*

4. Simulation studies

We conducted simulation studies to evaluate finite sample performance of the proposed estimator. We set the outcome model as $y_{ij} = 0.25 + 0.5x_{ij} + a_i + e_{ij}$ and the response model as $h\{P(\delta_{ij} = 1 \mid x_{ij}, a_i)\} = \gamma_0 + 0.6a_i + x_{ij})$, where $h(\cdot)$ being inverse of the logistic or complementary log-log link function and $\gamma_0 = 0.4$, or 1.0 for logistic or complementary log-log link function, respectively. Complementary log-log function is used to evaluate the effect of misspecified response model. We generated x_{ij} from $U(-0.5, 0.5)$, and e_{ij} and a_i from standard normal distribution. The number of cluster K was 400 or 200 and the maximum number of clusters, M, 20 or 10. Overall response probability was 71.4% or 74.4% for logistic or complementary log-log link function, respectively. We compared four estimators, (i) REML using full data, (ii) REML using observed data, (iii) the proposed estimator presented in Section 3.2, and (iv) the proposed estimator from the likelihood method with the modified E-step presented in

Section 3.3. We present bias, simulation mean squared error, and coverage probability based on 1,000 Monte Carlo replications. The REML based on the observed data is valid when data are missing at random.

Table 1 shows the results under CSNI missingness when $n_i = n$ for all i. Since $n_i = n$, the two proposed estimators are identical and we report results for the three estimators. All the estimators except the ones using observed data had negligible bias and nominal coverage probabilities as anticipated. The REML based on the observed records only had non-negligible bias in D, the variance component of random effects, and coverage probabilities were significantly different from the nominal value. The proposed estimator had negligible bias and coverage probabilities close to nominal value, which remained true when the true underlying response model was different from the working model.

Table 2 features results when n_i are generated from a binomial distribution. Since n_i varies across clusters, the two proposed estimators are not identical, and we report results for the four estimators. As in the Table 1, the proposed estimators had negligible bias and coverage probabilities close to nominal value even when the true underlying response model was different from the working model. The two proposed estimators showed similar performance, but the simulation variances of D from the likelihood based estimates in Section 3.3 was slightly smaller than those based on the estimating equation described in Section 3.2, especially when n is small. Interestingly, all the estimators exhibited negligible bias and coverage probabilities close to the nominal value for the variance of error term σ^2.

Table 3 shows the results when a_i is distributed as a Gaussian mixture distribution, the distribution function F given by $F(x) = \sum_{i=1}^{2} w_i P_i(x)$, where $w_1 = 1/3$, $w_2 = 2/3$ and $P_1(x)$, $P_2(x)$ are a univariate normal distribution function with mean $-10/3$ and $5/3$, and variance, 1. Both proposed methods produced estimators with negligible bias. This result is anticipated for the method proposed in Section (3.2) since it does not depend on normality of the random effects. The method proposed in Section (3.3) does depend on the normality, but the results were robust when normality of random effects was violated. The variance estimate for σ^2 depends on the assumption of the fourth moment, but the bias of the variance estimate seems small exhibiting coverage probabilities close to nominal.

Table 4 exhibits results when the covariates contained both continuous and discrete components. We set the outcome model as $y_{ij} = 0.25 + 0.25x_{1ij} + 0.25x_{2ij} + a_i + e_{ij}$ and the response model as $h\{P(\delta_{ij} = 1 \mid x_{1ij}, x_{2ij}, a_i)\} = \gamma_0 + 0.6a_i + 0.5x_{1ij} + 0.5x_{2ij}$, where $h(\cdot)$ being inverse of the logistic or complementary log-log link function and $\gamma_0 = 0.4$, or 1.0 for logistic or complementary log-log link function, respectively. We generated x_{1ij} from $U(-0.5, 0.5)$ and x_{2ij} from two supporting points $\{-0.5, 0.5\}$ and e_{ij} and a_i from standard normal distribution. The results show that the proposed estimators have negligible bias and coverage probabilities close to nominal when covariates are mixed with continuous and discrete.

5. The 2006 state inpatient database

As total health care spending in the United States soared to 17% of GDP, the cost of unscheduled rehospitalization within 1 month from previous discharge is a major healthcare problem, and identifying factors related to the cost of rehospitalization could be of great interest to policy makers (Kim et al., 2015). Kim et al. (2015) described the inpatient database in the state of California in year 2006, which is a part of the family of databases and software tools developed for the Healthcare Cost and Utilization Project. The state inpatient database includes inpatient discharge records with various demographic, socioeconomic, clinical variables. The subjects are patients aged 50 or older who were discharged alive from acute care hospitals between April and September during the year and who experienced unscheduled rehospitalizations within 30 days. Details on the data are available from the website (URL: https://www.hcup-us.ahrq.gov), Kim et al. (2015), and Kim et al. (2016).

In the database, 59,566 subjects are nested in 353 hospitals and the cluster size n_i varies from 1 to 930 ($\sum_{i=1}^{K} n_i = 59,566, K = 353$). The outcome of the analysis is the cost incurred from the rehospitalization in U.S. dollars($). The number of patients with observed outcome variable was 51,396, yielding overall missing rate of 13.8% and the missing proportions across the hospital levels ranged from 0% to 98.3%. Moreover, 327 over 353 hospitals had missing proportions less than 5% or greater than 95%. Figure 1 shows the heatmaps of mean of the log-transformed rehospitalization care cost in U.S. dollars($) and its missing rates according to counties of state Califormia.

We treated each hospital as a cluster and patients as analysis unit. We set the log-transformed rehospitalization care cost in U.S. dollars ($) as y, and Sex, Race, Age, Income status, and Insurance status as covariates, x. We fitted the response model using x and random effects in the logistic model. The likelihood ratio test at the boundary of parameter space for the variance component of the random effect being zero was significant, suggesting that data may not be missing at random. We assumed the linear mixed effect model (2.1) and used the working response model (2.3). We

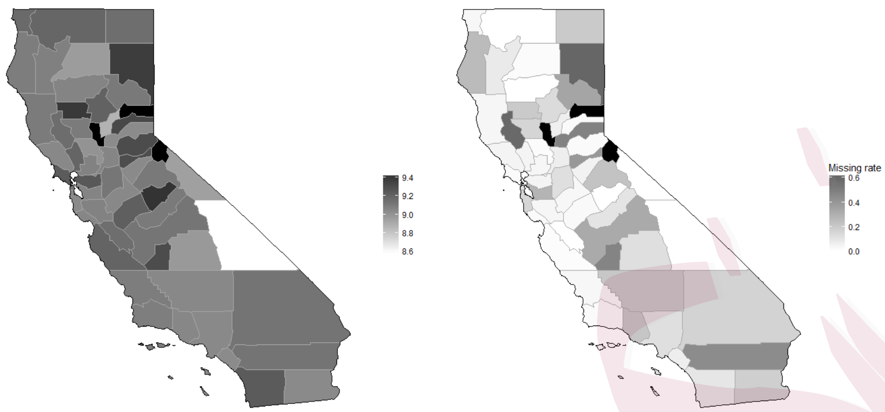

Figure 1: A heatmap of (Left) mean of the log-transformed rehospitalization care cost in U.S. dollars($) and (right) missing rate of the cost according to county. In the present data, any information was not recorded for Alpine, Sierra, and Sutter counties.

tried to fit the model under CSNI by maximizing the marginal likelihood (2.4) via Laplace approximation and the algorithm did not converge.

Table 5 shows results for analysis assuming missing at random and the proposed method under CSNI missingness. The proposed method under the assumption of CSNI changed the significance status of the factors such as low income and age, and the estimate for Black race. A careful examination of different models would be demanded to make recommendation for policy changes, and less computational burden can be a definite advantage in exploring various models.

6. Summary and discussion

In this study, we proposed a new approach to handle CSNI missingness in the context of linear mixed effects models using inverse probability weighting and calibration technique. The proposed method provides a consistent estimator with a weaker set of assumptions and simpler computation than previous works. This work can be extended to the case where conditional independence of e_{ij} is violated and the variance of y_i is not of compound symmetry form. The extension involves different calibration equation incorporating elements of inverse of the marginal variance. An extension of the proposed method to generalized linear mixed effects models is not obvious and calls for future research.

Supplementary Materials

In the supplementary material, we include the proof of the Lemma 1, equation (3.7) and (3.8), and Theorem 1.

Acknowledgements

Y. Kwon and M. C. Paik were supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (MSIP) (No. 2013R1A2A2A01067262).

References

Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973.

Follmann, D. and Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics*, 51(1):pp. 151–168.

Gao, S. (2004). A shared random effect parameter approach for longitudinal dementia data with non-ignorable missing data. *Statistics in Medicine*, 23(2):211–219.

Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *The Annals of Statistics*, 9(4):861–869.

Han, P. (2014). Multiply robust estimation in regression analysis with missing data. *Journal of the American Statistical Association*, 109(507):1159–1173.

Ibrahim, J. G., Chen, M.-H., and Lipsitz, S. R. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika*, 88(2):pp. 551–564.

Kang, J. D. Y. and Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4):523–539.

Kim, G.-S., Paik, M. C., and Kim, H. (2016). Causal inference with observational data under cluster-specific non-ignorable assignment mechanism. *Computational Statistics and Data Analysis*, http://dx.doi.org/10.1016/j.csda.2016.10.002.

Kim, H., Hung, W. W., Paik, M. C., Ross, J. S., Zhao, Z., Kim, G.-S., and Boockvar, K. (2015). Predictors and outcomes of unplanned readmission to a different hospital. *International Journal for Quality in Health Care*, 27(6):513–519.

Kim, J.-K., Kwon, Y., and Paik, M. C. (2016). Calibrated propensity score method for survey nonresponse in cluster sampling. *Biometrika*, 103(2):461–473.

Kim, J.-K. and Shao, J. (2013). *Statistical Methods for Handling Incomplete Data*. Chapman & Hall / CRC.

Molenberghs, G., Beunckens, C., Sotto, C., and Kenward, M. G. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(2):371–388.

Paik, M. C. (1997). The generalized estimating equation approach when data are not missing completely at random. *Journal of the American Statistical Association*, 92(440):1320–1329.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429):106–121.

Rotnitzky, A. and Robins, J. M. (1995). Semiparametric regression estimation in the presence of dependent censoring. *Biometrika*, 82(4):805–820.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–90.

Shao, J. and Zhang, J. (2015) A transformation approach in linear mixed-effects models with informative missing responses. *Biometrika*, 102(1):107–119.

Shao, J. and Wang, L. (2016). Semiparametric inverse propensity weighting for nonignorable missing data. *Biometrika*, 103(1):175–187.

Skinner, C. J. and D'Arrigo, J. (2011). Inverse probability weighting for clustered nonresponse. *Biometrika*, 98(4):953–966.

Wang, S., Shao, J., and Kim, J.-K. (2014). An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Statistica Sinica*, 24(3):1097–1116.

Yang, S., Kim, J.-K., and Zhu, Z. (2013). Parametric fractional imputation for mixed models with nonignorable missing data. *Statistics and Its Interface*, 6(3):339–347.

Yuan, Y. and Little, R. J. A. (2007). Model-based estimates of the finite population mean for two-stage cluster samples with unit non-response. *Journal of the Royal Statistical Society: Series C*, 56(1):79–97.

Zhang, H. and Paik, M. C. (2009). Handling missing responses in generalized linear mixed model without specifying missing mechanism. *Journal of Biopharmaceutical Statistics*, 19(6):1001–1017.

Yongchan Kwon, Myunghee Cho Paik

Department of Statistics, Seoul National University, Seoul 151-742, Korea

E-mail: ykwon0407@snu.ac.kr, myungheechopaik@snu.ac.kr

Jae Kwang Kim

Department of Statistics, Iowa State University, Ames, Iowa 50011, U.S.A.

E-mail: jkim@iastate.edu

Hongsoo Kim

Graduate School of Public Health, Seoul National University, Seoul 151-742, Korea

E-mail: hk65@snu.ac.kr

Table 1: Three estimators with their bias, mean squared error, and coverage probability based on 1,000 Monte Carlo samples under CSNI when the true response model function is logistic or complmentary log-log. The number of cluster and cluster sizes are in parentheses.

	Bias($\times10^2$)			MSE($\times10^3$)			CP		
	FUL	COM	EE	FUL	COM	EE	FUL	COM	EE
LOG(400, 20)									
INT	0.02	1.48	0.06	2.64	2.92	2.74	0.941	0.934	0.944
β	0.05	-0.80	0.05	1.47	2.18	2.26	0.955	0.947	0.963
D	-0.68	-4.86	-0.74	5.82	7.97	6.03	0.930	0.871	0.935
σ^2	-0.04	0.17	-0.06	0.25	0.39	0.40	0.953	0.944	0.950
LOG(400, 10)									
INT	0.13	2.93	0.11	2.60	3.56	2.82	0.953	0.923	0.954
β	0.16	-1.49	0.16	3.33	5.18	5.48	0.944	0.945	0.944
D	-0.51	-5.14	-0.54	5.96	8.56	6.65	0.943	0.878	0.942
σ^2	0.04	0.47	0.03	0.55	0.87	0.93	0.942	0.949	0.944
LOG(200, 20)									
INT	-0.35	1.14	-0.29	5.08	5.30	5.26	0.949	0.944	0.948
β	0.09	-0.55	0.29	3.00	4.50	4.92	0.956	0.951	0.948
D	-0.46	-4.75	-0.59	11.46	13.63	12.14	0.929	0.890	0.928
σ^2	-0.05	0.17	-0.07	0.51	0.77	0.80	0.946	0.947	0.945
LOG(200, 10)									
INT	0.27	3.13	0.30	5.67	6.72	5.90	0.943	0.938	0.940
β	-0.14	-1.67	0.05	6.46	9.76	10.48	0.957	0.949	0.952
D	-0.66	-5.42	-0.85	11.37	15.76	13.55	0.944	0.885	0.933
σ^2	-0.05	0.31	-0.15	1.15	1.83	1.91	0.940	0.944	0.941
CLL(400, 20)									
INT	-0.12	2.08	-0.14	2.66	3.08	2.73	0.945	0.930	0.948
β	-0.21	-1.67	-0.29	1.53	2.36	2.35	0.957	0.938	0.959
D	0.10	-6.64	0.12	5.31	9.68	5.88	0.954	0.806	0.953
σ^2	-0.00	0.33	-0.03	0.25	0.36	0.39	0.953	0.945	0.937
CLL(400, 10)									
INT	0.17	4.42	0.13	2.75	4.75	2.96	0.943	0.865	0.945
β	0.02	-2.73	-0.07	3.77	5.90	5.83	0.936	0.913	0.932
D	-0.11	-7.77	-0.22	6.32	12.40	7.78	0.945	0.788	0.940
σ^2	-0.10	0.61	-0.06	0.55	0.85	0.90	0.947	0.938	0.941
CLL(200, 20)									
INT	-0.33	1.88	-0.36	5.57	5.76	5.61	0.937	0.938	0.947
β	0.14	-1.29	0.10	3.24	4.58	4.88	0.938	0.942	0.955
D	-0.63	-7.32	-0.56	10.91	15.86	11.65	0.919	0.834	0.936
σ^2	-0.08	0.32	-0.02	0.50	0.71	0.82	0.955	0.956	0.945
CLL(200, 10)									
INT	0.22	4.57	0.25	5.38	7.49	5.67	0.950	0.922	0.949
β	-0.13	-2.75	-0.06	6.92	10.73	10.93	0.951	0.943	0.945
D	-0.49	-8.26	-0.50	11.92	18.19	13.84	0.936	0.845	0.949
σ^2	-0.08	0.61	-0.17	1.12	1.74	1.79	0.945	0.940	0.952

MSE, Mean squared error; CP, Coverage probability; FUL, Full; COM, Complete; EE, Estimating equation; EM, Expectation-Maximization; LOG, Logistic; CLL, Complementary log-log; INT, Intercept.

Table 2: Four estimators with their bias, mean squared error, and coverage probability with varying n_i based on 1,000 Monte Carlo samples under CSNI when the true response model function is logistic or complmentary log-log. The number of cluster and cluster sizes are in parentheses.

	Bias($\times 10^2$)				MSE($\times 10^3$)				CP			
	FUL	COM	EE	EM	FUL	COM	EE	EM	FUL	COM	EE	EM
LOG(400, 20)												
INT	0.18	1.72	0.15	0.15	2.60	2.92	2.68	2.68	0.945	0.932	0.945	0.945
β	0.03	-0.74	0.12	0.12	1.68	2.40	2.49	2.49	0.954	0.962	0.957	0.957
D	-0.34	-4.38	-0.30	-0.28	5.61	7.49	5.95	5.92	0.932	0.870	0.935	0.940
σ^2	-0.07	0.23	-0.06	-0.06	0.30	0.46	0.47	0.47	0.952	0.938	0.936	0.936
LOG(400, 10)												
INT	0.27	3.43	0.31	0.31	2.94	4.22	3.16	3.16	0.942	0.889	0.940	0.940
β	0.41	-1.16	0.57	0.57	3.58	5.33	5.68	5.68	0.952	0.953	0.952	0.952
D	-0.58	-5.25	-0.58	-0.53	5.41	8.61	6.46	6.48	0.963	0.886	0.955	0.955
σ^2	0.06	0.56	0.05	0.05	0.64	1.02	1.06	1.06	0.941	0.940	0.931	0.931
LOG(200, 20)												
INT	-0.15	1.49	-0.10	-0.10	4.92	5.13	4.97	4.97	0.950	0.945	0.952	0.952
β	0.14	-0.61	0.32	0.32	3.51	5.34	5.68	5.68	0.949	0.945	0.947	0.948
D	-0.38	-4.79	-0.48	-0.45	11.26	13.58	12.28	12.28	0.942	0.885	0.930	0.936
σ^2	-0.00	0.15	-0.10	-0.10	0.62	0.88	0.92	0.92	0.936	0.948	0.949	0.949
LOG(200, 10)												
INT	0.13	3.20	0.08	0.08	5.47	6.64	5.73	5.73	0.947	0.924	0.946	0.946
β	0.46	-1.24	0.49	0.49	7.64	10.76	12.09	12.09	0.940	0.949	0.945	0.945
D	-0.64	-5.29	-0.66	-0.70	13.15	16.76	15.08	15.04	0.929	0.887	0.922	0.922
σ^2	-0.06	0.24	-0.28	-0.28	1.33	1.95	2.03	2.03	0.945	0.943	0.940	0.943
CLL(400, 20)												
INT	0.20	2.68	0.20	0.20	2.51	3.23	2.61	2.61	0.947	0.922	0.950	0.950
β	0.01	-1.55	0.07	0.07	1.93	2.89	2.89	2.89	0.937	0.928	0.945	0.945
D	-0.51	-7.27	-0.50	-0.53	5.45	10.61	5.99	5.98	0.939	0.799	0.944	0.942
σ^2	0.09	0.49	0.09	0.09	0.26	0.39	0.40	0.40	0.960	0.956	0.959	0.959
CLL(400, 10)												
INT	0.25	5.00	0.25	0.25	2.92	5.43	3.14	3.14	0.946	0.841	0.943	0.943
β	-0.08	-2.99	0.21	0.21	3.88	6.33	5.94	5.94	0.942	0.914	0.947	0.947
D	-0.58	-8.50	-0.57	-0.60	6.68	13.55	8.17	8.09	0.927	0.768	0.929	0.928
σ^2	0.05	0.86	0.07	0.08	0.63	0.95	0.95	0.95	0.956	0.945	0.948	0.952
CLL(200, 20)												
INT	0.00	2.56	0.09	0.09	5.14	5.91	5.41	5.41	0.942	0.924	0.942	0.942
β	0.37	-1.42	0.10	0.10	3.57	4.98	5.54	5.54	0.947	0.945	0.949	0.949
D	-0.37	-7.29	-0.54	-0.52	10.53	15.51	12.09	12.06	0.943	0.847	0.929	0.929
σ^2	-0.13	0.25	-0.14	-0.14	0.58	0.82	0.93	0.94	0.956	0.950	0.938	0.942
CLL(200, 10)												
INT	0.37	5.18	0.43	0.43	5.61	8.36	6.01	6.01	0.947	0.886	0.954	0.954
β	0.22	-2.81	0.28	0.28	7.42	11.24	10.72	10.72	0.944	0.947	0.960	0.960
D	-0.78	-8.97	-0.88	-0.92	12.13	20.04	14.64	14.58	0.939	0.828	0.935	0.934
σ^2	-0.03	0.83	0.03	0.03	1.23	1.94	2.11	2.10	0.953	0.948	0.934	0.936

MSE, Mean squared error; CP, Coverage probability; FUL, Full; COM, Complete; EE, Estimating equation; EM, Expectation-Maximization; LOG, Logistic; CLL, Complementary log-log; INT, Intercept.

Table 3: Four estimators with their bias, mean squared error, and coverage probability with varying n_i based on 1,000 Monte Carlo samples under CSNI when the true response model function is logistic or complmentary log-log. The number of cluster and cluster sizes are in parentheses.

	Bias($\times 10^2$)				MSE($\times 10^3$)				CP			
	FUL	COM	EE	EM	FUL	COM	EE	EM	FUL	COM	EE	EM
LOG(400, 20)												
INT	-0.09	1.86	-0.08	-0.08	3.01	3.46	3.13	3.13	0.946	0.937	0.940	0.940
β	0.13	-0.83	0.25	0.25	1.66	2.45	2.64	2.64	0.957	0.949	0.953	0.953
σ^2	-0.11	0.88	-0.04	-0.04	0.29	0.50	0.46	0.46	0.950	0.933	0.947	0.946
LOG(400, 10)												
INT	0.24	4.18	0.26	0.26	3.17	5.01	3.32	3.32	0.939	0.885	0.944	0.944
β	-0.17	-2.37	-0.26	-0.26	3.59	5.95	5.63	5.63	0.947	0.939	0.957	0.957
σ^2	-0.13	1.76	-0.14	-0.14	0.62	1.30	1.03	1.02	0.948	0.917	0.943	0.945
LOG(200, 20)												
INT	-0.61	1.30	-0.65	-0.65	6.42	6.74	6.62	6.62	0.942	0.932	0.941	0.941
β	-0.11	-1.23	-0.18	-0.18	3.65	5.36	5.74	5.74	0.949	0.947	0.939	0.939
σ^2	-0.04	0.86	-0.12	-0.12	0.54	0.89	0.86	0.86	0.956	0.949	0.957	0.958
LOG(200, 10)												
INT	-0.05	4.06	0.11	0.11	6.25	8.25	6.73	6.73	0.953	0.909	0.946	0.946
β	-0.27	-2.53	-0.41	-0.41	8.10	12.01	12.68	12.68	0.943	0.936	0.947	0.947
σ^2	0.04	1.93	0.08	0.09	1.34	2.46	2.13	2.12	0.940	0.930	0.929	0.932
CLL(400, 20)												
INT	-0.06	3.16	-0.09	-0.09	2.97	4.06	3.06	3.06	0.953	0.893	0.952	0.952
β	-0.22	-1.88	-0.00	-0.00	1.86	2.97	3.02	3.02	0.941	0.928	0.949	0.949
σ^2	-0.03	1.29	-0.09	-0.09	0.30	0.59	0.48	0.48	0.949	0.909	0.943	0.943
CLL(400, 10)												
INT	0.18	6.51	0.16	0.16	3.00	7.53	3.27	3.27	0.954	0.777	0.948	0.948
β	0.21	-3.40	0.19	0.19	3.59	6.32	5.80	5.80	0.961	0.925	0.956	0.956
σ^2	0.04	2.77	0.01	0.01	0.64	1.71	0.98	0.98	0.943	0.862	0.951	0.956
CLL(200, 20)												
INT	-0.10	3.17	-0.08	-0.08	5.96	7.00	6.04	6.04	0.951	0.930	0.946	0.946
β	-0.02	-1.79	0.02	0.02	3.42	5.19	5.70	5.70	0.951	0.942	0.944	0.944
σ^2	-0.05	1.37	-0.03	-0.03	0.55	1.00	0.91	0.91	0.954	0.929	0.947	0.947
CLL(200, 10)												
INT	0.23	6.56	0.19	0.19	6.17	10.69	6.56	6.56	0.951	0.864	0.951	0.951
β	0.45	-3.19	0.46	0.46	7.59	12.01	11.87	11.87	0.943	0.926	0.953	0.953
σ^2	-0.04	2.77	0.09	0.10	1.28	2.76	2.09	2.09	0.941	0.910	0.943	0.943

MSE, Mean squared error; CP, Coverage probability; FUL, Full; COM, Complete; EE, Estimating equation; EM, Expectation-Maximization; LOG, Logistic; CLL, Complementary log-log; INT, Intercept.

Table 4: Four estimators with their bias, mean squared error, and coverage probability with varying n_i based on 1,000 Monte Carlo samples under CSNI when the true response model function is logistic or complmentary log-log. The number of cluster and cluster sizes are in parentheses.

	Bias($\times 10^2$)				MSE($\times 10^3$)				CP			
	FUL	COM	EE	EM	FUL	COM	EE	EM	FUL	COM	EE	EM
LOG(400, 20)												
INT	-0.07	1.47	-0.09	-0.09	2.55	2.80	2.64	2.64	0.952	0.938	0.947	0.947
β_1	-0.01	-0.55	-0.03	-0.03	1.78	2.59	2.70	2.70	0.953	0.948	0.952	0.952
β_2	-0.00	-0.27	0.19	0.19	0.54	0.84	0.90	0.90	0.950	0.941	0.946	0.946
D	0.08	-1.68	0.06	0.08	5.50	6.01	5.95	5.99	0.959	0.939	0.947	0.950
σ^2	-0.09	-0.03	-0.12	-0.12	0.29	0.39	0.43	0.43	0.947	0.959	0.958	0.958
LOG(400, 10)												
INT	0.03	3.16	0.09	0.09	2.70	3.79	2.87	2.87	0.945	0.904	0.945	0.945
β_1	-0.25	-1.17	-0.42	-0.42	3.84	5.57	6.07	6.07	0.943	0.940	0.939	0.938
β_2	0.05	-0.90	-0.08	-0.08	1.20	1.80	1.96	1.96	0.953	0.953	0.953	0.953
D	-0.32	-3.37	-0.34	-0.34	6.20	7.83	7.35	7.29	0.938	0.912	0.945	0.945
σ^2	-0.08	0.06	-0.14	-0.14	0.63	0.88	0.94	0.94	0.948	0.948	0.952	0.954
LOG(200, 20)												
INT	-0.16	1.46	-0.09	-0.09	5.28	5.47	5.35	5.35	0.947	0.945	0.947	0.947
β_1	0.22	-0.11	0.35	0.35	3.67	5.35	5.72	5.72	0.946	0.947	0.947	0.947
β_2	0.15	-0.40	0.05	0.05	1.23	1.75	1.86	1.86	0.938	0.938	0.939	0.940
D	-0.30	-2.18	-0.42	-0.41	10.75	11.75	11.91	11.78	0.943	0.925	0.932	0.934
σ^2	-0.16	-0.17	-0.21	-0.21	0.54	0.81	0.87	0.87	0.954	0.953	0.949	0.949
LOG(200, 10)												
INT	0.15	3.30	0.22	0.22	6.15	7.33	6.44	6.44	0.937	0.915	0.944	0.944
β_1	0.33	-0.71	0.02	0.02	7.91	10.88	11.88	11.88	0.930	0.947	0.946	0.946
β_2	-0.33	-1.42	-0.62	-0.62	2.47	3.82	3.92	3.92	0.946	0.939	0.948	0.948
D	-0.37	-3.53	-0.40	-0.43	11.83	13.88	13.79	13.66	0.929	0.910	0.925	0.930
σ^2	-0.13	0.12	-0.07	-0.07	1.31	1.92	2.09	2.09	0.938	0.948	0.933	0.938
CLL(400, 20)												
INT	0.03	2.45	0.06	0.06	2.60	3.15	2.62	2.62	0.952	0.932	0.952	0.952
β_1	-0.19	-1.05	-0.32	-0.32	1.76	2.48	2.70	2.70	0.955	0.952	0.961	0.961
β_2	-0.17	-1.04	-0.30	-0.30	0.55	0.90	0.88	0.88	0.960	0.935	0.957	0.957
D	-0.48	-3.69	-0.40	-0.43	5.69	7.04	6.32	6.29	0.940	0.896	0.945	0.944
σ^2	0.03	0.16	0.03	0.03	0.27	0.38	0.42	0.42	0.965	0.955	0.956	0.956
CLL(400, 10)												
INT	0.18	4.72	0.11	0.11	2.97	5.21	3.13	3.13	0.934	0.844	0.937	0.937
β_1	-0.19	-1.61	-0.20	-0.20	3.77	5.50	5.97	5.97	0.942	0.938	0.944	0.944
β_2	-0.02	-1.51	-0.10	-0.10	1.39	2.09	2.05	2.05	0.937	0.919	0.947	0.947
D	-0.49	-5.76	-0.31	-0.28	6.54	10.31	8.28	8.24	0.943	0.841	0.919	0.923
σ^2	-0.11	0.24	-0.11	-0.12	0.60	0.90	0.96	0.96	0.948	0.944	0.944	0.945
CLL(200, 20)												
INT	-0.31	2.12	-0.29	-0.29	5.48	5.82	5.58	5.58	0.941	0.936	0.939	0.939
β_1	-0.12	-0.76	0.04	0.04	3.79	5.31	5.88	5.88	0.943	0.942	0.943	0.943
β_2	0.16	-0.64	0.08	0.07	1.23	1.66	1.83	1.83	0.951	0.950	0.951	0.951
D	0.67	-2.69	0.58	0.60	11.36	12.08	12.53	12.44	0.946	0.913	0.942	0.942
σ^2	-0.08	0.03	-0.06	-0.06	0.64	0.81	0.89	0.89	0.937	0.948	0.949	0.949
CLL(200, 10)												
INT	-0.14	4.44	-0.15	-0.15	5.22	7.27	5.58	5.58	0.955	0.914	0.957	0.957
β_1	-0.20	-1.55	-0.01	-0.01	7.32	10.11	11.19	11.19	0.954	0.951	0.950	0.949
β_2	0.05	-1.59	-0.23	-0.23	2.54	3.88	4.00	4.00	0.947	0.940	0.940	0.940
D	-1.21	-6.50	-1.19	-1.23	12.18	16.96	15.28	15.17	0.933	0.881	0.924	0.930
σ^2	-0.20	0.07	-0.37	-0.36	1.24	1.83	1.98	1.98	0.948	0.954	0.948	0.950

MSE, Mean squared error; CP, Coverage probability; FUL, Full; COM, Complete; EE, Estimating equation; EM, Expectation-Maximization; LOG, Logistic; CLL, Complementary log-log; INT, Intercept.

Table 5: Factors predicting rehospitalization cost (logarithm in US dollar) with estimates, standard errors and p-value using 2006 California inpatient database.

	Estimates			Standard error			p-value		
	COMP	EE	EM	COMP	EE	EM	COMP	EE	EM
Intercept	9.121	9.145	9.144	0.0207	0.0327	0.0322	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
Sex									
Male	.	.	.	.	.	.	.	.	.
Female	-0.050	-0.045	-0.045	0.0083	0.0127	0.0127	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
Race									
White	.	.	.	.	.	.	.	.	.
Black	0.057	-0.003	-0.004	0.0172	0.0340	0.0340	0.001	0.465	0.454
Hispanic	0.001	0.001	0.001	0.0126	0.0245	0.0245	0.489	0.476	0.476
Others	0.046	0.103	0.102	0.0162	0.0438	0.0439	0.002	0.010	0.010
Age									
50-59	.	.	.	.	.	.	.	.	.
60-69	0.047	0.015	0.015	0.0139	0.0260	0.0258	$< 10^{-3}$	0.286	0.285
70-79	0.007	-0.022	-0.022	0.0156	0.0265	0.0262	0.324	0.203	0.202
> 80	-0.079	-0.128	-0.128	0.0158	0.0332	0.0329	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
Income									
High	-0.001	0.009	0.008	0.0115	0.0249	0.0250	0.458	0.358	0.378
Medium	.	.	.	.	.	.	.	.	.
Low	-0.027	-0.012	-0.011	0.0123	0.0252	0.0253	0.013	0.313	0.328
Insurance									
Medicare	.	.	.	.	.	.	.	.	.
Medicaid	0.019	0.010	0.010	0.0161	0.0187	0.0186	0.122	0.301	0.302
Private	-0.113	-0.068	-0.067	0.0146	0.0269	0.0270	$< 10^{-3}$	0.006	0.006
Others	-0.065	-0.079	-0.079	0.0264	0.0299	0.0298	0.007	0.004	0.004

COMP, Complete; EE, Estimating equation; EM, Expectation-Maximization; Others include self-pay, no-charge, county indigent programs, charity care, etc.