Statistica Sinica **14**(2004), 361-376

TWO-STEP ESTIMATION FOR A GENERALIZED LINEAR MIXED MODEL WITH AUXILIARY COVARIATES

Jianwei Chen[1], Jianwen Cai[2] and Haibo Zhou[2]

[1]*University of Rochester and* [2]*University of North Carolina at Chapel Hill*

Abstract: Generalized linear mixed models (GLMM) are useful in a variety of applications. With surrogate covariate data, existing methods of inference for GLMM are usually computationally intensive. We propose a two-step inference procedure for GLMM with missing covariate data. It is shown that the proposed estimator is consistent and asymptotically normal with covariance matrix that can be easily estimated. Simulation studies show that the proposed method outperforms those ignoring random effects or only using the validation data. We illustrate the proposed method with a data set from an environmental epidemiology study on the maternal serum DDE level in relationship to male birth defects.

Key words and phrases: Generalized linear mixed model, male birth defects, missing covariate, random effects, surrogate variables, two-step estimation.

1. Introduction

Generalized linear mixed models (GLMM) are popular statistical models for analyzing data collected from different clusters or from longitudinal studies. They provide a flexible likelihood framework under which population characteristics can be modeled as fixed effects and individual variations can be modeled as random effects. The traditional likelihood approach to GLMM usually involves high dimensional integrations which are computationally intensive. Alternative methods for inference have been proposed. One such method is based on estimation of the random effects via maximization of the joint density of the observations and random effects (McGilchrist (1994)). The idea can be traced back to the best linear unbiased prediction (BLUP) approach of Henderson (1950) for the linear mixed model, which is reviewed in Robinson (1991). Other approaches similar in principle include the penalized quasi-likelihood approach of Breslow and Clayton (1993), the approximate Bayes method of the Stiratelli, Laird and Ware (1984), the algorithms proposed in Schall (1991), the pseudo likelihood approach in Wolfinger and Connell (1993), the hierarchical likelihood of Lee and Nelder (1996), and the simulated moments method of Jiang (1998).

Auxiliary covariate problems arise frequently in biomedical studies when the primary exposure variable is only assessed on a subset of the study subjects. Reasons for observing the exposure variable partially may involve design or practical issues. In many of these studies, however, subjects with no exposure measurements do have some auxiliary information about the exposure variable. There is an extensive literature dealing with nonparametric structure for missing data in fixed effects models. Pepe and Fleming (1991) and Carroll and Wand (1991) considered a semiparametric approach where they modeled the conditional distribution of the missing covariate nonparametrically from the validation sample. This approach has been successfully used in the failure time regression analysis settings (Zhou and Pepe (1995) and Zhou and Wang (2000)); Reilly and Pepe (1995) considered a mean-score method for dealing with the missing data problem; Robins, Rotnitzky and Zhao (1994) proposed a general class of semiparametric estimators, based on a set of inverse selection probability weighted estimating equations when the data are missing at random.

Generally, existing approaches with respect to random effects in the generalized linear model are parametric. Normality for the random effects is commonly assumed, although the hierarchical generalized linear models (Lee and Nelder (1996)) allow a broader class of parametric models for the random effects. Verbeke and Lesaffre (1996) show that misspecification of the random effects model will cause it to be badly estimated. In this paper, we propose a two-step method for a generalized linear mixed model with auxiliary covariate data. We do not assume any parametric distribution for the random effects and allow the distribution of the exposure variable conditional on the auxiliary covariates to vary across centers.

The rest of the paper is organized as follows. In Section 2, we introduce the model and propose a center-specific estimator for the regression coefficients based on an estimated likelihood function. We then construct a two-step estimator using an optimal weighted version of the first step estimators. The large sample properties of the proposed two-step estimator are given in Section 3. We propose a consistent estimator for the variance of the proposed estimator. In Section 4, we present results from simulation studies comparing the proposed estimator with alternatives. The method is demonstrated with a data set from the Collaborative Perinatal Projects in Section 5. Final remarks are given in Section 6.

2. Estimation Method

2.1. The model

Suppose there are K independent centers in a study and each center has n_k subjects, where $k = 1, \ldots, K$. The total sample size is $n = \sum_{k=1}^{K} n_k$. For a given center k, let $(X_{ki}^T, Z_{ki}^T)^T$ be the covariate vector of length p, where X_{ki} is the exposure variable that may be missing, while Z_{ki} is always observed. Let W_{ki} denote an auxiliary measurement for X_{ki}. We further assume that there is a random effect u_k in each center, where u_k, $k = 1, \ldots, K$, are independent random variables with mean zero and a finite variance σ_k^2. The conditional density of Y_{ki} given the random effects u_k and the covariates $\{X_{ki}, Z_{ki}\}$ is assumed to belong to a canonical exponential family, i.e.,

$$f(Y_{ki}|X_{ki}, Z_{ki}, W_{ki}, u_k) = \exp\{[Y_{ki}\eta_{ki} - b(\eta_{ki})]/a(\phi) + c(Y_{ki}, \phi)\}, \tag{1}$$

where $a(\cdot)$, $b(\cdot)$ and $c(\cdot, \cdot)$ are known functions, ϕ is a dispersion parameter and η_{ki} is related to the random effects by

$$\eta_{ki} = \alpha + \beta_x^T X_{ki} + \beta_z^T Z_{ki} + u_k. \tag{2}$$

Let $\theta = (\alpha, \beta_x^T, \beta_z^T)^T \equiv (\alpha, \beta^T)^T$ be the regression parameter vector to be estimated. Note that (1) and (2) imply that, given $\{X_{ki}, Z_{ki}\}$, W_{ki} provides no extra information about the regression model.

For a given center k, we assume that there is a simple random validation sample with size n_{V_k}, denoted by V_k, such that individuals belonging to V_k will have their $\{X, Z, W\}$ measured. Similarly, we let $\bar{V}_k$ denote the remaining individuals in center k, the non-validation set, and assume that individuals in $\bar{V}_k$ will only have their $\{Z, W\}$ measured. Note that $n_{\bar{V}_k} = n_k - n_{V_k}$. Hence the observed data structure for individual i in center k is $\{Y_{ki}, Z_{ki}, W_{ki}, X_{ki}\}$ if $i \in V_k$, and $\{Y_{ki}, Z_{ki}, W_{ki}\}$, if $i \in \bar{V}_k$, where $k = 1, \ldots, K$ and $i = 1, \ldots, n_k$.

2.2. Center-specific estimates

Mimicking the idea of stratified analysis in fixed effects generalized linear models for a given center, we treat the random effect u_k as an unknown parameter which needs to be estimated. Thus, the conditional density of $Y_{ki}|u_k$, for $k = 1, \ldots, K$, can be written as

$$f_{\theta^{(k)}}(Y_{ki}|X_{ki}, Z_{ki}) = \exp\{[Y_{ki}\eta_{ki}(\theta^{(k)}) - b(\eta_{ki}(\theta^{(k)}))]/a(\phi) + c(Y_{ki}, \phi)\}, \tag{3}$$

where $\eta_{ki}(\theta^{(k)}) = \alpha_{u_k} + \beta_x^T X_{ki} + \beta_z^T Z_{ki}$, with $\theta^{(k)} = (\alpha_{u_k}, \beta^T)^T$ and $\alpha_{u_k} = u_k + \alpha$.

Conditional on the random effects u_k, we can write the joint density function of $\{Y_{ki}, i = 1, \ldots, n_k\}$ as a function of $\theta^{(k)} = (\alpha_{u_k}, \beta^T)^T$,

$$L_k(\theta^{(k)}) = \prod_{i \in V_k} f_{\theta^{(k)}}(Y_{ki}|X_{ki}, Z_{ki}) \prod_{j \in \bar{V}_k} f_{\theta^{(k)}}(Y_{kj}|Z_{kj}, W_{kj}). \tag{4}$$

The contribution from a nonvalidation set member is given by

$$f_{\theta^{(k)}}(Y_{kj}|Z_{kj}, W_{kj}) = \int f_{\theta^{(k)}}(Y_{kj}|x, Z_{kj}, W_{kj}) dP(x|Z_{kj}, W_{kj}),$$

which involves an unspecified distribution function $P(X|Z_{kj}, W_{kj})$. Hence (4) is not directly computable without some assumption on $P(X|Z_{kj}, W_{kj})$.

Following Pepe and Fleming (1991) and Zhou and Wang (2000), we propose to estimate $f_{\theta^{(k)}}(Y_{kj}|Z_{kj}, W_{kj})$ empirically from observations in the validation set and then to estimate $\theta^{(k)}$ from the resulting estimated likelihood function. Specifically, we assume that $P(X_{kj}|Z_{kj}, W_{kj}) \equiv P(X_{kj}|W_{kj})$, i.e., the relationship between the auxiliary variable W_{kj} and the corresponding covariate X_{kj} does not depend on the covariate Z_{kj}. In other words, given the auxiliary information W_{kj}, Z_{kj} provides no extra information on X_{kj}. Denote by $\widehat{P}(X_{kj}|W_{kj})$ the empirical density function of $P(X_{kj}|W_{kj})$ based on observations from the validation set in the kth center. We estimate $f_{\theta^{(k)}}(Y_{kj}|Z_{kj}, W_{kj})$ for a nonvalidation sample member in the kth center as follows:

$$\widehat{f}_{\theta^{(k)}}(Y_{kj}|Z_{kj}, W_{kj}) = \frac{\sum\limits_{i\in V_k} f_{\theta^{(k)}}(Y_{kj}|X_{ki}, Z_{kj}) I_{[W_{ki}=W_{kj}]}}{\sum\limits_{i\in V_k} I_{[W_{ki}=W_{kj}]}}, \tag{5}$$

where $I_{[.]}$ is an indicator function. Note that $\widehat{f}_{\theta^{(k)}}(Y_{kj}|Z_{kj}, W_{kj})$ is an unbiased estimator for the conditional distribution $f_{\theta^{(k)}}(Y|Z, W)$. The center-specific estimator of $\theta^{(k)}$ is therefore defined as the maximizer of the following estimated likelihood function,

$$EL_k(\theta^{(k)}) = \prod_{i\in V_k} f_{\theta^{(k)}}(Y_{ki}|X_{ki}, Z_{ki}) \prod_{j\in \bar{V}_k} \widehat{f}_{\theta^{(k)}}(Y_{kj}|Z_{kj}, W_{kj}), \tag{6}$$

for $k = 1, \ldots, K$. Let $\widehat{\theta}^{(k)} = (\widehat{\alpha}_{u_k}, \widehat{\beta}^{(k)^T})^T$ be the center-specific estimator of $\theta^{(k)}$ from the kth center, where $k = 1, \ldots, K$. It can be computed by implementing the Newton-Raphson iterative procedure.

2.3. A two-step weighted estimation

Clearly, the center-specific estimates from step one depend on the random effects in each center. They are inefficient since they only use information from their corresponding centers. We propose a more efficient refined estimator $\widehat{\theta} = (\widehat{\alpha}, \widehat{\beta})^T$, a weighted average of the center-specific estimators, where

$$\widehat{\theta} = \sum_{k=1}^{K} w_k \widehat{\theta}^{(k)}, \tag{7}$$

and $\sum_{k=1}^{K} w_k = 1$. We call the $\widehat{\theta}$ a two-step estimator of θ. To accommodate different center sizes, a simple way to assign weight is to let it be proportional to the center size n_k, $k = 1, \ldots, K$, $w_k = n_k/n$ with $n = \sum_{k=1}^{K} n_k$. We propose to use optimal weights in our estimator, optimality in the sense that weights minimize the variance of the estimated parameters. Using the Lagrange principle, we solve for the optimal weights by minimizing:

$$\begin{aligned}
&M(w_1,\ldots,w_K;\lambda)\\
&= E\left[(\widehat{\theta}-E\widehat{\theta})^T(\widehat{\theta}-E\widehat{\theta})\right]-\lambda\left(1-\sum_{k=1}^{K}w_k\right)\\
&= E\left[\sum_{k=1}^{K}w_k(\widehat{\theta}^{(k)}-E\widehat{\theta}^{(k)})^T\sum_{j=1}^{K}w_j(\widehat{\theta}^{(j)}-E\widehat{\theta}^{(j)})\right]-\lambda\left(1-\sum_{k=1}^{K}w_k\right),
\end{aligned}$$

where λ is the Lagrange multiplier. After some simple calculation, the weights w_k, $k=1,\ldots,K$, can be determined by the equations

$$\frac{\partial M}{\partial w_k}=2\sum_{j=1}^{K}w_jE\left[(\widehat{\theta}^{(k)}-E\widehat{\theta}^{(k)})^T(\widehat{\theta}^{(j)}-E\widehat{\theta}^{(j)})\right]+\lambda=0.$$

Note that $(\widehat{\theta}^{(k)}-\theta)$ are independent for $k=1,\ldots,K$. By using the constraint $\sum_{k=1}^{K}w_k=1$, we derive the expressions of the weight as

$$w_k=\frac{1/E\left[(\widehat{\theta}^{(k)}-E\widehat{\theta}^{(k)})^T(\widehat{\theta}^{(k)}-E\widehat{\theta}^{(k)})\right]}{\sum_{j=1}^{K}\left(1/E\left[(\widehat{\theta}^{(j)}-E\widehat{\theta}^{(j)})^T(\widehat{\theta}^{(j)}-E\widehat{\theta}^{(j)})\right]\right)},\tag{8}$$

for $k=1,\ldots,K$. These weights correspond to the contribution of each center-specific estimator to the proposed two-step estimator. Obviously the proposed estimator has the smallest variance among the class of estimators that are linear combinations of the center-specific estimators. Formula (8) reveals the fact that when the variance of a center-specific estimator is large, i.e., when a center-specific estimator does not fit the data well, the assigned weight for that center-specific estimator will be small. We also note that the proposed estimator is more efficient than any of the first step estimators. This can be seen from the fact that if the weights are taken to be $w_i=1$ and $w_k=0$, $k\neq i$, then the refined estimator $\widehat{\theta}$ will reduce to the center-specific estimator $\widehat{\theta}^{(i)}$.

3. Asymptotic Properties

We explore the asymptotic properties of the two-step estimator and propose a consistent variance estimator. For convenience of expression, we introduce the following notation. For $k=1,\ldots,K$, let $\rho_k^v=\lim_{n_k\to\infty}n_{v_k}/n_k$,

$$I_k(\theta^{(k)})=\rho_k^vE_{\widehat{\theta}^{(k)}|u_k}\left[-\frac{d^2}{d\theta^{(k)2}}\log f_{\theta^{(k)}}(Y|X,Z)\right]$$

$$+(1-\rho_k^v)E_{\widehat{\theta}^{(k)}|u_k}\left[-\frac{d^2}{d\theta^{(k)2}}\log f_{\theta^{(k)}}(Y|Z,W)\right],$$

$$\Sigma_k(\theta^{(k)}) = \mathrm{Var}_{\widehat{\theta}^{(k)}|u_k}\left\{E\left[\frac{d}{d\theta^{(k)}}\log f_{\theta^{(k)}}(Y|Z)|X,W\right]\right\},$$

$$\Sigma_k^*(\theta^{(k)}) = I_k^{-1}(\theta^{(k)}) + \frac{(1-\rho_k^v)^2}{\rho_k^v}I_k^{-1}(\theta^{(k)})\Sigma_k(\theta^{(k)})I_k^{-1}(\theta^{(k)}),$$

$$S_{X_{ki},W_{ki}}(\theta^{(k)}) = \sum_{j\in\bar{V}_k}\left[\frac{df_{\theta^{(k)}}(Y_{kj}|X_{ki},Z_{kj})/d\theta^{(k)}}{f_{\theta^{(k)}}(Y_{kj}|Z_{kj},W_{kj})} - \frac{df_{\theta^{(k)}}(Y_{kj}|Z_{kj},W_{kj})/d\theta^{(k)}}{(f_{\theta^{(k)}}(Y_{kj}|Z_{kj},W_{kj}))^2}\right.$$
$$\left.\times f_{\theta^{(k)}}(Y_{kj}|X_{ki},Z_{kj})\right]\frac{I_{[W_{kj}=W_{ki}]}}{\sum_{l\in\bar{V}_k}I_{[W_{kl}=W_{ki}]}}.$$

To obtain asymptotic properties, we assume the following regularity conditions.

(a) Observations across centers are independent and, conditional on the unobserved random effects, individuals within a center are independent.
(b) As $K\to\infty$ and $n_k\to\infty$, $K^{1/2}\sum_{k=1}^K w_k/n_k^{1/2} = O(1)$ and $K\sum_{k=1}^K w_k^2 = O(1)$.
(c) $K\sum_{k=1}^K w_k^2\sigma_k^2 e_1 e_1^T \to V_1$ and $K\sum_{k=1}^K w_k^2\Gamma_k \to V_2$ as $K\to\infty$ with $V_1+V_2<\infty$, where $\Gamma_k = E[\mathrm{Var}_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)})]$.
(d) For some $\delta>0$, $E|\mathrm{tr}((\widehat{\theta}^{(k)} - E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)}))^{\otimes 2})|^{(1+\delta)}$ and $E|u_k|^{2(1+\delta)}$ exist, $K^{1+\delta}\sum_{k=1}^K w_k^{2(1+\delta)}E|\mathrm{tr}((\widehat{\theta}^{(k)} - E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)}))^{\otimes 2})|^{(1+\delta)}\to 0$, and $K^{1+\delta}\sum_{k=1}^K w_k^{2(1+\delta)}E|u_k|^{2(1+\delta)}\to 0$ as $K\to\infty$.
(e) For some $\delta>0$, $K^{1+\delta}\sum_{k=1}^K w_k^{2+2\delta}\sigma_k^{2(1+\delta)}\to 0$ and $K^{1+\delta}\sum_{k=1}^K w_k^{2+2\delta}\,|(\Gamma_k)_{ij}|^{1+\delta}\to 0$, as $K\to\infty$, where $(\Gamma_k)_{ij}$ denotes the (i,j) element of Γ_k.

Following the theory of Pepe and Fleming (1991), the score function from (6) can be expressed asymptotically as

$$\frac{1}{\sqrt{n_k}}\frac{d\log EL_k(\theta^{(k)})}{d\theta^{(k)}}$$
$$=\frac{1}{\sqrt{n_k}}\frac{d\log L_k(\theta^{(k)})}{d\theta^{(k)}} + \frac{1}{\sqrt{n_k}}\left\{\frac{1-\rho_k^v}{\rho_k^v}\sum_{i\in V_k}S_{X_{ki},W_{ki}}(\theta^{(k)})\right\} + O_p\left(\frac{1}{\sqrt{n_k}}\right),$$

where the first term, $d\log L_k(\theta^{(k)})/d\theta^{(k)}$, would be the score function if $f_{\theta^{(k)}}(X|Z,W)$ were known. Note that $I_k(.)$ is the conditional expected information of $\theta^{(k)}$ based on the likelihood for the observed data if the form of $f_{\theta^{(k)}}(X|Z,W)$ is completely known and $\Sigma_k(.)$ is the conditional variance induced by estimating the likelihood component for nonvalidation data within each center k. Furthermore, it follows that

$$E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)})=\theta^{(k)}+o_p\left(\frac{1}{\sqrt{n_k}}\right),\ \mathrm{Var}_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)})=\frac{1}{n_k}\Sigma_k^*(\theta^{(k)})+o_p\left(\frac{1}{n_k}\right),\quad (9)$$

uniformly for $k = 1, \ldots, K$.

Theorem 1. *Under Conditions* (a)−(d)*, as* $K \to \infty$*, the two-step estimator* $\widehat{\theta} = (\widehat{\alpha}, \widehat{\beta}^T)^T$ *is a consistent estimator of* $\theta = (\alpha, \beta^T)^T$ *and* $K^{1/2}\left(\widehat{\theta} - \theta\right)$ *has an asymptotic normal distribution* $N(0, V)$*. Here* $V = V_1 + V_2$ *with*

$$V_1 = \lim_{K\to\infty} K \sum_{k=1}^{K} w_k^2 \sigma_k^2 e_1 e_1^T \quad \textit{and} \quad V_2 = \lim_{K\to\infty} K \sum_{k=1}^{K} w_k^2 E_{u_k}\left[\mathrm{Var}_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)})\right],$$

where e_1 *is a vector with* 1 *in the* 1^{st} *position and* 0 *elsewhere.*

Remark. In fact, we have the asymptotic expansions

$$\mathrm{Var}\left[K^{1/2}(\widehat{\alpha}-\alpha)\right] = \lim_{K\to\infty} K \sum_{k=1}^{K} w_k^2\sigma_k^2 + \lim_{K\to\infty} K \sum_{k=1}^{K} w_k^2 E_{u_k}\left(\mathrm{Var}_{\widehat{\alpha}^{(k)}|u_k}(\widehat{\alpha}^{(k)})\right),$$

$$\mathrm{Var}\left[K^{1/2}(\widehat{\beta}-\beta)\right] = \lim_{K\to\infty} K \sum_{k=1}^{K} w_k^2 E_{u_k}\left(\mathrm{Var}_{\widehat{\beta}^{(k)}|u_k}(\widehat{\beta}^{(k)})\right).$$

In particular, consider the situation that $\sigma_1^2 = \cdots = \sigma_K^2$, then the asymptotic variance of $\widehat{\alpha}$ reduces to

$$\mathrm{Var}\left[K^{1/2}(\widehat{\alpha}-\alpha)\right] = \Big(\lim_{K\to\infty} K \sum_{k=1}^{K} w_k^2\Big)\sigma^2 + \lim_{K\to\infty} K \sum_{k=1}^{K} w_k^2 E_{u_k}\left(\mathrm{Var}_{\widehat{\alpha}^{(k)}|u_k}(\widehat{\alpha}^{(k)})\right).$$

This result can be used to estimate the variance σ^2 of the random effects.

The proof of Theorem 1 is outlined in the Appendix. From the expression for the variance of the estimator, it is clear that the direct estimator $\widehat{V}_{DE}(\widehat{\beta})$ is a consistent estimator of the variance of the estimated $\widehat{\beta}$ where

$$\widehat{V}_{DE}(\widehat{\beta}) = \sum_{k=1}^{K} w_k^2 \widehat{\mathrm{Var}}_{\widehat{\beta}^{(k)}|u_k}(\widehat{\beta}^{(k)}), \quad (10)$$

$$\widehat{\mathrm{Var}}_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)}) = \frac{1}{n_k}\Big(\widehat{I}_k^{-1}(\widehat{\theta}^{(k)}) + \frac{(1-\widehat{\rho}_k^v)^2}{\widehat{\rho}_k^v}\widehat{I}_k^{-1}(\widehat{\theta}^{(k)})\widehat{\Sigma}_k(\widehat{\theta}^{(k)})\widehat{I}_k^{-1}(\widehat{\theta}^{(k)})\Big), \quad (11)$$

$\widehat{\rho}_k^v = n_k^v/n_k$, $\widehat{I}_k(\widehat{\theta}^{(k)}) = -d^2 \log EL_k(\theta^{(k)})/d\theta^{(k)^2}|_{\theta^{(k)}=\widehat{\theta}^{(k)}}$ and $\widehat{\Sigma}_k(\theta^{(k)}) = \widehat{\mathrm{Var}}\{\widehat{S}_{X_{ki},W_{ki}}(\theta^{(k)}),\ i \in V_k,\}$. Here $\widehat{\mathrm{Var}}\{\widehat{S}_{X_{ki},W_{ki}}(\theta^{(k)}), i \in V_k,\}$ is the sample variance-covariance matrix of $\{\widehat{S}_{X_{ki},W_{ki}}(\theta^{(k)}),\ i \in V_k,\}$ with

$$\widehat{S}_{X_{ki},W_{ki}}(\theta^{(k)}) = \sum_{j\in\bar{V}_k}\left[\frac{df_{\theta^{(k)}}(Y_{kj}|X_{ki},Z_{kj})/d\theta^{(k)}}{\widehat{f}_{\theta^{(k)}}(Y_{kj}|Z_{kj},W_{kj})} - \frac{d\widehat{f}_{\theta^{(k)}}(Y_{kj}|Z_{kj},W_{kj})/d\theta^{(k)}}{(\widehat{f}_{\theta^{(k)}}(Y_{kj}|Z_{kj},W_{kj}))^2}\right.$$
$$\left.\times f_{\theta^{(k)}}(Y_{kj}|X_{ki},Z_{kj})\right]\frac{I_{[W_{kj}=W_{ki}]}}{\sum\limits_{l\in\bar{V}_k} I_{[W_{kl}=W_{ki}]}}.$$

The direct variance estimator for $\widehat{\alpha}$, $\widehat{V}_{DE}(\widehat{\alpha}) = \sum_{k=1}^K w_k^2 \widehat{\mathrm{Var}}_{\widehat{\alpha}^{(k)}|u_k}(\widehat{\alpha}^{(k)})$, is not a consistent estimator of variance in this situation: the variance of $\widehat{\alpha}$ will be larger than $\widehat{V}_{DE}(\widehat{\alpha})$ which assumes independence of observation within a center. The following theorem gives a consistent estimator for the variance-covariance matrix of the proposed two-step estimator $\widehat{\theta} = (\widehat{\alpha}, \widehat{\beta})$.

Theorem 2. *Under Conditions* (a)$-$(e), $K \cdot \widehat{V}_{PE}(\widehat{\theta})$ *is a consistent estimator of the variance-covariance matrix* V *of* $K^{1/2}(\widehat{\theta} - \theta)$, *where* $\widehat{V}_{PE}(\widehat{\theta}) = (K/(K-1)) \sum_{k=1}^K w_k^2 (\widehat{\theta} - \widehat{\theta}^{(k)})^{\otimes 2}$ *and* $b^{\otimes 2} = bb^T$ *for a vector* b.

The proof of Theorem 2 is also outlined in the Appendix. From the results of the Remark and Theorem 2, we see that both variance estimators $\widehat{V}_{DE}(\widehat{\beta})$ and $\widehat{V}_{PE}(\widehat{\beta})$ are consistent estimators for the variance-covariance matrix of $\widehat{\beta}$. When it comes to estimate the intercept, the variance estimator $\widehat{V}_{PE}(\widehat{\alpha})$ is a consistent estimator while $\widehat{V}_{DE}(\widehat{\alpha})$ will under-estimate the variance of $\widehat{\alpha}$.

The proposed two-step estimator and the estimated variance depend on the selection of the weights. Based on the results in (8), the optimal selection of the weights can be expressed as

$$w_k = \frac{1/\mathrm{tr}[\mathrm{Var}_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)})]}{\sum\limits_{k=1}^K \Big(1/\mathrm{tr}[\mathrm{Var}_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)})]\Big)}, \tag{12}$$

for $k=1,\ldots,K$, where $\mathrm{tr}(A)$ denote the trace of the matrix A and $\mathrm{Var}_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)})$ can be consistently estimated by $\widehat{\mathrm{Var}}_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)})$ defined in (11).

Furthermore, when the variance of the random effect is the same across centers, i.e., $\sigma_1^2 = \cdots = \sigma_K^2$, the variance of the random effect can be estimated by

$$\widehat{\sigma}^2 = \frac{\widehat{V}_{PE}(\widehat{\alpha}) - \widehat{V}_{DE}(\widehat{\alpha})}{\sum_{k=1}^K \widehat{w}_k^2}$$

$$= \frac{K}{K-1}\sum_{k=1}^{K}\frac{\widehat{w}_k^2}{\sum_{l=1}^{K}\widehat{w}_l^2}\left(\widehat{\alpha}-\widehat{\alpha}^{(k)}\right)^2 - \sum_{k=1}^{K}\frac{\widehat{w}_k^2}{\sum_{l=1}^{K}\widehat{w}_l^2}\widehat{\mathrm{Var}}_{\widehat{\alpha}^{(k)}|u_k}(\widehat{\alpha}^{(k)}).$$

4. Simulation Studies

We conduct extensive Monte Carlo simulation studies to evaluate the performance of the proposed two-step estimator. To illustrate the effectiveness of the proposed estimator, we compare it with two competing estimators: the naive estimator ($\widehat{\theta}_N$) and the complete-case two-step estimator ($\widehat{\theta}_V$) that is based only on the validation data. The naive estimator considers the auxiliary covariate information but not the random effects, and the complete-case two-step estimator using only the validation data considers the random effects but not the auxiliary covariates information in the nonvalidation set. For the proposed estimator ($\widehat{\theta}$), we used an estimator with the optimal weight in (12). The data are generated according to the following logistic regression model:

$$P_\beta(Y_{ki}=1|X_{ki},u_k) = \frac{\exp\{\alpha + X_{ki}\beta + u_k\}}{1+\exp\{\alpha + X_{ki}\beta + u_k\}}, \tag{13}$$

for $i = 1,\ldots,n_k$, $k = 1,\ldots,K$, where $X_{ki} \sim N(0,1), u_k \sim N(0,\sigma_k^2)$. The auxiliary variable W is defined as $W_{ki} = I(X_{ki} + \varepsilon_k > 0)$, where $\varepsilon_k \sim N(0,\lambda_k^2)$ is a random error and λ_k^2 is the parameter that controls the strength of the association between X and W. The parameter values used in our simulation studies are $\alpha = 0$, $\beta = \log 2 = 0.693$, and $\lambda_k = \mathrm{Unif}[0, 0.5]$. The sample sizes studied are $n = 1000$, 2000 and 4000, and the corresponding numbers of centers are $K = 10$, 20 and 40. We set the number of individuals in each center to be equal, $n_1 = \cdots = n_K = 100$. The validation fraction in each center is chosen to be 50%. For a given sample size, the average of the estimates (Mean), sample standard deviation (SE), the square root of the average of the mean squared error ($\sqrt{\mathrm{MSE}}$), the average of estimated standard errors ($\widehat{\mathrm{SE}}$) and 95% confidence intervals coverage rate (C.I.) are obtained from 500 independent runs. The simulation results are summarized in Tables 1−3.

Table 1 presents the results for cases with common variance of the random effects across centers for $(n,K) = (2000,20)$. We consider $\sigma = 0.50, 0.75$ and 1.0. Table 2 presents the results for cases with varying variance of the random effects across centers for $(n,K) = (1000,10), (2000,20)$ and $(4000,40)$. We generated σ_k from the uniform distribution on $[0,1.5]$. The results of the center-specific estimators for each center are listed in Table 3.

Under the models studied, we make the following observations. (i) Both the complete-case two-step estimator ($\widehat{\theta}_V$) and the proposed two-step estimated likelihood estimator ($\widehat{\theta}$) are valid. The naive estimator ($\widehat{\theta}_N$) yields biased estimates when there are random effects in the model (see Table 1). The bias in $\widehat{\beta}_N$ could be severe when there are strong random effects, i.e., when σ^2 is large. (ii) The estimator $\widehat{\theta}$ is more efficient than $\widehat{\theta}_V$. The relative efficiency calculated as the ratio of the corresponding mean square errors (MSE) of the estimators shows that $\widehat{\theta}_{VE}$ is only about 50%-80% as efficient as $\widehat{\theta}$. (iii) The proposed variance estimates, $\widehat{V}_{PE}(\widehat{\beta})$ and $\widehat{V}_{DE}(\widehat{\beta})$, provide good estimation of the variation of $\widehat{\beta}$. $\widehat{V}_{PE}(\widehat{\alpha})$ provides a good estimate for the variance of the intercept while $\widehat{V}_{DE}(\widehat{\alpha})$ is biased. The 95% confidence intervals using either $\widehat{V}_{DE}(\widehat{\beta})$ or $\widehat{V}_{PE}(\widehat{\beta})$ provide good coverage for the cases studied. The variance estimator $\widehat{V}_{PE}(\widehat{\alpha})$ also gives good coverage for the 95% confidence intervals of α. On the contrary, the variance estimator for $\widehat{\alpha}_N$ underestimates the true variance, especially for large σ_k^2 cases, and $\widehat{\beta}_N$ is biased. Consequently, the 95% confidence interval coverage is very poor. (iv) Examining the results in Tables 2 and 3, it is clear that the proposed two-step estimator has better finite sample properties than the center-specific estimators both in term of bias and variance. We also note that results from some additional simulations with sparse data in individual centers show that $\widehat{V}_{DE}(\widehat{\beta})$ is more stable than $\widehat{V}_{PE}(\widehat{\beta})$ (results not shown). We use $\widehat{V}_{DE}(\widehat{\beta})$ in the presentation of our data analysis.

Table 1. Simulation study results for $\sigma_k^2 \equiv \sigma^2$, $k = 1, \ldots, K$, and $\theta = (\alpha, \beta) = (0, \ln 2)$ with $K = 20$ and $n_k = 100$, $k = 1, \ldots, K$.

							95 % C.I. w/	
σ	Estimator	Mean	SE	$\sqrt{\text{MSE}}$	$\widehat{\text{SE}}_{PE}$	$\widehat{\text{SE}}_{DE}$	SE_{PE}	SE_{DE}
0.50	$\widehat{\alpha}_N$	-0.007	0.118	0.119		0.047		0.558
	$\widehat{\alpha}_V$	-0.008	0.121	0.122	0.121	0.072	0.928	0.762
	$\widehat{\alpha}$	-0.006	0.115	0.116	0.113	0.051	0.928	0.590
	$\widehat{\beta}_N$	0.651	0.058	0.071		0.058		0.878
	$\widehat{\beta}_V$	0.687	0.077	0.077	0.075	0.078	0.924	0.948
	$\widehat{\beta}$	0.677	0.057	0.059	0.059	0.061	0.940	0.950
0.75	$\widehat{\alpha}_N$	-0.010	0.161	0.161		0.047		0.432
	$\widehat{\alpha}_V$	-0.010	0.159	0.159	0.155	0.076	0.934	0.647
	$\widehat{\alpha}$	-0.008	0.154	0.155	0.150	0.052	0.930	0.471
	$\widehat{\beta}_N$	0.616	0.061	0.099		0.057		0.711
	$\widehat{\beta}_V$	0.685	0.078	0.078	0.077	0.080	0.916	0.945
	$\widehat{\beta}$	0.677	0.059	0.061	0.061	0.063	0.934	0.948
1.00	$\widehat{\alpha}_N$	-0.011	0.200	0.200		0.047		0.367
	$\widehat{\alpha}_V$	-0.010	0.189	0.190	0.185	0.077	0.931	0.583
	$\widehat{\alpha}$	-0.009	0.187	0.188	0.181	0.055	0.942	0.429
	$\widehat{\beta}_N$	0.575	0.065	0.135		0.056		0.465
	$\widehat{\beta}_V$	0.682	0.079	0.080	0.080	0.083	0.933	0.955
	$\widehat{\beta}$	0.675	0.059	0.061	0.063	0.065	0.953	0.959

Table 2. Simulation study results for $\sigma_k = \text{Unif}[0, 1.5]$, $k = 1, \ldots, K$. The true parameter $(\alpha, \beta) = (0, \ln 2)$ and $n_1 = \cdots = n_K = 100$.

							95 % C.I. w/	
K	Estimator	Mean	SE	$\sqrt{\text{MSE}}$	$\widehat{\text{SE}}_{PE}$	$\widehat{\text{SE}}_{DE}$	SE_{PE}	SE_{DE}
10	$\widehat{\alpha}_N$	-0.007	0.206	0.206		0.067		0.484
	$\widehat{\alpha}_V$	-0.004	0.194	0.194	0.189	0.103	0.921	0.706
	$\widehat{\alpha}$	-0.002	0.179	0.179	0.177	0.073	0.910	0.583
	$\widehat{\beta}_N$	0.632	0.084	0.104		0.081		0.858
	$\widehat{\beta}_V$	0.687	0.108	0.108	0.109	0.113	0.919	0.957
	$\widehat{\beta}$	0.678	0.086	0.088	0.086	0.103	0.917	0.995
20	$\widehat{\alpha}_N$	-0.011	0.158	0.158		0.047		0.410
	$\widehat{\alpha}_V$	-0.010	0.144	0.144	0.142	0.074	0.936	0.669
	$\widehat{\alpha}$	-0.009	0.138	0.138	0.134	0.052	0.938	0.519
	$\widehat{\beta}_N$	0.620	0.060	0.095		0.057		0.740
	$\widehat{\beta}_V$	0.687	0.077	0.077	0.077	0.080	0.940	0.959
	$\widehat{\beta}$	0.677	0.056	0.058	0.060	0.063	0.949	0.959
40	$\widehat{\alpha}_N$	-0.002	0.105	0.105		0.033		0.464
	$\widehat{\alpha}_V$	-0.002	0.096	0.096	0.093	0.052	0.941	0.695
	$\widehat{\alpha}$	-0.002	0.091	0.091	0.087	0.036	0.949	0.582
	$\widehat{\beta}_N$	0.631	0.044	0.075		0.040		0.672
	$\widehat{\beta}_V$	0.688	0.054	0.055	0.054	0.056	0.949	0.960
	$\widehat{\beta}$	0.680	0.042	0.044	0.042	0.044	0.938	0.937

Table 3. Simulation study results for the center-specific estimators within each center with $\sigma_k = \text{Unif}[0, 1.5]$, $k = 1, \ldots, K$, for $K = 10$ and $n = 1000$. The true parameter $(\alpha, \beta) = (0, \ln 2)$.

	$\widehat{\alpha}_V^{(k)}$			$\widehat{\beta}_V^{(k)}$			$\widehat{\alpha}^{(k)}$			$\widehat{\beta}^{(k)}$		
Center	Mean	SE	$\widehat{\text{SE}}$	Mean	SE	$\widehat{\text{SE}}$	Mean	SE	$\widehat{\text{SE}}$	Mean	SE	$\widehat{\text{SE}}$
1	0.031	0.454	0.313	0.717	0.353	0.359	0.019	0.380	0.220	0.698	0.281	0.272
2	-0.059	1.424	0.435	0.756	0.426	0.463	-0.059	1.361	0.312	0.736	0.352	0.347
3	-0.020	0.386	0.312	0.767	0.374	0.375	-0.012	0.315	0.220	0.741	0.302	0.280
4	0.061	0.580	0.322	0.755	0.389	0.386	0.062	0.504	0.225	0.721	0.273	0.280
5	-0.011	0.329	0.312	0.780	0.395	0.395	0.001	0.222	0.218	0.747	0.285	0.276
6	-0.032	0.850	0.340	0.754	0.389	0.388	-0.015	0.777	0.238	0.729	0.297	0.293
7	-0.027	1.285	0.432	0.767	0.441	0.463	-0.043	1.195	0.296	0.736	0.350	0.332
8	0.052	0.485	0.316	0.749	0.407	0.403	-0.048	0.445	0.223	0.728	0.317	0.282
9	0.005	1.076	0.372	0.755	0.426	0.401	-0.004	1.014	0.258	0.728	0.321	0.307
10	-0.005	0.573	0.319	0.744	0.380	0.369	-0.014	0.523	0.224	0.704	0.296	0.283

5. Risk of Male Birth Defects and DDE Exposure

DDE are ubiquitous environmental contaminants. It has been hypothesized in a recent study that in-utero exposure to the androgen antagonist DDE could be related to the frequency of male birth defects among boys (Longnecker, Klebnoff, Zhou and Brock (2002)). The investigators tested this hypothesis on a population with relatively high serum DDE levels consisting of male offsprings of those enrolled in the Collaborative Perinatal Project (CPP). We illustrate our proposed method by evaluating the effect of DDE exposure on the risk of male birth defects (cryptochidism, hypospadias and polythelia) using a subset data set from the Longnecker et al. (2002) study.

The CPP was designed to identify determinants of neurological disorders and other conditions in children (Niswander (1972)). Pregnant women were recruited from 12 US medical centers from 1959 to 1966. The method of subject selection varied across study centers. Approximately 42,000 women were enrolled, resulting in 55,000 children in the study. The children were systematically assessed for the presence of birth defects and other outcomes through age 7. The third trimester serum was assayed for DDE at the center for Disease Control and Prevention (Longnecker et al., 2002).

We define the binary outcome variable to be one if at least one of the three birth defects is reported. It is well documented that race (black or white) is a good auxiliary information for the DDE level (e.g., Stehr-Green (1989)). In our analysis, there are 1,474 children in all from 11 medical centers, with 712 having DDE measured. The smallest center has 46 children with 29 having DDE measured; the largest one is 302 with 155 having DDE measured. The possible confounding variables are maternal age in years (MAGE), family's socio-economic index (SEINDX), infant race (RACE), smoking during pregnancy (SMK), maternal height (MHGT), Triglycerides (TRIGLYC), maternal pre-pregnancy body mass index (MBMIPP) and preterm birth (PRETERM). We fit model (13) to this subset of CPP data with center-specific random effects.

Table 4 shows the results from fitting the random effect model based on the validation set only, and those of the proposed method. Both methods show no statistically significant effect of DDE on the risk of birth defects. The proposed method is more efficient since it uses more data than the validation set analysis: the proposed method has a smaller variance and hence a narrower confidence interval. The improvement in efficiency is more significant among other covariates. The standard errors from the proposed method is about one-third to one-half of that from the validation set analysis. Among the confounding variables in the model, the body mass index is significant with the additional information from the non-validation set. Our results agree with those of Longnecker et al. (2002).

Table 4. Analysis of CPP data set on DDE in relationship to male birth defects.

	Validation Method		Proposed Method	
Factors	Coefficient	S.E.	Coefficient	S.E.
Intercept	-3.9751	2.5356	-3.3837	1.8136
DDE	-0.0113	0.0071	-0.0094	0.0065
MAGE	0.0032	0.0170	-0.0067	0.0122
SEINDX	0.0478	0.0615	0.0207	0.0445
RACE	-0.0165	0.3406	-0.2138	0.2518
MBMIPP	0.0133	0.0209	0.0365*	0.0165
TRIGLYC	-0.0005	0.0014	-0.0014	0.0010
SMK	0.0368	0.2041	0.1427	0.1469
MHGT	0.0368	0.0382	0.0338	0.0269
PRETERM	0.2007	0.2906	-0.1021	0.1985

6. Discussion

The derivation of the asymptotic properties for our proposed two-step method requires that the number of observations within each center goes to infinity. The simulation results indicate that when the cluster size is reasonably large, the asymptotic approximation works well for the finite samples. Although the requirement for large cluster size is easily satisfied in many large scale multicenter or longitudinal studies, the proposed method is not applicable for the type of data where limited observations are available within each cluster, such as in the eye-disease or family studies.

Acknowledgements

This research was partially supported by National Institutes of Health grants CA-79949 (H. Z. Zhou and J. Chen) and HL-57444 (J. Cai). We are thankful to the Co-Editor, an associate editor and a referee for helpful comments.

Appendix. Proofs of Theorems 1 and 2

Proof of Theorem 1. It follows from the results of Pepe and Fleming (1991) that $E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)}) = \theta^{(k)} + o_p\left(1/\sqrt{n_k}\right)$, uniformly for $k = 1, \ldots, K$. By the expression of the two-step estimator and Condition (b), we have

$$\begin{aligned}
&\sqrt{K}(\widehat{\theta} - \theta)\\
&= \sqrt{K}\sum_{k=1}^{K} w_k \left\{\left(\widehat{\theta}^{(k)} - E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)})\right) + \left(\theta^{(k)} - \theta\right) + \left(E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)}) - \theta^{(k)}\right)\right\}\\
&= \sqrt{K}\sum_{k=1}^{K} w_k \left\{\left(\widehat{\theta}^{(k)} - E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)})\right) + u_k e_1\right\} + \sqrt{K}\sum_{k=1}^{K} w_k o_p\left(1/\sqrt{n_k}\right)\\
&= U_K(\widehat{\theta}) + o_p(1), \qquad \text{(A.1)}
\end{aligned}$$

with $U_K(\widehat{\theta}) = \sqrt{K} \sum_{k=1}^{K} w_k \left\{ \left(\widehat{\theta}^{(k)} - E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)})\right) + u_k e_1 \right\}$, where e_1 is a vector with 1 on the 1^{st} position and 0 elsewhere.

Now we establish the asymptotic normality of the two-step estimator. Because $w_k\{(\widehat{\theta}^{(k)} - E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)})) + u_k e_1\}$, $k = 1, \ldots, K$, are independent, we only need to compute the first two moment and check Liapounov's condition. The expectation of $U_K(\widehat{\theta})$ is given by

$$E\left[U_K(\widehat{\theta})\right] = \sqrt{K} \sum_{k=1}^{K} w_k \left\{ E\left[\widehat{\theta}^{(k)} - E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)})\right] + E\left[u_k e_1\right] \right\} = 0. \quad \text{(A.2)}$$

Using Condition (c), the variance-covariance matrix of $U_K(\widehat{\theta})$ is

$$\begin{aligned} \text{Var}\left[U_K(\widehat{\theta})\right] &= K \sum_{k=1}^{K} w_k^2 \Big\{ \text{Var}\left[E_{\widehat{\theta}^{(k)}|u_k}\left(\widehat{\theta}^{(k)} - E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)})\right)\right] \\ &\quad + E\left[\text{Var}_{\widehat{\theta}^{(k)}|u_k}\left(\widehat{\theta}^{(k)} - E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)})\right)\right] + \sigma_k^2 e_1 e_1^T \Big\} \\ &= K \sum_{k=1}^{K} w_k^2 \left\{ \sigma_k^2 e_1 e_1^T + E\left[\text{Var}_{\widehat{\theta}^{(k)}|u_k}\left(\widehat{\theta}^{(k)}\right)\right] \right\} \\ &= V_1 + V_2 + o(1). \end{aligned} \quad \text{(A.3)}$$

By Condition (d) and the results in (A.2)−(A.3), it is clear that Liapounov's condition is satisfied. Therefore, from the Liapounov Central Limit Theorem, we have $\sqrt{K}\left(\widehat{\theta} - \theta\right) \to N\left(0, V_1 + V_2\right)$.

Proof of Theorem 2. First, we note that the estimator $K \cdot \widehat{V}_{PE}(\widehat{\theta})$ of the variance-covariance matrix $V_1 + V_2$ can be expressed as

$$\begin{aligned} \frac{K^2}{K-1}\sum_{k=1}^{K} w_k^2(\widehat{\theta}^{(k)} - \widehat{\theta})^{\otimes 2} &= \frac{K^2}{K-1}\sum_{k=1}^{K} w_k^2(\widehat{\theta}^{(k)} - \theta)^{\otimes 2} + \frac{K^2}{K-1}\sum_{k=1}^{K} w_k^2(\widehat{\theta}^{(k)} - \theta)(\theta - \widehat{\theta})^T \\ &\quad + \frac{K^2}{K-1}\sum_{k=1}^{K} w_k^2(\theta - \widehat{\theta})^{\otimes 2} + \frac{K^2}{K-1}\sum_{k=1}^{K} w_k^2(\theta - \widehat{\theta})(\widehat{\theta}^{(k)} - \theta)^T \\ &\equiv I_1 + I_2 + I_3 + I_4. \end{aligned} \quad \text{(A.4)}$$

Furthermore, the first term can be written as

$$\begin{aligned} I_1 = &\frac{K^2}{K-1}\sum_{k=1}^{K} w_k^2(\widehat{\theta}^{(k)} - E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)}) + u_k e_1)^{\otimes 2} + \frac{K^2}{K-1}\sum_{k=1}^{K} w_k^2(E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)}) - \theta^{(k)})^{\otimes 2} \\ &+ \frac{K^2}{K-1}\sum_{k=1}^{K} w_k^2(\widehat{\theta}^{(k)} - E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)}) + u_k e_1)(E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)}) - \theta^{(k)})^T \end{aligned}$$

$$+\frac{K^2}{K-1}\sum_{k=1}^{K} w_k^2 \left(E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)}) - \theta^{(k)}\right)\left(\widehat{\theta}^{(k)} - E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)}) + u_k e_1\right)^T$$
$$\equiv I_{11} + I_{12} + I_{13} + I_{14}. \tag{A.5}$$

Let $T_k = K^2 w_k^2 \left(\widehat{\theta}^{(k)} - E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)}) + u_k e_1\right)^{\otimes 2}$, for $k = 1, \ldots, K$. Note that T_k are independent with $ET_k = K^2 w_k^2(\Gamma_k + \sigma_k^2 e_1 e_1^T)$. Hence by Condition (c), $(1/K)\sum_{k=1}^K ET_k \to V_1 + V_2$, as $K \to \infty$. Let $(B)_{ij}$ denote the $(ij)^{th}$ element of the matrix B. Under Conditions (d) and (e) and the C_r inequality (see Sen and Singer (1993)), for each (i,j) we have

$$K^{-(1+\delta)}\sum_{k=1}^{K} E\left(\mid (T_k - ET_k)_{ij} \mid^{1+\delta}\right)$$
$$\leq K^{(1+\delta)}\sum_{k=1}^{K} w_k^{2+2\delta} E\left|2\mathrm{tr}\left((\widehat{\theta}^{(k)} - E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)}))^{\otimes 2}\right) + 2u_k^2 + \sigma_k^2 + (\Gamma_k)_{ij}\right|^{1+\delta}$$
$$\leq K^{1+\delta}4^{\delta}\sum_{k=1}^{K} w_k^{2+2\delta}\left\{2^{1+\delta}E\left|\mathrm{tr}\left((\widehat{\theta}^{(k)} - E_{\widehat{\theta}^{(k)}|u_k}(\widehat{\theta}^{(k)}))^{\otimes 2}\right)\right|^{1+\delta}\right.$$
$$\left. +2^{1+\delta}E\,|u_k|^{2+2\delta} + \sigma_k^{2(1+\delta)} + \mid (\Gamma_k)_{ij} \mid^{1+\delta}\right\}$$
$$\to 0, \quad \text{as} \quad K \to \infty.$$

Therefore, by The Markov Weak Law of Large Mumber, it follows that $(1/K)$ $\sum_{k=1}^K T_k \xrightarrow{P} V_1 + V_2$ as $K \to \infty$. Thus we obtain $I_{11} \xrightarrow{P} V_1 + V_2$. Furthermore, by Condition (b) and the convergence of $(1/K)\sum_{k=1}^K T_k$, it can be shown that $I_{1i} \xrightarrow{P} 0$, for $i = 2, 3, 4$. Hence we have

$$I_1 = I_{11} + I_{12} + I_{13} + I_{14} \xrightarrow{P} V_1 + V_2. \tag{A.6}$$

On the other hand, by Condition (b) and Theorem 1, we have

$$I_3 = \frac{K}{K-1} \times K\sum_{k=1}^{K} w_k^2 \times \left(\theta - \widehat{\theta}\right)^{\otimes 2} \xrightarrow{P} 0, \quad \text{as} \quad K \to \infty. \tag{A.7}$$

Using the Cauchy-Schwartz Inequality and the results in (A.6) and (A.7), it can be shown that as $K \to \infty$, $I_2 \xrightarrow{P} 0$ and $I_4 \xrightarrow{P} 0$. Based on the above results, we have that $K \cdot \widehat{V}_{PE}(\widehat{\theta})$ is a consistent estimator for $V_1 + V_2$.

References

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference to generalized linear mixed models. *J. Amer. Statist. Assoc.* **88**, 9-25.

Cai, J., Sen, P. K. and Zhou, H. B. (1999). A random effects model for multivariate failure time data from multicenter clinical trials. *Biometrics* **55**, 182-189.

Carroll, R. J. and Wand, M. P. (1991). Semiparametric estimation in logistic measurement error models. *J. Roy. Statist. Soc. Ser. B* **53**, 573-585.

Henderson, C. R. (1950). Estimation of genetic parameters (abstract). *Ann. Math. Statist.* **21** 309-310.

Jiang, J. (1998). Consistent estimators in generalized linear mixed model. *J. Amer. Statist. Assoc.* **93**, 720-729.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *J. Roy. Statist. Soc. Ser. B* **58**, 619-678.

Longnecker, M. P., Klebnoff, M. A., Brock, J. W., Zhou, H., Gary, K. A., Needham, L. L. and Wilcox, A. J. (2002). Maternal serum level of 1, 1-dichloro-2, 2-bis (p-chloropheny) ethylene and risk of cryptorchidism, hypospadias and polythelia among male offspring. *Amer. J. Epidemiology* **155**, 313-322.

McGilchrist, C. A. (1994). Estimation in generalized mixed models. *J. Roy. Statist. Soc. Ser. B* **56**, 61-69.

Pepe, M. S. and Fleming, T. R. (1991). A nonparametric method for dealing with mismeasured covariate data. *J. Amer. Statist. Assoc.* **86**, 108-113.

Reilly, M. and Pepe, M. S. (1995). A mean score method for missing and auxiliary covariate data in regression models. *Biometrika* **82**, 299-314.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regression are not always observed. *J. Amer. Statist. Assoc.* **89**, 846-866.

Robinson, G. K. (1991). That Blup is a good thing: the estimation of random effects. *Statist. Sci.* **6**, 15-51.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika* **78**, 719-727.

Stehr-Green, P. A. (1989). Demographic and seasonal influences on human serum pesticide residue levels. *J. Toxicol Environ Health* **27**, 405-421.

Stiratelli, R., Laird, N. and Ware, J. H. (1984). Random-effects models for serial observations with binary response. *Biometrics* **40**, 961-971.

Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effect model with heterogeneity in the random-effects population. *J. Amer. Statist. Assoc.* **91**, 217-221.

Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models: a pseudo-likelihood approach. *J. Statist. Comput. Simulation* **48**, 233-243.

Zhou, H. and Pepe, M. S. (1995). Auxiliary covariate data in failure time regression. *Biometrika* **83**, 139-149.

Zhou, H. and Wang, C.Y. (2000). Failure time regression analysis with measurement error in covariates. *J. Roy. Statist. Soc. Ser. B* **62**, 657-665.

Department of Biostatistics and Computational Biology, University of Rochester, School of Medicine and Dentistry 601 Elmwood Avenue, Box 630, Rochester, NY 14642, U.S.A

E-mail: jchen@bst.rochester.edu

Department of Biostatistics, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, U.S.A.

E-mail: cai@bios.unc.edu

Department of Biostatistics, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, U.S.A.

E-mail: zhou@bios.unc.edu

(Received August 2002; accepted September 2003)