Biometrics 65, 361–368
June 2009

DOI: 10.1111/j.1541-0420.2008.01103.x

# Diagnosis of Random-Effect Model Misspecification in Generalized Linear Mixed Models for Binary Response

**Xianzheng Huang**

Department of Statistics, University of South Carolina, Columbia, South Carolina 29208, U.S.A.
*email:* huang@stat.sc.edu

Summary. Generalized linear mixed models (GLMMs) are widely used in the analysis of clustered data. However, the validity of likelihood-based inference in such analyses can be greatly affected by the assumed model for the random effects. We propose a diagnostic method for random-effect model misspecification in GLMMs for clustered binary response. We provide a theoretical justification of the proposed method and investigate its finite sample performance via simulation. The proposed method is applied to data from a longitudinal respiratory infection study.

Key words: Clustered binary response; Generalized linear mixed models; Random effects.

## 1. Introduction

Generalized linear mixed models (GLMMs) are frequently used to analyze data from a wide range of applications. They are flexible models for nonnormal responses, repeated measurements, and other forms of clustered data. This class of models can easily account for multiple sources of variation and address various correlation structures in correlated data. A natural concern in using GLMMs is misspecifying the model for the random effects. For computational convenience, random effects in GLMMs are almost routinely assumed to be normal. However, the normality assumption may be unrealistic in some applications. Moreover, to decide which covariates in the model have random coefficient is also a difficult question.

Early investigation to address this concern suggested that misspecifying the models for the random effects usually only results in a small amount of bias in the maximum likelihood estimators (MLEs) for the fixed effects (Neuhaus, Hauck, and Kalbfleisch, 1992). However, more recently, many authors have found that likelihood-based inference can be severely affected if the random-effect model is misspecified. For example, Heagerty and Kurland (2001) computed the asymptotic bias in the MLEs for the parameters in a logistic mixed model in four instances of random-effect model misspecification. They concluded that incorrect assumptions on the random effects can lead to substantial bias in the MLEs for the fixed effects. Agresti, Caffo, and Ohman-Strickland (2004) conducted empirical studies on the impact of model misspecification for the random effects in GLMMs, showing that the MLEs for the fixed effects can be very sensitive to the assumed random-effect model. Finally, Litière, Alonso, and Molenberghs (2007) used simulation to show that the type I and type II errors of tests for the mean structure in a logistic mixed model can be seriously affected by violations of the random-effect model assumptions.

There are many inferential methods developed to avoid invalid inference due to random-effects model misspecification. For instance, Chen, Zhang, and Davidian (2002) developed a seminonparametric approach to model random effects using a smooth density representation. Nonparametric approaches (Heckman and Singer, 1984) and use of normal mixtures (Magder and Zeger, 1996) have been proposed to circumvent making restrictive parametric assumptions on the random effects. These nonparametric and semiparametric methods typically involve intensive computation with a potential loss in efficiency. Unlike parametric approaches, the aforementioned methods often lack a natural likelihood function, which is useful for model selection, hypothesis testing, and variance estimation. Moreover, in some applications the characteristics of random effects are of scientific interest in their own right, which may not be explicitly explored if nonparametric or semiparametric methods are used.

Until now, there has been no diagnostic procedure developed to detect random-effect model misspecification in GLMMs. White (1981, 1982) studied the properties of MLE resulting from a misspecified model for the observed data and proposed an information matrix test for general model misspecification. Although White's method is applicable for any model in principle, it is complicated to implement and it does not provide direction of model correction when misspecification is detected. Agresti et al. (2004) suggested comparing results from both parametric and nonparametric methods, arguing that a substantial discrepancy between the two analyses indicates model misspecification.

The difficulty in detecting model misspecification for the random effects is mainly due to the obvious fact that there is no data realization or surrogate observation for the random effects. Consequently, none of the traditional diagnostic techniques that rely solely on the observed data can justify model assumptions for the random effects. In this article, we propose a novel parametric diagnostic method that makes use of both the observed data and a reconstructed data set induced from the observed data, with computational complexity comparable with that of GLMMs. The observed data, the construction

of the reconstructed data, and the models for these two data types are given in Section 2. In Section 3, test statistics are defined to assess the adequacy of the assumed random-effect model. In Section 4, we study the operating characteristics of the proposed test statistics via simulation. In Section 5, we investigate the impact of additional misspecification for the fixed-effect part of the model on the proposed test statistics. In Section 6, we apply our diagnostic technique to a data set from a longitudinal respiratory infection study. In Section 7, we provide a summary discussion and address future research topics.

## 2. Data and Models

### 2.1 *Observed Data*

Herein we focus on clustered data with binary response. Extensions of the proposed method to other types of response are discussed in Section 7. Denote by $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^T$ the $n_i \times 1$ vector of binary responses for cluster $i, i = 1, \ldots, m$. Consider the conditional mean model defined by

$$E(Y_{ij} \mid \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \boldsymbol{b}_i) = h(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\boldsymbol{b}_i), \tag{1}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects, $p < m$, $\boldsymbol{b}_i$ is a $q \times 1$ mean-zero vector of random effects, $\mathbf{X}_{ij}$ is the $j$th row of the $n_i \times p$ design matrix $\mathbf{X}_i$ for the fixed effects, $\mathbf{Z}_{ij}$ is the $j$th row of the $n_i \times q$ design matrix $\mathbf{Z}_i$ for the random effects, for $i = 1, \ldots, m, j = 1, \ldots, n_i$, and $h(\cdot)$ is a monotonic differentiable inverse link function. Our assumptions on the model in equation (1) are that $E(Y_{ij} \mid \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \boldsymbol{b}_i) = E(Y_{ij} \mid \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{b}_i)$, for $i = 1, \ldots, m$, and that the link function is appropriate for the data, thus the main concern is the choice of an assumed random-effect model. Furthermore, the $m$ clusters are independent, and within cluster $i$, $\{Y_{ij}\}_{j=1}^{n_i}$ are independent given $\boldsymbol{b}_i$. Define $\boldsymbol{\Omega}$ as the $r \times 1$ parameter vector that includes $\boldsymbol{\beta}$ and the parameters in the assumed model for $\boldsymbol{b}_i$, denoted by $\boldsymbol{\tau}$. The contribution to the observed-data likelihood from cluster $i$ is given by

$$f_{\mathbf{Y}_i}(\mathbf{Y}_i \mid \mathbf{X}_i, \mathbf{Z}_i; \boldsymbol{\Omega}) = \int f_{\boldsymbol{b}_i}(\boldsymbol{b}_i; \boldsymbol{\tau}) \prod_{j=1}^{n_i} h(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\boldsymbol{b}_i)^{Y_{ij}} \times \{1 - h(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\boldsymbol{b}_i)\}^{1-Y_{ij}}\, d\boldsymbol{b}_i, \tag{2}$$

where $f_{\boldsymbol{b}_i}(\boldsymbol{b}_i; \boldsymbol{\tau})$ is the density function associated with the assumed model for $\boldsymbol{b}_i$. It is clear from equation (2) that the quality of the MLE for $\boldsymbol{\beta}$, denoted by $\widehat{\boldsymbol{\beta}}$, usually relies on the assumed model for $\boldsymbol{b}_i$. Correct specification of $f_{\boldsymbol{b}_i}(\boldsymbol{b}_i; \boldsymbol{\tau})$ is a sufficient condition for $\widehat{\boldsymbol{\beta}}$ being consistent. In this article, where asymptotic properties are concerned, we refer to the properties when $m \to \infty$ and the cluster size is bounded.

### 2.2 *Reconstructed Data*

Based on the observed data, we form a reconstructed data set by partitioning the $n_i$ subjects within cluster $i$ into $G_i$ subgroups, for $i = 1, \ldots, m$; then we define the new clustered binary response for cluster $i$ as $\mathbf{Y}_i^* = (Y_{i1}^*, \ldots, Y_{iG_i}^*)^T$, where $Y_{ig}^* = 0$ if all the $Y_{ij}$'s in subgroup $g$ of cluster $i$ equal zero, and $Y_{ig}^* = 1$ otherwise, for $g = 1, \ldots, G_i$. Deduced from equation (1), the conditional mean model for $Y_{ig}^*$ is given by, for $i = 1, \ldots, m$ and $g = 1, \ldots, G_i$,

$$E\big(Y_{ig}^* \,\big|\, \mathbf{X}_{ij}, \mathbf{Z}_{ij}, \boldsymbol{b}_i, j \in \text{group } g\big) = 1 - \prod_{j \in \text{group } g} \{1 - h(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\boldsymbol{b}_i)\},$$

where "$\prod_{j \in \text{group } g}$" refers to the product taken over all subjects in subgroup $g$ of cluster $i$. It follows that the contribution of cluster $i$ to the reconstructed-data likelihood is

$$f_{\mathbf{Y}_i^*}\big(\mathbf{Y}_i^* \,\big|\, \mathbf{X}_i, \mathbf{Z}_i; \boldsymbol{\Omega}\big) = \int f_{\boldsymbol{b}_i}(\boldsymbol{b}_i; \boldsymbol{\tau}) \prod_{g=1}^{G_i} \left[1 - \prod_{j \in \text{group } g} \big\{1 - h(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\boldsymbol{b}_i)\big\}\right]^{Y_{ig}^*} \times \left[\prod_{j \in \text{group } g} \{1 - h(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\boldsymbol{b}_i)\}\right]^{1-Y_{ig}^*} d\boldsymbol{b}_i. \tag{3}$$

The subgroup composition will be detailed in Section 4.

## 3. Diagnostic Method

### 3.1 *Test Statistics*

Denote by $\widehat{\boldsymbol{\Omega}}$ and $\widehat{\boldsymbol{\Omega}}^*$ the MLEs for $\boldsymbol{\Omega}$ based on the observed data and the reconstructed data, respectively. Both estimators are consistent when the random-effect model is correctly specified, but not necessarily so otherwise. More importantly, as will be shown in Section 3.2, in the presence of model misspecification, the asymptotic means of $\widehat{\boldsymbol{\Omega}}$ and $\widehat{\boldsymbol{\Omega}}^*$, denoted by $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}^*$, respectively, can differ. This motivates an indicator of model misspecification, which is defined as

$$T^2 = \frac{m - r}{r(m-1)} (\widehat{\boldsymbol{\Omega}}^* - \widehat{\boldsymbol{\Omega}})^T \widehat{\mathbf{V}}^{-1} (\widehat{\boldsymbol{\Omega}}^* - \widehat{\boldsymbol{\Omega}}), \tag{4}$$

where $\widehat{\mathbf{V}}$ is an estimator for the variance-covariance matrix of $\widehat{\boldsymbol{\Omega}}^* - \widehat{\boldsymbol{\Omega}}$. The derivation of $\widehat{\mathbf{V}}$ is given in Web Appendix A, where we show that $r(m-1)(m-r)^{-1}T^2$ is a Hotelling's $T^2$ statistic, and $T^2 \sim F(r, m-r)$ asymptotically under the null hypothesis $H_0 : \widetilde{\boldsymbol{\Omega}} = \widetilde{\boldsymbol{\Omega}}^*$. Strong evidence against $H_0$ implies model misspecification. When model misspecification exists, the MLEs for different parameters in $\boldsymbol{\Omega}$ are affected differently, depending on how the assumed model compares to the true model. To study how different ways of misspecifying the model can influence different parameters, we define another test statistic to compare two MLEs for any one parameter based on two data types,

$$t_\theta = (\widehat{\theta}^* - \widehat{\theta})\widehat{\nu}^{-1}, \tag{5}$$

where $\theta$ denotes any one element in $\boldsymbol{\Omega}$, $\widehat{\theta}$ and $\widehat{\theta}^*$ are the MLEs for $\theta$ based on the observed data and the reconstructed data, respectively, and $\widehat{\nu}^2$ is the diagonal element of $\widehat{\mathbf{V}}$ corresponding to $\theta$. By the construction of $t_\theta$ and $\widehat{\nu}^2$, $t_\theta$ follows a Student's $t$ distribution with $m - r$ degrees of freedom asymptotically under the null hypothesis $H_0^\theta : \widetilde{\theta} = \widetilde{\theta}^*$, where $\widetilde{\theta}$ and $\widetilde{\theta}^*$ are the elements in $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}^*$ corresponding to $\theta$, respectively.

In summary, $T^2$ in equation (4) is a global test statistic that assesses the overall discrepancy between the MLEs for $\boldsymbol{\Omega}$ based on two data types, and $t_\theta$ in equation (5) is

an individual test statistic that evaluates the disparity between the MLEs for a particular parameter computed from two data types. As demonstrated in Section 4, the global test can provide evidence of model misspecification, and the individual test can suggest which type of misspecification occurs.

An alternative global test statistic for $H_0$ can be constructed by combining the $r$ individual $t_\theta$'s following the approach described in Wu, Genton, and Stefanski (2006), which is referred to as the pooled component test (PCT) statistic. PCT is motivated by and designed for testing the equality of two mean vectors when $r > m$. It was shown that when $m > r$, as in the scenarios we consider here, PCT performs similarly as the Hotelling's $T^2$ test.

### 3.2 *Theoretical Justification*

Under the conditions given in White (1982), $\widehat{\boldsymbol{\Omega}}$ converges almost surely to $\widetilde{\boldsymbol{\Omega}}$ as $m \to \infty$, where $\widetilde{\boldsymbol{\Omega}}$ minimizes the Kullback–Leibler information criterion defined by

$$\lim_{m\to\infty} E_{\mathbf{Y}\,|\,\mathbf{X},\mathbf{Z}}\left\{\log\frac{g_{\mathbf{Y}}(\mathbf{Y}\,|\,\mathbf{X},\mathbf{Z};\boldsymbol{\Omega}_0)}{f_{\mathbf{Y}}(\mathbf{Y}\,|\,\mathbf{X},\mathbf{Z};\boldsymbol{\Omega})}\right\},$$

in which $\mathbf{Y} = \{\mathbf{Y}_i\}_{i=1}^m, \mathbf{X} = \{\mathbf{X}_i\}_{i=1}^m, \mathbf{Z} = \{\mathbf{Z}_i\}_{i=1}^m, g_{\mathbf{Y}}(\mathbf{Y}\,|\,\mathbf{X},\mathbf{Z};\boldsymbol{\Omega}_0)$ is the true density of $\mathbf{Y}$ given $\mathbf{X}$ and $\mathbf{Z}, \boldsymbol{\Omega}_0$ is the parameter vector associated with this true model, and the expectation is taken with respect to the true distribution. Equivalently, $\widetilde{\boldsymbol{\Omega}}$ solves

$$\lim_{m\to\infty} E_{\mathbf{Y}\,|\,\mathbf{X},\mathbf{Z}}\{(\partial/\partial\boldsymbol{\Omega})\log f_{\mathbf{Y}}(\mathbf{Y}\,|\,\mathbf{X},\mathbf{Z};\boldsymbol{\Omega})\} = \mathbf{0}. \tag{6}$$

Similarly, $\widehat{\boldsymbol{\Omega}}^*$ converges almost surely to $\widetilde{\boldsymbol{\Omega}}^*$ as $m \to \infty$, and $\widetilde{\boldsymbol{\Omega}}^*$ is uniquely determined by

$$\lim_{m\to\infty} E_{\mathbf{Y}^*\,|\,\mathbf{X},\mathbf{Z}}\{(\partial/\partial\boldsymbol{\Omega})\log f_{\mathbf{Y}^*}(\mathbf{Y}^*\,|\,\mathbf{X},\mathbf{Z};\boldsymbol{\Omega})\} = \mathbf{0}, \tag{7}$$

where $\mathbf{Y}^* = \{\mathbf{Y}_i^*\}_{i=1}^m$. We next compute $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}^*$ to justify the theoretical motivation of the proposed test statistics.

There are many possible ways that one may misspecify the model for the random effects $\boldsymbol{b}_i$. To provide a concrete presentation of the impact of random-effect model misspecification on $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}^*$, we focus on the logistic model considered by Heagerty and Kurland (2001) with conditional mean model

$$\begin{aligned} E(Y_{ij}\,|\,\mathbf{X}_{ij},\, b_{ij}) \;&= \{1+\exp(-\beta_0-\beta_1X_{ij,1}-\beta_2X_{ij,2} \\ &\qquad -\beta_3X_{ij,1}X_{ij,2}-b_{ij})\}^{-1}, \end{aligned} \tag{8}$$

where $X_{ij,1} = x_i$ represents a between-cluster covariate that takes values either 0 or 1, $X_{ij,2} = (j-1)/(n_i-1)$ is a within-cluster covariate, and $b_{ij}$ is the random effect, for $i = 1,\ldots,m$ and $j = 1,\ldots,n_i$. The true regression parameter values are, $\beta_0 = -2, \beta_1 = 1, \beta_2 = 0.5$, and $\beta_3 = -0.25$. Suppose that one always assumes $b_{ij} = b_{i0}$, where $b_{i0} \sim N(0,\, \sigma_0^2)$, for $i = 1,\ldots,m$ and $j = 1,\ldots,n_i$. For the truth regarding $b_{ij}$, we consider the following four cases, (I): $b_{ij} = b_{i0} = \sigma(a_i - \lambda)/\sqrt{\lambda}$, where $a_i \sim$ gamma$(\lambda,\, 1), \sigma = 3$, and $\lambda = 1$; (II): $b_{ij} = b_{i0}$, where $[b_{i0}\,|\,x_i = 0] \sim N(0,\, \sigma_{00}^2)$ and $[b_{i0}\,|\,x_i = 1] \sim N(0,\, \sigma_{01}^2), \sigma_{00} = 3$ and $\sigma_{01} = 0.5$; (III): $b_{ij} = b_{i0} + b_{i1}X_{ij,2}$, where $b_{i0} \sim N(0,\, \sigma_0^2)$ is independent of $b_{i1} \sim N(0,\, \sigma_1^2), \sigma_0 = 0.5, \sigma_1 = 2$; and (IV): $\mathrm{cov}(b_{ij},\, b_{ik}) = \sigma^2\rho^{|j-k|}$, where $\sigma = 3$ and $\rho = 0.5$. That is, we assume a normal random-intercept logistic model whereas the truth is that, the random intercept is nonnormal, or it depends on a covariate, or there is a random slope in addition to the random intercept, or the random effects are autocorrelated. We follow the settings used in Heagerty and Kurland (2001) and choose specific parameter values in each case to create situations where moderate to severe bias is observed in some elements in $\widehat{\boldsymbol{\Omega}}$. It is possible in practice that more than one type of misspecification occurs, but we choose herein to study cases (I)–(IV) individually. The results from our case-by-case investigations can shed light on the more complex cases. For ease of exposition, we assume in this subsection that $n_i = n$ and $G_i = G$, for $i = 1,\ldots,m$.

To solve equations (6) and (7), we exploit the approach of using artificial sample described in Rotnitzky and Wypij (1994) and also used by Heagerty and Kurland (2001). For example, in solving equation (6), the artificial sample consists of $2^n$ distinct combinations, indexed by $l$, of zeros and ones in an $n \times 1$ binary response vector. For the logistic model in equation (8), the distinct fixed-effect design matrices include $\mathbf{X}^{(1)} = [\mathbf{1}\ \mathbf{0}\ \mathbf{S}\ \mathbf{0}]$ and $\mathbf{X}^{(2)} = [\mathbf{1}\ \mathbf{1}\ \mathbf{S}\ \mathbf{S}]$, where $\mathbf{1}$ is the $n \times 1$ vector of ones, $\mathbf{0}$ is the $n \times 1$ vector of zeros, and $\mathbf{S} = (0,\ 1/(n-1),\ 2/(n-1),\ \cdots,\ 1)^T$. Assuming equal proportions of clusters with design matrices $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$, the solution to equation (6) maximizes the following weighted log likelihood over $\boldsymbol{\Omega}$,

$$\begin{aligned} \sum_{l=1}^{2^n} \big\{ &\pi^{(1)}(\mathbf{Y}_l)\log f_{\mathbf{Y}_l}\left(\mathbf{Y}_l \,\middle|\, \mathbf{X}^{(1)};\boldsymbol{\Omega}\right) \\ &+ \pi^{(2)}(\mathbf{Y}_l)\log f_{\mathbf{Y}_l}\left(\mathbf{Y}_l \,\middle|\, \mathbf{X}^{(2)};\boldsymbol{\Omega}\right)\big\}, \end{aligned} \tag{9}$$

where $f_{\mathbf{Y}_l}(\mathbf{Y}_l\,|\,\mathbf{X}^{(k)};\boldsymbol{\Omega})$ is given by equation (2), and $\pi^{(k)}(\mathbf{Y}_l) = g(\mathbf{Y}_l\,|\,\mathbf{X}^{(k)};\, \boldsymbol{\Omega}_0)$, for $k = 1, 2$.

Because it is extremely tedious to analytically derive the $2^{n+1}$ probabilities, $\pi^{(k)}(\mathbf{Y}_l)$, for $k = 1, 2$ and $l = 1,\ldots,2^n$, we resort to the Monte Carlo method described in Heagerty and Kurland (2001). This method estimates each of the two sets of probabilities, $\{\pi^{(1)}(\mathbf{Y}_l)\}_{l=1}^{2^n}$ and $\{\pi^{(2)}(\mathbf{Y}_l)\}_{l=1}^{2^n}$, via a random sample of size $Q$. For instance, to estimate $\pi^{(1)}(\cdot)$, we generate $Q$ vectors of clustered response from the true GLMM evaluated at $\mathbf{X}^{(1)}$, and use the sample proportion of $\mathbf{Y}_l$ to estimate $\pi^{(1)}(\mathbf{Y}_l)$, for $l = 1,\ldots,2^n$. In the subsequent results, we set $n = 8$ and $Q = 10^9$. To perform the necessary integration, we use a 50-point Gauss–Hermite quadrature to approximate equation (2) when computing $f_{\mathbf{Y}_l}(\mathbf{Y}_l\,|\,\mathbf{X}^{(k)};\boldsymbol{\Omega})$. In order to determine the size of $Q$ to ensure desired precision in estimating $\pi^{(k)}(\cdot)$, and to determine the number of quadrature points needed to achieve a reasonable approximation to the integral, we experiment on a fifth case, case (V), where the true and the assumed random-effect models coincide, that is, $b_{ij} = b_{i0} \sim N(0,\, \sigma_0^2)$, where $\sigma_0 = 3$. Obviously, $\widetilde{\boldsymbol{\Omega}} \equiv \boldsymbol{\Omega}_0$ in case (V). When $n = 8$, with $Q = 10^9$ and with 50 quadrature points in the Gauss–Hermite quadrature approximation, we find the difference between the $\widetilde{\boldsymbol{\Omega}}$ obtained from this algorithm and $\boldsymbol{\Omega}_0$ to be virtually negligible, suggesting that the algorithm produces very accurate and precise solutions to equation (6).

Applying the same algorithm on the reconstructed data with $G = 2$, we compute $\widetilde{\boldsymbol{\Omega}}^*$ by maximizing

$$\sum_{l=1}^{2^G} \left\{ \pi^{(1)}\left(\mathbf{Y}_l^*\right) \log f_{\mathbf{Y}_l^*}^*\left(\mathbf{Y}_l^* \,\middle|\, \mathbf{X}^{(1)}; \boldsymbol{\Omega}\right) + \pi^{(2)}\left(\mathbf{Y}_l^*\right) \log f_{\mathbf{Y}_l^*}\left(\mathbf{Y}_l^* \,\middle|\, \mathbf{X}^{(2)}; \boldsymbol{\Omega}\right) \right\}, \tag{10}$$

where $f_{\mathbf{Y}_l^*}^*(\mathbf{Y}_l^* \,|\, \mathbf{X}^{(k)}; \boldsymbol{\Omega})$, for $k = 1, 2$, is defined in equation (3). Strictly speaking, because the true probabilities that specify the distribution of $\mathbf{Y}$ given $\mathbf{X}, \pi^{(k)}(\cdot)$, are estimated, $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}^*$ so obtained are still estimators instead of the limiting MLEs that solve equations (6) and (7). But as reinforced by our findings from case (V), the algorithm yields only very little variability, thus the solutions to equations (6) and (7) found from this algorithm are close enough to the true values of $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}^*$ to truly reflect the impact of model misspecification.

Table 1 presents $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}^*$ under cases (I)–(V). According to Table 1, except for case (V) where $\widetilde{\boldsymbol{\Omega}} = \widetilde{\boldsymbol{\Omega}}^* = \boldsymbol{\Omega}_0$, most of the elements in $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}^*$ exhibit moderate to large bias. More importantly, in each case of model misspecification, $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}^*$ are affected differently. The test statistics defined in equations (4) and (5) are constructed to assess how much $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}^*$ differ, overall or elementwise, and by so doing, detect model misspecification.

## 4. Finite Sample Performance

### 4.1 *Simulation Study*

We now present the finite-sample performance of the proposed test statistics via simulation. In the simulation, 300 Monte Carlo replicated data sets are generated from the logistic model in equation (8) with the random effects generated according to cases (I)–(V). Each data set consists of $m = 300$ clusters, each of size $n = 8$. To create the reconstructed data, we first sort the subjects within a cluster by the values of $X_{ij,2}$, then we divide the sorted data in each cluster into $G$ equal subgroups. This subgroup composition, referred to as homogeneous composition henceforth, maximizes the between-subgroup variation and yields more efficient MLEs based on the reconstructed data. We set $G = 2$ for all cases except for case (IV), where we set $G = 4$. The choice of $G$ will be elaborated in the next subsection.

Based on each Monte Carlo replicate, we compute $\widehat{\boldsymbol{\Omega}}$, $\widehat{\boldsymbol{\Omega}}^*$, and the proposed test statistics. The Monte Carlo averages of $\widehat{\boldsymbol{\Omega}}$ and $\widehat{\boldsymbol{\Omega}}^*$ (not shown) resemble the results in Table 1. The empirical powers and sizes of the test statistics are presented in Table 2, with significance level equal to 0.05. The results from case (V) suggest that the test statistics have sizes close to the nominal level. When misspecification occurs, $T^2$ shows promising power, and at least one of the $t_\theta$'s tends to be significant. Furthermore, combining the results in Tables 1 and 2, it appears that $t_\theta$ tends to be significant more often when $\widetilde{\theta}$ deviates further from the truth.

**Table 1**

*The limiting MLEs based on the observed data and the reconstructed data when assuming normal random intercept, $b_{ij} = b_{i0} \sim N(0, \sigma_0^2)$, in the logistic model in equation* (8)*. Five cases, (I)–(V), of the true random-effect distribution are considered. The numbers in parentheses are the associated relative bias defined by $100 \times (\breve{\theta} - \theta_0)/\theta_0$, where $\breve{\theta}$ denotes the limiting MLE for a parameter in the assumed random-intercept logistic model, and $\theta_0$ is the true value. The true values of the regression parameters are $\beta_0 = -2, \beta_1 = 1$, $\beta_2 = 0.5$, and $\beta_3 = -0.25$.*

| | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\sigma_0$ |
|---|---|---|---|---|---|
| (I) | | | | | |
| Observed data | −2.70 (35.20) | 1.03 (3.19) | 0.48 (−3.40) | −0.23 (−7.47) | 3.16 (5.47) |
| Reconstructed | −2.61 (30.39) | 0.91 (−8.54) | 0.44 (−11.10) | −0.19 (−23.45) | 2.34 (−21.90) |
| (II) | | | | | |
| Observed data | −1.48 (−26.20) | 0.27 (−72.66) | 0.39 (−21.30) | −0.10 (−58.70) | 1.58 (–) |
| Reconstructed | −1.85 (−7.42) | 1.08 (7.97) | 0.32 (−36.90) | 0.06 (−123.31) | 1.52 (–) |
| (III) | | | | | |
| Observed data | −2.31 (15.41) | 1.15 (14.79) | 1.22 (144.69) | −0.60 (139.96) | 1.02 (–) |
| Reconstructed | −1.99 (−0.67) | 1.17 (17.17) | 0.50 (0.87) | −0.76 (205.64) | 0.84 (–) |
| (IV) | | | | | |
| Observed data | −0.98 (−50.80) | 0.49 (−50.54) | 0.31 (−38.31) | −0.16 (−37.86) | 0.74 (−25.67) |
| Reconstructed | −1.06 (−46.78) | 0.46 (−54.41) | 0.25 (−50.19) | −0.15 (−41.16) | 0.60 (−79.99) |
| (V) | | | | | |
| Observed data | −2.00 (0.00) | 1.00 (0.00) | 0.50 (0.00) | −0.25 (0.00) | 3.00 (0.00) |
| Reconstructed | −2.00 (0.00) | 1.00 (0.00) | 0.50 (0.00) | −0.25 (0.00) | 3.00 (0.00) |

(I): $b_{ij} = b_{i0} = \sigma(a_i - \lambda)/\sqrt{\lambda}$, where $a_i \sim$ gamma$(\lambda, 1), \sigma = 3$, and $\lambda = 1$.
(II): $b_{ij} = b_{i0}$, where $[b_{ij} \,|\, X_{ij,1} = 0] \sim N(0, \sigma_{00}^2)$ and $[b_{ij} \,|\, X_{ij,1} = 1] \sim N(0, \sigma_{01}^2)$, with $\sigma_{00} = 3$ and $\sigma_{01} = 0.5$.
(III): $b_{ij} = b_{i0} + b_{i1} X_{ij,2}$, where $b_{i0} \sim N(0, \sigma_0^2), b_{i1} \sim N(0, \sigma_1^2), \sigma_0 = 0.5, \sigma_1 = 2$, and $b_{i0}$ is independent of $b_{i1}$.
(IV): Random effects are autocorrelated with $\text{cov}(b_{ij}, b_{ik}) = \sigma^2 \rho^{|j-k|}$, where $\sigma = 3$ and $\rho = 0.5$.
(V): $b_{ij} = b_{i0} \sim N(0, \sigma_0^2)$, where $\sigma_0 = 3$.

**Table 2**

*The empirical powers and sizes of $t_\theta$ and $T^2$ from 300 Monte Carlo replicated data sets, each with $m = 300$ clusters, and cluster size $n = 8$. Assume a normal random-intercept logistic model. The true random-effect models are specified in cases (I)–(V).*

| | $t_{\beta_0}$ | $t_{\beta_1}$ | $t_{\beta_2}$ | $t_{\beta_3}$ | $t_{\sigma_0}$ | $T^2$ |
|---|---|---|---|---|---|---|
| (I) | 0.04 | 0.03 | 0.05 | 0.06 | **0.72** | **0.66** |
| (II) | **0.70** | **0.87** | 0.07 | 0.06 | 0.06 | **1** |
| (III) | **0.61** | 0.04 | **0.76** | 0.08 | **0.19** | **0.95** |
| (IV) | **0.22** | 0.05 | 0.05 | 0.04 | **0.64** | **0.99** |
| (V) | 0.03 | 0.05 | 0.05 | 0.06 | 0.04 | 0.03 |

(I): $b_{ij} = b_{i0} = \sigma(a_i - \lambda)/\sqrt{\lambda}$, where $a_i \sim$ gamma($\lambda$, 1), where $\sigma = 3$, and $\lambda = 1$.

(II): $b_{ij} = b_{i0}$, where $[b_{ij} \mid X_{ij,1} = 0] \sim N(0, \sigma_{00}^2)$ and $[b_{ij} \mid X_{ij,1} = 1] \sim N(0, \sigma_{01}^2)$, with $\sigma_{00} = 3$ and $\sigma_{01} = 0.5$.

(III): $b_{ij} = b_{i0} + b_{i1}X_{ij,2}$, where $b_{i0} \sim N(0, \sigma_0^2), b_{i1} \sim N(0, \sigma_1^2)$, $\sigma_0 = 0.5, \sigma_1 = 2$, and $b_{i0}$ is independent of $b_{i1}$.

(IV): Random effects are autocorrelated with $\text{cov}(b_{ij}, b_{ik}) = \sigma^2\rho^{|j-k|}$, where $\sigma = 3$ and $\rho = 0.5$.

(V): $b_{ij} = b_{i0} \sim N(0, \sigma_0^2)$, where $\sigma_0 = 3$.

Note: Powers greater than 0.10 are in boldface.

It is not meaningful to compare the MLEs or the test statistics among different types of model misspecification because the influences of different model misspecifications are not always comparable. For instance, one cannot conclude, by comparing the results from cases (I) and (II) in Table 2, that the test statistics have more power to detect the second type of misspecification than the first type. Within each of the four types of misspecification, we monitor the changes in $t_\theta$ and $T^2$ as the misspecification becomes more severe. In what follows, we report in detail the results for case (III), when there is a random coefficient $b_{i1}$ for the within-cluster covariate in the true model. Fixing the variance component for $b_{i0}$ at $\sigma_0^2 = 0.25$, we raise the variance component for $b_{i1}$ by increasing $\sigma_1$ from 0.5 to 3 so that the assumed random-intercept model deviates further from the true model gradually. The observed empirical powers of the test statistics when $m = 100$ and 300 are given in Table 3. It is evident from Table 3 that, as the misspecification becomes more severe, the power of $T^2$ increases quickly, so do the powers of $t_{\beta_0}$ and $t_{\beta_2}$, even when the sample size is moderate. On the other hand, when the misspecification has only a small effect on the estimators, the test statistics are much less significant. We have observed the same phenomenon for the other three cases, except that the pattern of $t_\theta$'s differs from case to case, where the pattern is in terms of which $t_\theta$ tends to be more significant. Overall, the study on the power suggests that $T^2$ can have success in detecting random-effect model misspecification, and $t_\theta$ can distinguish among different types of misspecification.

**Table 3**

*The empirical powers of $t_\theta$ and $T^2$ from 300 Monte Carlo replicated data sets, each with $m = 100$ or 300 clusters, and cluster size $n = 8$. Assume normal random-intercept logistic model. The true random-effect model is given by $b_{ij} = b_{i0} + b_{i1}X_{ij,2}$, where $b_{i0} \sim N(0, 0.5^2), b_{i1} \sim N(0, \sigma_1^2)$, and $b_{i0}$ is independent of $b_{i1}$.*

| $m$ | $\sigma_1$ | $t_{\beta_0}$ | $t_{\beta_1}$ | $t_{\beta_2}$ | $t_{\beta_3}$ | $t_{\sigma_0}$ | $T^2$ |
|---|---|---|---|---|---|---|---|
| 100 | 0.5 | 0.04 | 0.05 | 0.06 | 0.06 | 0.14 | 0.14 |
| | 1 | 0.08 | 0.06 | 0.09 | 0.06 | 0.11 | 0.15 |
| | 2 | 0.31 | 0.05 | 0.35 | 0.04 | 0.05 | 0.49 |
| | 3 | 0.56 | 0.04 | 0.69 | 0.06 | 0.12 | 0.84 |
| 300 | 0.5 | 0.06 | 0.06 | 0.07 | 0.06 | 0.08 | 0.13 |
| | 1 | 0.13 | 0.03 | 0.19 | 0.04 | 0.03 | 0.23 |
| | 2 | 0.57 | 0.04 | 0.74 | 0.07 | 0.20 | 0.94 |
| | 3 | 0.94 | 0.04 | 0.98 | 0.09 | 0.32 | 1 |

### 4.2 *Implementation Details*

The results in Table 2 reveal that the pattern regarding the magnitude of $t_\theta$'s depends on the nature of the model misspecification. When the distribution family of a random effect is misspecified, the $t_\theta$ for the corresponding variance component tends to be significant. If the variance of the random intercept depends on a covariate, or a random slope for a covariate is missing from the assumed model, then the $t_\theta$ associated with that covariate will stand out as being significant. Lastly, significant $t_\theta$'s for the fixed intercept and the variance components can be evidence of misspecifying the correlation structure of the random effects. Based on such knowledge, we propose a two-step diagnostic method to detect random-effect model misspecification. In the first step, one tests globally the existence of model misspecification via $T^2$. If $T^2$ is not significant, then one may conclude lack of sufficient evidence of model misspecification. Otherwise, one executes the second step, where individual $t_\theta$ is inspected. The pattern of the $t_\theta$ values will provide clues regarding the type of model misspecification.

The operating characteristics of the proposed test statistics are affected by the subgroup composition. Because the data reconstruction causes loss in information, which can lead to unreliable inference and degrade the proposed testing procedure, we recommend use of homogeneous composition whenever possible to minimize the information loss. As for the number of subgroups in a cluster, $G$, we find in cases (I)–(IV) that smaller $G(\geq 2)$ results in larger discrepancy between $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}^*$. Therefore, with moderate or large samples, we suggest $G = 2$ in order to magnify the effect of model misspecification. One exception is that, when the random effects are autocorrelated, larger $G(<n)$ leads to higher power especially with small or moderate samples. This is expected in finite samples because to detect special correlation structure within a cluster requires more information per cluster. In summary, when creating the reconstructed data, unless one suspects autocorrelated random effects, or when $m$ or $n$ is small, one should set $G = 2$ to maximize the power of the tests, and homogeneous composition is always preferable. To preserve the nominal size of the tests depends more on the size of $m$ than the choice of $G$. Under the current simulation setting, the type I errors of the tests remain close to the nominal level when $m > 50$. As a preliminary check of whether $m$ is large enough for the testing procedure to be reliable, one can estimate the variance of $\widehat{\boldsymbol{\Omega}}$ and $\widehat{\boldsymbol{\Omega}}^*$ first to see whether or not they are terribly variable.

We explore an alternative strategy of creating reconstructed data where we keep a fraction of the data within each cluster. Note that the likelihood for the reconstructed data so obtained has the same functional form as that for the observed data. In contrast, the likelihood function for our

reconstructed data and the likelihood for the observed data have different functional forms. Such nontrivial difference results in $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}^*$ far more distinct than those resulting from the alternative strategy when there exists model misspecification. And substantial distinction between $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}^*$ is the key to detecting model misspecification. Simulation studies (not shown) show that the power of the test statistics under the alternative strategy is much lower than that under the strategy used in this article.

## 5. Additional Fixed-Effect Misspecification

Besides misspecifying the random effects, one may as well misspecify the fixed-effect part of the GLMMs. To investigate the characteristics of the test statistics in the presence of both sources of misspecification, we design experiments where different types of fixed-effect misspecification interact with the random-effect specification given by cases (I)–(V). Examples of fixed-effect misspecification considered in our experiments include misspecifying the functional form of a covariate, missing a between- or within-cluster covariate, and missing the interaction of two covariates. In all the interactions we have studied, we observe amazing robustness of the test statistics to the additional fixed-effect misspecification. One noteworthy phenomenon is that, when the random effect is correctly specified as in case (V), $T^2$ and all $t_\theta$'s remain mostly insignificant despite the fixed-effect misspecification.

Figure 1 depicts the results from one of these experiments, where we consider the random-effect specification in cases

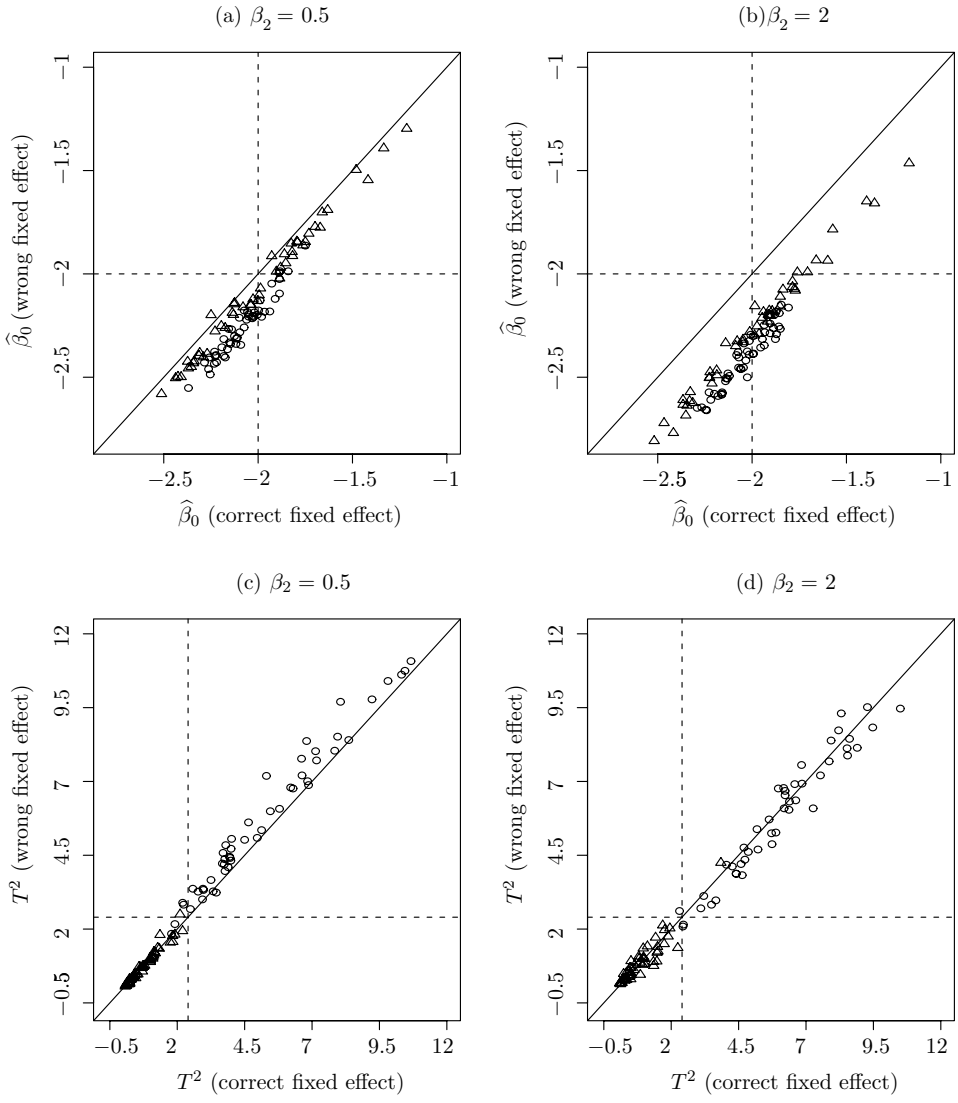


**Figure 1.** Plots (a) and (b) are $\widehat{\beta}_0$ under the wrong fixed-effect specification versus $\widehat{\beta}_0$ under the correct fixed-effect specification for case (III) [○] and case (V) [△]. Plots (c) and (d) are $T^2$ for the wrong fixed-effect specification versus $T^2$ for the correct fixed-effect specification under case (III) [○] and case (V) [△]. The dashed reference lines in (a) and (b) are at the true value of $\beta_0$. The dashed reference lines in (c) and (d) are at the 95th percentile of $F(4, 296)$. The solid diagonal lines in all plots are the lines with slope one and intercept zero. The plotted results are from 50 Monte Carlo replications randomly selected from a total of 300 Monte Carlo replications.

(III) and (V), and assume the conditional mean model given by

$$E(Y_{ij} \,|\, \mathbf{X}_{ij},\, b_{ij}) = \{1 + \exp(-\beta_0 - \beta_1 X_{ij,1} - \beta_2 X_{ij,2} - b_{ij})\}^{-1},$$

whereas the truth is $E(Y_{ij} \,|\, \mathbf{X}_{ij}, b_{ij}) = \{1 + \exp(-\beta_0 - \beta_1 X_{ij,1} - \beta_2 X_{ij,2}^2 - b_{ij})\}^{-1}$, where $\beta_0 = -2, \beta_1 = 1$, and $\beta_2 = 0.5$ or 2. Figure 1a and b show that, whether or not the random-effect model is misspecified, misspecifying the fixed effect leads to biased MLEs. However, Figure 1c and d suggest that fixed-effect misspecification has little impact on $T^2$. More interestingly, under the correct model for the random effect, almost all $T^2$ fall below the critical value even when the fixed effect is misspecified.

A closer look at $\widetilde{\boldsymbol{\Omega}}$ and $\widetilde{\boldsymbol{\Omega}}^*$ reveals that the impact of random-effect model misspecification on $\widetilde{\boldsymbol{\Omega}}^* - \widetilde{\boldsymbol{\Omega}}$ dominates that of fixed-effect misspecification. That is, if the fixed effect is misspecified, $\widehat{\boldsymbol{\Omega}}$ and $\widehat{\boldsymbol{\Omega}}^*$ change more similarly than the way they change due to random-effect model misspecification. Consequently, the test statistics are usually robust to fixed-effect misspecification. This allows one to test the random-effect specification and fixed-effect specification separately by first using our diagnostic method to check the random-effect assumptions, then applying other tests on the fixed effects. The diagnosis in the first step is fairly robust to the fixed-effect specification yet to be justified in the second step.

## 6. Application to Respiratory Infection Data

We now apply the proposed diagnostic method to the data analyzed using semiparametric regression in Lin and Carroll (2001). The data are from a study where preschool children were examined every three months for 18 months for the presence of respiratory infection, which recorded each child's age at the beginning of the study, gender, height, season when the examination took place, presence of respiratory infection, etc. The subsequent analyses use a subset of the data from 192 children who were examined at least four times in the study. The response variable of interest is the binary variable, $Y_{ij}$, which equals one if child $i$ had symptoms of respiratory infection during examination $j$, and zero otherwise, for $i = 1, \ldots, 192$ and $j = 1, \ldots, n_i$, where $n_i$ ranges from 4 to 6.

Based on our preliminary analysis and the analysis presented in Lin and Carroll (2001), we first posit a conditional mean model given by, for $i = 1, \ldots, 192$ and $j = 1, \ldots, n_i, E(Y_{ij} = 1 \,|\, b_{i0}) = \{1 + \exp(-\beta_0 - \beta_1 \text{bslage}_i - \beta_2 \text{season}_{ij} - b_{i0})\}^{-1}$, where "bslage" is defined as $[\{\text{baseline age (in months)} - 36\}/12]^3$, and $b_{i0}$ is the normal random intercept with variance $\sigma_0^2$. The reconstructed data have $G_i = 2$ subgroups within each child, for $i = 1, \ldots, 192$, created using homogeneous composition according to season. Depending on $n_i$, the subgroup size can be 2 or 3. The MLEs, $\widehat{\boldsymbol{\Omega}}$ and $\widehat{\boldsymbol{\Omega}}^*$, and the test statistics are computed for this normal random-intercept logistic model. The $p$-value for the resultant $T^2$ is 0.01, indicating strongly that there exists model misspecification. Moreover, the values of $t_{\beta_0}$ and $t_{\beta_2}$ are highly significant, with $p$-values 0.001 and 0.008, respectively. This leads us to consider another logistic model with a random slope for "season" given by

$$E(Y_{ij} = 1 \,|\, b_{ij}) = [1 + \exp\{-\beta_0 - \beta_1 \text{bslage}_i - (\beta_2 + b_{i1})\text{season}_{ij} - b_{i0}\}]^{-1}, \tag{11}$$

where $b_{i1} \sim N(0, \sigma_1^2)$. We find $\widehat{\sigma}_0^2$ under equation (11) nearly zero. Hence we drop $b_{i0}$ from equation (11) and fit a normal random-slope logistic model. This model results in a $T^2$ with $p$-value 0.09. Using 0.05 as the significance level, we conclude that there is not sufficient evidence of model misspecification for the normal random-slope model. We also conduct the variance component test developed by Lin (1997). Lin's test for the normal random-slope model suggests that the variance component for the random slope is highly significant, with $p$-value less than 0.001. Table 4 presents the MLEs and the test statistics from the analyses on the normal random-intercept model and the normal random-slope model.

It is worth pointing out that, in Table 4, the values of $t_{\beta_0}$ and $t_{\beta_2}$ in the normal random-slope model are much less significant than their counterparts in the normal random-intercept model, but they still exceed in absolute value the critical value ($\approx$1.97) at 0.05 significance level. However, we do not view this as sufficient evidence of model misspecification because the global test statistic $T^2$ for the normal random-slope model is not significant. If one intends to conclude model misspecification if at least one $t_\theta$ is significant, then one is conducting multiple comparisons. To control the familywise type I error in multiple comparisons at 0.05, the critical value for $t_\theta$ should be higher (in absolute value) than 1.97. We do not pursue the issue of multiple comparisons in this article.

If one is concerned about the significant $t_{\beta_0}$ and $t_{\beta_2}$, or the nearly significant $T^2$ for the normal random-slope model, one may continue to search for more appropriate assumed model of the random coefficient for "season." For instance, we explore the assumed model for $b_{i1}$ specified by the first-order seminonparametric density (Chen et al., 2002), which is given by $f_{b_{i1}}(b_{i1}; \boldsymbol{\tau}) = \{a_0 + a_1\eta^{-1}(b_{i1} - \xi)\}^2\eta^{-1} \times \phi\{\eta^{-1}(b_{i1} - \xi)\}$, where $\phi(\cdot)$ is the standard normal density function, $a_0 = \sin(\omega), a_1 = \cos(\omega)$, $\omega \in (-\pi/2, \pi/2], \eta > 0$, $\boldsymbol{\tau} = (\omega, \eta)^T$, and lastly, $\xi = -2\eta a_0 a_1$ so that $E(b_{i1}) = 0$. The estimator for $\sigma_1^2 = \text{var}(b_{i1})$ is a function of the estimated $\boldsymbol{\tau}$, and we use the Delta method to obtain a variance estimator for $\widehat{\sigma}_1^2$, and also the variance estimator $\widehat{\nu}^2$ needed in $t_{\sigma_1^2}$. As shown in Table 4, none of the $t_\theta$'s is significant when the assumed model for the random slope is more flexible.

## 7. Discussion

We focus on clustered binary response data and propose a two-step diagnostic method to detect random-effect model misspecification in GLMMs. This method utilizes both the observed data and a reconstructed data created from the observed data. It will fail if $\widehat{\boldsymbol{\Omega}}$ and $\widehat{\boldsymbol{\Omega}}^*$ are both inconsistent due to model misspecification yet they converge to the same limit as $m \to \infty$. We have not encountered such a case so far. We have investigated a wide range of GLMMs relevant in practice, and found that the proposed method can be very effective in detecting random-effect model misspecification, and moreover, in directing model selection.

**Table 4**

*The MLEs in the logistic mixed model for the indicator of presence of respiratory infection and the test statistics. The numbers in parentheses next to the MLEs are the sandwich-type estimated standard errors for the MLEs. The numbers in parentheses next to the test statistics are the associated p-values. The notation "SNP" refers to the first-order seminonparametric.*

| | | Observed data | Reconstructed data | $t_\theta$ | $T^2$ |
|---|---|---|---|---|---|
| Normal | $\beta_0$ | −2.57 (0.33) | −1.88 (0.40) | 3.20 (0.001) | 3.24 (0.01) |
| Random intercept | $\beta_1$ | −0.04 (0.02) | −0.04 (0.02) | −1.08 (0.281) | |
| Logistic model | $\beta_2$ | −0.09 (0.10) | −0.38 (0.16) | 2.66 (0.008) | |
| | $\sigma_0^2$ | 0.89 (0.45) | 0.90 (0.59) | 0.01 (0.995) | |
| Normal | $\beta_0$ | −2.07 (0.27) | −1.47 (0.39) | 2.34 (0.02) | 2.08 (0.09) |
| Random slope | $\beta_1$ | −0.04 (0.02) | −0.04 (0.02) | 1.24 (0.21) | |
| Logistic model | $\beta_2$ | −0.29 (0.14) | −0.59 (0.21) | −2.06 (0.04) | |
| | $\sigma_1^2$ | 0.12 (0.07) | 0.18 (0.12) | 0.66 (0.51) | |
| SNP | $\beta_0$ | −2.05 (0.29) | −1.40 (0.47) | 1.70 (0.09) | 1.81 (0.11) |
| Random slope | $\beta_1$ | −0.04 (0.02) | −0.04 (0.02) | 0.98 (0.33) | |
| Logistic model | $\beta_2$ | −0.34 (0.22) | −0.70 (0.40) | −1.10 (0.27) | |
| | $\sigma_1^2$ | 0.22 (0.26) | 0.37 (1.01) | 0.15 (0.88) | |

The intriguing robustness of the proposed test statistics to fixed-effect misspecification calls for more thorough exploration on the property of MLE under model misspecification with different data structure. Better understanding of this may suggest ways to improve the testing procedure, and even lead to a more sophisticated way to detect random-effect model misspecification *and* fixed-effect misspecification separately in a unified framework.

We are currently investigating generalization of the proposed method to the nonlinear mixed models for other types of nonnormal response. The use of reconstructed data is a novel idea that we have not seen being studied in the literature. This idea has the potential to test statistical assumptions that have been claimed to be "not testable" due to lack of observed data on latent or missing quantities, such as assumptions on missing data mechanism. In conclusion, the use of reconstructed data is a topic worth further investigation.

## 8. Supplementary Materials

Web Appendix A referenced in Section 3 is available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

## Acknowledgements

The author is grateful to the editor, the associate editor, the referee, Dr Joshua M. Tebbs, and Dr Barry K. Moser for their constructive and insightful comments.

## References

Agresti, A., Caffo, B., and Ohman-Strickland, P. (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Computational Statistics and Data Analysis* **47,** 639–653.

Chen, J., Zhang, D., and Davidian, M. (2002). A Monte Carlo EM algorithm for generalized linear models with flexible random effects distribution. *Biostatistics* **3,** 347–360.

Heagerty, P. J. and Kurland, B. F. (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* **88,** 973–985.

Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distribution assumptions in econometric models for duration data. *Econometrica* **52,** 271–320.

Lin, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika* **84,** 309–326.

Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data. *Biometrika* **88,** 1179–1185.

Litière, S., Alonso, A., and Molenberghs, G. (2007). Type I and type II error under random-effects misspecification in generalized linear mixed models. *Biometrics* **63,** 1038–1044.

Magder, L. S. and Zeger, S. L. (1996). A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians. *Journal of the American Statistical Association* **91,** 1141–1151.

Neuhaus, J. M., Hauck, W. W., and Kalbfleisch, J. D. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika* **79,** 755–762.

Rotnitzky, A. and Wypij, D. (1994). A note on the bias of estimators with missing data. *Biometrics* **50,** 1163–1170.

White, H. (1981). Consequence and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association* **76,** 419–433.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50,** 1–25.

Wu, Y., Genton, M. G., and Stefanski, L. A. (2006). A multivariate two-sample mean test for small sample size and missing data. *Biometrics* **62,** 877–885.

*Received October* 2007. *Revised April* 2008.
*Accepted April* 2008.