Biometrics 61, 1000–1009
December 2005
DOI: 10.1111/j.1541-0420.2005.00365.x

Goodness-of-Fit Methods for Generalized Linear Mixed Models

Zhiying Pan and D. Y. Lin*

Department of Biostatistics, University of North Carolina, CB 7420, McGavran-Greenberg Hall, Chapel Hill, North Carolina 27599-7420, U.S.A.

**email:* lin@bios.unc.edu

Summary. We develop graphical and numerical methods for checking the adequacy of generalized linear mixed models (GLMMs). These methods are based on the cumulative sums of residuals over covariates or predicted values of the response variable. Under the assumed model, the asymptotic distributions of these stochastic processes can be approximated by certain zero-mean Gaussian processes, whose realizations can be generated through Monte Carlo simulation. Each observed process can then be compared, both visually and analytically, to a number of realizations simulated from the null distribution. These comparisons enable one to assess objectively whether the observed residual patterns reflect model misspecification or random variation. The proposed methods are particularly useful for checking the functional form of a covariate or the link function. Extensive simulation studies show that the proposed goodness-of-fit tests have proper sizes and are sensitive to model misspecification. Applications to two medical studies lead to improved models.

Key words: Linear mixed models; Longitudinal data; Model checking; Model misspecification; Random effects; Regression diagnostics; Repeated measures; Residual plots.

1. Introduction

Generalized linear mixed models (GLMMs) (Breslow and Clayton, 1993) are obtained from generalized linear models (GLMs) (McCullagh and Nelder, 1989) by incorporating random effects into the linear predictors, and include the well-known linear mixed models (LMMs) for normal responses (Laird and Ware, 1982) as a special case. These models are useful for modeling the dependence among response variables inherent in longitudinal or repeated measures studies, for accommodating overdispersion among binomial or Poisson responses, and for producing shrinkage estimators in multiparameter problems, such as the construction of maps of small area disease rates (Breslow and Clayton, 1993). Recent developments in model fitting algorithms and the implementation of methods in statistical packages (e.g., SAS PROC NLMIXED) have greatly facilitated the applications of these models. Although model misspecification can have adverse effects on the statistical inference under GLMMs, model-checking techniques have not been carefully explored.

The most commonly used method for selecting fixed effects is to test for the significance of additional terms in embedded models. Other methods for model selection include Akaike's (1974) information criterion (AIC) and the Bayesian information criterion (Schwartz, 1978). Vonesh, Chinchilli, and Pu (1996) proposed the model concordance correlation coefficient, which is a generalization of the R^2 measure for linear models, to assess the overall adequacy of the response function.

Residual plots are routinely used to assess the adequacy of regression models for independent responses (e.g., Cook and Weisberg, 1994). It is often difficult to determine whether the observed pattern reflects model misspecification or random fluctuation. This kind of graphical assessment is even more challenging with dependent responses due to the correlatedness of the residuals. Furthermore, such plots are uninformative for binary data because all the points lie on one of the two curves according to the two possible values of the response.

In this article, we develop objective and informative procedures for assessing the adequacy of GLMMs by taking the cumulative sums of residuals over covariates or predicted values. We approximate the null distributions of these cumulative-sum processes with certain zero-mean Gaussian processes, whose distributions can be generated by Monte Carlo simulation. One may then compare the observed residual pattern with a few realizations from the null distribution, so as to assess how unusual the observed pattern is. To make the inspection even more objective, one may calculate the P-value for the Kolmogorov-type supremum test based on a large number of realizations from the null distribution. Similar methods have been developed for GLMs with independent responses (Su and Wei, 1991), for the proportional hazards model with censored survival data (Lin, Wei, and Ying, 1993), and for marginal GLMs with dependent responses (Lin, Wei, and Ying, 2002), and these methods have recently been implemented in SAS. The developments of the model-checking procedures for GLMMs are challenging because the existence of random effects not only complicates the theoretical derivations, such as the proofs for the asymptotic distributions of the cumulative-sum processes and for the consistency of the tests, but also imposes computational challenges.

The structure of the article is as follows. We provide a brief overview of GLMMs in Section 2. We describe the proposed

model-checking techniques in Section 3 while relegating the proofs of the theoretical results to the Appendices. We report some simulation results in Section 4 and provide applications to two medical studies in Section 5. We conclude with a few remarks in Section 6.

2. Generalized Linear Mixed Models

We consider longitudinal studies, in which repeated measures of a response variable are taken on a random sample of n subjects. For $i = 1, \ldots, n$, and $j = 1, \ldots, t_i$, let y_{ij} be the response of the ith subject on the jth occasion, and let $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$ be the corresponding $p \times 1$ and $q \times 1$ vectors of covariates associated with the fixed effects and random effects, respectively. The GLMM takes the form

$$g\{\mathrm{E}(y_{ij} \mid \mathbf{b}_i)\} = \mathbf{X}'_{ij}\boldsymbol{\beta} + \mathbf{Z}'_{ij}\mathbf{b}_i, \quad i = 1, \ldots, n; \quad j = 1, \ldots, t_i, \tag{1}$$

where $g(\cdot)$ is a known differentiable link function, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression parameters, and $\mathbf{b}_i$ is a $q \times 1$ vector of unobservable random effects for the ith subject.

The following assumptions are made: (i) the conditional distribution of y_{ij} given $\mathbf{b}_i$ follows a distribution from the exponential family with density function $f_{y|b}(y \mid \mathbf{b})$; (ii) conditional on $\mathbf{b}_i$, the repeated measures $y_{i1}, \ldots, y_{it_i}$ are independent; and (iii) the $\mathbf{b}_i$ are independently and identically distributed with density function $f_b(\mathbf{b})$. It is usually assumed that $\mathbf{b}_i$ follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{D}$ depending on a vector parameter $\boldsymbol{\alpha}$. Write $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}')'$.

The likelihood function for $\boldsymbol{\theta}$ is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int \prod_{j=1}^{t_i} f_{y|b}(y_{ij} \mid \mathbf{b}_i) f_b(\mathbf{b}_i)\, \mathrm{d}\mathbf{b}_i.$$

In general, a full maximum likelihood analysis requires numerical integration for calculating the log likelihood, score function, and information matrix. Various approximations and Gibbs sampler methods (Zeger and Karim, 1991; Breslow and Clayton, 1993) have also been proposed to fit such models. The maximum likelihood estimator of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}$, is the solution to $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}$, where

$$\mathbf{U}(\boldsymbol{\theta}) = \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \mathbf{U}_i(\boldsymbol{\theta}),$$

and $\mathbf{U}_i$ involves only the observations from the ith subject. Under mild conditions,

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \boldsymbol{\Omega}^{-1} n^{-1/2} \mathbf{U}(\boldsymbol{\theta}) + o_p(1), \tag{2}$$

where $\boldsymbol{\Omega} = \lim_{n\to\infty} \mathcal{I}(\boldsymbol{\theta})$, and $\mathcal{I}(\boldsymbol{\theta}) = -n^{-1}\partial^2 \log L(\boldsymbol{\theta}) \big/ \partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$. In addition, $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically zero-mean normal with covariance matrix $\boldsymbol{\Omega}^{-1}$.

3. Model-Checking Techniques

Under model (1), the marginal means of $y_{ij} (i = 1, \ldots, n;\ j = 1, \ldots, t_i)$ are given by

$$\mathrm{E}(y_{ij}) = \mathrm{E}\{\mathrm{E}(y_{ij} \mid \mathbf{b}_i)\} = \int g^{-1}(\mathbf{X}'_{ij}\boldsymbol{\beta} + \mathbf{Z}'_{ij}\mathbf{b}_i) f_b(\mathbf{b}_i)\, \mathrm{d}\mathbf{b}_i. \tag{3}$$

We denote the above quantities by $m_{ij}(\boldsymbol{\theta})$ or m_{ij}. Given $\hat{\boldsymbol{\theta}}$, the predicted values of the responses are $\hat{m}_{ij} \equiv m_{ij}(\hat{\boldsymbol{\theta}})$, and the residuals are then defined as $e_{ij} = y_{ij} - \hat{m}_{ij}$ $(i = 1, \ldots, n;\ j = 1, \ldots, t_i)$. We obtain the residuals by numerical integration.

As mentioned earlier, it is difficult to interpret the plots of individual residuals because their variabilities are unknown. By contrast, it is possible to ascertain the distributions of certain aggregates of residuals. In this article, we consider the cumulative sums of residuals with respect to covariate values or predicted values. These cumulative sums are special cases of the following two classes of stochastic processes:

$$W(\mathbf{x}) = n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{t_i} I(\mathbf{X}_{ij} \le \mathbf{x}) e_{ij}, \tag{4}$$

$$W_g(r) = n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{t_i} I(\hat{m}_{ij} \le r) e_{ij}, \tag{5}$$

where $\mathbf{x} = (x_1, \ldots, x_p)' \in \Re^p$, $r \in \Re$, $I(\mathbf{X}_{ij} \le \mathbf{x}) = I(X_{1ij} \le x_1, \ldots, X_{pij} \le x_p)$, and X_{kij} is the kth component of $\mathbf{X}_{ij}$ $(k = 1, \ldots, p)$.

Define the null hypothesis H_0 as the correct specification of model (1). We focus on the deterministic components of the model and assume that the random components, including the error distribution and the random effects, are correctly specified. Under H_0, the aforementioned cumulative-sum processes are expected to fluctuate around 0, so that an unusually large value of $\sup_{\mathbf{x}}|W(\mathbf{x})|$ or $\sup_r |W_g(r)|$ would indicate model misspecification.

In Section 3.1, we show how to approximate the null distributions of W and W_g. In Sections 3.2–3.4, we take on the specific tasks of checking the functional form of a fixed effect, the link function, and the overall mean response function. In Section 3.5, we consider the special case of LMMs.

3.1 *Null Distributions of W and W_g*

It is shown in Appendix A that the cumulative-sum process $W(\mathbf{x})$ converges in distribution to a zero-mean Gaussian process under H_0. Define

$$\hat{W}(\mathbf{x}) = n^{-1/2} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{t_i} I(\mathbf{X}_{ij} \le \mathbf{x}) e_{ij} + \boldsymbol{\eta}'(\mathbf{x}; \hat{\boldsymbol{\theta}}) \mathcal{I}^{-1}(\hat{\boldsymbol{\theta}}) \mathbf{U}_i(\hat{\boldsymbol{\theta}}) \right\} G_i, \tag{6}$$

where $(G_1, \ldots, G_n)$ are independent standard normal variables that are independent of the data $(y_{ij}, \mathbf{X}_{ij}, \mathbf{Z}_{ij})$ $(i = 1, \ldots, n;\ j = 1, \ldots, t_i)$, and

$$\boldsymbol{\eta}(\mathbf{x}; \boldsymbol{\theta}) = -n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{t_i} I(\mathbf{X}_{ij} \le \mathbf{x}) \partial m_{ij}(\boldsymbol{\theta}) \big/ \partial\boldsymbol{\theta}. \tag{7}$$

It is shown in Appendix A that the conditional distribution of $\hat{W}(\mathbf{x})$ given the data $(y_{ij}, \mathbf{X}_{ij}, \mathbf{Z}_{ij})$ $(i = 1, \ldots, n; j = 1, \ldots, t_i)$ is the same in the limit as the unconditional distribution of $W(\mathbf{x})$ under H_0. To approximate the null distribution of $W(\mathbf{x})$, we obtain a large number of realizations of $\hat{W}(\mathbf{x})$ by repeatedly generating the normal random sample $(G_1, \ldots, G_n)$ while fixing the data $(y_{ij}, \mathbf{X}_{ij}, \mathbf{Z}_{ij})$ $(i = 1, \ldots, n; j = 1, \ldots, t_i)$ at their observed values. It is also shown in Appendix A that, for large samples, the distribution of $W_g(r)$ can be approximated by that of $\hat{W}_g(r)$, where $\hat{W}_g(r)$ is obtained from $\hat{W}(\mathbf{x})$ by replacing $I(\mathbf{X}_{ij} \le \mathbf{x})$ in (6) and (7) with $I(\hat{m}_{ij} \le r)$.

3.2 *Checking the Functional Form of a Fixed Effect*

To check the functional form of the kth component of the covariate vector $\mathbf{X}$, we consider the following process

$$W_k(x) = n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{t_i} I(X_{kij} \le x) e_{ij}, \quad k = 1, \ldots, p,$$

where $x \in \Re$. Note that $W_k(x)$ is a special case of $W(\mathbf{x})$ given in (4) with $x_l = \infty$ for all $l \ne k$. Thus, the null distribution of $W_k(x)$ can be approximated by the $\hat{W}_k(x)$ process, which is a special case of $\hat{W}(\mathbf{x})$ given in (6) with $x_l = \infty$ for all $l \ne k$.

To access visually how abnormal the observed process $w_k(\cdot)$ of $W_k(\cdot)$ is, one may plot $w_k(\cdot)$ along with a few realizations from the $\hat{W}_k(\cdot)$ process. To enhance the objectivity of this graphical procedure, one may complement it with the Kolmogorov-type supremum statistic $S_k \equiv \sup_x |W_k(x)|$. Denote the observed value of S_k by s_k. The P-value, $\Pr(S_k \ge s_k)$, can be approximated by $\Pr(\hat{S}_k \ge s_k)$, where $\hat{S}_k = \sup_x |\hat{W}_k(x)|$. The probability $\Pr(\hat{S}_k \ge s_k)$ can be estimated by generating a large number of realizations from $\hat{W}_k(x)$. As shown in Appendix B, the S_k test is generally consistent against misspecification of the functional form for X_k.

If the above procedure indicates misspecification of the functional form, then it would be desirable to obtain a more appropriate functional form. To this end, it is helpful to ascertain what the cumulative sum of residuals would look like when the functional form is misspecified in a given way. Lin et al. (2002) showed the mean functions of the cumulative sums of residuals for various patterns of misspecification of the functional form under GLMs. Those prototype residual patterns can be used for GLMMs as well.

3.3 *Checking the Link Function*

To check the link function, we consider the process $W_g(r)$ defined in (5). As indicated in Section 3.1, the null distribution of $W_g(r)$ can be approximated by the conditional distribution of $\hat{W}_g(r)$ given the data. The observed process $w_g(\cdot)$ may be plotted along with a few realizations of $\hat{W}_g(\cdot)$, and this graphical display may be supplemented with an estimated P-value for the supremum test $S_g \equiv \sup_r |W_g(r)|$, which can again be estimated by simulation. As shown in Appendix B, S_g is generally consistent against the misspecification of the link function. In fact, this procedure is sensitive to any alternative that leads to incorrect specification of the marginal mean, including the link function, the functional form of the response variable, and the conditional linear predictor in (1).

3.4 *Omnibus Test*

A global goodness-of-fit test of the mean response function can be constructed from the supremum statistic $L \equiv \sup_{\mathbf{x}} |W(\mathbf{x})|$, where $W(\mathbf{x})$ is defined in (4). The approach to simulating the null distribution and examining the residual pattern is the same as in the case of checking the function form of a fixed effect, although it is difficult to visualize the plot when $\mathbf{x}$ is of high dimension. As shown in Appendix B, the supremum test is consistent against any misspecification of the marginal mean structure in (3). If the random components are correctly specified, misspecification of (3) implies misspecification of the conditional mean given in equation (1).

3.5 *Model-Checking Techniques for Linear Mixed Models*

In the special case of LMMs, the marginal means have the simple form $m_{ij} = \mathbf{X}'_{ij}\boldsymbol{\beta}$, and the residuals e_{ij} become $y_{ij} - \mathbf{X}'_{ij}\hat{\boldsymbol{\beta}}$ $(i = 1, \ldots, n; j = 1, \ldots, t_i)$. One may incorporate some weight functions into the processes defined in (4) and (5). Specifically, we consider the following weighted processes:

$$W^*(\mathbf{x}) = n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{t_i} C_{ij} I(\mathbf{X}_{ij} \le \mathbf{x}) e_{ij}, \tag{8}$$

$$W^*_g(r) = n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{t_i} C_{ij} I(\mathbf{X}'_{ij}\hat{\boldsymbol{\beta}} \le r) e_{ij}, \tag{9}$$

where $\mathbf{C}_i \equiv (C_{i1}, \ldots, C_{it_i})' = \hat{\mathbf{V}}_i^{-1}\mathbf{1}$, and $\hat{\mathbf{V}}_i$ is the estimated marginal covariance matrix of $(y_{i1}, \ldots, y_{it_i})'$. Clearly, these weighted processes at $\mathbf{x} = \infty$ and $r = \infty$ correspond to the score function for the intercept term, and thus are expected to be more efficient than the unweighted processes defined in (4) and (5). To approximate the null distributions of the weighted processes, we replace $I(\mathbf{X}_{ij} \le \mathbf{x})$ with $C_{ij} I(\mathbf{X}_{ij} \le \mathbf{x})$ in $\hat{W}(\mathbf{x})$, and $I(\hat{m}_{ij} \le r)$ with $C_{ij} I(\mathbf{X}'_{ij}\hat{\boldsymbol{\beta}} \le r)$ in $\hat{W}_g(r)$. Although similar weighted processes could be defined for GLMMs, they are not pursued here due to the lack of closed forms of the score function and $\hat{\mathbf{V}}_i$ as well as the computational difficulties in evaluating them.

For LMMs, the proposed supremum tests are robust in that they do not require correct specification of $\mathrm{Var}(\mathbf{Y}_i)$. This is due to the fact that the least-squares method provides a consistent estimator of $\boldsymbol{\beta}$ given only correct specification of $\mathrm{E}(\mathbf{Y}_i)$ (Zeger, Liang, and Albert, 1988).

4. Simulation Studies

We carried out extensive simulation studies to investigate the finite-sample behavior of the supremum tests described in Section 3. We present here some of the results for the linear and logistic regression models with Gaussian random effects. We considered three repeated measures for each subject. To estimate the sizes of the tests, we generated 5000 replicates of data and used 5000 realizations from the null distributions in calculating the P-values; for the power assessment, we used 1000 replicates and 1000 realizations.

To evaluate the empirical sizes of the proposed tests, we generated repeated measures from the following null model:

$$g\{\mathrm{E}(y_{ij} \mid b_i)\} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2ij} + b_i,$$

where b_i is a normal random variable, X_{1i} is a Bernoulli variable with 0.4 success probability, and the conditional

distribution of $\mathbf{X}_{2i} \equiv (X_{2i1}, X_{2i2}, X_{2i3})'$ given X_{1i} follows a multivariate normal distribution with means X_{1i} and variances 1 for X_{2ij} $(j = 1, 2, 3)$, and covariances 0.5 between X_{2ij} and $X_{2ik} (j \neq k)$. For the logistic regression model, $g(\cdot)$ is the logit function, and $b_i \sim N(0, 1)$. For the linear model, $g(\cdot)$ is the identity function, $b_i \sim N(0, 2)$, and the within-cluster errors $e_{ij} (j = 1, 2, 3)$ are independent standard normal. We set $\boldsymbol{\beta} \equiv (\beta_0, \beta_1, \beta_2)' = (1, 1, 1)'$.

To investigate the power of the functional form test, we considered the following models:

- Model 1. $g\{\mathrm{E}(y_{ij} \mid b_i)\} = 1 + X_{1i} + \beta_2 \log(X_{2ij}) + b_i$,
- Model 2. $g\{\mathrm{E}(y_{ij} \mid b_i)\} = 1 + X_{1i} + X_{2ij} + \beta_3 X_{2ij}^2 + b_i$,

and tested the functional form of X_2. The setup for Model 1 is the same as that of the null model except that X_{2ij} is log normal. For Model 2, $X_{2ij} = j - 2$ for $j = 1, 2, 3$, which may be viewed as the centralized observation time.

To assess the power of the link function test, we considered the following two models:

- Model 3. $Y_{ij} = \exp(1 + X_{1i} + X_{2ij} + b_i) + e_{ij}$,
- Model 4. $Y_{ij} = (1 + X_{1i} + X_{2ij} + b_i)^2 + e_{ij}$.

These two models deviate from the null linear mixed-effects model in that, instead of the identity function, the logarithmic and square root link functions are assumed for Models 3 and 4, respectively.

The power of the omnibus test was evaluated for each of the alternative models. The power of the link function test was also assessed for Models 1 and 2, so as to determine how sensitive the test is to misspecification of the linear predictor.

The results of these simulation studies are summarized in Tables 1–4. The proposed supremum tests appear to maintain their sizes near the nominal levels. The functional form test has very good power, and the link function and omnibus tests also have reasonable power against misspecification of the functional forms of covariates. Not surprisingly, it requires a fairly large sample to detect misspecification of the link function.

There is obviously no need to use the proposed methods to test for the quadratic term. The Wald test or the likelihood-ratio test are asymptotically efficient in testing against such alternatives. The latter tests, however, are only applicable to nested alternatives. By contrast, the proposed supremum tests can be used to assess which functional form is more appropriate. The results in Tables 2 and 4 suggest that the proposed tests are only slightly less powerful than the Wald tests, at least for the situations considered here.

Table 1
Empirical sizes of the supremum tests for the linear and logistic models with normal random effects

Nominal level	Model	n	Functional form test	Link function test	Omnibus test
0.10	Linear	50	0.100	0.089	0.095
		100	0.099	0.095	0.100
		200	0.106	0.114	0.104
		300	0.104	0.096	0.104
	Logistic	100	0.101	0.098	0.095
		150	0.104	0.112	0.108
		200	0.107	0.099	0.101
		300	0.100	0.102	0.104
0.05	Linear	50	0.042	0.041	0.041
		100	0.046	0.041	0.044
		200	0.051	0.054	0.049
		300	0.049	0.045	0.049
	Logistic	100	0.042	0.043	0.046
		150	0.051	0.054	0.055
		200	0.049	0.045	0.049
		300	0.046	0.050	0.047
0.01	Linear	100	0.007	0.005	0.006
		200	0.010	0.009	0.010
		300	0.009	0.008	0.008
		400	0.010	0.009	0.009
	Logistic	100	0.008	0.008	0.008
		150	0.010	0.010	0.012
		200	0.006	0.009	0.005
		300	0.008	0.011	0.010

5. Examples

We consider CD4 data from a clinical trial to evaluate the efficacy of zidovudine (AZT) in treating patients with mild symptomatic HIV infection (Fischl et al., 1990). The study enrolled 711 patients with 361 randomized to AZT and 350 to placebo. The CD4 counts were measured repeatedly over the course of the study. The 4328 measurements taken during the first 40 weeks of the study are included in our analysis. The same data set was used by Lin et al. (2002) to illustrate their model-checking methods for marginal linear models.

For the fixed effects of the initial model, we include the linear term and quadratic term of time as well as their interactions with treatment, which was shown by Lin et al. (2002) to be a reasonable model to describe the marginal relationships of time and treatment with CD4 counts. To choose a covariance structure, we compare a model with only a random intercept with a model with both a random intercept and a random slope of time. Homogeneous independent within-subject errors are assumed for both models. The AICs for the two models are 54,699.3 and 54,649.9, the difference in -2 log likelihood being 53.4. These results suggest that the model with a random intercept and a random slope describes the random variations better. Thus, we first consider the following model:

$$Y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 T_{ij}^2 + \beta_3 R_i T_{ij} + \beta_4 R_i T_{ij}^2 + b_{i1} + b_{i2} T_{ij} + \epsilon_{ij}, \tag{10}$$

where T_{ij} is the time (in weeks) when the kth measurement is made on the ith patient, Y_{ij} is the CD4 count at T_{ij}, R_i is the indicator of AZT for the ith patient, b_{i1} and b_{i2} are the random intercept and random slope of time for the ith subject, and ϵ_{ij} is the within-subject error at T_{ij}. Because no treatment difference at baseline can be assumed by randomization, R_i as a main effect is not included in the model. The estimation results for the fixed effects and covariance parameters for model (10) are summarized in Panel A of Tables 5 and 6, respectively.

Table 2
Empirical powers of the functional form test and omnibus test at the 0.05 *significance level against incorrect functional forms of covariates in linear mixed models*

Model	Parameter	n	Functional form test	Omnibus test
1	$\beta_2 = 1.0$	50	0.820	0.815
		100	0.990	0.989
		200	1.000	1.000
		300	1.000	1.000
	$\beta_2 = 0.5$	50	0.325	0.318
		100	0.613	0.614
		200	0.889	0.885
		300	0.981	0.977
2	$\beta_3 = 0.5$	50	0.819 (0.839)	0.671
		100	0.981 (0.984)	0.938
		200	1.000 (1.000)	0.997
		300	1.000 (1.000)	1.000
	$\beta_3 = 0.3$	50	0.420 (0.450)	0.290
		100	0.699 (0.724)	0.514
		200	0.940 (0.940)	0.846
		300	0.984 (0.987)	0.957

The numbers in parentheses pertain to the Wald statistic for testing $H_0 : \beta_3 = 0$.

Table 4
Empirical powers of the supremum tests at the 0.05 *significance level against incorrect functional forms of covariates in logistic regression models with random effects*

Model	Parameter	n	Functional form test	Link function test	Omnibus test
1	$\beta_2 = 1.0$	100	0.409	0.381	0.411
		150	0.612	0.557	0.602
		200	0.787	0.729	0.788
		300	0.929	0.899	0.921
	$\beta_2 = 0.5$	100	0.211	0.197	0.218
		150	0.323	0.284	0.327
		200	0.436	0.403	0.431
		300	0.620	0.569	0.609
2	$\beta_3 = 1.0$	100	0.769 (0.783)	0.535	0.655
		150	0.905 (0.912)	0.695	0.828
		200	0.965 (0.968)	0.826	0.922
		300	0.998 (0.999)	0.957	0.990
	$\beta_3 = 0.5$	100	0.298 (0.308)	0.180	0.226
		150	0.424 (0.433)	0.255	0.317
		200	0.528 (0.543)	0.329	0.414
		300	0.704 (0.719)	0.477	0.591

See the footnote of Table 2.

Figure 1a and 1b displays the cumulative residuals against time and predicted values under model (10) based on the weighted processes defined in (8) and (9), respectively. The P-value for testing the functional form of time is 0.284. The P-value of the S_g test is 0.028, indicating misspecification of the link function or the functional form of the response. The observed residual pattern in Figure 1b resembles the black curve in Figure 2b of Lin et al. (2002), which, in the case of the S_g test, suggests a square root transformation of either Y or $E(Y)$.

Panel B of Tables 5 and 6 shows the results with the square root of CD4 count as the response variable in model (10). The residual analysis reveals no obvious systematic pattern under the new model, with the P-value for the functional form test of time being 0.405 and the P-value of the S_g test being 0.18; see Figure 1c and 1d. The square root transformation of CD4 counts indeed results in a much better model.

Table 3
Empirical powers of the link function test and omnibus test at the 0.05 *significance level against incorrect link functions in linear mixed-effects models*

Model	n	Link function test	Omnibus test
3	50	0.078	0.077
	100	0.213	0.182
	200	0.457	0.428
	300	0.609	0.568
4	50	0.238	0.170
	100	0.552	0.424
	200	0.903	0.809
	300	0.990	0.967

For the placebo group, the mean CD4 curve declines monotonically over the study period, whereas for the AZT group, it rises for the first few weeks and then declines over time. The highly significant covariance parameters for the random effects suggest that there is definitely subject-to-subject variability. Each patient's response curve deviates from the average curve in both the intercept and the slope of time.

As a second example, we examine a subset of 275 children from the cohort of the Indonesian children's health study on respiratory infection (Sommer, Katz, and Tarwotjo, 1984). This data set was used by Diggle et al. (2002, p. 182–184) to address the question of how a child's risk of respiratory infection would change if their vitamin A status were to change. This was accomplished by fitting a random-effects model which allows each child to have a separate intercept to

Table 5
Estimates of the fixed effects of the linear mixed model for the HIV study

Parameter	Estimate	SE	Est/SE	P-value
A. Response = CD4 count				
time	−1.6994	0.6037	−2.82	0.0050
$time^2$	0.0202	0.0161	1.25	0.2108
treatment × *time*	4.3972	0.8206	5.36	<0.0001
treatment × $time^2$	−0.1029	0.0219	−4.69	<0.0001
B. Response = (CD4 count)$^{1/2}$				
time	−0.0510	0.0142	−3.59	0.0003
$time^2$	0.0006	0.0004	1.52	0.1294
treatment × *time*	0.1084	0.0194	5.60	<0.0001
treatment × $time^2$	−0.0024	0.0005	−4.79	<0.0001

Table 6
Estimates of the covariance parameters of the linear mixed model for the HIV study

Parameter	Estimate	SE	Est/SE	*P*-value
A. Response = CD4 count				
σ_1^2	18,830.00	1239.56	15.19	<0.0001
σ_{12}	49.90	27.32	1.83	0.0677
σ_2^2	4.92	1.03	4.79	<0.0001
Residual	11,497.00	289.89	39.66	<0.0001
B. Response = (CD4 count)$^{1/2}$				
σ_1^2	10.1712	0.6704	15.17	<0.0001
σ_{12}	0.0420	0.0154	2.72	0.0065
σ_2^2	0.0036	0.0006	5.71	<0.0001
Residual	6.2044	0.1570	39.51	<0.0001

σ_1^2, σ_2^2, and σ_{12} are the variances and the covariance for the random intercept and random slope.

represent their propensity to infection. Panel A of Table 7 provides the regression estimates for the following model:

$$\begin{aligned}\text{logit}\{\mathrm{E}(y_{ij}\,|\,b_i)\} &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2ij} + \beta_3 X_{3ij} \\ &\quad + \beta_4 X_{4ij} + \beta_5 X_{5ij} + \beta_6 X_{6ij} \\ &\quad + \beta_7 {X_{6ij}}^2 + b_i, \end{aligned} \tag{11}$$

where y_{ij} is the jth measurement of the ith subject, X_{1i} is gender, X_{2ij} is height for age as a percentage of the United States National Center for Health Statistics standards, X_{3ij} and X_{4ij} are annual cosine and sine variables to adjust for seasonality, X_{5ij} is an indicator for suffering xerophthalmia, an ocular manifestation of chronic vitamin A deficiency, X_{6ij} is the centered age, and b_i is the random intercept for the ith subject which is assumed to follow a zero-mean normal distribution with variance D.

Figure 2a and 2b displays the cumulative residuals against the season variable and the predicted values under model (11) based on the processes defined in (4) and (5), respectively. The *P*-values for the functional form test of season and the supremum test S_g are 0.018 and 0.064. The *P*-values for the functional form tests S_2 and S_6 for height and age are 0.866 and 0.144. Thus, the functional form of season, which is expressed as seasonal cosine and sine in the model, is the most problematic. Therefore, we replace the cosine and sine variables with three indicators for the four seasons in equation (11). The estimates of this revised model are given in Panel B of Table 7. For this model, the *P*-value of the supremum test S_g is 0.094; the *P*-value of the supremum test S_6 is 0.101. These results indicate that the functional form of X_6, although reasonable, is not entirely satisfactory. The observed residual pattern in Figure 2c resembles the black curve in Figure 2c of Lin et al. (2002), which suggests inclusion of the cubic term of X_6.

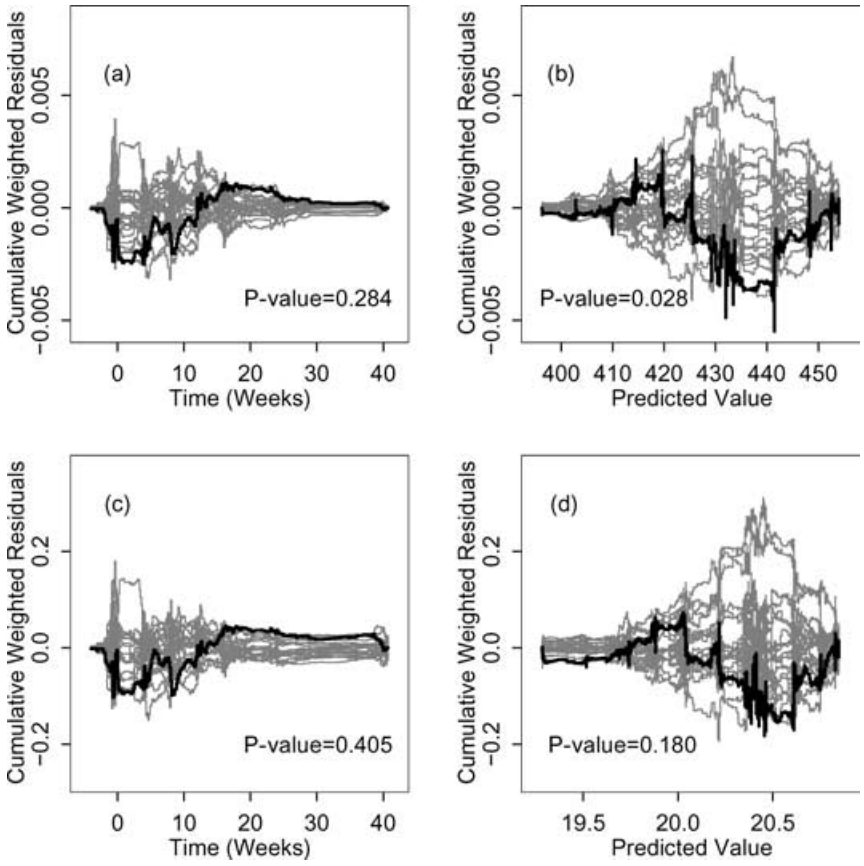

Figure 1. Cumulative residual plots for the HIV study: (a) and (b) pertain to the model with CD4 count as the response, and (c) and (d) pertain to the model with the square root of CD4 count as the response. In each panel, 20 simulated realizations are shown in gray curves along with the black observed curve; the *P*-value pertains to the supremum test with 1000 realizations.

Table 7
Estimates of the fixed effects and the variance parameter of the random-effects models for the Indonesian study on respiratory infection

Parameter	Estimate	SE	Est/SE	*P*-value
A. Model 1: cosine and sine season variables with quadratic age effect model				
Intercept	−2.230	0.243	−9.19	<0.0001
Gender	−0.514	0.253	−2.03	0.044
Height for age	−0.044	0.022	−2.01	0.046
Seasonal cosine	−0.613	0.174	−3.52	0.0005
Seasonal sine	−0.166	0.173	−0.96	0.338
Xerophthalmia	0.541	0.482	1.12	0.262
Age	−0.031	0.008	−3.90	0.0001
Age^2	−0.0011	0.0004	−2.70	0.007
D	0.521	0.333	1.57	0.118
B. Model 2: indicator season variables with quadratic age effect model				
Intercept	−2.274	0.369	−6.17	<0.0001
Gender	−0.525	0.256	−2.05	0.040
Height for age	−0.046	0.022	−2.06	0.040
Xerophthalmia	0.525	0.487	1.08	0.280
Age	−0.030	0.008	−3.78	0.0002
Age^2	−0.0011	0.0004	−2.75	0.006
$I(season = 1)$	−0.717	0.497	−1.44	0.149
$I(season = 2)$	0.788	0.340	2.32	0.021
$I(season = 3)$	0.033	0.364	0.09	0.928
D	0.549	0.342	1.61	0.110
C. Model 3: indicator season variables with cubic age effect model				
Intercept	−2.243	0.368	−6.10	<0.0001
Gender	−0.549	0.255	−2.16	0.031
Height for age	−0.043	0.022	−1.96	0.050
Xerophthalmia	0.533	0.487	1.09	0.275
$Age/10$	−0.520	0.133	−3.90	0.0001
$(age/10)^2$	−0.141	0.042	−3.36	0.0008
$(age/10)^3$	0.034	0.016	2.13	0.033
$I(season = 1)$	−0.740	0.498	−1.49	0.136
$I(season = 2)$	0.801	0.341	2.35	0.019
$I(season = 3)$	0.020	0.365	0.05	0.960
D	0.517	0.336	1.54	0.124

Panel C of Table 7 shows the estimates of the new model, with ${X_6}^3$ added to the above model with three season indicator variables. It is evident that ${X_6}^3$ is indeed required. Note that X_6 (i.e., age) is scaled by a factor of 10 to avoid the nonconvergence problem. The residual analysis reveals no obvious systematic pattern under the new model, with the *P*-value for the S_6 test being 0.605 and the *P*-value for the S_g test being 0.408; see Figure 2d. The nonquadratic age effect was also found by Zhang and Lin (2003), who developed a testing procedure for the functional form of a fixed effect in a GLMM.

6. Discussion

We have developed new model-checking techniques for GLMMs based on the cumulative sums of residuals. These methods can be used to examine individual components of the deterministic part of a model as well as the overall fit. The proposed methods do not require the specification of an alternative hypothesis and are informative about the nature of model misspecification.

Although different tests are proposed to assess different aspects of model (1), all of these tests are actually checking the fit of the entire model, including the random components. This is because a faulty functional form of a fixed covariate may cause the cumulative residuals against the predicted values to exhibit systematic tendency, and vice versa. Furthermore, the misspecification of the random effects may affect the estimation of the fixed-effects parameters (Heagerty and Kurland, 2001). Nevertheless, W_k tends to be the most sensitive to the misspecification of the functional form of X_k, and W_g to the misspecification of the linear predictor.

We have focused on checking the deterministic components of GLMMs. For random components of LMMs, several graphical techniques are available, including the empirical semivariogram (Diggle, 1988) and the draftman's display and parallel axis plots (Dawson, Gennings, and Carter, 1997) for checking the covariance structure, and the generalized weighted normal plot (Lange and Ryan, 1989) for checking the distributional assumption of the random effects. It would be worthwhile to explore graphical and numerical methods for assessing the random components of GLMMs.

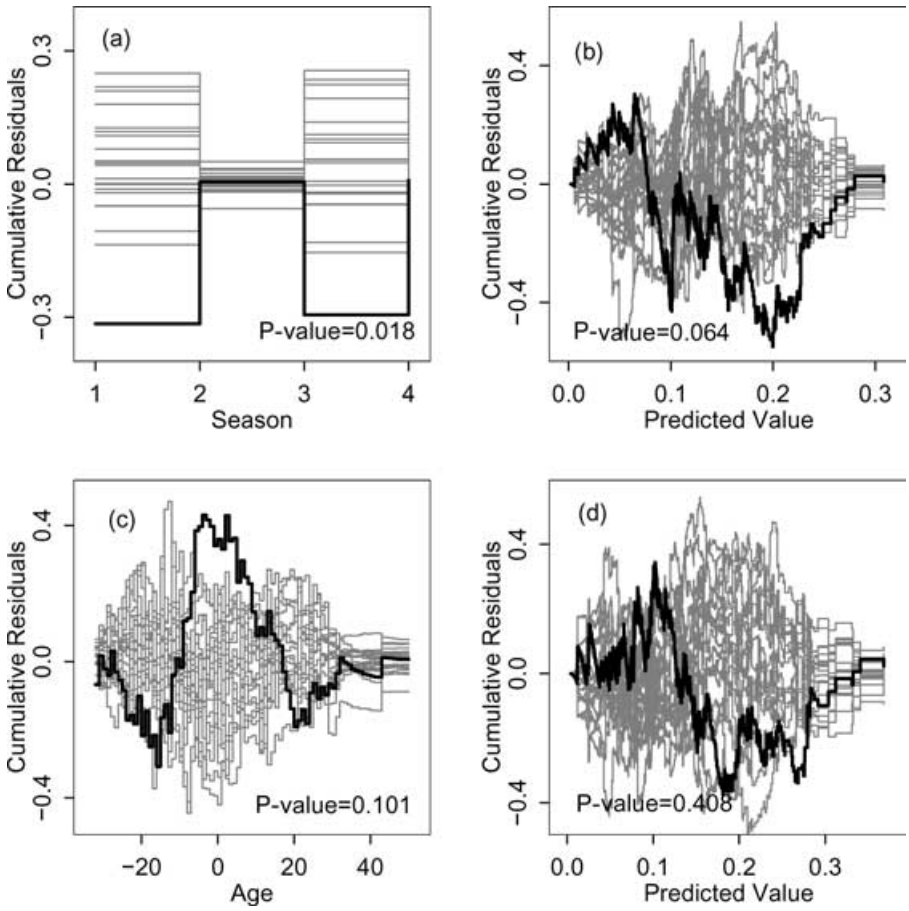

Figure 2. Cumulative residual plots for the Indonesian study on respiratory infection: (a) and (b) pertain to Model 1, (c) pertains to Model 2, and (d) pertains to Model 3. In each panel, 20 simulated realizations are shown in gray curves along with the black observed curve; the *P*-value pertains to the supremum test with 1000 realizations.

Acknowledgements

This research was supported by the National Institutes of Health. The authors thank an associate editor for helpful comments.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **AC-19,** 716–723.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88,** 9–25.

Cook, D. R. and Weisberg, S. (1994). *An Introduction to Regression Graphics*. New York: Wiley.

Dawson, K. S., Gennings, C., and Carter, W. H. (1997). Two graphical techniques useful in detecting correlation structure in repeated measure data. *American Statistician* **51,** 275–283.

Diggle, P. J. (1988). An approach to the analysis of repeated measurements. *Biometrics* **40,** 959–971.

Diggle, P. J., Heagerty, P. J., Liang, K. Y., and Zeger, S. L. (2002). *Analysis of Longitudinal Data*, 2nd edition. Oxford: Oxford University Press.

Fischl, M. A., Richman, D. D., Nansen, N., et al. (1990). The safety and efficacy of zidovudine (AZT) in the treatment of subjects with mildly symptomatic human immunodeficiency virus type 1 (HIV) infection. *Annals of Internal Medicine* **112,** 727–737.

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

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38,** 963–974.

Lange, N. and Ryan, L. (1989). Assessing normality in random effects models. *Annals of Statistics* **17,** 624–642.

Lin, D. Y., Wei, L. J., and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80,** 557–572.

Lin, D. Y., Wei, L. J., and Ying, Z. (2002). Model-checking techniques based on cumulative residuals. *Biometrics* **58,** 1–12.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edition. London: Chapman and Hall.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Hayward, California: Institute of Mathematical Statistics.

Schwartz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics* **6,** 461–464.

Sommer, A., Katz, J., and Tarwotjo, I. (1984). Increased risk of respiratory infection and diarrhea in children with preexisting mild vitamin A deficiency. *American Journal of Clinical Nutrition* **40,** 1090–1095.

Su, J. Q. and Wei, L. J. (1991). A lack-of-fit test for the mean function in a generalized linear model. *Journal of the American Statistical Association* **86,** 420–426.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.

Vonesh, E. F., Chinchilli, V. M., and Pu, K. (1996). Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics* **52,** 572–587.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association* **86,** 79–86.

Zeger, S. L., Liang, K. Y., and Albert, P. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44,** 1049–1060.

Zhang, D. and Lin, X. (2003). Hypothesis testing in semiparametric additive mixed models. *Biostatistics* **4,** 57–74.

Received August 2004. *Revised December* 2004.
Accepted January 2005.

Appendix A

Weak Convergence of $W(\boldsymbol{x})$, $W_g(r)$, $\hat{W}(\boldsymbol{x})$, $\hat{W}_g(r)$, $W^*(\boldsymbol{x})$, *and* $W^*_g(r)$

We first establish the weak convergence of $W(\mathbf{x})$. Assume that the first and second derivatives of $m_{ij}(\boldsymbol{\theta})$ are bounded. Clearly,

$$W(\mathbf{x}) = n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{t_i} I(\mathbf{X}_{ij} \le \mathbf{x})(y_{ij} - m_{ij}) - n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{t_i} I(\mathbf{X}_{ij} \le \mathbf{x})(\hat{m}_{ij} - m_{ij}).$$

By the Taylor series expansion, the process $W(\mathbf{x})$ is asymptotically equivalent to the process

$$n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{t_i} I(\mathbf{X}_{ij} \le \mathbf{x})(y_{ij} - m_{ij}) + \boldsymbol{\eta}'(\mathbf{x};\boldsymbol{\theta})n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}),$$

where $\boldsymbol{\eta}(\mathbf{x};\boldsymbol{\theta})$ is defined in (7). It then follows from (2) that

$$W(\mathbf{x}) = n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{t_i} I(\mathbf{X}_{ij} \le \mathbf{x})(y_{ij} - m_{ij}) + \boldsymbol{\eta}'(\mathbf{x};\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}n^{-1/2}\mathbf{U}(\boldsymbol{\theta}) + o_p(1). \quad (A.1)$$

Thus,

$$W(\mathbf{x}) = n^{-1/2}\sum_{i=1}^{n} H_i(\mathbf{x}) + o_p(1),$$

where

$$H_i(\mathbf{x}) = \sum_{j=1}^{t_i} I(\mathbf{X}_{ij} \le \mathbf{x})(y_{ij} - m_{ij}) + \boldsymbol{\eta}'(\mathbf{x};\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}\mathbf{U}_i(\boldsymbol{\theta}).$$

By the uniform law of large numbers (Pollard, 1990, p. 39), $\boldsymbol{\eta}(\mathbf{x};\boldsymbol{\theta})$ converges to a nonrandom function uniformly in $\mathbf{x}$. Thus, for any fixed $\mathbf{x}$, the right-hand side of (A.1) is essentially a normalized sum of n independent zero-mean random variables. It then follows from the multivariate central limit theorem that $W(\mathbf{x})$ converges in finite-dimensional distributions to a zero-mean Gaussian process. Because it consists of sums of monotone functions, the tightness of the first term in (A.1) can be obtained by the functional central limit theorem (Pollard, 1990, p. 38). The tightness of the second term in (A.1) follows from the asymptotic normality of $n^{-1/2}\mathbf{U}(\boldsymbol{\theta})$ and the uniform convergence of $\boldsymbol{\eta}(\mathbf{x};\boldsymbol{\theta})$. Hence, $W(\mathbf{x})$ converges weakly to a zero-mean Gaussian process with covariance function between $\mathbf{x}_1$ and $\mathbf{x}_2$ being

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} H_i(\mathbf{x}_1)H_i(\mathbf{x}_2). \quad (A.2)$$

Conditional on the data (y_{ij}, $\mathbf{X}_{ij}$, $\mathbf{Z}_{ij}$) ($i = 1,\ldots,n$; $j = 1,\ldots,t_i$), the only random components in $\hat{W}(\mathbf{x})$ are the standard normal variables $G_1,\ldots,G_n$. As shown earlier, $n^{-1/2}\sum_{i=1}^{n} H_i(\mathbf{x})$ converges weakly to the zero-mean Gaussian process with covariance function given in (A.2). Thus, it follows from the conditional multiplier central limit theorem (van der Vaart and Wellner, 1996, Theorem 2.9.6) that the conditional distribution of $\hat{W}(\mathbf{x})$ given the data (y_{ij}, $\mathbf{X}_{ij}$, $\mathbf{Z}_{ij}$) ($i = 1,\ldots,n$; $j = 1,\ldots,t_i$) converges weakly to the same limiting distribution as the unconditional null distribution of $W(\mathbf{x})$.

Now, we establish the weak convergence of $W_g(r)$ and $\hat{W}_g(r)$. By assuming that

$$\mathrm{E}\left\{\sum_{j=1}^{t_i} I(m(\boldsymbol{\zeta}_1) \le r_1)(y_{ij} - m_{ij}) - \sum_{j=1}^{t_i} I(m(\boldsymbol{\zeta}_2) \le r_2)(y_{ij} - m_{ij})\right\}^2 \to 0$$

as $\boldsymbol{\zeta}_1 \to \boldsymbol{\zeta}_2$ and $r_1 \to r_2$ for all $m(\boldsymbol{\zeta}_2)$ sufficiently close to $\boldsymbol{\theta}$, we have

$$W_g(r) = \tilde{W}_g(r) + o_p(1),$$

where

$$\tilde{W}_g(r) = n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{t_i} I(m_{ij} \le r)e_{ij}.$$

The arguments for the weak convergence of $W(\mathbf{x})$ can then be used to show the weak convergence of $\tilde{W}_g(r)$. The proof

for the weak convergence of $\hat{W}_g(r)$ is exactly the same as that for $\hat{W}(\mathbf{x})$.

To prove the weak convergence of $W^*(\mathbf{x})$, we define

$$\tilde{W}^*(\mathbf{x}) = n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{t_i} c_{ij} I(\mathbf{X}_{ij} \le \mathbf{x}) e_{ij},$$

where $\mathbf{c}_i \equiv (c_{i1}, \ldots, c_{it_i})' = \mathbf{1}'\mathbf{V}_i^{-1}$. The weak convergence of $\tilde{W}^*(\mathbf{x})$ follows from that of $W(\mathbf{x})$. Because $\mathbf{C}_i$ converges in probability to $\mathbf{c}_i$, $W^*(\mathbf{x})$ converges to the same limit as $\tilde{W}^*(\mathbf{x})$. This establishes the weak convergence of $W^*(\mathbf{x})$. Similar arguments can be used to show the weak convergence of $W_g^*(r)$.

Appendix B

Consistency of Supremum Tests

Let $f_{x_{ij}}(\mathbf{x})$ be the density function of $\mathbf{X}_{ij}$. Define the general alternative as that there does not exist a vector $\boldsymbol{\beta}$ such that the true marginal mean of y_{ij} given $\mathbf{x}$, denoted by $m_{ij}(\mathbf{x})$, satisfies equation (3) for all $\mathbf{x}$ in the sample space generated by $\mathbf{X}_{ij}$ $(i = 1, \ldots, n;\ j = 1, \ldots, t_i)$. Under the alternative, $\hat{\boldsymbol{\beta}} \xrightarrow{\mathcal{P}} \boldsymbol{\beta}^*$ and $\hat{\boldsymbol{\alpha}} \xrightarrow{\mathcal{P}} \boldsymbol{\alpha}^*$, where $\boldsymbol{\beta}^*$ and $\boldsymbol{\alpha}^*$ are some constants. By extending the argument of Su and Wei (1991) and Lin et al. (1993), we can show that $n^{-1/2}W(\mathbf{x})$ converges in probability to

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\left[\sum_{j=1}^{t_i}\int_{\mathbf{v}\le\mathbf{x}}\left\{m_{ij}(\mathbf{v}) - \int g^{-1}\left(\mathbf{v}'\boldsymbol{\beta}^* + \mathbf{Z}_{ij}'\mathbf{b}_i\right) \times f_b(\mathbf{b}_i;\boldsymbol{\alpha}^*)\,\mathrm{d}\mathbf{b}_i\right\} f_{x_{ij}}(\mathbf{v})\,\mathrm{d}\mathbf{v}\right],$$

which is nonzero for some $\mathbf{x}$ under the alternative. Thus, the test L is consistent. For the special case of LMMs, the misspecification of the marginal mean in (3) is equivalent to the misspecification of the conditional mean structure in (1). For general GLMMs, the above alternative implies the misspecification of the conditional mean in equation (1) given the correct specification of the random components of the model, including the distribution and the design matrix of the random effects. Thus, L is tailored to checking the deterministic components of the model, including the functional form of the fixed-effects and the link function.

As a special case of the omnibus test, S_k is consistent against any alternative under which $W_k(x)$ is not centered at 0 for all x. It is unlikely for $W_k(x)$ to be centered at 0 for all x when the functional form of the kth covariate is misspecified. For example, if $f_b(\mathbf{b})$ is the Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{D}$, we can show that, for standard link functions, S_k is consistent against misspecification of the functional form of X_k provided that $\hat{\boldsymbol{\alpha}}$ and the components of $\hat{\boldsymbol{\beta}}$ for the remainder of $\mathbf{X}$ converge to the true values. This can be seen from the simplified expressions of the marginal mean for standard link functions. For the identity link, $m_{ij} = \mathbf{X}_{ij}'\boldsymbol{\beta}$; for the log link, $m_{ij} = \exp(\mathbf{X}_{ij}'\boldsymbol{\beta} + \mathbf{Z}_{ij}'\mathbf{D}\mathbf{Z}_{ij}/2)$; for the probit link, $m_{ij} = \Phi(a(\mathbf{D}) \cdot \mathbf{X}_{ij}'\boldsymbol{\beta})$, where $a(\mathbf{D}) = |\mathbf{D}\mathbf{Z}_{ij}\mathbf{Z}_{ij}' + \mathbf{I}|^{-q/2}$; for the logit link, $m_{ij} \approx \exp(c_{ij}\mathbf{X}_{ij}'\boldsymbol{\beta})/\{1 + \exp(c_{ij}\mathbf{X}_{ij}'\boldsymbol{\beta})\}$, where $c_{ij} = |c^2\mathbf{D}\mathbf{Z}_{ij}\mathbf{Z}_{ij}' + \mathbf{I}|^{-q/2}$ and $c = 16(3)^{1/2}/(15\pi)$ (Zeger et al., 1988). The consistency of S_k is then a byproduct of the omnibus test L.

For the S_g test, a rigorous proof of its consistency for the general GLMMs is more delicate. Here, we provide a heuristic justification. A crude, first-order approximation to the marginal mean given in (3) is $\mathrm{E}(y_{ij}) = g^{-1}(\mathbf{X}_{ij}'\boldsymbol{\beta})$, which is valid in the limit as the components of dispersion approach 0 (Breslow and Clayton, 1993). It then follows from the arguments of Lin et al. (1993) that $n^{-1/2}\,W_g(r)$ converges in probability to

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\left[\sum_{j=1}^{t_i}\int_{\mathbf{v}'\boldsymbol{\beta}^*\le r}\{h^{-1}(\mathbf{v}'\boldsymbol{\beta}^*) - g^{-1}(\mathbf{v}'\boldsymbol{\beta}^*)\} f_{x_{ij}}(\mathbf{v})\,\mathrm{d}\mathbf{v}\right].$$

Thus, the S_g test is consistent against incorrect link function in the form of $h(\mathbf{X}'\boldsymbol{\beta}^*)$, where h is not equal to g.