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 \( i \)th subject on the \( j \)th occasion, and let \( X_{ij} \) and \( Z_{ij} \) 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

\[
g\{E(y_{ij} | b_i)\} = X_{ij}'\beta + Z_{ij}'b_i, \quad i = 1, \ldots, n; \quad j = 1, \ldots, t_i, \tag{1}
\]

where \( g(\cdot) \) is a known differentiable link function, \( \beta \) is a \( p \times 1 \) vector of unknown regression parameters, and \( b_i \) is a \( q \times 1 \) vector of unobservable random effects for the \( i \)th subject.

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

The likelihood function for \( \theta \) is

\[
L(\theta) = \prod_{i=1}^{n} \prod_{j=1}^{t_i} f_{y | b}(y_{ij} | b_i) f_b(b_i) \, db_i.
\]

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 \( \theta \), denoted by \( \hat{\theta} \), is the solution to \( U(\theta) = 0 \), where

\[
U(\theta) = \frac{\partial \log L(\theta)}{\partial \theta} = \sum_{i=1}^{n} U_i(\theta),
\]

and \( U_i \) involves only the observations from the \( i \)th subject.

Under mild conditions,

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

where \( \Omega = \lim_{n \to \infty} \mathcal{I}(\theta) \), and \( \mathcal{I}(\theta) = -n^{-1} \partial^2 \log L(\theta) / \partial \theta \partial \theta' \).

In addition, \( n^{1/2}(\hat{\theta} - \theta) \) is asymptotically zero-mean normal with covariance matrix \( \Omega^{-1} \).

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

\[
E(y_{ij}) = E(E(y_{ij} | b_i)) = \int g^{-1}(X_{ij}'\hat{\beta} + Z_{ij}'b_i) f_b(b_i) \, db_i. \tag{3}
\]

We denote the above quantities by \( m_0(\theta) \) or \( m_\theta \). Given \( \hat{\theta} \), the predicted values of the responses are \( m_\hat{\theta} \equiv m_\theta(\hat{\theta}) \), 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:

\[
W(x) = n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{t_i} I(X_{ij} \leq x) e_{ij}, \tag{4}
\]

\[
W_g(r) = n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{t_i} I(m_{ij} \leq r) e_{ij}, \tag{5}
\]

where \( x = (x_1, \ldots, x_p) \in \mathbb{R}^p, \ r \in \mathbb{R} \), \( I(X_{ij} \leq x) = I(X_{ij} \leq x_1, \ldots, X_{ij} \leq x_p) \), and \( X_{ij} \) is the \( k \)th component of \( X_{ij} \) (\( k = 1, \ldots, p \)).

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 \( \sup_k |W(x)| \) or \( \sup_r |W_g(r)| \) would indicate model mis specification.

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 \( W(x) \) converges in distribution to a zero-mean Gaussian process under \( H_0 \). Define

\[
\hat{W}(x) = n^{-1/2} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{t_i} I(X_{ij} \leq x) e_{ij} + \eta'(x; \hat{\theta}) \mathcal{I}^{-1}(\hat{\theta}) U_i(\hat{\theta}) \right\} G_i, \tag{6}
\]

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

\[
\eta(x; \theta) = -n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{t_i} I(X_{ij} \leq x) \partial m_{ij}(\theta) / \partial \theta. \tag{7}
\]
It is shown in Appendix A that the conditional distribution of \( W(x) \) given the data \((y_i, \mathbf{x}_i, \mathbf{z}_i)\) \((i = 1, \ldots, n; j = 1, \ldots, t_i)\) is the same as the limit of the unconditional distribution of \( W(x) \) under \( H_0 \). To approximate the null distribution of \( W(x) \), we obtain a large number of realizations of \( W(x) \) by repeatedly generating the normal random sample \((G_1, \ldots, G_n)\) while fixing the data \((y_i, \mathbf{x}_i, \mathbf{z}_i)\) \((i = 1, \ldots, n; j = 1, \ldots, t_i)\) 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 \( W_g(r) \), where \( W_g(r) \) is obtained from \( W(x) \) by replacing \( I(\mathbf{x}_j \leq \mathbf{x}) \) in (6) and (7) with \( I(\tilde{m}_ij \leq r) \).

### 3.2 Checking the Functional Form of a Fixed Effect

To check the functional form of the \( k \)th component of the covariate vector \( \mathbf{X} \), we consider the following process

\[
W_k(x) = \frac{n^{1/2}}{t} \sum_{i=1}^n \sum_{j=1}^{t_i} I(X_{ij} \leq x) e_{ij}, \quad k = 1, \ldots, p,
\]

where \( x \in \mathbb{R} \). Note that \( W_k(x) \) is a special case of \( W(x) \) given in (4) with \( x_i = \infty \) for all \( i \neq k \). Thus, the null distribution of \( W_k(x) \) can be approximated by the \( W_k(x) \) process, which is a special case of \( W(x) \) given in (6) with \( x_i = \infty \) for all \( i \neq k \).

To access visually how abnormal the observed process \( \theta_k(\cdot) \) of \( W_k(\cdot) \) is, one may plot \( \theta_k(\cdot) \) along with a few realizations from the \( \theta_k(\cdot) \) process. To enhance the objectivity of this graphical procedure, one may complement it with the Kolmogorov-type supremum statistic \( S_k = \sup_x |W_k(x)| \). Denote the observed value of \( S_k \) by \( s_k \). The \( P \)-value, \( \text{Pr}(S_k \geq s_k) \), can be approximated by \( \text{Pr}(S_k \geq s_k) \), where \( S_k = \sup_x |W_k(x)| \). The probability \( \text{Pr}(S_k \geq s_k) \) can be estimated by generating a large number of realizations from \( W_k(x) \).

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 \( W(r) \) defined in (5). As indicated in Section 3.1, the null distribution of \( W(r) \) can be approximated by the conditional distribution of \( W_g(r) \) given the data. The observed process \( \theta(\cdot) \) may be plotted along with a few realizations of \( \theta(\cdot) \), and this graphical display may be supplemented with an estimated \( P \)-value for the supremum test \( S_g = \sup_x |W_g(r)| \), 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 \( L = \sup_x |W(x)| \), where \( W(x) \) 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 \( x \) 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 \( m_y = \mathbf{X}_y \beta \), and the residuals \( e_y \) become \( y_i - \mathbf{X}_i \beta \). One may incorporate some weight functions into the processes defined in (4) and (5). Specifically, we consider the following weighted processes:

\[
W^*(x) = \frac{n^{1/2}}{t} \sum_{i=1}^n \sum_{j=1}^{t_i} C_{ij} I(X_{ij} \leq x) e_{ij},
\]

\[
W_g^*(r) = \frac{n^{1/2}}{t} \sum_{i=1}^n \sum_{j=1}^{t_i} C_{ij} I(X_{ij} \beta \leq r) e_{ij},
\]

where \( C_i \equiv (C_{i1}, \ldots, C_{it})' = \tilde{V}_i^{-1} \), and \( \tilde{V}_i \) is the estimated marginal covariance matrix of \((y_{i1}, \ldots, y_{it})'\). Clearly, these weighted processes at \( x = \infty \) 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 \( I(\mathbf{x}_j \leq \mathbf{x}) \) with \( C_{ij} I(\mathbf{X}_j \leq \mathbf{x}) \) in \( W(x) \), and \( I(\tilde{m}_ij \leq r) \) with \( C_{ij} I(\mathbf{X}_j \beta \leq r) \) in \( W_g(r) \). 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 \( \tilde{V}_i \) 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 \( \text{Var}(Y_i) \). This is due to the fact that the least-squares method provides a consistent estimator of \( \beta \) given only correct specification of \( \text{E}(Y_i) \) (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:

\[ g\{y_{ij} \mid b_i \} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + b_i, \]

where \( b_i \) is a normal random variable, \( X_{i1} \) is a Bernoulli variable with 0.4 success probability, and the conditional
distribution of \( X_{2ij} \equiv (X_{2i1}, X_{2i2}, X_{2i3})' \) given \( X_{1i} \) follows a multivariate normal distribution with means \( X_{1i} \) and variances 1 for \( X_{2i1} \) (\( j = 1, 2, 3 \)), 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 \( \beta \equiv (\beta_0, \beta_1, \beta_2)' = (1, 1, 1)' \).

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

- Model 1. \( g(\{E(y_{ij} \mid b_i}\}) = 1 + X_{1i} + \beta_2 \log(X_{2ij}) + b_i \),
- Model 2. \( g(\{E(y_{ij} \mid b_i}\}) = 1 + X_{1i} + X_{2ij} + \beta_1 X_{2ij}^2 + b_i \),

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.

### 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.

### Table 1

<table>
<thead>
<tr>
<th>Functional form test</th>
<th>Link function test</th>
<th>Omnibus test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal level</td>
<td>Model n</td>
<td></td>
</tr>
<tr>
<td>0.10 Linear 50 0.100 0.089 0.095</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.099 0.104 0.104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.106 0.096 0.104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.107 0.099 0.101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.100 0.102 0.104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logistic 100 0.101 0.098 0.095</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.104 0.102 0.104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.107 0.099 0.101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.100 0.102 0.104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.042 0.041 0.041</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.046 0.044 0.044</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.051 0.054 0.049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.049 0.045 0.049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logistic 100 0.042 0.043 0.046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.051 0.054 0.055</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.049 0.045 0.049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.046 0.050 0.047</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.007 0.005 0.006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.010 0.009 0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.009 0.008 0.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.010 0.009 0.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logistic 100 0.008 0.008 0.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.010 0.010 0.012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.006 0.009 0.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.008 0.011 0.010</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 \(-2 \log \text{likelihood} = 53.4\). 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 \( k \)-th measurement is made on the \( i \)-th patient, \( Y_{ij} \) is the CD4 count at \( T_{ij} \), \( R_i \) is the indicator of AZT for the \( i \)-th patient, \( b_{i1} \) and \( b_{i2} \) are the random intercept and random slope of time for the \( i \)-th subject, and \( \epsilon_{ij} \) 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

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>n</th>
<th>Functional form test</th>
<th>Omnibus test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\beta_2 = 1.0$</td>
<td>50</td>
<td>0.820</td>
<td>0.815</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.990</td>
<td>0.989</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>$\beta_2 = 0.5$</td>
<td>50</td>
<td>0.325</td>
<td>0.318</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.613</td>
<td>0.614</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.889</td>
<td>0.885</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>0.981</td>
<td>0.977</td>
</tr>
<tr>
<td>2</td>
<td>$\beta_1 = 0.5$</td>
<td>50</td>
<td>0.819 (0.839)</td>
<td>0.671</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.981 (0.984)</td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>1.000 (1.000)</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>1.000 (1.000)</td>
<td>1.000</td>
</tr>
</tbody>
</table>

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

Table 3

<table>
<thead>
<tr>
<th>Model</th>
<th>n</th>
<th>Link function test</th>
<th>Omnibus test</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>50</td>
<td>0.078</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.213</td>
<td>0.182</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.457</td>
<td>0.428</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>0.609</td>
<td>0.568</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>0.238</td>
<td>0.170</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.552</td>
<td>0.424</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.903</td>
<td>0.809</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>0.990</td>
<td>0.967</td>
</tr>
</tbody>
</table>

Table 4

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>n</th>
<th>Functional form test</th>
<th>Omnibus test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\beta_2 = 1.0$</td>
<td>100</td>
<td>0.409</td>
<td>0.381</td>
</tr>
<tr>
<td></td>
<td></td>
<td>150</td>
<td>0.612</td>
<td>0.557</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.787</td>
<td>0.729</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>0.929</td>
<td>0.899</td>
</tr>
<tr>
<td>2</td>
<td>$\beta_2 = 0.5$</td>
<td>100</td>
<td>0.211</td>
<td>0.197</td>
</tr>
<tr>
<td></td>
<td></td>
<td>150</td>
<td>0.323</td>
<td>0.284</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200</td>
<td>0.436</td>
<td>0.403</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>0.620</td>
<td>0.569</td>
</tr>
</tbody>
</table>

See the footnote of Table 2.

Table 5

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>Est/SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{A. Response} = \text{CD4 count}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>-1.6994</td>
<td>0.6037</td>
<td>-2.82</td>
<td>0.0050</td>
</tr>
<tr>
<td>$\text{time}^2$</td>
<td>0.0202</td>
<td>0.0161</td>
<td>1.25</td>
<td>0.2108</td>
</tr>
<tr>
<td>$\text{treatment} \times \text{time}$</td>
<td>4.3972</td>
<td>0.8206</td>
<td>5.36</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\text{treatment} \times \text{time}^2$</td>
<td>-0.1029</td>
<td>0.0219</td>
<td>-4.69</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\text{B. Response} = (\text{CD4 count})^{1/2}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>-0.0510</td>
<td>0.0142</td>
<td>-3.59</td>
<td>0.0003</td>
</tr>
<tr>
<td>$\text{time}^2$</td>
<td>0.0006</td>
<td>0.0004</td>
<td>1.52</td>
<td>0.1294</td>
</tr>
<tr>
<td>$\text{treatment} \times \text{time}$</td>
<td>0.1084</td>
<td>0.0194</td>
<td>5.60</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\text{treatment} \times \text{time}^2$</td>
<td>-0.0024</td>
<td>0.0005</td>
<td>-0.47</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
Table 6
Estimates of the covariance parameters of the linear mixed model for the HIV study

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>Est/SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_1$</td>
<td>18,830.00</td>
<td>1239.56</td>
<td>15.19</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$\sigma^2_{12}$</td>
<td>49.90</td>
<td>27.32</td>
<td>1.83</td>
<td>0.0677</td>
</tr>
<tr>
<td>$\sigma^2_2$</td>
<td>6,930.00</td>
<td>478.75</td>
<td>14.56</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Residual</td>
<td>11,471.00</td>
<td>289.89</td>
<td>39.66</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Panel A: Response = CD4 count

Panel B: Response = (CD4 count)$^{1/2}$

$\sigma^2_1, \sigma^2_2, \text{ and } \sigma_{12}$ are the variances and the covariance for the random intercept and random slope.

The regression estimates for the following model:

$$\logit\{E(y_{ij} | b_i)\} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 X_{4ij} + \beta_5 X_{5ij} + \beta_6 X_{6ij} + b_i,$$

where $y_{ij}$ is the $j$th measurement of the $i$th 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 $i$th subject which is assumed to follow a zero-mean normal distribution with variance $D$.

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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>Est/SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Model 1: cosine and sine season variables with quadratic age effect model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>−2.230</td>
<td>0.243</td>
<td>−9.19</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Gender</td>
<td>−0.514</td>
<td>0.253</td>
<td>−2.03</td>
<td>0.044</td>
</tr>
<tr>
<td>Height for age</td>
<td>−0.044</td>
<td>0.022</td>
<td>−2.01</td>
<td>0.046</td>
</tr>
<tr>
<td>Seasonal cosine</td>
<td>−0.613</td>
<td>0.174</td>
<td>−3.52</td>
<td>0.0005</td>
</tr>
<tr>
<td>Seasonal sine</td>
<td>−0.166</td>
<td>0.173</td>
<td>−0.96</td>
<td>0.338</td>
</tr>
<tr>
<td>Xerophthalmia</td>
<td>0.541</td>
<td>0.482</td>
<td>1.12</td>
<td>0.262</td>
</tr>
<tr>
<td>Age</td>
<td>−0.031</td>
<td>0.008</td>
<td>−3.90</td>
<td>0.0001</td>
</tr>
<tr>
<td>Age²</td>
<td>−0.0011</td>
<td>0.0004</td>
<td>−2.70</td>
<td>0.007</td>
</tr>
<tr>
<td>D</td>
<td>0.521</td>
<td>0.333</td>
<td>1.57</td>
<td>0.118</td>
</tr>
<tr>
<td><strong>B. Model 2: indicator season variables with quadratic age effect model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>−2.274</td>
<td>0.369</td>
<td>−6.17</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Gender</td>
<td>−0.525</td>
<td>0.256</td>
<td>−2.05</td>
<td>0.040</td>
</tr>
<tr>
<td>Height for age</td>
<td>−0.046</td>
<td>0.022</td>
<td>−2.06</td>
<td>0.040</td>
</tr>
<tr>
<td>Xerophthalmia</td>
<td>0.525</td>
<td>0.487</td>
<td>1.08</td>
<td>0.280</td>
</tr>
<tr>
<td>Age</td>
<td>−0.030</td>
<td>0.008</td>
<td>−3.78</td>
<td>0.0002</td>
</tr>
<tr>
<td>Age²</td>
<td>−0.0011</td>
<td>0.0004</td>
<td>−2.75</td>
<td>0.006</td>
</tr>
<tr>
<td>I(season = 1)</td>
<td>−0.717</td>
<td>0.497</td>
<td>−1.44</td>
<td>0.149</td>
</tr>
<tr>
<td>I(season = 2)</td>
<td>0.788</td>
<td>0.340</td>
<td>2.32</td>
<td>0.021</td>
</tr>
<tr>
<td>I(season = 3)</td>
<td>0.033</td>
<td>0.364</td>
<td>0.09</td>
<td>0.928</td>
</tr>
<tr>
<td>D</td>
<td>0.549</td>
<td>0.342</td>
<td>1.61</td>
<td>0.110</td>
</tr>
<tr>
<td><strong>C. Model 3: indicator season variables with cubic age effect model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>−2.243</td>
<td>0.368</td>
<td>−6.10</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Gender</td>
<td>−0.549</td>
<td>0.255</td>
<td>−2.16</td>
<td>0.031</td>
</tr>
<tr>
<td>Height for age</td>
<td>−0.043</td>
<td>0.022</td>
<td>−1.96</td>
<td>0.050</td>
</tr>
<tr>
<td>Xerophthalmia</td>
<td>0.533</td>
<td>0.487</td>
<td>1.09</td>
<td>0.275</td>
</tr>
<tr>
<td>Age/10</td>
<td>−0.520</td>
<td>0.133</td>
<td>−3.90</td>
<td>0.0001</td>
</tr>
<tr>
<td>(age/10)²</td>
<td>−0.141</td>
<td>0.042</td>
<td>−3.36</td>
<td>0.0008</td>
</tr>
<tr>
<td>(age/10)³</td>
<td>0.034</td>
<td>0.016</td>
<td>2.13</td>
<td>0.033</td>
</tr>
<tr>
<td>I(season = 1)</td>
<td>−0.740</td>
<td>0.498</td>
<td>−1.49</td>
<td>0.136</td>
</tr>
<tr>
<td>I(season = 2)</td>
<td>0.801</td>
<td>0.341</td>
<td>2.35</td>
<td>0.019</td>
</tr>
<tr>
<td>I(season = 3)</td>
<td>0.020</td>
<td>0.365</td>
<td>0.05</td>
<td>0.960</td>
</tr>
<tr>
<td>D</td>
<td>0.517</td>
<td>0.336</td>
<td>1.54</td>
<td>0.124</td>
</tr>
</tbody>
</table>

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
Thus, 

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

where

\[ H_i(x) = \sum_{j=1}^{t_i} I(X_{ij} \leq x)(y_{ij} - m_{ij}) + \eta'(x; \theta)\Omega^{-1/2}U_i(\theta). \]

By the uniform law of large numbers (Pollard, 1990, p. 39), \( \eta(x; \theta) \) converges to a nonrandom function uniformly in \( x \). Thus, for any fixed \( x \), 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 \( W(x) \) 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 \( n^{-1/2}U(\theta) \) and the uniform convergence of \( \eta(x; \theta) \). Hence, \( W(x) \) converges weakly to a zero-mean Gaussian process with covariance function between \( x_1 \) and \( x_2 \) being

\[ \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} H_i(x_1)H_i(x_2). \]  

Conditional on the data \( (y_{ij}, X_{ij}, Z_{ij}) \) \( (i = 1, \ldots, n; j = 1, \ldots, t_i) \), the only random components in \( \hat{W}(x) \) are the standard normal variables \( G_{1i}, \ldots, G_{ni} \). As shown earlier, \( n^{-1/2} \sum_{i=1}^{n} H_i(x) \) 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 \( \hat{W}(x) \) given the data \( (y_{ij}, X_{ij}, Z_{ij}) \) \( (i = 1, \ldots, n; j = 1, \ldots, t_i) \) converges weakly to the same limiting distribution as the unconditional null distribution of \( W(x) \). 

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

\[ E\left\{ \sum_{j=1}^{t_i} I(m(\zeta_i) \leq r_1)(y_{ij} - m_{ij}) \right\} \left\{ \sum_{j=1}^{t_i} I(m(\zeta_i) \leq r_2)(y_{ij} - m_{ij}) \right\}^2 \to 0 \]

as \( \zeta_1 \to \zeta_2 \) and \( r_1 \to r_2 \) for all \( m(\zeta_i) \) sufficiently close to \( \theta \), we have

\[ \hat{W}_g(r) = \hat{W}_g(r) + o_p(1), \]

where

\[ \hat{W}_g(r) = n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{t_i} I(m_{ij} \leq r)e_{ij}. \]

The arguments for the weak convergence of \( \hat{W}(x) \) can then be used to show the weak convergence of \( \hat{W}_g(r) \). The proof
for the weak convergence of \( \hat{W}_g(r) \) is exactly the same as that for \( W(x) \).

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

\[
\hat{W}^*(x) = n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{l_i} c_{ij} I(X_{ij} \leq x) e_{ij},
\]

where \( c_{ij} = \{c_{i1}, \ldots, c_{it_i}\} \) is \( 1 \times V^{-1} \). The weak convergence of \( \hat{W}^*(x) \) follows from that of \( W(x) \). Because \( C_t \) converges in probability to \( c_t \), \( W^*(x) \) converges to the same limit as \( W^*(x) \). This establishes the weak convergence of \( W^*(x) \). Similar arguments can be used to show the weak convergence of \( \hat{W}_g(r) \).

\section*{Appendix B}

\textit{Consistency of Supremum Tests}

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

\[
\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{l_i} \int_{v \leq x} \{ m_{yj}(v) - \int g^{-1}(v'\beta^* + Z_{ij}^* b_i) df_{x_i}(v) \} dv, \]

which is nonzero for some \( x \) 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 \( k \)th covariate is misspecified. For example, if \( f_k(b) \) is the Gaussian distribution with mean 0 and covariance matrix \( D \), we can show that, for standard link functions, \( S_k \) is consistent against misspecification of the functional form of \( X_k \) provided that \( \alpha \) and the components of \( \beta \) for the remainder of \( X \) 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, \( m_{yj} = X_{ij}^* \alpha_{i} \); for the log link, \( m_{yj} = \exp(X_{ij}^* \beta + Z_{ij}^* D Z_{ij})/2 \); for the probit link, \( m_{yj} = \Phi(\alpha(D) \cdot X_{ij}^* \beta), \) where \( \alpha(D) = \{D Z_{ij}^* D Z_{ij} + I^{1/2}\} \); for the logit link, \( m_{yj} \approx \exp(c_{ij} X_{ij}^* \beta) \{1 + \exp(c_{ij} X_{ij}^* \beta)\} \), where \( c_{ij} = |c^2 D Z_{ij}^* D Z_{ij} + I^{1/2}| \) 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 \( S_k \) 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 \( E(y_{ij}) \approx g^{-1}(X_{ij}^* \beta), \) 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 \( n^{-1/2} W_k(r) \) converges in probability to

\[
\lim_{n \to \infty} \sum_{i=1}^{n} \sum_{j=1}^{l_i} \int_{v' \leq r} \{ h^{-1}(v' \beta^*) - g^{-1}(v' \beta^*) \} f_{x_i}(v) dv.
\]

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