

Published in final edited form as:

*Stat Med.* 2011 September 10; 30(20): 2562–2572. doi:10.1002/sim.4265.

## On Fitting Generalized Linear Mixed-effects Models for Binary Responses using Different Statistical Packages

Hui Zhang<sup>1</sup>, Naiji Lu<sup>2,3</sup>, Changyong Feng<sup>2</sup>, Sally W. Thurston<sup>2</sup>, Yinglin Xia<sup>2,3</sup>, and Xin M. Tu<sup>2,3,4</sup>

<sup>1</sup>Department of Biostatistics, St. Jude Children's Research Hospital Memphis, TN 38105

<sup>2</sup>Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642

<sup>3</sup>Department of Psychiatry, University of Rochester, Rochester, NY 14642

<sup>4</sup>VA Center of Excellence at Canandaigua, Canandaigua VAMC, Canandaigua, NY 14424

### Summary

The generalized linear mixed-effects model (GLMM) is a popular paradigm to extend models for cross-sectional data to a longitudinal setting. When applied to modeling binary responses, different software packages and even different procedures within a package may give quite different results. In this report, we describe the statistical approaches that underlie these different procedures and discuss their strengths and weaknesses when applied to fit correlated binary responses. We then illustrate these considerations by applying these procedures implemented in some popular software packages to simulated and real study data. Our simulation results indicate a lack of reliability for most of the procedures considered, which carries significant implications for applying such popular software packages in practice.

### Keywords

Integral Approximation; Linearization; GLIMMIX; lme4; NLMIXED; R; SAS; ZELIG

## 1 Introduction

Longitudinal study designs have become increasingly popular in clinical trials and observational studies across all disciplines, as they capture both between-individual differences and within-subject dynamics. They also offer the opportunity to study more complex biological, psychological and behavioral hypotheses, especially those involving changes over time such as causal treatment effects. Since longitudinal study designs create serial, or within-subject, correlations over repeated assessments, statistical methods for cross-sectional data analysis must be generalized to account for such correlated observations.

---

### Conflict of Interests Statement

The authors have declared no conflict of interest.

One popular paradigm to handle this extension is the generalized linear mixed-effects model (GLMM). This approach explicitly models the within-subject correlation using random effects, or latent random variates [1]. Since the likelihood function arising from GLMM is generally quite complex, it is difficult to directly maximize this function except for the special linear models case for continuous outcomes. As a result, different estimates have been proposed and implemented by major statistical packages to provide inference about model parameters. For example, in SAS [2], which is arguably the most popular software package for data analysis, there are two major procedures that provide support for fitting GLMM. In R [3], another popular package for data analysis, especially among academic investigators, there are at least three packages for inference from GLMM. Thus, for many practitioners, it is quite confusing to select an appropriate procedure to use for their data at hand. A more important question is whether these procedures all provide valid inference.

In this paper, we discuss the major algorithms that drive the different procedures within as well as across different packages and investigate the performance of these procedures when applied to binary responses. We chose to examine the binary response since our experiences indicate this response type seems to mostly accentuate the differences between the different procedures, although our considerations apply equally well to other response types such as count responses. In Section 2, we give a brief review of GLMM for modeling binary responses. In Section 3, we discuss major software packages for fitting GLMM. In Section 4, we use simulated data and real data to examine the performance of these procedures/packages. In Section 5, we give our concluding remarks.

## 2 Generalized Linear Mixed-effects Models

We start with a brief review of the generalized linear mixed-effects model (GLMM) for modeling longitudinal binary outcomes and then discuss two popular approaches for inference from GLMM.

### 2.1 GLMM for Binary Outcome

Consider a sample of  $n$  subjects and let  $y_i$  denote a binary response and  $\mathbf{x}_i$  a vector of explanatory variables (predictors/covariates) of interest from the  $i$ th subject ( $i = 1, 2, \dots, n$ ). The logistic regression model is given by:

$$y_i | \mathbf{x}_i \sim \text{Bernoulli}(\mu_i), \quad \mu_i = \text{logit}^{-1}(\mathbf{x}_i^T \boldsymbol{\beta}) \quad (1)$$

where *i.i.d.* denotes *independently distributed*,  $\text{Bernoulli}(\mu_i)$  the *Bernoulli* distribution with mean  $\mu_i$ , and  $\boldsymbol{\beta}$  the vector of model parameters. Inference about  $\boldsymbol{\beta}$  using either maximum likelihood or estimating equations has been extensively discussed in the literature [e.g., 3, Chapters 4 and 9; 4, Chapter 2] and thus is not repeated for space consideration.

For a longitudinal study, each subject will be repeatedly assessed over time. For ease of exposition, we assume a design with  $m$  fixed assessment points and let  $y_{it}$  and  $\mathbf{x}_{it}$  denote the binary response and explanatory variables of interest from the  $i$ th subject at the  $t$ th assessment time ( $t = 1, 2, \dots, M$ ). Jointly modeling the repeated outcomes  $y_{it}$  is difficult since no valid joint distribution can be constructed either in general, or for a particular set of

parameter values. For example, if  $\mathbf{y}_i = (y_{i1}, \dots, y_{im})^\top$  is an equicorrelated vector of binary responses with the same success probability and a symmetric distribution, then the lower bound on the correlation parameter depends on  $n$  and  $\mu_i$  [6]. If further restrictions are placed on the joint distribution, as in Bahadur [7], the correlation is also restricted from above [8,9].

Given such complexity, a popular approach to extend the model in (1) to account for dependence across repeated  $y_{it}$  is to use random effects [5,10,11,12,13,14]. More precisely, consider the following extension of (1):



$$y_{it} | \mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{b}_i \sim \text{Bernoulli}(\text{logit}^{-1}(\mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{z}_{it}^\top \boldsymbol{\gamma} + \mathbf{b}_i^\top \boldsymbol{\delta})) \quad (2)$$

where  $N(\boldsymbol{\mu}, \Sigma)$  denotes a multivariate normal with mean  $\boldsymbol{\mu}$  and variance  $\Sigma$ , and  $\mathbf{z}_{it}$  a vector of predictors/covariates. Although  $\mathbf{b}_i$  is typically assumed to follow a multivariate normal as in (2), other types of distributions may also be considered [11]. In addition, even conditioning on  $\mathbf{x}_{it}$ ,  $\mathbf{z}_{it}$ , and  $\mathbf{b}_i$ , the  $y_{it}$ 's may still be correlated. However, we focus on the normal random-effects and independent  $y_{it}$ 's (conditional on  $\mathbf{x}_{it}$ ,  $\mathbf{z}_{it}$ ,  $\mathbf{b}_i$ ) as specified in (2) in this report, as it is the most popular in applications.

## 2.2 Inference for GLMM

For the standard logistic model in (1), the log-likelihood is readily evaluated and an objective function for estimation of the parameters is simple to construct based on the independence of the observations. For the GLMM-based extension in (2), the log-likelihood is also easily derived based on the independent  $y_{it}$ 's upon conditioning on  $\mathbf{x}_{it}$ ,  $\mathbf{z}_{it}$  and  $\mathbf{b}_i$ :



$$\ell(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}) = \sum_i \sum_t y_{it} \log(\pi_{it}) + (1 - y_{it}) \log(1 - \pi_{it}) \quad (3)$$

where  $\phi(\cdot | \mathbf{0}, \Sigma_b)$  denotes the probability density function of a multivariate normal with mean  $\mathbf{0}$  and variance  $\Sigma_b$ . Although the log-likelihood seems simple in appearance, it is not a straightforward task to numerically compute the maximum likelihood estimate (MLE) because of the potentially high-dimensional integration when a large number of random effects (i.e., high dimension of  $\mathbf{b}_i$ ) is used.

The two most popular approaches for estimation in GLMM as implemented in major software packages such as SAS and R are (1) approximating the log-likelihood function and (2) approximating the model. Algorithms in the second category are developed using linearization of or Laplace approximation to the model. They employ some type of expansions such as the Taylor series expansion to approximate the model by one based on pseudo-data with fewer nonlinear components. The process of computing the linear approximation must be repeated several times until some criteria indicate lack of further progress [15,16]. The fitting methods based on linearization typically involve two levels of iterations. The GLMM is first approximated by a linear mixed-effects model based on current values of the covariance parameter estimates, and the resulting linear mixed-effects model is then fit, which is itself an iterative process. On convergence, the new parameter

estimates are used to update the linearization, which results in a new linear mixed-effects model. The iterative process terminates when the difference in parameter estimates between successive linear mixed model fits falls within a specified tolerance level. Note that the Laplace approximation is a special case of the Gauss-Hermite quadrature discussed below, when there exists only one quadrature point.

The advantages of the linearization approach include a relatively simple form of the linearized model that typically can be fit based only on the mean and variance in the linearized form. This approach can fit models with a large number of random effects, crossed random effects, multiple types of subjects, and even correlated  $y_{it}$ 's after conditioning on  $\mathbf{x}_{it}$ ,  $\mathbf{z}_{it}$  and  $\mathbf{b}_i$ . However, as it does not maximize the underlying log-likelihood function, this approach produces estimates with unknown asymptotic properties, except in the special linear mixed-effects model case for which this approach does produce the MLE [17]. In addition, algorithms implementing this approach can fail at both levels of the double iteration scheme.

In contrast, the integral approximation approach aims to directly approximate the log-likelihood in (3) and maximize the approximated function [13,14]. Various techniques have been proposed to compute the approximation including Newton and Gauss-Hermite quadratures, Monte Carlo integration, and Markov Chain Monte Carlo. The advantage of this alternative is to provide an actual objective function for optimization, albeit it is an approximated log-likelihood. Nonetheless, this enables likelihood ratio tests and likelihood-based fit statistics. Further, unlike the linearization approach whose approximation accuracy and thus the asymptotic properties of the estimates are limited by the type of models fit (e.g., linear, logistic, etc.), the approximation to the log-likelihood under this approach can be improved to any degree by increasing the precision of numerical integration, at least in principle (see Section 4.1 for such a dependent relationship in some procedures from simulated data). Thus, algorithms based on the integral approximation are expected to provide better estimates than their linearization-based alternatives. In particular, since they can get arbitrarily close to the MLE as the precision of numerical integration increases, estimates obtained under this approach have similar nice large sample properties such as consistency and asymptotic normality and efficiency. One drawback of the integral approximation is the difficulty in getting a good approximation when there are many random effects and/or nested random effects in the model. In Section 4, we compare the performance of the different approaches as implemented in the two most popular software, SAS and R, using simulated data.

### 3 Available Software for GLMM with Binary Response

In this section, we discuss major procedures in SAS and R for inference about the GLMM model in (2).

#### 3.1 SAS

Two major procedures in SAS for fitting GLMM are GLIMMIX and NLMIXED. The primary difference between the two lies in the estimation approaches employed. While GLIMMIX implements the linearization approach based on the Laplace approximation,

NLMIXED employs the integral approximation to the log-likelihood to provide inference. NLMIXED approximates the log-likelihood by integral approximation through an adaptive Gaussian quadrature when carrying out the dual quasi-Newton optimization. The objective function for the NLMIXED procedure is the marginal log likelihood in (2) obtained by integrating out the random effects from the joint distribution of responses and random effects using quadrature techniques. As noted earlier, this approach is expected to yield estimates that can be made arbitrarily close to the MLE as the approximation precision to the log-likelihood improves. The NLMIXED procedure not only generates true log-likelihood fit statistics that can be used to compare nested models, but also permits greater flexibility to accommodate user-defined likelihood functions to implement other statistical models. A major limitation of NLMIXED is that the  $y_{it}$ 's in (2) cannot be correlated after controlling for  $\mathbf{x}_{it}$ ,  $\mathbf{z}_{it}$  and  $\mathbf{b}_i$ .

GLIMMIX, on the other hand, fits GLMM based on the approach of linearization [18]. This allows multiple random effects, nested and crossed random effects, multiple cluster types, and within our context, a random structure to accommodate correlated  $y_{it}$ 's in (2) even after controlling for  $\mathbf{x}_{it}$ ,  $\mathbf{z}_{it}$  and  $\mathbf{b}_i$ . However, as noted earlier, the procedure only generates Wald-type test statistics and cannot perform likelihood-based tests such as the likelihood ratio statistic due to the absence of a true objective function for the overall optimization. GLIMMIX implements the restricted maximum likelihood (REML) method, which has been shown to produce better estimates than maximum likelihood (ML) when the number of higher-level units is small [19]. In addition, GLIMMIX also supports sandwich variance estimates, whereas NLMIXED supports only model-based standard errors. The latter option has been shown to provide more robust estimates in the special case of linear mixed-effects model in the presence of missing data when the assumed parametric distributions failed to adequately fit the data [21]. However, a major problem with GLIMMIX is that it does not maximize the underlying log-likelihood function, producing estimates with unknown asymptotic properties except in the special linear model case. This flaw is particularly problematic for modeling binary responses as we demonstrate using simulated data in Section 4.

Note that initial parameter values are specified differently between the two procedures. Initial values must be manually set for NLMIXED prior to running the procedure. Within our context of binary outcomes, such values may be generated by fitting the data first using generalized estimating equations [4,9] via PROC GENMOD. In contrast, GLIMMIX expects no such user input and uses a double iteration scheme to generate initial values from an iteratively derived approximated linear mixed-effects model.

### 3.2 R

There are also multiple procedures (or packages as they are called in R) available for fitting GLMM in R. One popular package is lme4 (version 0.99875-8 was used in the simulation study), which implements the Gauss-Hermite quadrature to approximate the log-likelihood using numerical integration. It defaults to the Laplace approximation when only one quadrature point is used (the default option of lme4) [22]. Another available package is ZELIG, which was proposed to serve as a common framework for statistical analysis and

software development built around the R language [23]. For binary responses, this function uses the same computing engine as in lme4 to provide estimates, and thus yields quite similar results as the latter package. A third option is the glmmML package [18]. Like lme4 and ZELIG, glmmML offers integral approximations to the log-likelihood using the Gauss-Hermite quadrature, in addition to the Laplace approximation. Thus, like its SAS counterpart NLMIXED, lme4, ZELIG and glmmML can generate true log-likelihood fit statistics to provide improved inference about model parameters. For small samples, the Wald statistic based on the MLE may not provide reliable inference as the sampling distribution is likely to be poorly approximated by a (multivariate) normal distribution. The likelihood statistic, however, may still provide good inference since it is computed based on the height rather than the entire distribution of the log-likelihood statistic. As in the case of NLMIXED, a primary limitation of the integral approximation approach is that the  $y_{it}$ 's in (2) must be uncorrelated after controlling for  $\mathbf{x}_{it}$ ,  $\mathbf{z}_{it}$  and  $\mathbf{b}_i$ , making it impossible to model correlated errors above and beyond those accounted for by the random effects.

## 4 Illustration

We illustrate our considerations with simulated as well as real study data. We investigate the performance of each procedure/package within R and SAS so we can compare the accuracy of estimates and inference across the different procedures/packages. All the codes for the simulation and real study data are available from the authors upon request. In all the examples, we set the type I error level at  $\alpha = 0.05$ . We start with simulated data.

### 4.1 Simulation Study

We simulated data from a GLMM model with a single explanatory variable within a longitudinal data setting containing 3 assessments. To evaluate the performance of the procedures, we repeated the analysis for a range of sample size,  $n = 50, 100, 500$ , representing small, moderate and large samples. To obtain reliable estimates of model parameters and type I errors, we perform 1,000 Monte Carlo (MC) replications for each simulation run.

We simulated the explanatory variable  $x_i$  and the response  $y_{it}$  based on the following GLMM:


(4)

where  $\beta_0 = \beta_1 = 1$  and  $\tau = 0.001, 0.5$  and  $2$ . For  $\tau = 0.001$ , the within-subject correlation was very small and thus negligible, making the  $y_{it}$  conditional on  $x_i$  follow approximately three independent Bernoulli distributions. Thus, for small  $\tau$ , the different procedures/packages should yield similar results. For  $\tau = 0.5$  and  $\tau = 2$ , the within-subject correlations were about  $0.2$  and  $0.5$  (estimated based on the simulated data), respectively. We did not consider higher correlations in our simulations since they are unrealistic for most applications in our



opinion. Further, our experience with the software packages seemed to produce unreliable estimates when fitting GLMM to data with extremely high correlations.

We considered the hypothesis,  $H_0: \beta_1 = 1$  vs.  $H_a: \beta_1 \neq 1$ , and simulated the  $y_{it}$ 's from the model in (4) with different  $\tau$ 's. For each model, we estimated the type I error based on the

Wald statistic,  $W = \frac{\hat{\beta}_1 - 1}{\sqrt{\hat{\Sigma}_{\beta}}}$  is the element of the estimated asymptotic variance  $\Sigma_{\beta}$  corresponding to  $\beta_1$ . This statistic at the  $m$ th MC replication ( $1 \leq m \leq 1000$ ) will be denoted by  $W_m$ . Then, the type I error rate for testing  $H_0$  was estimated by:

$\hat{\alpha} = \frac{1}{1000} \sum_{m=1}^{1000} I(W_m \geq q_{1,0.95})$ , where  $q_{1,0.95}$  is the 95th percentile of  $N(0,1)$ .

Shown in Table 1, 2, and 3 are the estimates of model parameters  $\beta$  and associated standard errors, averaged over 1,000 MC replications, and the type I error rates based on the different procedures (packages) in SAS (R) for the three sample sizes considered. For SAS GLIMMIX, both maximum likelihood (ML) and restricted maximum likelihood (REML) estimates were reported. For R, results were reported for both the linearization (labeled as "Laplace") and integral approximation (labeled as "Gauss-Hermite") approaches, with the latter results based on 20 quadrature points.

As expected, for small  $\tau = 0.001$ , all the procedures and packages gave similar estimates of  $\beta$  and standard errors across the sample sizes. As  $\tau$  increased to 0.5 and 2, results for the estimates and standard errors become quite discrepant across the different procedures (packages). Within SAS, there was downward bias in both the estimates of  $\beta$  and standard errors by GLIMMIX; the amount of bias reached almost 50% at  $\tau = 2$ . In both cases of  $\tau$ , NLMIXED provided good estimates.

As in the case of SAS, estimates of  $\beta$  and standard errors based on R packages began to differ as  $\tau$  increased. Results from lme4 and ZELIG were closer to each other than those from glmmML, which is not surprising, given that lme4 and ZELIG were based on the same computing algorithms as noted earlier. Estimates of  $\beta$  and standard errors from glmmML were downwardly biased, though the patterns of bias were more difficult to decipher for lme4 and ZELIG. The integral approach based on the Gauss-Hermite approximation seemed to reduce the amount of bias in the estimated  $\beta$ 's and the standard errors, especially for lme4 and ZELIG. It is interesting to note that although the Gauss-Hermite approximation in these packages employs the same integral approximation approach to the log-likelihood as its SAS NLMIXED counterpart, it does not seem to provide more accurate estimates (than its Laplace counterpart) for these packages.

Estimates of type I error rates were quite variable across the different procedures and functions as well. For  $\tau = 0.001$ , all type I error rates were closer to the nominal value 0.05, especially for  $n = 500$ , though those from SAS GLIMMIX and R glmmML has small downward bias, while the ones from lme4 and ZELIG showed some upward bias. For  $\tau = 0.5$  and 2, all type I error rates increased substantially, except for those from NLMIXED that remained close to 0.05. It is quite clear that with the exception of NLMIXED, none of the

procedures would provide acceptable type I error levels for inference about the null  $H_0 : \beta_1 = 1$  when the within-subjects correlation is 0.2 or greater.

Note that during the review of this manuscript, one reviewer suggested that we reran the simulation model within the same setting, except for increasing the magnitude of the  $\beta$ 's, to see if larger (absolute) values of  $\beta$ 's would play any role in the bias. We found that the biases and type I errors did seem to depend on the magnitude of  $\beta$ 's. For example, shown in Tables 4, 5 and 6 are the estimates, standard errors and type I errors for three different set values of the  $\beta$ 's,  $\beta_0 = \beta_1 = 0, 1.5$  and 2. As seen, the bias in the type I error does increase, as the values of  $\beta$ 's move further away from 0. The bias in the type I error showed a similar pattern of increase when  $\beta$ 's were shifted away from 0 to the left.

Note also that the standard error estimates are generally smaller for the SAS GLIMMIX and the R procedures compared to the SAS NLMIXED, especially for GLIMMIX. However, smaller standard error estimates may not be a good thing when parameter estimates are highly biased. As estimates from some procedures such as those based on the Laplace approximation have unknown asymptotic properties, it is more important to examine the bias in the parameter estimates and in the type I error. Highly biased estimates with smaller standard estimates are the worst combination since it not only leads to wrong conclusions, but does so with greater statistical power.

## 4.2 Real Data Example

We also applied the different procedures/packages to a multi-center randomized clinical trial comparing two treatments for a respiratory illness [24]. There were 111 subjects randomized to active (54) or placebo (57) treatment across two sites. The primary outcome was respiratory status (0=poor, 1=good), which was assessed at 4 visits. The study data was analyzed by Davis using distribution-free (or semi-parametric) longitudinal methods [24]. In addition to the treatment conditions, he also included baseline respiratory status, study site, sex and age as covariates. There was no missing data in the outcome, the predictor (treatment conditions), and the covariates. Our reanalysis using GLMM included the same predictor and covariates as the fixed effects, and a random intercept.

Shown in Table 7 are the estimates, standard errors and p-values from fitting the GLMM using the different procedures/packages. The results from the R packages lme4 and glmmML were based on the Gauss-Hermite approach for integral approximation with 3 quadrature points, since there is no significant improvement using more quadrature points for these procedures (see Discussion for more details). Since ZELIG yielded almost identical results as lme4, we only reported the results for the latter package. Likewise, we only reported the results from the REML option of SAS GLIMMIX, since they showed slightly less bias than those from the ML method as observed in the simulation study.

The estimates of  $\beta$  and standard errors showed similar patterns as those from the simulation study. In particular, the standard errors from SAS GLIMMIX and R glmmML all had almost 50% downward bias (with NLMIXED as a reference standard), consistent with the amount of bias observed for these procedures in the simulation study. The standard errors from lme4 also seemed to have some downward bias. The differences in the p-values between



NLMIXED and the two R packages are larger than those between NLMIXED and GLIMMIX. The downward bias in both the estimates and standard errors by GLIMMIX may have helped reduce the amount of upward bias in the p-values yielded by this SAS procedure.

## 5 Discussion

We examined the performances of procedures/packages for fitting generalized linear mixed-effects models (GLMM) for correlated binary responses using the popular SAS and R statistical software packages. Unlike the special case of a linear mixed-effects model, maximum likelihood estimates (MLE) are difficult to compute and approximate estimates have been proposed and implemented in most statistical software packages such as SAS and R. As some of these estimates have unknown large sample properties, it is important to evaluate their performances to determine if they can provide reasonably good estimates and inference for practical purposes.

We reviewed two popular approaches for providing such approximate estimates as implemented in SAS and R, and discussed the pros and cons between the two approaches. Although the linearization approach is flexible, and can accommodate a complex structure of random effects, it may not provide reliable estimates, since it is difficult to theoretically characterize the properties of such estimates. In particular, it is unclear whether such estimates are even consistent. In comparison, the integral-based approximation approach yields estimates that approximate the MLE, thereby providing not only consistent, but asymptotically normal estimates. The superior performance of the latter approach as demonstrated by the SAS NLMIXED in our simulation study confirms the theoretical behavior of the estimates from this approach. We are a bit surprised by the performance of the R lme4 and glmmML packages, as neither seems to yield comparable inference as its SAS NLMIXED counterpart given that it implements the same integral approximation approach, albeit using different algorithms.

To further investigate the robustness of inference against non-normal random effects, we also changed the bivariate normal random effects  $\mathbf{b}_j$  to a shifted bivariate-Gamma with the same variance structure, i.e.,


(5)

where  $SG(\boldsymbol{\mu}, \boldsymbol{\Phi})$  denotes a Shifted bivariate Gamma with mean  $\boldsymbol{\mu}$  and variance  $\boldsymbol{\Phi}$  [25].

Shown in Table 8 are the simulation results from this revised model for  $\tau = 0.5$  and 2 based on a sample size  $n = 100$  under 1,000 MC replications. Again, we used 3 quadrature points for the R procedures to save computing time. Although parameter estimates and type I errors were less affected for  $\tau = 0.5$ , a large bias was obvious for both when  $\tau = 2$ , even for NLMIXED. Along with the findings for normal random effects, these results seem to

indicate that GLMM for binary outcomes is not robust against misspecification of random effects.

Judging from the results of our simulation study, we conclude that the SAS NLMIXED procedure provides the most accurate parameter estimates and inference (type I error) under correct model assumptions. Among the remaining procedures, estimates are generally biased, especially the type I errors, with the level of accuracy depending not so much on the number of quadrature points (for the procedures based on the integral approximation), but rather on the magnitude of the values of the parameters. For example, we varied the number of for the latter in the range of 3 and 20 for the R procedures, but failed to observe any improvement in type I errors when increasing the quadrature points from 3 to 20, except for a significant increase in computing time, with the latter increased by 15 times when using 20 over 3 quadrature points. Thus, for all practical purposes, it seems that 3 or 4 quadrature points is sufficient when using these procedures. However, as noted in Section 2.2, we should limit NLMIXED to models with a relatively small number of random effects to ensure accurate intergral approximation to the log-likelihood.

Unexpectedly, though quite interestingly as well, the levels of accuracy of these remaining procedures seem to depend on the magnitude of the values of the parameters, with larger (absolute) values yielding higher biases. Further investigations are needed to determine the sources of such a relationship so improvement may be made to these procedures. If these procedures have to be used in a given application for reasons such as those discussed in Section 3, we recommend the use of bootstrapped standard errors to improve inference reliability [26].

## Acknowledgments

This research is supported in part by NIH grant U54 RR023480 and UL1-RR024160.

## References

1. Breslow NE, Clayton DC. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*. 1993; 88:9–25.
2. SAS Institute Inc. Enhancements in SAS/STAT 9.2 Software. Cary, NC: SAS Institute Inc.; 2008.
3. R Development Core Team. Vienna, Austria: R Foundation for Statistical Computing; 2010. R: A language and environment for statistical computing. ISBN 3-900051-07-0, URL <http://www.R-project.org>
4. McCullagh, P.; Nelder, JA. *Generalized Linear Models*. 2nd edition. London: Chapman and Hall; 1989.
5. Kowalski, J.; Tu, XM. *Modern Applied U Statistics*. New York: Wiley; 2008.
6. Gilliland D, Schabenberger O. Limits on Pairwise Association for Equi-Correlated Binary Variables. *Journal of Applied Statistical Sciences*. 2001; 10:279–285.
7. Bahadur, RR. *Studies in Item Analysis and Prediction*. Stanford, CA: Stanford University Press; 1961. “A Representation of the Joint Distribution of Responses to n Dichotomous Items”.
8. Prentice RL. Correlated binary regression with covariates specific to each binary observation. *Biometrics*. 1988; 44:321–327.
9. Shults J, Sun W, Tu XM, Kim H, Amsterdam J, Hilbe JM, Ten-Have T. A comparison of several approaches for choosing between working correlation structures in generalized estimating equation

analysis of longitudinal binary data. *Statistics in Medicine*. 2009; 28:2338–2355. [PubMed: 19472307]

10. Diggle, PJ.; Heagerty, PJ.; Liang, KY.; Zeger, SL. *Analysis of Longitudinal Data*. 2nd. edition. New York: Oxford University Press; 2002.
11. Davidian M, Gallant RA. The nonlinear mixed effects model with a smooth random effects density. *Biometrika*. 1993; 80:475–488.
12. Fitzmaurice, GM.; Laird, NM.; Ware, JH. *Applied Longitudinal Analysis*. New York: Wiley; 2004.
13. Demidenko, E. *Mixed Models: Theory and Applications*. Hoboken, New Jersey: Wiley; 2004.
14. Hedeker, D.; Gibbons, RD. *Longitudinal Data Analysis*. New York: Wiley; 2006.
15. Pinheiro, JC.; Bates, DM. *Mixed-effects models in S and S-PLUS*. New York: Springer; 2000.
16. Schabenberger O, Gregoire TG. Population-averaged and subject-specific approaches for clustered categorical data. *Journal of Statistical Computation and Simulation*. 1996; 54:231–253.
17. Lindstrom MJ, Bates DM. Newton–Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*. 1988; 83:1014–1022.
18. Broström, G. *Technical Reports*. Umeå, Sweden: Department of Statistics, Umeå University; 2003. Generalized linear models with random intercepts. SE-90187
19. Wolfinger R, O’Connell M. Generalized linear mixed models: a pseudolikelihood approach. *Journal of Statistical Computation and Simulation*. 1993; 48:233–243.
20. Flom PL, McMahon JM, Pouget ER. Using PROC NLMIXED and PROC GLMMIX to analyze dyadic data with binary outcomes. NESUG. 2006 URL <http://www.nesug.org/proceedings/nesug06/an/da08.pdf>.
21. Lu N, Tang W, He H, Yu Q, Crits-Christoph P, Zhang H, Tu XM. On the Impact of Parametric Assumptions and Robust Alternatives for Longitudinal Data Analysis. *Biometrical Journal*. 2009; 51:627–643. [PubMed: 19688758]
22. Bates D. Fitting mixed-effects models using the lme4 packages in R. 2007 URL <http://www.stat.wisc.edu/~bates/IMPS2008/lme4D.pdf>.
23. Imai K, King G, Lau O. Toward a common framework for statistical analysis and development. *Journal of Computational and Graphical Statistics*. 2008; 17:892–913.
24. Davis CS. Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. *Statistics in Medicine*. 1991; 10:1959–1980. [PubMed: 1805321]
25. Moran PAP. Statistical inference with bivariate gamma distributions. *Biometrika*. 1969; 56:627–634.
26. Field CA, Welsh AH. Bootstrapping clustered data. *Journal of the Royal Statistical Society: Series B*. 2007; 69:369–390.

Table 1

Estimates of parameters, standard errors, and type I errors for the simulation model with  $n = 50$  using different procedures/packages.

Comparison of estimates across procedures/packages for the simulation model with $n = 50$								
$\tau$	Package	Procedure	type I error	$\beta_0 = 1$	s.e. $\times 10$	$\beta_1 = 1$	s.e. $\times 10$	
0.001	SAS	NLMIXED	0.031	1.073	2.51	1.108	2.95	
		GLIMMIX	REML	0.036	1.026	2.22	1.048	2.57
			ML	0.038	1.025	2.20	1.044	2.53
	R	lme4	Laplace	0.070	1.084	2.27	1.121	2.63
		ZELIG	Gauss-Hermite	0.057	1.092	2.29	1.123	2.66
			Laplace	0.070	1.085	2.27	1.121	2.63
	glimmML	Gauss-Hermite	0.057	1.088	2.28	1.119	2.66	
		Laplace	0.026	1.046	2.30	1.063	2.57	
		Gauss-Hermite	0.025	1.050	2.33	1.067	2.60	
	0.5	SAS	NLMIXED	0.043	1.030	2.26	1.066	2.74
GLIMMIX			REML	0.147	0.939	1.82	0.935	2.14
			ML	0.150	0.935	1.79	0.929	2.08
R		lme4	Laplace	0.099	1.048	2.37	1.093	2.82
		ZELIG	Gauss-Hermite	0.103	1.060	2.36	1.081	2.81
			Laplace	0.099	1.048	2.37	1.093	2.82
glimmML		Gauss-Hermite	0.097	1.055	2.36	1.077	2.80	
		Laplace	0.072	0.996	2.44	0.982	2.66	
		Gauss-Hermite	0.066	1.004	2.50	0.990	2.72	
2		SAS	NLMIXED	0.040	1.049	4.36	1.052	5.91
	GLIMMIX		REML	0.523	0.601	2.33	0.532	2.83
			ML	0.549	0.599	2.25	0.526	2.70
	R	lme4	Laplace	0.179	1.165	2.37	1.252	2.82
		ZELIG	Gauss-Hermite	0.184	1.078	4.20	1.043	5.12
		Laplace	0.180	1.163	2.37	1.251	2.82	
		Gauss-Hermite	0.185	1.055	4.21	1.049	5.25	

Comparison of estimates across procedures/packages for the simulation model with $n = 50$								
$\tau$	Package	Procedure		type I error	$\beta_0 = 1$	s.e. $\times 10$	$\beta_1 = 1$	s.e. $\times 10$
		glmmML	Laplace	0.158	1.0572	2.44	0.8326	2.66
			Gauss-Hermite	0.118	0.9949	2.50	0.8221	2.72

**Table 2**

Estimates of parameters, standard errors, and type I errors for the simulation model with  $n = 100$  using different procedures/packages.

Comparison of estimates across procedures/packages for the simulation model with $n = 100$							
$\tau$	Package	Procedure	type I error	$\beta_0 = 1$	s.e. $\times 10$	$\beta_1 = 1$	s.e. $\times 10$
0.001	SAS	NLMIXED	0.041	1.053	1.70	1.075	1.98
		GLIMMIX	0.042	1.015	1.51	1.023	1.72
		ML	0.043	1.012	1.50	1.020	1.71
	R	lme4	0.076	1.060	1.56	1.084	1.79
		Gauss-Hermite	0.063	1.056	1.55	1.078	1.78
		Laplace	0.076	1.060	1.56	1.084	1.79
	ZELIG	Gauss-Hermite	0.062	1.055	1.55	1.078	1.78
		Laplace	0.034	1.032	1.59	1.035	1.76
		glnmML	0.032	1.036	1.61	1.039	1.78
	SAS	NLMIXED	0.050	1.022	1.31	1.021	1.56
0.5	SAS	GLIMMIX	0.150	0.935	1.05	0.904	1.22
		ML	0.159	0.933	1.04	0.902	1.21
		Laplace	0.121	1.041	1.62	1.052	1.91
	R	Gauss-Hermite	0.104	1.021	1.61	1.039	1.89
		Laplace	0.121	1.041	1.62	1.052	1.91
		Gauss-Hermite	0.099	1.020	1.61	1.038	1.89
	glnmML	Laplace	0.098	0.994	1.71	0.948	1.83
		Gauss-Hermite	0.094	1.004	1.75	0.958	1.88
	SAS	NLMIXED	0.044	1.013	2.43	0.999	3.25
		GLIMMIX	0.829	0.592	1.36	0.494	1.62
2	R	ML	0.844	0.585	1.34	0.489	1.59
		Laplace	0.161	1.113	2.99	1.186	3.96
		Gauss-Hermite	0.192	0.998	2.80	1.003	3.34
	ZELIG	Laplace	0.162	1.114	2.99	1.185	3.96
		Gauss-Hermite	0.182	0.995	2.79	1.024	3.44



Comparison of estimates across procedures/packages for the simulation model with $n = 100$							
$\tau$	Package	Procedure		type I error	$\beta_0 = 1$	s.e. $\times 10$	$\beta_1 = 1$
		glmmML	Laplace	0.167	1.0002	3.26	0.7982
			Gauss-Hermite	0.141	0.9837	3.52	0.7857
							s.e. $\times 10$
							3.28
							3.56

**Table 3**

Estimates of parameters, standard errors, and type I errors for the simulation model with  $n = 500$  using different procedures/packages.

Comparison of estimates across procedures/packages for the simulation model with $n = 500$							
$\tau$	software	Procedure/Package	type I error	$\beta_0 = 1$	s.e. $\times 10$	$\beta_1 = 1$	s.e. $\times 10$
0.001	SAS	NLMIXED	0.046	1.025	0.75	1.029	0.85
		GLIMMIX	0.034	1.004	0.67	1.004	0.75
		ML	0.036	1.005	0.67	1.003	0.75
		lme4	0.094	1.029	0.67	1.036	0.75
	R	Gauss-Hermite	0.098	1.022	0.67	1.029	0.75
		Laplace	0.094	1.029	0.67	1.036	0.75
		ZELJG	0.097	1.022	0.67	1.029	0.75
		glmmML	0.040	1.014	0.69	1.013	0.76
		Gauss-Hermite	0.044	1.016	0.69	1.015	0.77
		NLMIXED	0.049	1.003	0.69	1.002	0.82
0.5	SAS	REML	0.192	0.912	0.66	0.897	0.75
		ML	0.162	0.914	0.67	0.902	0.77
		lme4	0.117	1.010	0.70	1.013	0.82
		Gauss-Hermite	0.118	1.003	0.70	1.001	0.81
	R	Laplace	0.117	1.010	0.70	1.013	0.82
		ZELJG	0.117	1.003	0.70	1.001	0.81
		glmmML	0.152	0.977	0.76	0.932	0.80
		Laplace	0.126	0.988	0.78	0.943	0.82
		Gauss-Hermite	0.066	0.992	1.25	0.976	1.66
		NLMIXED	0.992	0.605	0.88	0.519	1.05
2	SAS	REML	0.992	0.605	0.88	0.519	1.05
		ML	0.993	0.586	0.87	0.467	1.01
		lme4	0.249	1.056	1.29	1.114	1.70
		Gauss-Hermite	0.256	0.983	1.23	0.912	1.41
	R	Laplace	0.247	1.056	1.29	1.113	1.70
		Gauss-Hermite	0.209	0.988	1.23	0.955	1.49

Comparison of estimates across procedures/packages for the simulation model with $n = 500$							
$\tau$	software	Procedure/Package		type I error	$\beta_0 = 1$	$\beta_1 = 1$	s.e. $\times 10$
		glmmML	Laplace				
			Gauss-Hermite				
				0.416	0.962	0.767	1.42
				0.385	0.953	0.758	1.53

Estimates of parameters, standard errors, and type I errors for the revised simulation model with  $\beta_0 = \beta_1 = 0$  and  $n = 500$  using different procedures/packages.

Table 4

Comparison of estimates for the revised simulation model with $\beta_0 = \beta_1 = 0$ across procedures/packages ( $n = 500$ )								
$\tau$	Package	Procedure	type I error	$\beta_0 = 0$	s.e. $\times 10$	$\beta_1 = 0$	s.e. $\times 10$	
0.5	SAS	NLMIXED	0.047	0.0005	0.57	0.003	0.6	
		GLIMMIX	REML	0.041	-0.001	0.64	0.001	0.71
			ML	0.043	0.0002	0.59	-0.0003	0.60
	R	lme4-Laplace		0.075	-0.001	0.58	-0.002	0.64
		ZELIG		0.075	-0.001	0.58	-0.002	0.64
		glmmML	Laplace	0.077	-0.0003	0.60	-0.001	0.60
			Gauss-Hermite	0.066	-0.0003	0.62	-0.001	0.63
	2	SAS	NLMIXED	0.054	0.002	1.39	-0.005	1.88
GLIMMIX			REML	0.053	0.002	0.87	-0.004	1.01
			ML	0.052	0.001	0.87	-0.001	1.00
R		lme4-Laplace		0.119	-0.0003	1.20	-0.001	1.54
		ZELIG		0.119	-0.0003	1.20	-0.001	1.54
		glmmML	Laplace	0.095	0.002	1.29	0.001	1.31
			Gauss-Hermite	0.070	0.002	1.41	0.001	1.43

Table 5

Estimates of parameters, standard errors, and type I errors for the revised simulation model with  $\beta_0 = \beta_1 = 1.5$  and  $n = 500$  using different procedures/packages.

Comparison of estimates for the revised simulation model with $\beta_0 = \beta_1 = 1.5$ across procedures/packages ( $n = 500$ )							
$\tau$	Package	Procedure		type I error	$\beta_0 = 1.5$	$\beta_1 = 1.5$	se. $\times 10$
0.5	SAS	NLMIXED		0.037	1.513	1.515	1.29
		GLIMMIX	REML	0.275	1.407	1.381	0.95
		ML		0.271	1.407	1.380	0.94
	R	lme4-Laplace		0.200	1.556	1.565	1.04
		ZELIG		0.200	1.556	1.565	1.04
		glmmML	Laplace	0.133	1.473	1.426	1.04
			Gauss-Hermite	0.119	1.482	1.434	1.05
		NLMIXED		0.055	1.520	1.507	2.58
2	SAS	GLIMMIX	REML	1	0.902	0.767	1.12
			ML	1	0.903	0.767	1.10
		lme4-Laplace		0.202	1.567	1.595	1.84
	R	ZELIG		0.200	1.569	1.596	1.84
		glmmML	Laplace	0.536	1.469	1.176	1.55
			Gauss-Hermite	0.560	1.419	1.144	1.66
		Gauss-Hermite					

Table 6

Estimates of parameters, standard errors, and type I errors for the revised simulation model with  $\beta_0 = \beta_1 = 2$  and  $n = 500$  using different procedures/packages.

Comparison of estimates for the revised simulation model with $\beta_0 = \beta_1 = 2$ across procedures/packages ( $n = 500$ )								
$\tau$	Package	Procedure		type I error	$\beta_0 = 2$	s.e. $\times 10$	$\beta_1 = 2$	s.e. $\times 10$
0.5	SAS	NLMIXED		0.042	2.022	1.46	2.022	1.67
		GLIMMIX	REML	0.240	1.894	1.04	1.867	1.22
			ML	0.248	1.896	1.07	1.867	1.26
	R	lme4-Laplace		0.308	2.145	1.16	2.160	1.36
		ZELIG		0.308	2.145	1.16	2.160	1.36
		glmmML	Laplace	0.081	1.984	1.26	1.940	1.35
			Gauss-Hermite	0.078	1.989	1.26	1.945	1.36
	SAS	NLMIXED		0.053	2.012	2.19	2.006	2.92
GLIMMIX		REML	1	1.204	0.98	1.047	1.22	
		ML	1	1.201	0.99	1.045	1.22	
2		lme4-Laplace		0.260	2.007	1.50	1.983	2.00
		ZELIG		0.263	2.010	1.51	1.984	2.00
		glmmML	Laplace	0.570	2.016	1.76	1.624	1.74
	Gauss-Hermite		0.639	1.896	1.87	1.560	1.86	



Estimates of parameters, standard errors, and type I errors for the the Respiratory Clinical Data using different prodcedures/packages.

Table 7

Comparison of estimates across procedures/packages for the Respiratory Clinical Data example							
Package/Procedure	Intercept	Center	Treatment	Sex	Age	Baseline	
R lme4 (Gauss-Hermite)	Estimate	2.44	0.92	-1.97	-0.28	-0.03	2.85
	s.e.	1.96	0.50	0.49	0.67	0.02	0.50
	p-value	0.21	0.07	< 0.001	0.67	0.08	< 0.001
R glimmML (Gauss-Hermite)	Estimate	1.70	0.85	-1.74	-0.08	-0.02	2.01
	s.e.	1.18	0.29	0.35	0.40	0.01	0.34
	p-value	0.15	0.003	< 0.001	0.83	0.09	< 0.001
SAS NL MIXED	Estimate	2.25	0.97	-2.10	-0.24	-0.03	2.94
	s.e.	2.07	0.56	0.56	0.69	0.02	0.60
	p-value	0.28	0.09	< 0.001	0.72	0.18	< 0.001
SAS GLIMMIX (REML)	Estimate	1.35	0.62	-1.32	-0.12	-0.02	1.83
	s.e.	1.32	0.35	0.34	0.44	0.01	0.35
	p-value	0.31	0.08	< 0.001	0.78	0.19	< 0.001

Table 8

Estimates of parameters, standard errors, and type I errors from the revised simulation model with random effects following a shifted bivariate Gamma and  $n = 100$  using different procedures/packages.

Comparison of estimates across procedures/packages for the revised simulation model with a shifted bivariate Gamma random effect						
$\tau$	Package	Procedure	type I error	$\beta_0 = 1$	$\beta_1 = 1$	$s.e. \times 10$
0.5	SAS	NLMIXED	0.065	0.994	1.004	2.31
		GLIMMIX	REML	0.937	0.910	1.87
			ML	0.935	0.909	1.87
	R	lme4	Laplace	1.017	1.023	1.93
			Gauss-Hermite	1.002	1.019	1.83
		ZELIG	Laplace	1.017	1.023	1.93
			Gauss-Hermite	1.002	1.019	1.83
	glmmML	Laplace	0.080	0.985	0.951	1.89
			Gauss-Hermite	0.994	0.960	1.94
		NLMIXED	0.198	0.676	0.659	3.80
2	SAS	GLIMMIX	REML	0.806	0.492	2.13
			ML	0.822	0.401	2.13
		lme4	Laplace	0.360	0.683	2.83
	R		Gauss-Hermite	0.394	0.635	2.61
		ZELIG	Laplace	0.360	0.673	2.82
			Gauss-Hermite	0.394	0.635	2.61
		glmmML	Laplace	0.564	0.501	2.44
			Gauss-Hermite	0.505	0.513	2.63