An assessment of estimation methods for generalized linear mixed models with binary outcomes

Marinela Capanu[*,1], **Mithat Gönen**[1], and **Colin B. Begg**[1]
[1]Memorial Sloan-Kettering Cancer Center, 307 E 63rd St, 3rd Floor, New York, NY 10021

Abstract

Two main classes of methodology have been developed for addressing the analytical intractability of generalized linear mixed models (GLMMs): likelihood-based methods and Bayesian methods. Likelihood-based methods such as the penalized quasi-likelihood approach have been shown to produce biased estimates especially for binary clustered data with small clusters sizes. More recent methods using adaptive Gaussian quadrature perform well but can be overwhelmed by problems with large numbers of random effects, and efficient algorithms to better handle these situations have not yet been integrated in standard statistical packages. Bayesian methods, though they have good frequentist properties when the model is correct, are known to be computationally intensive and also require specialized code, limiting their use in practice. In this article we introduce a modification of the hybrid approach of Capanu and Begg [1] as a bridge between the likelihood-based and Bayesian approaches by employing Bayesian estimation for the variance components followed by Laplacian estimation for the regression coefficients. We investigate its performance as well as that of several likelihood-based methods in the setting of GLMMs with binary outcomes. We apply the methods to three datasets and conduct simulations to illustrate their properties. Simulation results indicate that for moderate to large numbers of observations per random effect, adaptive Gaussian quadrature and the Laplacian approximation are very accurate, with adaptive Gaussian quadrature preferable as the number of observations per random effect increases. The hybrid approach is overall similar to the Laplace method, and it can be superior for data with very sparse random effects.

Keywords

PQL; generalized linear mixed models; pseudo-likelihood; Bayesian; Laplace; binary clustered data

1. Introduction

Generalized linear mixed models (GLMM) have become increasingly popular for the analysis of non-normal data with random effects commonly encountered in medical research and across many disciplines. This popularity has generated a lot of methodological research in the past two decades. Unlike linear mixed models for which the likelihood function can be expressed in closed form, thus making (restricted) maximum likelihood estimates readily available, the GLMM likelihood function is expressed as an integral with respect to the random effects and does not have a closed form. As a result, numerous approximation methods have been proposed with different degrees of accuracy, complexity of implementation, and computational time. Breslow and Clayton [2] used Laplace

[*]Correspondence to: Memorial Sloan-Kettering Cancer Center, 307 E 63rd St, 3rd Floor, New York, NY 10021, capanum@mskcc.org.

approximations and Wolfinger and O'Connell [3] used Taylor expansions to reduce the estimation of GLMM to that of an approximated linear mixed model which can be analyzed in the usual way. Though arrived at through difference expansions, the penalized quasi-likelihood (PQL) introduced by Breslow and Clayton [2] and the pseudo-likelihood of Wolfinger and O'Connell [3] are equivalent and have been shown to be biased especially in the case of clustered binary data with few observations per cluster. The bias in the PQL estimates has led to the development of a series of modifications and proposals: correction of PQL [4, 5], modified Laplace approximation [6, 7], adaptive Gaussian quadrature [8], higher order Laplace approximations [9]. The list is long and thorough reviews of these developments can be found in McCullagh and Searle [10], Demidenko [11], Hedeker and Gibbons [12], and Lee et al. [13].

To tackle the analytical intractability of GLMMs, an alternative to likelihood-based approximations is to pursue a Bayesian approach in which Markov chain Monte Carlo methods are used to make inferences based on the posterior distribution of the parameters. Zeger and Karim [14] investigated the use of Gibbs sampling to fit GLMMs. Other implementations that involved Monte Carlo methods were adopted by Gamerman [15], Booth and Hobert [16], and Natarajan and Kass [17], among others. The use of these methods in practice has been hindered by their longer computational time and the lack of implementation in comprehensive statistical packages.

Recent improvements in computing technology have facilitated further developments in numerical integration methods to estimate GLMMs. Pinheiro and Chao [18] built on the methods introduced in Pinheiro and Bates [8] and proposed efficient Laplacian and adaptive Gaussian quadrature algorithms that reduce computational complexity and memory usage for approximating multilevel GLMMs. Chan et al. [19] combined Bayesian and classical approaches by using Gibbs sampling to evaluate the marginal likelihood and then obtain maximum likelihood estimates through the Newton-Raphson method. Ng et al. [20] evaluated the use of simulated maximum likelihood in estimating GLMMs and showed that it produces results similar to those from quadrature methods, although it can be more computationally intensive. Tsai and Hsiao [21] employed a Bayesian approach with approximate Jeffreys priors for the covariance matrix of the random effects and showed that the Bayesian approach outperforms PQL for the applications investigated. However, none of these approaches have readily available user code integrated in conventional statistical packages.

In a recent paper, Rue and Martino [22] proposed using integrated nested Laplace approximations (INLA) to approximate posterior marginals for latent Gaussian models. This approach is a promising alternative to reduce the computational burden of Markov chain Monte Carlo algorithms and has been received with enthusiasm as a way of stimulating greater use of Bayesian analysis in practice. Models that require a large number of hyperparameters are a concern with this approach, and further research may be needed to tackle this challenge. Fong et al. [23] evaluated the use of INLA for GLMMs and concluded that the method is accurate in general but less accurate for binomial data with small denominators. Implementation of INLA is available using an R interface.

Though numerous methods have been proposed, in practice the applied statistician routinely faces the frustrating prospect of widely disparate results produced by the methods that are currently implemented in commercially available software. This is the case in the analysis of our dataset of head and neck cancer patients in which, as seen later in Section 4.1, depending on the method used, the variance component estimates can differ by a factor of 9 and odds ratio estimates by 50%. This article is motivated by this frustration and develops guidance as well as a new method that is computationally efficient and statistically reliable.

In Capanu and Begg [1] we introduced a hybrid pseudo-likelihood approach with Bayesian estimation of the variance component in settings where the data are very sparse for individual random effects. The motivating context was a two-stage hierarchical model that was designed for estimating the relative risks of rare genetic variants. Using simulations we have shown that this approach has superior properties than two competitor techniques investigated: pseudo-likelihood estimation and pseudo-likelihood estimation with Breslow and Lin correction for the variance component. Building upon these findings, in this paper we propose a modification of this hybrid approach that combines Bayesian estimation of the variance components with Laplace estimation of the regression coefficients, and examine it under the more general setting of generalized linear mixed models with binary outcomes rather than in the specialized context of hierarchical models for rare variants as studied in Capanu and Begg [1]. Combining the simplicity of implementation of the Laplace approximation with the desirable properties of the Bayesian approach, the hybrid approach enjoys good properties in the settings examined and is relatively fast and easily accomplished in the standard statistical package SAS.

Section 2 describes the motivating study in which radiologic methods were used to assess recurrence in head and neck squamous cell carcinoma patients. The crossed random effects analysis of the salamander mating experiment [24] is revisited. A brief description of the Guatemalan survey dataset [25] later used for simulations, is also included. Section 3 describes the methods investigated, while Section 4 applies the methods to the head and neck and the salamander datasets. Section 5 presents simulations based on the head and neck data structure as well as based on the salamander mating experiment. It also reports the results of simulations based on the Guatemalan survey dataset which involves a multilevel structure with a large number of random effects. Section 6 concludes with a Discussion.

2. Datasets

In this section we describe a motivating study which involved head and neck squamous cell carcinoma (HNSCC) patients. The analysis involves binary outcomes with a large number of random effects, a setting which has proven difficult to analyze since different proposed procedures can yield different estimates and it is unclear which one is the most appropriate to use. We also revisit the widely analyzed salamander mating experiment which involves binary clustered data with crossed random effects and has been shown in numerous publications to pose challenges to existing estimation methods for GLMMs. The analyses of these datasets will be presented in Section 4. This section also describes the Guatemalan survey dataset. Simulations based on this dataset will be presented in Section 5.3.

2.1. Head and neck cancer data

Head and neck squamous cell carcinoma is the eighth most common malignancy in the US. Surgical resection remains the best treatment option at diagnosis, but approximately 70 per cent of the patients recur within 5 years. For various clinical and biological reasons, detecting recurrent disease in a timely manner is a challenge. Late detection of recurrent HNSCC limits the treatment options available for recurrent disease and confers a poor prognosis.

The clinical study that motivated our work focused on the accuracy and prognostic value of imaging parameters as well as factors associated with relapse free and overall survival in patients with HNSCC treated at Memorial Sloan-Kettering Cancer Center between 1996 and 2001 [26]. The study investigated multiple questions, but here we focus on the prognostic value of several risk factors as independent predictors of recurrent disease at 6 months. Patients are typically followed for recurrence by a radiological examination every 3-6 months. A radiological exam is usually recorded as negative (no recurrent disease), positive

(recurrent disease), or equivocal. Since recurrent lesions can appear anywhere in the body, depending on the clinical history of the patient, there can be multiple outcomes from the radiological exam for each patient. Reflecting the corresponding treatment options, observations within a patient were classified into three sites: local (at or near site of the primary tumor), regional (lymph nodes near the site of the primary tumor) and distant (other organs). Radiologic data was available for 128 patients. Out of the 128 patients, 106 had a single record, while 22 had measurements from multiple sites for a total number of 157 observations. All equivocal results were treated as positive for this analysis. Potential predictors considered were clinical variables such as age, previous head and neck surgery, and previous exposure to chemotherapy, as well as results of the clinical examination including palpation of the neck area and interpretation of various blood tests.

This analysis can be handled with a binary regression model that allows for clustering since some patients can experience recurrence in different locations and thus an individual patient can contribute multiple outcomes to the analysis. These models are part of the family of GLMMs. In Section 4.1 we give a precise description of the model and the multiple methods we used for fitting. Without getting into details, we want to point out here the widely disparate results we obtained from different model-fitting procedures (Table 1). Estimates of the variance component can differ by nine-fold from one method to the other, and the odds ratio for one of the fixed effects shows 50% difference across methods.

We felt uncomfortable facing these widely disparate results and we were unable to find any guidance in the literature as to which methods give more reliable results. This article is in part motivated by this experience. On the basis of our simulations (Section 5) we provide explicit recommendations.

2.2. Salamander Experiment

The salamander dataset has been previously described and analyzed at length by McCullagh and Nelder [24], Breslow and Clayton [2], Drum and McCullagh [27], Karim and Zeger [28], Lin and Breslow [5], and others and we describe it only briefly here. Ten males and ten females from each of two populations of Appalachian salamanders with different habitats, Rough Butt (RB) and White Side (WS), were paired in a crossed design. Each salamander was paired to three salamanders of the opposite sex from each population and the occurrence of mating was recorded as a binary response. The experiment was conducted three times, the first two times using the same 40 salamanders (Summer and Fall 1), while a new set of 40 animals was utilized the third time (Fall 2). As seen later in Section 4.2, analysis of these data suffers from the same limitations encountered for the head and neck data with different methods leading to widely different results.

2.3. Guatemalan Survey Dataset

The 1987 Guatemalan National Survey of Maternal and Child Health [29] was conducted with the goal of better understanding the determinants of use of modern prenatal care (physician or trained nurse) during pregnancy. The study was based on a multistage sample of women aged 15-44 years living in clusters of communities and had a three-level structure: births within mothers within communities. Rodriguez and Goldman [25] studied a subsample of respondents consisting of 2449 births that pertained to 1558 mothers living in 161 communities. Among the 1558 mothers, 52.4% had one child, 38.2% had two children, 9.1% had three children, and 0.3% had four children. The number of children per community ranged from 1 to 50 with a mean of 15 children, and the maximum number of families per community was 26 with an average of 10 families per community. Simulations based on the complex multilevel structure of this dataset are presented in Section 5.3.

3. Methods

Assume that conditional on the random effects δ, the response vector Y is distributed as independent random variables that follow an exponential family with density $p(\cdot|\delta, a, G)$. Then a GLMM can be represented succinctly as follows:

$$E\left[Y|\delta\right] = g^{-1}\left(U\alpha + X\delta\right) = g^{-1}\left(\eta\right) = \mu, \quad \text{such that} \quad var\left[\delta\right] = G \quad var\left[Y|\delta\right] = R_\mu^{1/2} R R_\mu^{1/2}, \quad (1)$$

where U is a matrix of predictors for the fixed effects, X is a matrix of predictors for the random effects, a is a vector of regression coefficients for the fixed effects, while δ is a vector of regression coefficients for the random effects with δ assumed to be normally distributed with mean 0 and variance matrix G. In addition, $g(.)$ is a differentiable monotonic link function (for example the logit link for mixed logistic regression), R_μ is a diagonal matrix containing the variance functions of the model evaluated at μ, and R is unknown. In the settings investigated here, we assume G is a $q \times q$ diagonal matrix with the diagonal elements called random effects variance components, τ_i, while the residual variance component is the common diagonal element of the diagonal matrix R. When we have longitudinal or clustered data with a binary outcome, the logistic regression mixed model is given by: n by:

$$Y_{ij}|u_{ij}, x_{ij}, \delta_i \overset{i.d}{\sim} Bernoulli\left(\mu_{ij}\right), \quad logit\left(\mu_{ij}\right) = u_{ij}^T\alpha + x_{ij}^T\delta_i,$$

with $\delta_i \sim N(0, G)$ for subject (cluster) $i = 1, \cdots, q$, and $j = 1, \cdots, n_i$.

If $p(\cdot)$ is used as a generic term for the probability density function, then the marginal distribution of the data in a mixed model can be expressed as

$$p\left(y|\alpha, G\right) = \int p\left(y|\delta, \alpha, G\right) p\left(\delta\right) d\delta. \quad (2)$$

In general, this integral does not have a closed-form expression for GLMMs, and different methods have been proposed to estimate it. Note the dimension of this integral, q, plays a critical role in the accuracy of the numerical method employed to compute this integral. In this article, we will investigate the performance of several approaches for estimating GLMMs: PQL, Laplacian approximation, adaptive Gaussian quadrature, and our hybrid Bayesian Laplacian approach. The next section provides a brief review of these methods.

3.1. The Pseudo-Likelihood Method

The pseudo-likelihood method estimates the model parameters of GLLMs by using a linearization technique which employs Taylor expansions iteratively to approximate the initial generalized linear mixed model with a linear mixed model [3]. Fitting the resulting linear mixed model is itself an iterative process which upon convergence leads to new parameter estimates that are then used to update the linearization. This results in a new linear mixed model, and the process continues until the relative change between parameter estimates from successive cycles is sufficiently small. The predictors $\widehat{\delta}$ are the estimated best linear unbiased predictors (BLUPs), referred to as EBLUP in the approximated linear mixed model. More details of this algorithm and additional formulas are provided in Wolfinger and O'Connell [3], SAS Institute Inc. [30] and Schabenberger [31] and a description of its application to fit hierarchical models can be found in Capanu et al. [32].

As further detailed in Section 3.2.1, the penalized quasi-likelihood method (PQL) of Breslow and Clayton [2] estimates generalized linear mixed models using the Laplace

approximation via ignoring one term of the approximation. Though arrived at through different expansions, the PQL and the pseudo-likelihood methods produce identical parameter estimates since the objective functions minimized by the two methods differ only by a constant [3]. Note that under both PQL and the pseudo-likelihood estimation, you can choose to freely estimate the dispersion parameter with the rest of the parameters rather than assuming no dispersion (i.e. the dispersion parameter is 1) the way the binomial distribution does.

It is well known that the PQL estimates of the variance components are subject to bias especially for certain cases such as clustered binary data with clusters of small size. In Section 3.3 we describe a modification of the pseudo-likelihood method designed to improve estimation of the variance components and of the regression coefficients.

3.2. Integral Approximations

3.2.1. Laplace Approximation—There have been several Laplacian approximations proposed for estimating marginal posterior densities [33] or nonlinear mixed models [2, 34, 8]. The idea behind the Laplace approximation is to approximate an integral by expanding the logarithm of the integrand in Taylor series and then using the Gaussian distribution to evaluate the integral obtained after the expansion. In the context of GLMMs, following SAS Institute Inc. [30], the marginal distribution of the data can be written as

$$p(y|\alpha,\theta,\phi) = \int p(y|\delta,\alpha,\phi)\, p(\delta|\theta)\, d\delta = \int \exp\{ch(y,\alpha,\theta,\phi;\delta)\} \quad d\delta \propto \quad |G|^{-1/2} \int e^{h(\delta)} d\delta$$

where θ is the vector of the covariance parameters contained in the matrix G, ϕ is a possible scale parameter, and c is a constant. Substituting the conditional density functions in this expression and following the derivation in Breslow and Clayton [2], we obtain

$$p(y|\alpha,\theta,\phi) \approx |G|^{-1/2} - |h''(\widehat{\delta})|^{-1/2}$$

where h'' is the second derivative matrix

$$h''\left(y,\alpha,\theta,\phi;\widehat{\delta}\right) = \frac{\partial^2 h(y,\alpha,\theta,\phi;\delta)}{\partial \delta^2}\Big|_{\widehat{\delta}},$$

while $\widehat{\delta}$ satisfies

$$\frac{\partial h\left(y,\alpha,\theta,\phi;\widehat{\delta}\right)}{\partial \delta} = 0.$$

Noting $h''(\delta) \approx X^T R_\mu X + G^{-1}$, which follows from ignoring a remainder term with zero expectation, we arrive at

$$p(y,\alpha,\theta,\phi) = |G|^{-1/2} + |X^T R_\mu X + G^{-1}|^{-1/2} - h\left(\widehat{\delta}\right) \quad (3)$$

This is the form of the Laplace approximation for GLMMs that we use in this article. PQL follows from (3) via ignoring the middle term $|X^T R_\mu X + G^{-1}|^{-1/2}$, which amounts to

assuming that R_μ varies very slowly as a function of μ. As a result of this additional approximating step, the PQL has less precision than the Laplace approximation.

Wolfinger [34] assumed a flat prior for α and expanded the integrand around α and δ, using the optimization only for the covariance parameters. Then, for given variance components $\widehat{\theta}$, the fixed effects and random effects α and δ were determined in a suboptimization that involved a linear mixed model with pseudo-data. In contrast, the Laplacian approximation implemented in PROC GLIMMIX (METHOD=LAPLACE) in SAS involves a suboptimization which for given values of $\widehat{\alpha}$ and $\widehat{\theta}$ determines the random effects vector $\widehat{\delta}$ that maximizes $h(y, \alpha, \theta, \varphi, \delta)$. Another difference between the implementation of the Laplacian approximation in the GLIMMIX procedure and the approximations proposed by Wolfinger [34] and Pinheiro and Bates [8] is that the latter use an approximation of the

second derivative $h'' \left(y, \alpha, \theta, \phi; \widehat{\delta} \right)$ whereas the SAS PROC GLIMMIX computes this analytically. We have employed the SAS PROC GLIMMIX to carry out the Laplacian approximation in our simulations and data analyses.

3.2.2. Adaptive Gaussian Quadrature—A quadrature method approximates a given integral by a weighted sum over predefined abscissas for the random effects. A good approximation can usually be obtained with an adequate number of quadrature points (nodes) as well as appropriate centering and scaling of the abscissas. Adaptive Gaussian quadrature as described by Pinheiro and Bates [8] centers and scales the quadrature points by using the empirical Bayes estimates (EBE) of the random effects and the Hessian matrix from the EBE suboptimization. This centering and scaling improves the likelihood approximation by placing the abscissas according to the density function of the random effects. Furthermore, the number of quadrature points can be adaptively selected by evaluating the log-likelihood function at the starting values of the parameters at a successively larger number of nodes until a tolerance is met [see 8, 30, 35, for more technical details]. The approximation to the log-likelihood can be improved by increasing the precision of the numerical integration, and thus adaptive Gaussian quadrature methods are expected to perform better than linearization-based alternatives [8, 36].

Adaptive Gaussian quadrature methods are currently implemented in SAS PROC GLIMMIX with "METHOD=QUAD" option and PROC NLMIXED. Note that in our analyses and simulations we have allowed the number of quadrature points to be selected adaptively (the default method in both PROC GLIMMIX and PROC NLMIXED). One can choose to specify a fixed number of quadrature points by using the "QPOINTS" option. However there are restrictions to which models these procedures can handle. The class of models that can be currently estimated by adaptive Gaussian quadrature in PROC GLIMMIX is considerably smaller than that handled with the Laplacian approximation described in Section 3.2.1 (PROC GLIMMIX METHOD=LAPLACE). For instance, crossed random effects models (such as that in the salamander data analysis) or models with non-nested subjects can not be currently fit using the quadrature approach (one needs to be able to specify the SUBJECT option in the RANDOM statement in order to use PROC GLIMMIX METHOD=QUAD). One limitation of PROC NLMIXED is that it only allows models with a single variance component and thus models with multiple variance components (such as that in the salamander data) or multilevel nonlinear mixed models (such as that in the Guatemalan survey described in Section 5.3) are not accommodated. Moreover, as the number of random effects increases, the quadrature approach becomes computationally infeasible due to the high dimensionality of the integral and this further limits the use of the NLMIXED or GLIMMIX procedures to fit adaptive Gaussian quadrature methods. Pinheiro and Chao [18] proposed efficient adaptive Gaussian

quadrature algorithms for approximating multilevel models. However their implementation is not available with standard statistical packages as of yet.

3.3. Hybrid Bayesian Estimation

In earlier work, we investigated a hybrid approach that combines the desirable properties of the Markov chain Monte Carlo methods with the simplicity (in terms of speed and implementation) of the pseudo-likelihood method [1]. This involved a three-step procedure with Bayesian estimation of the variance components of the random effects followed by pseudo-likelihood estimation of the fixed and random effects using the Bayesian variance component estimators. In this article, we have explored a modified version of this approach that replaces the pseudo-likelihood estimation in the third step with a Laplacian approximation to obtain the parameter estimates for the fixed and random effects once the Bayesian estimates of the variance components have been obtained. Specifically:

1. Step 1: Apply the pseudo-likelihood method to obtain an approximated linear mixed model, as described in Section 3.1.

2. Step 2: Use a Markov chain Monte Carlo algorithm to estimate the variance components of the approximated linear mixed model obtained in Step 1.

3. Step 3: Re-estimate the regression parameters using the Laplace method for the original GLMM with the variance components pre-specified at the Bayesian estimates obtained in Step 2.

Step 1 is carried out using the SAS macro %glimmix. To achieve the Bayesian estimation in step 2, we generate a posterior sample from the marginal posterior densities of the variance components using a random walk Metropolis-Hastings algorithm with a non-informative Inverse-Gamma prior distribution with shape and scale parameters both set at 0.01, and use the posterior mean as the estimator. This step is carried out using SAS PROC MIXED with the "PRIOR" option. Finally, step 3 is achieved using SAS PROC GLIMMIX with the option "METHOD=LAPLACE". A more detailed algorithm describing our method is provided as Supplementary Material.

4. Applications

We apply the methods outlined in Section 3 to the head and neck cancer data and the salamander mating experiment.

4.1. Head and Neck Cancer Data

We fit the following model

$$\text{logit} \quad Pr\left(y_{ij}=1|b_i\right)=u_{ij}^T\alpha+b_i, \quad (4)$$

where y_{ij} is the presence or absence of recurrence as assessed by the radiology imaging for the ith patient at the jth site ($i = 1, \cdots, 128, j = 1, \cdots, 3$), and $b_i \sim N(0, \sigma^2)$ are the independent random effects corresponding to the patients. The fixed effects covariates u_{ij} included an intercept, age, suspicious findings in clinical exam (lesion specific), previous head and neck surgery, and previous exposure to chemotherapy.

We fitted this model using currently available techniques: PQL, Laplace approximation, as well as the adaptive Gaussian quadrature as implemented by PROC NLMIXED and PROC GLIMMIX in SAS using a ridge-stabilized Newton-Raphson algorithm as the optimization technique (TECH=NRIDG). We also employed the Gibbs sampling algorithm [14] with diffuse priors for the hyperparameters (an Inverse Gamma distribution with both shape and

scale parameters set at 0.001 for the variance component and normal distributions with mean 0 and variance 1000 for the other hyperparameters). This was implemented using the WinBUGS software based on 20000 burn-ins followed by 20000 iterations. We further carried out the analysis using the hybrid Bayesian with Laplacian approximation (denoted "Hybrid BL").

Table 1 presents estimates for the fixed effects as well as for the variance component from the methods investigated. Methods yield widely different estimates for the patient to patient variability, with the smallest and largest estimates differing by a factor of 9. In fact, based on estimates from PQL with estimated dispersion parameter ($P\ QL_2$ in the table) and Laplace approximation one could even question the need of a random effects model. Though not as disparate as the estimates for the variance component, there are still substantial differences in the estimates for fixed effects and their standard errors among the different methods. For example, depending on which method is used, the odds ratio for surgery is either 0.38 (=exp(−0.98)) or 0.55 (=exp(−0.6)). As seen at the bottom of Table 1, the hybrid BL produces results that are close to adaptive quadrature and Gibbs sampling.

To summarize the results, if one is to make conclusions based only on significance of the factors, the different analyses roughly agree that a positive finding on the clinical examination increases the risk of recurrence (although if $P\ QL_2$ was used the p-value does not reach the commonly used 5% threshold) and that previous surgery and chemotherapy are not significantly associated with recurrence. Some methods suggest that older age may also contribute to an elevated risk, although not statistically significant by most of the methods. However, an examination of point estimates and standard errors reveal substantial disagreement between the methods: variance component estimate varies nine-fold and fixed-effects odds ratios up to 50%. There is a clear need for guidance on which of these methods is most appropriate.

4.2. Salamander Experiment

Following Lin and Breslow [5], consider a model that pools the data across the three experiments [called Model A in 2]:

$$\text{logit } Pr\left(y_{ij}=1 \mid b_i^f, b_j^m\right) = u_{ij}^T \alpha + b_i^f + b_j^m, \quad (5)$$

where y_{ij} is the mating occurrence for the ith female with the jth male ($i, j = 1, \cdots, 60$), and $b_i^f \sim N\left(0, \sigma_f^2\right)$ and $b_j^m \sim N\left(0, \sigma_m^2\right)$ are the female random effects and male random effects assumed independent of each other. Note that the model assumes that different animals had been used in each experiment (as it is atypical to use the same exact animals in two separate experiments) [27, 5]. The fixed effects covariates u_{ij} contain an intercept, an indicator WS_f for whether the female was WS or RB, an indicator WS_m for whether the male was WS or RB, and a term for their interaction. We also fitted a model with separate fixed and random effects for each of the three experiments [called Model C in 2].

Lin and Breslow [5] applied their proposed first- and second-order PQL correction to estimate this model and compared the results to those previously obtained by various other methods applied to the salamander data: the method of moments [24], Gibbs sampler [28], PQL [2], and REML [27]. They noted unsatisfactory performance of the PQL correction which they conjectured was likely due to large variability in the random effects. After conducting simulations, they concluded that REML had superior properties with almost unbiased estimates, though its applicability is limited as it requires a balanced design with marginal mean expressed as a linear function of the parameters of interest. For those who

view the problem from a Bayesian standpoint, the Gibbs sampler would be the standard Bayesian approach [14].

We analyzed the salamander dataset using the hybrid BL approach. We present results for the variance components separately for each experiment and from the pooled data in Table 2, while results for the fixed effects based on the pooled data analysis are reported in Supplementary Table 1. For comparison purposes, we reproduce the results reported in Table 1 and Table 2 of Lin and Breslow [5] for several of the methods they investigated: first-order and second order PQL correction ("CPQL"), PQL, PQL with the Breslow and Lin correction to the variance components (denoted "$\mathrm{PQL}\left(\hat{\sigma}^2_{CP}\right)$"), Laplace approximation, REML and Gibbs sampler. We also report the results from Shun [7] obtained using his two proposed corrections: using one correction term in the Laplace expansion (denoted "Shun Corrected (1)"), and using an exponentiated correction term (denoted "Shun Corrected (2)"). Results for the integrated nested Laplace approximation ("INLA") presented by Fong et al. [23] for Model C are also included for the variance components estimates by experiment in Table 2. The pooled data analysis with INLA was not reported by Fong et al. [23] and thus the results were not available for presentation here. Chan et al. [19] and Tsai and Hsiao [21] also performed Bayesian analysis on the salamander data but the former conducted the analysis on the summer experiment only while the latter assumed correlated random effects and these results are not included here.

As seen in Table 2, the estimates for the variance components are quite different among the various methods investigated, pointing to a clear need for simulation studies to distinguish the properties of the different methods. In terms of fixed effects, among the methods considered, the hybrid approach and the Laplace approximation on its own produce estimates that are similar to each other and are close to the REML and Gibbs sampling estimates (Supplementary Table 1).

5. Simulations and Data Analysis

We evaluate the properties of the methods described in Section 3 with simulations based on three examples: the head and neck data which has a very sparse and unbalanced structure, the widely-analyzed salamander experiment based on a crossed design, and the Guatemalan Child Health study which involves a multilevel model with large number of random effects which can be overwhelming to fit for some methods and/or statistical packages. Using simulations that follow the structure of these datasets, we study the properties of the following candidate methods: Laplace, adaptive Gaussian quadrature, and the hybrid approach. As a benchmark, PQL is also included. We have also examined the properties of the PQL implementation that freely estimates the dispersion parameter, rather than fixing it at 1, as well as the properties of the hybrid Bayesian pseudo-likelihood approach. However neither of these demonstrated any advantage relative to the other methods considered and for simplicity we did not include these results in our presentation. Note that for the salamander and Guatemalan datasets we could not evaluate with simulations the adaptive Gaussian quadrature method implemented by SAS: the NLMIXED procedure is limited to analyses involving a single variance component and both datasets require two variance components, while the GLIMMIX procedure can not handle crossed random effects (salamander simulations) and was computationally infeasible in the Guatemalan survey.

5.1. Head and Neck Data Simulations

As seen in Section 4.1, the analysis of the head and neck data yielded widely-different results by the different methods employed. We conducted simulations to investigate the finite-sample properties of the methods and generated data according to model (4) with 128

random effects (note that the dimension of the integral in (2) is $q = 128$). We studied the effect of the unbalanced structure and of the cluster size by looking at two scenarios: 1) balanced structure: each of the 128 patients had exactly 3, 5, or 10 measurements per patient; 2) unbalanced structure: the exact structure of the original head and neck data was used (83% of patients had a single measurement); to make the data more dense we doubled the data (i.e. 83% of patients now contribute two measurements), and also multiplied it by 5 times, each time resimulating the outcomes. The true $a^T = (-0.69, 1.1)$ corresponding to the intercept and to a Bernoulli covariate with success probability of 0.4 (mimicking the clinical exam predictor), were set to be the estimates obtained with the adaptive Gaussian quadrature method for the actual data. We varied σ^2 to be 0.16, 0.36, and 0.65, the latter one being the adaptive Gaussian quadrature estimate of the variance component of the actual data. We conducted 1000 simulations based on these scenarios and compared the methods considered in terms of the mean values, mean squared error, and coverage probabilities of the 95% confidence intervals.

5.1.1. Balanced Head and Neck Data Simulations—Table 3 presents the mean estimates for the fixed effects and variance component for different magnitudes of dispersion and increasing cluster sizes. As expected PQL has most bias compared to the other methods considered. In terms of the fixed effects, the hybrid BL approach has little bias, similar to the Laplace and the adaptive quadrature methods. The bias improves with larger numbers of observations within a random effect, with the adaptive quadrature methods having negligible bias both for fixed effects and the variance component for the denser data such as with 5 or more observations per patient. For the more sparse scenario of 3 records per patient, the Laplace and the hybrid BL produce the least bias for the variance component for smaller degrees of dispersion ($\sigma^2 = 0.16$) but lead to a somewhat underestimated variance component for the larger values of σ^2. This improves as the cluster size increases. All methods have coverage rates close to the nominal level of 95% (see Table 4) and comparable mean squared errors (Supplementary Table 4).

To study how the number of random effects impacts the performance of the methods, we simulated balanced data from model (4) as described above but now assuming 25 and 50 random effects ($q = 25$ and 50, respectively), instead of 128. These results are reported in Supplementary Table 2 and Supplementary Table 3 respectively. With smaller number of random effects, the performance is worse and depends on the magnitude of the σ^2 and the number of observations within the random effect: for very sparse data per random effect (3 observations) in certain situations one method does better than the others, but there is no one method that overall performs best in all of these configurations. The performance improves for larger numbers of observations per random effect with the adaptive quadrature methods exhibiting negligible bias. Small average absolute deviations from the AGQ estimates indicate that the Laplace and the hybrid BL methods overall yield estimates close to the adaptive quadrature estimates (data not shown).

5.1.2. Unbalanced Head and Neck Data Simulations—The results from the simulations based on the very sparse configuration of the original head and neck data are included in Table 5 along with the results for the denser configurations obtained by multiplying the original data by 2 or 5 times. For the sparse structure of the original head and neck data all methods have poorer performance, in particular for the variance component for which the estimates can be severely biased, with the hybrid Bayesian approach producing more stable variance component estimates compared to the other methods. The methods become more stable and the bias drastically improves as the sparseness becomes attenuated.

5.2. Salamander Experiment Simulations

We reproduce the simulation study conducted by Lin and Breslow [5] in which 360 binary observations were generated according to model (5). The true $a^T = (1.06, -3.05, -0.72, 3.77)$ corresponding to the intercept, WS_f, WS_m, and $WS_f \times WS_M$, respectively, were set to be the REML estimates fitted to the actual data. Lin and Breslow [5] studied two configurations for the variance components $\left(\sigma_f^2, \sigma_m^2\right) = (1.67, 1.5)$ corresponding to the actual REML variance component estimates (simulation results not shown in Lin and Breslow [5]) and $\left(\sigma_f^2, \sigma_m^2\right) = (0.5, 0.5)$, representing a smaller amount of dispersion. In addition to these configurations, we considered an intermediate level of dispersion $\left(\sigma_f^2, \sigma_m^2\right) = (1, 1)$. We carried out 1000 simulations corresponding to these configurations and evaluated the performance of the different methods in terms of mean values, mean squared error, and coverage probabilities of the 95% confidence intervals obtained with the different methods investigated. In their simulation study of the salamander data, Lin and Breslow [5] found that their proposed correction to PQL [4, 5] performed better than PQL. We have also evaluated the performance of the Breslow and Lin correction. However, besides the improvements over PQL it did not demonstrate superiority over the other methods considered and for simplicity of presentation we do not present this method in the tables.

Table 6 presents the mean values of the different parameter estimates for the different configurations of variance components studied. The PQL method underestimates substantially the variance components regardless of the degree of dispersion studied, with worst bias for the larger variance components scenarios. The regression coefficients are also biased, with bias also increasing with increasing variability in the random effects. This is not surprising in light of the results obtained from PQL with variance components prespecified at the true value, which show that even when the variance components are correctly estimated, the PQL regression coefficients are still biased (data not shown). The same behavior was observed in the numerical studies of PQL conducted by Jang and Lim [37] in which they show that for binary outcomes, the PQL regression coefficients are biased even when the variance components are precisely estimated. In contrast, the Laplacian approximation and the hybrid approach have little bias for the fixed effects coefficients, regardless of the magnitude of the variance components and despite the bias in the variance components that both methods exhibit for the larger variance components. Note that the hybrid approach yields less bias in the variance components for smaller amounts of dispersion, while the Laplacian approximation has less bias in the variance components for larger variability in the random effects.

Coverage probabilities of 95% confidence intervals formed using asymptotic normality of the parameter estimates are reported in Supplementary Table 6. For all scenarios considered, the Laplacian approximation has coverage rates close to the nominal level of 95%. The hybrid approach also has good coverage probabilities but tends to be anti-conservative for a_1 and a_3 parameters. This anti-conservativeness for these two parameters is even more pronounced for the other methods and becomes worse for the scenarios with larger variance components. The Laplacian approximation and the hybrid approach have similar mean squared errors though it appears that the hybrid approach has lower mean squared errors for the variance components estimates (see Supplementary Table 5).

Overall, both the Laplacian approximation and the hybrid approach have better properties than the different variations or corrections of PQL investigated (estimates shown only for PQL), with the Laplacian approximation being somewhat superior in terms of coverage probabilities, but with the hybrid approach providing least biased variance components estimates for models with moderate variability in the random effects.

5.3. Guatemalan Survey Simulations

Rodriguez and Goldman [25] used the hierarchical structure of the 1987 Guatemalan Survey to simulate 100 datasets to investigate the performance of several approaches for GLMMs available at that time. The random effects logistic regression model that they studied was

$$
\begin{aligned}
(y_{ijk}|p_{ijk}) &\sim Bernoulli\,(p_{ijk}) \quad \text{for} \quad k \\
&=1,\ldots,n_{ij}, \text{independent} \quad \text{with} \quad logit\,(p_{ijk}) \\
&=\alpha_0+\alpha_1 u_{1i}+\alpha_2 u_{2ij}+\alpha_3 u_{3ijk}+b_i^{(1)}+b_{ij}^{(2)},
\end{aligned}
$$

where y_{ijk} is a binary indicator with value 1 if modern prenatal care was received for the kth child within the jth family within the ith community, and n_{ij} is the number of children for the jth family within the ith community. Here $b_i^{(1)} \sim N\left(0,\sigma_1^2\right)$ and $b_{ij}^{(2)} \sim N\left(0,\sigma_2^2\right)$ are the independent community random effects and the independent family random effects, assumed independent of each other. The community, family, and child covariates, u_{1i}, u_{2ij}, and u_{3ijk}, respectively, are composite scales created by Rodriguez and Goldman [25] to simplify the original model used by Pebley-Goldman which contained many covariates at each level. The true parameter values used in the simulations were $\alpha_0 = 0.665$ and $\alpha_1=\alpha_2=\alpha_3=\sigma_1^2=\sigma_2^2=1$. The values of the variance components were chosen to result in a fairly large variability in the responses.

Pinheiro and Chao [18] illustrated their adaptive Gaussian quadrature methods for multilevel GLMMs using the 100 simulated datasets created by Rodriguez and Goldman [25]. Browne and Draper [38] also used the Guatemalan survey study to compare the performance of marginal and penalized quasi-likelihood methods with that of a Bayesian approach involving Markov chain Monte Carlo estimation with adaptive hybrid Metropolis-Gibbs sampling. They chose improper Uniform priors for the regression coefficients and two prior distributions for the variance components, Inverse Gamma, and an improper Uniform prior on $(0, \infty)$. Their analysis is based on 500 simulated datasets which they generated themselves to replicate the same data structure as used by Rodriguez and Goldman [25].

We applied the hybrid approach for the 100 datasets simulated by Rodriguez and Goldman [25] and compared the results against those already reported by Pinheiro and Chao [18] and Browne and Draper [38]. In Table 7 we report mean estimates for the fixed effects and variance components, while the coverage of nominal 95% intervals for the fixed effects is reported in Supplementary Table 7.

PQL has biased regression coefficients and severe bias in the family variance component. Though better than PQL, Laplace on its own also exhibits bias in the regression coefficients and severe bias in the family variance component. In contrast, the hybrid approach has low bias for the fixed effects, improved estimation for the family variance component as compared to the PQL and Laplace, though still underestimating the community variance component. The adaptive Gaussian quadrature methods have almost unbiased regression coefficients and the variance component estimates are closer to the true values than for the other methods investigated. In terms of coverages, PQL and Laplace have poor coverage especially for the family and child fixed effects (α_2 and α_3) while the hybrid approach and the adaptive Gaussian quadrature have all coverage close to the nominal level (coverages were not reported by Pinheiro and Chao [18] but were mentioned in the text). Similar to adaptive Gaussian quadrature, the Bayesian implementations of Browne and Draper [38] had good performance in terms of bias and coverage (data not shown).

Though adaptive Gaussian quadrature and the MCMC methods lead to more accurate variance components estimates than the hybrid approach in this simulation, they require specialized software and code that is not available in standard statistical packages as of yet. In contrast, the hybrid approach is easy to implement with three calls to widely used SAS procedures and a macro (macro %glimmix, PROC MIXED and PROC GLIMMIX), and though it underestimates the variance components it has little bias and good coverage for the fixed effects, improving on the Laplacian approximation.

6. Discussion

The analysis of binary clustered data with small cluster sizes continues to pose challenges to the available GLMM methods, with many of the proposed methods producing biased regression coefficients and variance components. Adaptive Gaussian quadrature methods [8] have been shown to perform well for these scenarios but can be overwhelmed for datasets with large numbers of random effects. Moreover, their implementation in commercially available software is limited: for example, SAS PROC GLIMMIX can not currently accommodate crossed random effects or models with non-nested subjects. SAS NLMIXED procedure which also implements adaptive Gaussian quadrature methods has been shown to have good properties [36]. However it is restricted to analyses involving a single variance component. Another popular package, R, also implements adaptive Gaussian quadrature (see packages lme4, ZELIG, glmmML) but simulations [36] indicated that for binary data these R packages did not enjoy the same performance as that of their SAS counterpart PROC NLMIXED. Bayesian implementations involving Gibbs sampling also have good performance but are burdened by long computational times, they require a wise choice of priors, and the creation of specialized code for implementation. Furthermore, to the best of our knowledge, the performance of these methods has not been studied for the extremely sparse structures such as the one in the head and neck data in which the majority of the random effects contribute a single observation each. We have compared the performance of several likelihood-based methods (PQL, Laplace, and adaptive Gaussian quadrature) with that of a hybrid Bayesian Laplacian approach for difficult scenarios such as these. Our simulation results indeed indicate that for reasonably large numbers of observations per random effect the adaptive quadrature methods work extremely well when implementation and computational time are feasible. In these scenarios, both implementations in SAS (GLIMMIX or NLMIXED) yield very similar results and the choice between the two procedures can be made based on the user's familiarity with the corresponding syntax. PQL does not demonstrate superiority, in fact there are situations in which it is severely biased, as already known.

The Laplacian approximation has received a lot of attention as a fast computational method for estimation of GLMMs, and numerous corrections and variations of it have been proposed based on its perceived weaknesses. However, our simulations show surprisingly good results for the Laplacian approximation in most situations, reasonably close to the adaptive Gaussian quadrature, although not superior. This confirms the findings of Joe [39] who also evaluated the accuracy of the Laplace approximation and was also surprised with its performance for the scenarios investigated. It appears that the general poor perception of the Laplace approximation in the statistical community is unjustified. Based on our findings, the Laplacian approximation is a viable approach with good properties in many settings.

The hybrid Bayesian Laplacian approach also shows overall good performance, comparable to that of the Laplace method. However, for the more sparse scenarios such as that of the head and neck and the Guatemalan survey datasets, the hybrid approach outperforms the Laplacian approximation, particularly in terms of reducing bias for the variance components for which the Laplacian approximation can yield severely biased estimates. Thus, in the

presence of sparseness and especially if variance component estimation is important, the hybrid approach is preferred to the Laplace method or even to the adaptive Gaussian quadrature methods as implemented by SAS.

Zhang et al. [36] also conducted a simulation study which concluded that adaptive Gaussian quadrature as implemented in SAS NLMIXED in the single variance component setting performed well. Our findings complement theirs in three important aspects: 1) we examined the properties of the hybrid method and observed that it works better than the other methods investigated for datasets with sparse random effects; 2) we included in our evaluation the Laplacian approximation and found that it performed well in many settings; 3) we evaluated various more complex designs such as crossed effects and nested effects.

Based on these findings, if interest lies in accurate estimation of the fixed effects of a GLLM for binary clustered data adaptive Gaussian quadrature and the Laplacian approximation are good choices with excellent coverage probabilities and low bias. Adaptive quadrature provides even more accurate results for denser data. However for situations where computation and implementation are an issue, the Laplace method is a preferred alternative. If one is interested in precise estimation of the variance components, the hybrid approach would be preferred particularly for datasets with very sparse numbers of observations per random effect. Another advantage of the hybrid approach is that it always produces a positive estimate for the variance components, whereas the Laplacian approximation can lead to negative estimates, or to confidence intervals with negative values.

An attractive feature that both the hybrid approach and the Laplacian approximation possess is that they are easy to apply in SAS with minimal model specification while other methods can be more complicated to implement. They both run fast: for the salamander experiment for example, they converge and produce results within seconds (Laplace method: 0.65 seconds, hybrid BL: 20.95 seconds) while adaptive quadrature cannot be implemented by either SAS PROC GLIMMIX or NLMIXED due to the random effects structure. SAS code is available at mskcc.org/marinelacapanu.

Adaptive Gaussian quadrature or Bayesian Gibbs sampling are attractive tools for models with software readily available for use and for which their implementation is computationally feasible. However for more complex models that require specialized code or substantial computing resources, the Laplacian approximation is a worthy alternative that is fast and easy to implement with widely used standard procedures. For very sparse random effects, the hybrid approach can provide superior performance, without increased difficulty of implementation or computational time.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

[1]. Capanu M, Begg CB. Hierarchical modeling for estimating relative risks of rare genetic variants: Properties of the pseudo-likelihood method. Biometrics. 2011; 67(2):371–380. [PubMed: 20707869]

[2]. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. Journal of the American Statistical Association. 1993; 88(421):9–25.

[3]. Wolfinger R, O'Connell M. Generalized linear mixed models: a pseudo-likelihood approach. Journal of Statistical Computation and Simulation. 1993; 48:233–243.

[4]. Breslow NE, Lin X. Bias correction in generalised linear mixed models with a single component of dispersion. Biometrika. 1995; 82(1):81–91.

[5]. Lin X, Breslow NE. Bias correction in generalized linear mixed models with multiple components of dispersion. Journal of the American Statistical Association. 1996; 91(435):1007–1016.

[6]. Shun Z, McCullagh P. Laplace approximation of high-dimensional integrals. Journal of the Royal Statistical Society. Series B. Methodological. 1995; 57:749–760.

[7]. Shun Z. Another look at the salamander mating data: a modified Laplace approximation approach. Journal of the American Statistical Association. 1997; 92:341–349.

[8]. Pinheiro JC, Bates DM. Approximations to the log-likelihood function in the nonlinear mixed-effects model. Journal of Computational and Graphical Statistics. 1995; 4:12–35.

[9]. Raudenbush SW, Yang ML, Yosef M. Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. Journal of Computational and Graphical Statistics. 2000; 9:141–157.

[10]. McCullagh, P.; Searle, SR. Generalized, linear, and mixed models. Wiley; New York: 2001.

[11]. Demidenko, E. Mixed models: Theory and Applications. Wiley; New York: 2004.

[12]. Hedeker, D.; Gibbons, RD. Longitudinal data analysis. Wiley; New York: 2006.

[13]. Lee, Y.; Nelder, JA.; Pawitan, Y. Chapman and Hall; London: 2006. Generalized linear models with random effect: unified analysis with H-likelihood.

[14]. Zeger SL, Karim RM. Generalized linear models with random effects: A Gibbs sampling approach. Journal of the American Statistical Association. 1991; 86:79–86.

[15]. Gamerman D. Sampling from the posterior distribution in generalized linear mixed models. Statistics and Computing. 1997; 7:57–68.

[16]. Booth JG, Hobert JP. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. Journal of the Royal Statistical Society. Series B. Methodological. 1999; 61:265–285.

[17]. Natarajan R, Kass RE. Reference Bayesian methods for generalized linear mixed models. Journal of the American Statistical Association. 2000; 95:227–337.

[18]. Pinheiro JC, Chao EC. Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. Journal of Computational and Graphical Statistics. 2006; 15:58–81.

[19]. Chan JSK, Kuk AYC, Yam CHK. Monte Carlo approximation through Gibbs output in generalized linear mixed models. Journal of Multivariate Analysis. 2006; 94:300–312.

[20]. Ng ESW, Carpenter JR, Goldstein H, Rasbash J. Estimation in generalized linear mixed models with binary outcomes by simulated maximum likelihood. Statistical Modelling. 2006; 6:23–42.

[21]. Tsai MY, Hsiao CK. Computation of reference Bayesian inference for variance components in longitudinal studies. Computational Statistics. 2008; 23:587–604.

[22]. Rue H, Martino S. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society. Series B. Methodological. 2009; 71:319–392.

[23]. Fong Y, Rue H, Wakefield J. Sampling from the posterior distribution in generalized linear mixed models. Biostatistics. 2010; 11:397–412. [PubMed: 19966070]

[24]. McCullagh, P.; Nelder, JA. Generalized linear models. Chapman and Hall; London: 1989.

[25]. Rodriguez G, Goldman N. An assessment of estimation procedures for multilevel models with binary responses. Journal of the Royal Statistical Society. Series A. 1995; 158:73–89.

[26]. Wong RJ, Lin DT, Schöder H, Patel SG, Gönen M, Wolden S, Pfister DG, Shah JP, Larson SM, Kraus DH. Diagnostic and prognostic value of [18f]fluorodeoxyglucose positron emission tomography for recurrent head and neck squamous cell carcinoma. Journal of Clinical Oncology. 2002; 20:4199–4208. [PubMed: 12377963]

[27]. Drum M, McCullagh P. REML estimation with exact covariance in the logistic mixed model. Biometrics. 1993; 49:677–689.

[28]. Karim RM, Zeger SL. Generalized linear models with random effects: salamander mating revisited. Biometrics. 1992; 48:631–644. [PubMed: 1637985]

[29]. Pebley AR, Goldman N. Family, community, ethnic identity and the use of formal health care services in Guatemala. Working Paper. 1992

[30]. SAS Institute Inc.. CN. SAS/STAT 9.2 Users Guide. 2008. Chapter 38: The GLIMMIX procedure

[31]. Schabenberger, O.; SAS Institute Inc.. Introducing the GLIMMIX procedure for generalized linear mixed models. SUGI 30; Cary, NC: 2005.

[32]. Capanu M, Orlow I, Berwick M, Hummer AJ, Thomas DC, Begg CB. The use of hierarchical models for estimating relative risks of individual genetic variants: an application to a study of melanoma. Statistics in Medicine. 2008; 27:1973–1992. [PubMed: 18335566]

[33]. Tierney L, Kadane JB. Accurate approximations for posterior moments and densities. Journal of the American Statistical Association. 1986; 81:82–86.

[34]. Wolfinger R. Laplace's approximations for nonlinear mixed models. Biometrika. 1993; 80:791–795.

[35]. SAS Institute Inc.. CN. SAS/STAT 9.2 Users Guide. 2008. Chapter 61: The NLMIXED Procedure

[36]. Zhang X, Lu N, Feng C, Thurston SW, Xia Y, Zhu L, Tu XM. On fitting generalized linear mixed-effects models for binary responses using different statistical packages. Statistics in Medicine. 2011; 30:2562–2572.

[37]. Jang W, Lim J. A numerical study of PQL estimation biases in generalized linear mixed models under heterogeneity of random effects. Communications in Statistics-Simulation and Computation. 2009; 38:692–702.

[38]. Browne WJ, Draper D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Bayesian Analysis. 2006; 1:473–514.

[39]. Joe H. Accuracy of Laplace approximation for discrete response mixed models. Computational Statistics and Data Analysis. 2008; 52:5066–5074.

Table 1

Estimated fixed effects and variance component for the head and neck squamous cell carcinoma data.

Method		Intercept	Age	Clinical exam	Surgery	Chemo	Variance component[h]
PQL$_1$[a]	Estimate (s.e.)	−0.58 (0.88)	0.02 (0.01)	1.02 (0.45)	−0.60 (0.48)	0.36 (0.50)	0.377 (0.51)
	p-value	0.51	0.09	0.03	0.23	0.48	-
PQL$_2$[b]	Estimate (s.e.)	−0.81 (1.13)	0.04 (0.02)	0.94 (0.48)	−0.98 (0.55)	0.22 (0.58)	3.321 (0.93)
	p-value	0.47	0.04	0.06	0.09	0.71	-
Laplace[c]	Estimate (s.e.)	−0.62 (0.93)	0.02 (0.02)	1.08 (0.47)	−0.64 (0.51)	0.39 (0.52)	0.371 (0.76)
	p-value	0.51	0.12	0.03	0.22	0.46	-
AGQ NLMIXED[d]	Estimate (s.e.)	−0.69 (1.00)	0.03 (0.02)	1.11 (0.49)	−0.67 (0.54)	0.39 (0.54)	0.644 (1.13)
	p-value	0.49	0.12	0.03	0.22	0.47	-
AGQ GLIMMIX[e]	Estimate (s.e.)	−0.69 (1.00)	0.03 (0.02)	1.11 (0.49)	−0.67 (0.54)	0.39 (0.54)	0.646 (1.14)
	p-value	0.49	0.12	0.03	0.22	0.47	-
Bayesian IG prior[f]	Median (95%CI)	−0.73 (−4.06, 1.31)	0.03 (0,0.11)	1.20 (0.19,2.62)	−0.76 (−2.62, 0.34)	0.39 (−0.89, 1.75)	0.677 (0.001, 17.03)
Hybrid BL[g]	Estimate (s.e.)	−0.67 (0.95)	0.03 (0.01)	1.10 (0.48)	−0.66 (0.51)	0.39 (0.53)	0.558 (0.55)
	p-value	0.48	0.08	0.03	0.21	0.47	-

[a]"PQL" stands for penalized quasi-likelihood (PQL) with dispersion parameter fixed at 1.

[b]"PQL2" stands for PQL with dispersion parameter allowed to be estimated.

[c]"Laplace" stands for the Laplacian approximation as implemented by SAS GLIMMIX (METHOD=LAPLACE).

[d]"AGQ NLMIXED" stands for adaptive Gaussian quadrature as implemented by SAS NLMIXED. The number of quadrature points was adaptively estimated to be 5.

[e]"AGQ GLIMMIX" stands for adaptive Gaussian quadrature as implemented by SAS GLIMMIX (METHOD=QUAD). The number of quadrature points was adaptively estimated to be 9.

[f]Median posterior estimate along with 95% credible intervals.

[g]"Hybrid BL" stands for hybrid Bayesian Laplacian approach.

[h]p-values for testing the variance component are not available.

Table 2

Estimated variance components for the salamander dataset; " – " indicates data not available.

Method	Summer		Fall 1		Fall 2		Pooled	
	σ_f^2	σ_m^2	σ_f^2	σ_m^2	σ_f^2	σ_m^2	σ_f^2	σ_m^2
REML[a]	1.68	0.34	2.46	1.44	0.69	2.40	1.67	1.50
Gibbs[a]	2.35	0.14	2.99	1.42	0.33	2.89	1.50	1.36
PQL[a]	1.41	0.09	1.26	0.62	0.26	1.50	0.72	0.63
CPQL[a,b]	1.71	0.40	1.64	1.00	0.63	1.86	0.99	0.91
Shun Corrected(1)[c]	1.71	0.18	2.10	1.10	0.46	2.07	-	-
Shun Corrected(2)[c]	1.80	0.25	2.53	1.37	0.55	2.25	-	-
INLA[d]	1.66	0.61	1.91	1.00	0.64	2.13	-	-
Laplace[a]	1.58	0.07	1.81	0.92	0.35	1.85	1.17	1.04
Hybrid BL	2.09	0.35	1.90	0.91	0.54	2.04	1.13	1.07

[a] Results reported in Table 1 of Breslow and Clayton [2].

[b] Variance components estimated using the Breslow and Lin correction.

[c] Results reported in Table 3 of Shun [7].

[d] Results reported in Web Table 4 of Fong et al. [23].

Table 3

Mean values of parameter estimates from the balanced head and neck data simulations:128 random effects.

(a) σ^2 = 0.16

	3 observations			5 observations			10 observations		
Method[a]	a_0	a_1	σ^2	a_0	a_1	σ^2	a_0	a_1	σ^2
True value	**−0.69**	**1.10**	**0.16**	**−0.69**	**1.10**	**0.16**	**−0.69**	**1.10**	**0.16**
PQL	−0.67	1.07	0.13	−0.67	1.07	0.13	−0.67	1.07	0.14
Laplace	−0.69	1.10	0.14	−0.69	1.10	0.15	−0.69	1.10	0.15
AGQ NLMIXED	−0.69	1.08	0.22	−0.67	1.07	0.17	−0.69	1.10	0.16
AGQ GLIMMIX	−0.71	1.12	0.22	−0.69	1.10	0.17	−0.69	1.11	0.16
Hybrid BL	−0.69	1.10	0.19	−0.69	1.10	0.16	−0.69	1.10	0.15

(b) σ^2 = 0.36

	3 observations			5 observations			10 observations		
Method[a]	a_0	a_1	σ^2	a_0	a_1	σ^2	a_0	a_1	σ^2
True value	**−0.69**	**1.10**	**0.36**	**−0.69**	**1.10**	**0.36**	**−0.69**	**1.10**	**0.36**
PQL	−0.65	1.04	0.25	−0.65	1.04	0.28	−0.66	1.05	0.31
Laplace	−0.68	1.08	0.29	−0.70	1.10	0.31	−0.69	1.10	0.34
AGQ NLMIXED	−0.69	1.10	0.39	−0.69	1.09	0.36	−0.69	1.10	0.36
AGQ GLIMMIX	−0.69	1.12	0.39	−0.69	1.09	0.36	−0.69	1.10	0.35
Hybrid BL	−0.69	1.11	0.31	−0.68	1.09	0.30	−0.69	1.10	0.34

(c) σ^2 = 0.65

	3 observations			5 observations			10 observations		
Method[a]	a_0	a_1	σ^2	a_0	a_1	σ^2	a_0	a_1	σ^2
True value	**−0.69**	**1.10**	**0.65**	**−0.69**	**1.10**	**0.65**	**−0.69**	**1.10**	**0.65**
PQL	−0.62	1.00	0.40	−0.63	1.00	0.48	−0.65	1.03	0.54
Laplace	−0.68	1.10	0.49	−0.68	1.09	0.56	−0.69	1.10	0.62
AGQ NLMIXED	−0.70	1.10	0.69	−0.69	1.10	0.66	−0.69	1.10	0.65
AGQ GLIMMIX	−0.70	1.10	0.69	−0.69	1.10	0.66	−0.69	1.10	0.65
Hybrid BL	−0.67	1.09	0.51	−0.69	1.10	0.56	−0.68	1.10	0.60

[a]See legend of Table 1 for notations.

Table 4

Coverage probability of nominal 95% intervals from the balanced head and neck data simulations.

(a) $\sigma^2 = 0.16$

Method[a]	3 observations		5 observations		10 observations	
	a_0	a_1	a_0	a_1	a_0	a_1
True value	−0.69	1.10	−0.69	1.10	−0.69	1.10
PQL	0.94	0.95	0.94	0.96	0.94	0.94
Laplace	0.94	0.95	0.95	0.95	0.94	0.94
AGQ NLMIXED	0.96	0.94	0.95	0.93	0.95	0.94
AGQ GLIMMIX	0.97	0.95	0.96	0.95	0.95	0.95
Hybrid BL	0.96	0.96	0.93	0.94	0.95	0.96

(b) $\sigma^2 = 0.36$

Method[a]	3 observations		5 observations		10 observations	
	a_0	a_1	a_0	a_1	a_0	a_1
True value	−0.69	1.10	−0.69	1.10	−0.69	1.10
PQL	0.95	0.94	0.93	0.94	0.94	0.94
Laplace	0.95	0.93	0.95	0.96	0.94	0.95
AGQ NLMIXED	0.95	0.95	0.95	0.96	0.95	0.95
AGQ GLIMMIX	0.96	0.96	0.95	0.96	0.95	0.95
Hybrid BL	0.93	0.95	0.95	0.94	0.93	0.96

(c) $\sigma^2 = 0.65$

Method[a]	3 observations		5 observations		10 observations	
	a_0	a_1	a_0	a_1	a_0	a_1
True value	−0.69	1.10	−0.69	1.10	−0.69	1.10
PQL	0.94	0.93	0.92	0.93	0.94	0.92
Laplace	0.93	0.95	0.94	0.95	0.94	0.94
AGQ NLMIXED	0.95	0.95	0.95	0.94	0.95	0.95
AGQ GLIMMIX	0.95	0.95	0.95	0.94	0.95	0.96
Hybrid BL	0.93	0.94	0.93	0.94	0.94	0.95

[a]See legend of Table 1 for notations.

Table 5

Mean values of parameter estimates from the unbalanced head and neck data simulations.

(a) $\sigma^2 = 0.16$

Method[a]	Head and neck data 1 ×[b]			Head and neck data 2 ×[b]			Head and neck data 5 ×[b]		
	a_0	a_1	σ^2	a_0	a_1	σ^2	a_0	a_1	σ^2
True value	**−0.69**	**1.10**	**0.16**	**−0.69**	**1.10**	**0.16**	**−0.69**	**1.10**	**0.16**
PQL	−0.68	1.07	0.13	−0.66	1.07	0.14	−0.68	1.08	0.14
Laplace	−0.72	1.13	0.74	−0.70	1.12	0.16	−0.69	1.10	0.15
AGQ NLMIXED	−0.74	1.19	0.65	−0.67	1.07	0.25	−0.68	1.08	0.16
AGQ GLIMMIX	−0.76	1.23	0.76	−0.70	1.12	0.25	−0.69	1.11	0.16
Hybrid BL	−0.76	1.20	0.43	−0.71	1.14	0.22	−0.69	1.11	0.15

(b) $\sigma^2 = 0.36$

Method[a]	Head and neck data 1×[b]			Head and neck data 2×[b]			Head and neck data 5×[b]		
	a_0	a_1	σ^2	a_0	a_1	σ^2	a_0	a_1	σ^2
True value	**−0.69**	**1.10**	**0.36**	**−0.69**	**1.10**	**0.36**	**−0.69**	**1.10**	**0.36**
PQL	−0.65	1.04	0.17	−0.64	1.02	0.23	−0.65	1.04	0.29
Laplace	−0.69	1.10	0.22	−0.68	1.09	0.28	−0.68	1.11	0.33
AGQ NLMIXED	−0.73	1.17	0.81	−0.68	1.08	0.41	−0.69	1.11	0.36
AGQ GLIMMIX	−0.75	1.20	0.93	−0.69	1.10	0.41	−0.69	1.11	0.36
Hybrid BL	−0.72	1.14	0.49	−0.69	1.09	0.31	−0.68	1.09	0.32

(c) $\sigma^2 = 0.65$

Method[a]	Head and neck data 1×[b]			Head and neck data 2×[b]			Head and neck data 5×[b]		
	a_0	a_1	σ^2	a_0	a_1	σ^2	a_0	a_1	σ^2
True value	**−0.69**	**1.10**	**0.65**	**−0.69**	**1.10**	**0.65**	**−0.69**	**1.10**	**0.65**
PQL	−0.62	0.99	0.22	−0.62	0.99	0.36	−0.63	1.02	0.50
Laplace	−0.68	1.07	2.14	−0.68	1.08	0.48	−0.69	1.09	0.59
AGQ NLMIXED	−0.74	1.19	1.19	−0.71	1.12	0.71	−0.69	1.11	0.65
AGQ GLIMMIX	−0.76	1.22	1.38	−0.71	1.13	0.74	−0.69	1.11	0.65
Hybrid BL	−0.70	1.10	0.56	−0.69	1.10	0.51	−0.69	1.09	0.58

[a] See legend of Table 1 for notations.

[b] 1×, 2×, and 5× indicate that the head and neck data structure was used as in the original head and neck data, was doubled, or was multiplied by 5 times, respectively.

Table 6

Mean values of parameter estimates from the salamander data simulations; " – " indicates data not available.

$(a)\sigma_f^2 = \sigma_m^2 = 0.50$

Method[a]	σ_f^2	σ_m^2	a_0	a_1	a_2	a_3
True value	0.50	0.50	1.06	−3.05	−0.72	3.77
PQL	0.33	0.32	0.94	−2.73	−0.64	3.38
Laplace	0.41	0.42	1.05	−3.05	−0.71	3.76
AGQ NLMIXED	-	-	-	-	-	-
AGQ GLIMMIX	-	-	-	-	-	-
Hybrid BL	0.50	0.49	1.09	−3.15	−0.74	3.89

$(b)\sigma_f^2 = \sigma_m^2 = 1.00$

Method[a]	σ_f^2	σ_m^2	a_0	a_1	a_2	a_3
True value	1.00	1.00	1.06	−3.05	−0.72	3.77
PQL	0.55	0.54	0.87	−2.52	−0.6	3.11
Laplace	0.83	0.84	1.06	−3.02	−0.75	3.76
AGQ NLMIXED	-	-	-	-	-	-
AGQ GLIMMIX	-	-	-	-	-	-
Hybrid BL	0.87	0.88	1.07	−3.05	−0.74	3.79

$(c)\sigma_f^2 = 1.67, \sigma_m^2 = 1.50$

Method[a]	σ_f^2	σ_m^2	a_0	a_1	a_2	a_3
True value	1.67	1.50	1.06	−3.05	−0.72	3.77
PQL	0.79	0.68	0.81	−2.34	−0.55	2.87
Laplace	1.46	1.28	1.07	−3.03	−0.74	3.78
AGQ NLMIXED	-	-	-	-	-	-
AGQ GLIMMIX	-	-	-	-	-	-
Hybrid BL	1.32	1.18	1.02	−2.99	−0.71	3.73

[a]See legend of Table 1 for notations.

Table 7

Estimated fixed effects and variance components for the Guatemalan simulated datasets; " – " indicates data not available.

Method/True Values	Intercept a_0	Community a_1	Family a_2	Child a_3	Community σ_1^2	Family σ_2^2	Community σ_1	Family σ_2
	0.665	1.000	1.000	1.000	1.000	1.000	1.000	1.000
PQL[a]	0.537	0.831	0.806	0.808	0.633	0.204	0.792	0.443
Laplace[a,b]	0.621	0.951	0.909	0.903	0.783	0.353	0.881	0.583
AGQ3[b]	0.661	1.014	0.969	0.964	-	-	0.944	0.898
AGQ5[b]	0.676	1.037	0.989	0.982	-	-	0.972	0.975
AGQ7[b]	0.677	1.039	0.990	0.984	-	-	0.974	0.979
Hybrid BL[a]	0.656	1.009	0.973	0.970	0.704	0.821	0.833	0.900

[a] See legend of Table 1 for notations.

[b] "AGQ3", "AGQ5", and "AGQ7" stand for adaptive Gaussian quadrature approximation with 3, 5, and 7 quadrature points as implemented by Pinheiro and Chao [18]. Results reported in Table 1 of Pinheiro and Chao [18]. Maximum dimension of the integrals involved in the computations for the adaptive Gaussian quadrature was 50.