

*Memorial Sloan-Kettering Cancer Center*  
Memorial Sloan-Kettering Cancer Center, Dept. of Epidemiology  
& Biostatistics Working Paper Series

---

*Year 2011*

*Paper 22*

---

A Hybrid Bayesian Laplacian Approach for  
Generalized Linear Mixed Models

Marinela Capanu\*

Mithat Gonen†

Colin B. Begg‡

\*Memorial Sloan-Kettering Cancer Center, [capanum@mskcc.org](mailto:capanum@mskcc.org)

†Memorial Sloan-Kettering Cancer Center, [gonenm@mskcc.org](mailto:gonenm@mskcc.org)

‡Memorial Sloan-Kettering Cancer Center, [beggc@mskcc.org](mailto:beggc@mskcc.org)

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/mskccbiostat/paper22>

Copyright ©2011 by the authors.

# A Hybrid Bayesian Laplacian Approach for Generalized Linear Mixed Models

Marinela Capanu, Mithat Gonen, and Colin B. Begg

## Abstract

The analytical intractability of generalized linear mixed models (GLMMs) has generated a lot of research in the past two decades. Applied statisticians routinely face the frustrating prospect of widely disparate results produced by the methods that are currently implemented in commercially available software. This article is motivated by this frustration and develops guidance as well as new methods that are computationally efficient and statistically reliable. Two main classes of approximations have been developed: likelihood-based methods and Bayesian methods. Likelihood-based methods such as the penalized quasi-likelihood approach of Breslow and Clayton (1993) have been shown to produce biased estimates especially for binary clustered data with small clusters sizes. More recent methods such as the adaptive Gaussian quadrature approach 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. Similarly, 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 build on our previous method (Capanu and Begg 2010) and propose a hybrid approach that provides 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 with the goal of obtaining good statistical properties, with relatively good computing speed, and using widely available software. The hybrid approach is shown to perform well against the other competitors considered. Another important finding of this research is the surprisingly good performance of the Laplacian approximation in the difficult case of binary clustered data with small clusters sizes. We apply the methods to a real study of head and neck squamous cell carcinoma and illustrate

their properties using simulations based on a widely-analyzed salamander mating dataset and on another important dataset involving the Guatemalan Child Health survey.

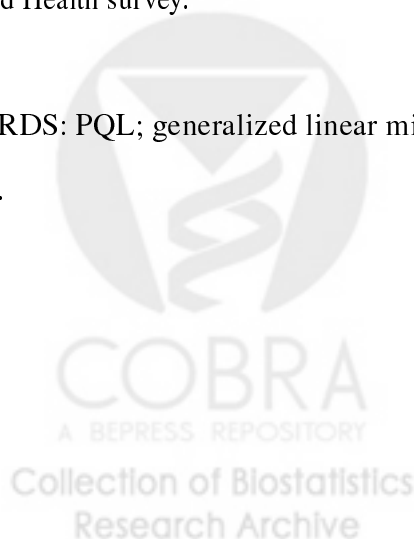
# A hybrid Bayesian Laplacian approach for generalized linear mixed models

MARINELA CAPANU, MITHAT G. ÖNEN, COLIN B. BEGG  
*Memorial Sloan-Kettering Cancer Center*

## Abstract

The analytical intractability of generalized linear mixed models (GLMMs) has generated a lot of research in the past two decades. Applied statisticians routinely face the frustrating prospect of widely disparate results produced by the methods that are currently implemented in commercially available software. This article is motivated by this frustration and develops guidance as well as new methods that are computationally efficient and statistically reliable. Two main classes of approximations have been developed: likelihood-based methods and Bayesian methods. Likelihood-based methods such as the penalized quasi-likelihood approach of Breslow and Clayton (1993) have been shown to produce biased estimates especially for binary clustered data with small clusters sizes. More recent methods such as the adaptive Gaussian quadrature approach 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. Similarly, 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 build on our previous method (Capanu and Begg 2010) and propose a hybrid approach that provides 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 with the goal of obtaining good statistical properties, with relatively good computing speed, and using widely available software. The hybrid approach is shown to perform well against the other competitors considered. Another important finding of this research is the surprisingly good performance of the Laplacian approximation in the difficult case of binary clustered data with small clusters sizes. We apply the methods to a real study of head and neck squamous cell carcinoma and illustrate their properties using simulations based on a widely-analyzed salamander mating dataset and on another important dataset involving the Guatemalan Child Health survey.

KEY WORDS: PQL; generalized linear mixed models; pseudo-likelihood; Bayesian; Laplace; binary clustered data.



<sup>1</sup>Supported by the National Cancer Institute Grant CA131010.  
Supplementary material available at <http://imstat.org/aoas/supplements>.

# 1 Introduction

Generalized linear mixed models (GLMM) have generated a lot of 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, GLMMs' 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 (1993) used Laplace approximations and Wolfinger and O'Connell (1993) 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 different expansions, the penalized quasi-likelihood (PQL) introduced by Breslow and Clayton (1993) and the pseudo-likelihood of Wolfinger and O'Connell (1993) 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 (Breslow and Lin 1995; Lin and Breslow 1996), modified Laplace approximation (Shun and McCullagh 1995; Shun 1997), adaptive Gaussian quadrature (Pinheiro and Bates 1995), higher order Laplace approximations (Raudenbush et al. 2000). The list is long and thorough reviews of these developments can be found in McCullagh and Searle (2001), Demidenko (2004), Hedeker and Gibbons (2006), and Lee et al. (2006).

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 (1991) investigated the use of Gibbs sampling to fit GLMMs. Other implementations that involved Monte Carlo methods were adopted by Gamerman (1997), Booth and Hobert (1999), and Natarajan and Kass (2000), 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 (2006) built on the methods introduced in Pinheiro and Bates (1995) and proposed efficient Laplacian and adaptive Gaussian quadrature algorithms that reduce computational complexity and memory usage for approximating multilevel GLMMs. Chan et al. (2006) 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. (2006) 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 (2008) 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 (2009) 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. (2010) 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 still faces situations in which available methods implemented in commercially statistical software produce widely disparate results. This is the case in the analysis of our motivating 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%.

In Capanu and Begg (2010) 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 as compared to 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 hybrid approach that combines Bayesian estimation of the variance components with Laplace estimation of the regression coefficients, and examine it under the general setting of generalized linear mixed models (GLLMs). Combining the simplicity of implementation of PQL and Laplace approximations with the desirable properties of the Bayesian approach, the hybrid approach enjoys good properties in the settings examined and is fast and easily accomplished in the standard statistical package SAS. The original hybrid pseudo-likelihood approach is also evaluated.

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 (McCullagh and Nelder 1989) is revisited. 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 salamander mating experiment. It also report the results of simulations based on a Guatemalan survey dataset (Rodriguez and Goldman 1995) which involves a multilevel structure with a large number of random effects. Results of the analysis of the head and neck data and the salamander data are interpreted in light of the simulation findings. Section 6 concludes with a Discussion.

## 2 Datasets

In this section we describe the motivating study which involved head and neck squamous cell carcinoma (HNSCC) patients. The analysis involves binary outcome 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.

### 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 (Wong et al. 2002). 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 records from the radiological exam for each patient. Reflecting the corresponding treatment options, observations within a patient were limited to up to 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. 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 contribute multiple data points to each 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 (top part of 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 motivated by such data analyses. We provide explicit recommendations and present a hybrid method that is fast and reliable.

## 2.2 Salamander Experiment

The salamander dataset has been previously described and analyzed at length by McCullagh and Nelder (1989), Breslow and Clayton (1993), Drum and McCullagh (1993), Karim and Zeger (1992), Lin and Breslow (1996), 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 of the same limitations encountered for the head and neck data with different methods leading to widely different results.

## 3 Methods

Consider a GLMM represented succinctly as follows:

$$\begin{aligned}
 Y &= \mu + e, & \text{such that} \\
 g(\mu) &= U\alpha + X\delta = \eta & (1) \\
 \text{cov}(\delta) &= G \\
 E(e|\mu) &= 0, & \text{cov}(e|\mu) = R_\mu^{1/2} R R_\mu^{1/2},
 \end{aligned}$$



where  $\alpha$  is a vector of fixed effects,  $\delta$  is a vector of random effects normally distributed with mean 0 and variance matrix  $G$ , and  $g(\cdot)$  is a differentiable monotonic link function (for example the logit link for mixed logistic regression). Here  $e$  is a vector of unobserved residuals,  $R_\mu$  is a diagonal matrix containing the variance functions of the model evaluated at  $\mu$ , and  $R$  is unknown. In the settings investigated here, we assume  $G$  is a diagonal matrix with the diagonal elements called random effects variance components, while the residual variance component is the common diagonal element of the diagonal matrix  $R$ .

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

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

In general, this integral does not have a closed-form expression for GLMMs, and different methods have been proposed to estimate it. In this article, we will investigate the performance of several approaches for estimating GLMMs: PQL, Breslow and Lin correction to PQL, Laplacian approximation, and two versions of our proposed hybrid approach, pseudo-likelihood and Laplacian estimation of the regression coefficients. 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 iteratively using a linearization technique which employs Taylor expansions to approximate the initial generalized linear mixed model with a linear mixed model (Wolfinger and O'Connell 1993). 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  $\hat{\delta}$  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 (1993) and Schabenberger (2005) and a description of its application to fit hierarchical models can be found in Capanu et al. (2008).

The penalized quasi-likelihood method (PQL) of Breslow and Clayton (1993) estimates generalized linear mixed models using Laplace approximations. 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 (Wolfinger and O'Connell 1993). 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 3.2 and 3.4 we describe two modifications of PQL designed to improve estimation of the variance components and of the regression coefficients.

### 3.2 Lin and Breslow Correction for the Variance Component

Breslow and Lin (1995) and Lin and Breslow (1996) proposed a correction of the PQL estimates of the variance components as well as first-order and second-order corrections for the PQL regression coefficients. The proposed correction matrices are derived based on a series of linear expansions and depend on the fixed effects and random effects design matrices and on the variance functions associated with the generalized mixed model. Using simulations, Breslow and Lin showed that the corrections can improve the performance of the PQL method for models with moderate values of variance components (between 0.5 and 1 on the log relative risk scale) but they still have severe bias for larger variance components. For general use they recommend applying the correction only for the variance components and recalculating the PQL regression coefficients based on the corrected PQL variance components.

### 3.3 Integral Approximations

#### 3.3.1 Laplace Approximation

There have been several Laplacian approximations proposed for estimating marginal posterior densities (Tierney and Kadane 1986) or nonlinear mixed models (Breslow and Clayton 1993; Wolfinger 1993; Pinheiro and Bates 1995). 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 Schabenberger (2005), the marginal distribution of the data can be written as

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

where  $\theta$  is the vector of the covariance parameters contained in the matrix  $G$ ,  $\phi$  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 (1993), we obtain

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

where  $h''$  is the second derivative matrix

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

while  $\hat{\delta}$  satisfies

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

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

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

This is the general form of the Laplace approximation for GLMMs. PQL follows from (3) via ignoring the middle term  $|X^T R_\mu X + G^{-1}|^{-1/2}$ , which amounts to assuming that  $R_\mu$  varies very slowly as a function of  $\mu$ .

Wolfinger (1993) assumed a flat prior for  $\alpha$  and expanded the integrand around  $\alpha$  and  $\delta$ , using the optimization only for the covariance parameters. Then, for given variance components  $\hat{\theta}$ , the fixed effects and random effects  $\alpha$  and  $\delta$  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  $\hat{\alpha}$  and  $\hat{\theta}$  determines the random effects vector  $\hat{\delta}$  that maximizes  $h(y, \alpha, \theta, \phi; \delta)$ . Another difference between the implementation of the Laplacian approximation in the GLIMMIX procedure and the approximations proposed by Wolfinger (1993) and Pinheiro and Bates (1995) is that the latter use an approximation of the second derivative  $h''(y, \alpha, \theta, \phi; \hat{\delta})$  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.

### 3.3.2 Adaptive Gaussian Quadrature

Introduced by Pinheiro and Bates (1995), adaptive Gaussian quadrature methods approximate 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 as well as appropriate centering and scaling of the abscissas. The name of “adaptive” comes from the fact that the number of quadrature points (nodes) is adaptively selected by evaluating the log-likelihood function at the starting values of the parameters until a tolerance is met (see Pinheiro and Bates 1995; SAS Institute Inc. 2008a,b, 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 (Pinheiro and Bates 1995; Zhang et al. 2011).

Adaptive Gaussian quadrature methods are currently implemented in SAS PROC GLIMMIX with

“METHOD=QUAD” option and PROC NLMIXED. 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 3.3.1 (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 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 5.2) 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 (2006) proposed efficient adaptive Gaussian quadrature algorithms for approximating multilevel models however their implementation is not available with standard statistical packages as of yet.

### 3.4 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 (Capanu and Begg 2010). This involves 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. 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 pseudo-likelihood 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 can be implemented using the %glimmix macro with the option “GDATA” which allows one to prespecify the variance-covariance matrix of the random effects. Details of the implementation can be found in the Appendix.

In this article, we have also explored a modified version of the hybrid approach that replaces the pseudo-likelihood estimation in Step 3 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 (as described in Step 2). Under this version, Step 3 is achieved using SAS PROC GLIMMIX with the option “METHOD=LAPLACE”. Both versions of the hybrid approach (with pseudo-likelihood and Laplace estimation of the regression parameters) have been evaluated in our simulations.

## 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

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

where  $y_{ij}$  is the presence or absence of recurrence as assessed by the radiology imaging for the  $i^{th}$  patient at the  $j^{th}$  site ( $i = 1, \dots, 128, j = 1, \dots, 3$ ), and  $b_i \sim N(0, \sigma^2)$  are the independent random effects corresponding to the patients. The fixed effects covariates  $x_{ij}$  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 in SAS with the NRRIDG as the optimization technique. Note that implementation of the adaptive quadrature method using the GLIMMIX procedure with METHOD=QUAD option leads to identical results. We also employed the Gibbs sampling algorithm (Zeger and Karim 1991) 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.

The top part of Table 1 presents estimates for the fixed effects as well as for the variance component from the four 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 ( $PQL_1$  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 ( $=\exp(-0.6)$ ).

We further carried out the analysis using the two versions of the hybrid approach: the hybrid Bayesian pseudo-likelihood approach (denoted “Hybrid BPL” in the tables) and the hybrid Bayesian with Laplacian approximation (denoted “Hybrid BL”), and these results are shown at the bottom of Table 1. Hybrid BPL which, just as  $PQL_1$  allows the dispersion parameter to be estimated, yields variance component estimates about 8 fold higher than the Laplace and the PQL version in which the dispersion parameter is fixed at 1 ( $PQL_2$ ). In contrast, the hybrid BL, produce results that are closer to the adaptive quadrature and Gibbs sampling. The estimates for the fixed effects are clustered similarly, with methods yielding smaller variance component estimates resulting in fixed effect estimates closer to each other than for the methods yielding larger variance component estimates.

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  $PQL_2$  or Hybrid BLP were 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 conducting simulations to establish which of these methods is more appropriate than the others.

## 4.2 Salamander Experiment

Following Lin and Breslow (1996), consider a model that pools the data across the three experiments (called Model A in Breslow and Clayton 1993):

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

where  $y_{ij}$  is the mating occurrence for the  $i^{\text{th}}$  female with the  $j^{\text{th}}$  male ( $i, j = 1, \dots, 60$ ), and  $b_i^f \sim N(0, \sigma_f^2)$  and  $b_j^m \sim N(0, \sigma_m^2)$  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) (Drum and McCullagh 1993; Lin and Breslow 1996). The fixed effects covariates  $x_{ij}$  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 Breslow and Clayton 1993).

Lin and Breslow (1996) 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 (McCullagh and Nelder 1989), Gibbs sampler (Karim and Zeger 1992), PQL (Breslow and Clayton 1993), and REML (Drum and McCullagh 1993). 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 (Zeger and Karim 1991).

We analyzed the salamander dataset using the two proposed versions of the hybrid approach: the hybrid Bayesian pseudo-likelihood approach (denoted “Hybrid BPL” in the tables) and the hybrid Bayesian with Laplacian approximation (denoted “Hybrid BL”). We present results for the variance components separately for each experiment and from the pooled data in Table 2, while in Table 3 we report results for the fixed effects based on the pooled data analysis. For comparison purposes, we reproduce the results reported in Table 1 and Table 2 of Lin and Breslow (1996) 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 “PQL( $\hat{\sigma}_{CP}^2$ )”), Laplace approximation, REML and Gibbs sampler. We also report the results from Shun (1997) 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. (2010) 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. (2010) and thus the results were not available for presentation here. Chan et al. (2006) and Tsai and Hsiao (2008) 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 compared in Table 3, the hybrid approach with Laplace approximation for the regression coefficients 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.

## 5 Simulations and Data Analysis

We evaluate the properties of the methods described in Section 3 with simulations based on two examples: the salamander experiment 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 proposed hybrid approach and compare its performance with those of other methods that have been used to analyze these datasets. Note that we could not evaluate with simulations the adaptive Gaussian quadrature method implemented by SAS as the NLMIXED procedure is limited to analyses involving a single variance component, while the GLIMMIX procedure can not handle crossed random effects (salamander simulations) and was computationally infeasible in the Guatemalan survey.

### 5.1 Salamander Experiment Simulations

To evaluate the finite-sample properties of the hybrid approach, we reproduce the simulation study conducted by Lin and Breslow (1996) in which 360 binary observations were generated according to model (5). The true  $\alpha^T = (1.06, -3.05, -0.72, 3.77)$  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 (1996) studied two configurations for the variance components  $(\sigma_f^2, \sigma_m^2) = (1.67, 1.5)$  corresponding to the actual REML variance component estimates (simulation results not shown in Lin and Breslow (1996)) and  $(\sigma_f^2, \sigma_m^2) = (0.5, 0.5)$ , representing a smaller amount of dispersion. In addition to these configurations, we considered an intermediate level of dispersion  $(\sigma_f^2, \sigma_m^2) = (1, 1)$ . 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. For reasons described in Section 3.2, the Breslow and Lin correction evaluated in our simulations performs the correction for the variance components and then uses PQL to recalculate the regression coefficients using the corrected variance components (denoted “PQL( $\hat{\sigma}_{CP}^2$ )” in the tables). As a benchmark, we also include the results from PQL with the variance components prespecified at the true values (denoted “PQL( $\sigma^2$ )”).

Table 4 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. For moderate amount of dispersion  $[(\sigma_f^2, \sigma_m^2) = (0.5, 0.5)]$ , both versions of the hy-



brid Bayesian approach produce almost unbiased estimates for the variance components. Nevertheless, the hybrid Bayesian approach with pseudo-likelihood estimation of the regression coefficients (BPL) still has bias in the regression coefficients. This is not surprising in light of the results presented for PQL in which the variance components are prespecified at the true value, which show that even when the variance components are correctly estimated, the PQL regression coefficients are still biased. The same behavior was observed in the numerical studies of PQL conducted by Jang and Lim (2009) 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 with Laplacian approximation of the regression coefficients have nearly unbiased estimates 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.

Table 5 reports the coverage probabilities of 95% confidence intervals formed using asymptotic normality of the parameter estimates. For all scenarios considered, the Laplacian approximation has coverage rates close to the nominal level of 95%. The hybrid approach with Laplacian approximation also has good coverage probabilities but tends to be anti-conservative for  $\alpha_1$  and  $\alpha_3$  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 with Laplacian approximation 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).

Overall, both the Laplacian approximation and the hybrid approach with Laplacian approximation have better properties than the different variations of PQL investigated, 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.

## 5.2 Guatemalan Survey Simulations

The 1987 Guatemalan National Survey of Maternal and Child Health (Pebley and Goldman 1992) 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 clustered 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 (1995) studied a subsample of respondents consisting of 2449 births that pertained to 1558 mothers living in 161 communities. The sample sizes per community ranged from 1 to 30 mothers with a mean of 15 children. Rodriguez and Goldman (1995) used this hierarchical structure 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

$$(y_{ijk}|p_{ijk}) \sim \text{Bernoulli}(p_{ijk}) \quad \text{independent with}$$

$$\text{logit}(p_{ijk}) = \beta_0 + \beta_1 x_{1ijk} + \beta_2 x_{2jk} + \beta_3 x_{3k} + u_{jk} + v_k,$$

where  $y_{ijk}$  is a binary indicator with value 1 if modern prenatal care was received for the  $k^{\text{th}}$  child within the  $j^{\text{th}}$  family within the  $i^{\text{th}}$  community and where  $v_k \sim N(0, \sigma_1^2)$  and  $u_{jk} \sim N(0, \sigma_2^2)$  are the independent community random effects and the independent family random effects, assumed independent of each other. Starting from the lowest level of the hierarchy, the child, family, and community covariates,  $x_{1ijk}$ ,  $x_{2jk}$ , and  $x_{3k}$ , respectively, are composite scales created by Rodriguez and Goldman (1995) 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  $\beta_0 = 0.665$  and  $\beta_1 = \beta_2 = \beta_3 = \sigma_1^2 = \sigma_2^2 = 1$ . The values of the variance components were chosen to result in a fairly large variability in the responses.

Recently, Pinheiro and Chao (2006) illustrated their adaptive Gaussian quadrature methods for multi-level GLMMs using the 100 simulated datasets created by Rodriguez and Goldman (1995). Browne and Draper (2006) 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 (1995).

We applied the hybrid approach with Laplace estimation for the 100 datasets simulated by Rodriguez and Goldman (1995) and compared the results against those already reported by Pinheiro and Chao (2006) and Browne and Draper (2006). In Table 6 we report mean estimates for the fixed effects and variance components, while in Table 7 we present the coverage of nominal 95% intervals for the fixed effects.

In their simulations, Pinheiro and Chao (2006) studied the properties of PQL with dispersion parameter allowed to be freely estimated with the rest of the parameters (denoted “PQL<sub>1</sub>” in our tables). For comparison we also investigated the performance of PQL with dispersion parameter fixed at 1 (denoted “PQL<sub>2</sub>”). Both versions of PQL have 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 with Laplacian estimation has almost unbiased estimates for the fixed effects, improved estimate for the family variance component as compared to the PQL and Laplace, though still underestimating the community variance component.

The adaptive Gaussian quadrature methods and the Bayesian implementations of Browne and Draper (2006) 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 ( $\beta_2$  and  $\beta_3$ ) while the hybrid approach, the adaptive Gaussian quadrature, and the adaptive Metropolis-Gibbs sampling have all coverage close to the nominal level (coverages were not reported by Pinheiro and Chao (2006) but were mentioned in the text).

Though the 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 almost no bias and good coverage for the fixed effects, improving on the Laplacian approximation.

### 5.3 Revisiting the Applications

Results of Section 5 can be used to interpret the findings of data analyses reported in Section 4. For example, in the head and neck cancer application, we have seen that the variance component estimates across the methods cluster in two groups: high estimates ( $PQL_2$  and Hybrid BPL) and low ones ( $PQL_1$ , Laplace, adaptive quadrature, and Hybrid BL). There was a 9-fold difference between the highest and lowest estimates (Table 1).

Table 4 suggests that this pattern of a large difference is repeated across the simulated data sets and it is the same methods that cluster together; generally speaking, PQL methods yield much higher estimates than Laplace-based methods. In our simulations the latter group of methods provided estimates closer to the true value. In addition, in Table 1, both the results from NLMIXED and WinBugs, two methods that can be considered somewhat of a gold standard but difficult to recommend for routine use due to numerical difficulties, are much closer to Laplace and Hybrid BL than they are to  $PQL_2$  and Hybrid BPL. In light of these two findings the true value of the variance component in the head and neck example is likely to be in the lower group.

## 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 as implemented in the SAS PROC GLIMMIX (Pinheiro and Bates 1995) perform well for these scenarios but can be overwhelmed for datasets with large numbers of random effects and can not be currently applied to crossed random effects or models without clustering. SAS NLMIXED procedure which also implements adaptive Gaussian quadrature methods has been shown to have good properties (Zhang et al. 2011) however it is only limited to analyses involving a single variance component. Efficient adaptive Gaussian quadrature methods have been proposed (Pinheiro and Chao 2006) but they are not available for use with standard statistical packages. Bayesian implementations involving Gibbs sampling also have good performances but are burdened by long computational times, they require a wise choice of priors, and the creation of specialized code for implementation.

The Laplacian approximation has received a lot of interest as a fast computational method for estimation of GLMMs, and numerous corrections and variations of it have been proposed based on the assumption that the standard Laplace approximation on its own does not have good performance. However, our simulations involving the salamander mating data show surprisingly good results for the Laplacian approximation exactly in this challenging setting of binary clustered data with small cluster sizes. This confirms the findings of Joe (2008) who also evaluated the accuracy of the Laplace approximation and was also surprised with its performance for the scenarios he investigated. It appears that the general feeling in the statistical community is that the Laplace approximation on its own has poor properties, but based on these findings, the Laplacian approximation in fact exceeds our expectations and certainly deserves more attention for future use. As Joe (2008) points out “I think other authors, who seem to recommend some of the other methods, also did not expect Laplacian and adaptive Gauss-Hermite quadrature to do so well”. The proposed hybrid approach is designed to improve on the performance of the Laplacian approximation by making use of Bayesian estimation of the variance components.

Similarly to the Laplace method, the hybrid Bayesian Laplacian approach also shows good performance in the difficult scenario of binary clustered data with small denominators, sometimes with slightly anti-conservative coverage for models with more heterogeneous random effects, but with more accurate estimation of the variance components for the models with smaller variability in the random effects. The hybrid approach has better performance than the Laplacian approximation for the complex hierarchical structure of the Guatemalan survey data, with almost unbiased regression coefficients, but somewhat underestimated

variance components. The analyses of our two datasets in Section 4 indicate that the Laplace method and the hybrid Bayesian Laplacian approach produce results closer to Gibbs sampling and/or adaptive Gaussian quadrature, which supports our simulation findings that these two methods have better statistical properties than the other competitors considered.

Based on these findings, if interest lies in accurate estimation of the fixed effects of a GLLM for binary clustered data, the Laplacian approximation is a good choice with excellent coverage probabilities and low bias. If however, one is interested in precise estimation of the variance components, the hybrid Bayesian with Laplacian approximation approach would be preferred for settings with moderate variability in the responses. For multilevel models with large numbers of random effects, the hybrid Bayesian Laplacian approach is superior to the Laplacian approximation on its own. Another advantage of the hybrid approach is that it always produces a positive estimate for the variance components, whereas the Laplacian approximation (just as PQL) can lead to negative or zero estimates, or to confidence intervals with negative values which may not be appropriate in certain settings.

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 are more complicated to implement. They both run fast: for the salamander experiment for example, they converge and produce results within seconds while adaptive quadrature implemented by SAS PROC GLIMMIX is overwhelmed with the number of evaluations and runs out of memory without producing any results. SAS code is available at [mskcc.org/marinelacapanu](http://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 are heavily time consuming, the hybrid approach and the Laplacian approximation are worthy alternatives that are fast and easy to implement with widely used standard procedures.

## References

- Booth, J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society. Series B. Methodological* **61**, 265–285.
- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.
- Breslow, N. E. and Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika* **82**, 81–91.

- Browne, W. J. and Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis* **1**, 473–514.
- Capanu, M. and Begg, C. B. (2010). Hierarchical modeling for estimating relative risks of rare genetic variants: Properties of the pseudo-likelihood method. *Biometrics* in press.
- Capanu, M., Orlow, I., Berwick, M., Hummer, A. J., Thomas, D. C., and Begg, C. B. (2008). The use of hierarchical models for estimating relative risks of individual genetic variants: an application to a study of melanoma. *Statistics in Medicine* **27**, 1973–1992.
- Chan, J. S. K., Kuk, A. Y. C., and Yam, C. H. K. (2006). Monte Carlo approximation through Gibbs output in generalized linear mixed models. *Journal of Multivariate Analysis* **94**, 300–312.
- Demidenko, E. (2004). *Mixed models: Theory and Applications*. Wiley, New York.
- Drum, M. and McCullagh, P. (1993). REML estimation with exact covariance in the logistic mixed model. *Biometrics* **49**, 677–689.
- Fong, Y., Rue, H., and Wakefield, J. (2010). Sampling from the posterior distribution in generalized linear mixed models. *Biostatistics* **11**, 397–412.
- Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing* **7**, 57–68.
- Hedeker, D. and Gibbons, R. D. (2006). *Longitudinal data analysis*. Wiley, New York.
- Jang, W. and Lim, J. (2009). A numerical study of PQL estimation biases in generalized linear mixed models under heterogeneity of random effects. *Communications in Statistics-Simulation and Computation* **38**, 692–702.
- Joe, H. (2008). Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics and Data Analysis* **52**, 5066–5074.
- Karim, R. M. and Zeger, S. L. (1992). Generalized linear models with random effects: salamander mating revisited. *Biometrics* **48**, 631–644.
- Lee, Y., Nelder, J. A., and Pawitan, Y. (2006). *Generalized linear models with random effect: unified analysis with H-likelihood*. Chapman and Hall, London.
- Lin, X. and Breslow, N. E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association* **91**, 1007–1016.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*. Chapman and Hall, London.
- McCullagh, P. and Searle, S. R. (2001). *Generalized, linear, and mixed models*. Wiley, New York.
- Natarajan, R. and Kass, R. E. (2000). Reference Bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association* **95**, 227–337.
- Ng, E. S. W., Carpenter, J. R., Goldstein, H., and Rasbash, J. (2006). Estimation in generalized linear mixed models with binary outcomes by simulated maximum likelihood. *Statistical Modelling* **6**, 23–42.
- Pebley, A. R. and Goldman, N. (1992). Family, community, ethnic identity and the use of formal health care services in Guatemala. Working Paper.
- Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* **4**, 12–35.

- Pinheiro, J. C. and Chao, E. C. (2006). Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics* **15**, 58–81.
- Raudenbush, S. W., Yang, M. L., and Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics* **9**, 141–157.
- Rodriguez, G. and Goldman, N. (1995). An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society. Series A* **158**, 73–89.
- Rue, H. and Martino, S. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series B. Methodological* **71**, 319–392.
- SAS Institute Inc., C. N. (2008a). SAS/STAT 9.2 Users Guide. Chapter 38: The GLIMMIX procedure.
- SAS Institute Inc., C. N. (2008b). SAS/STAT 9.2 Users Guide. Chapter 61: The NLMIXED Procedure.
- Schabenberger, O. (2005). Introducing the GLIMMIX procedure for generalized linear mixed models. SAS Institute Inc., Cary, NC, SUGI 30.
- Shun, Z. (1997). Another look at the salamander mating data: a modified Laplace approximation approach. *Journal of the American Statistical Association* **92**, 341–349.
- Shun, Z. and McCullagh, P. (1995). Laplace approximation of high-dimensional integrals. *Journal of the Royal Statistical Society. Series B. Methodological* **57**, 749–760.
- Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and densities. *Journal of the American Statistical Association* **81**, 82–86.
- Tsai, M. Y. and Hsiao, C. K. (2008). Computation of reference Bayesian inference for variance components in longitudinal studies. *Computational Statistics* **23**, 587–604.
- Wolfinger, R. (1993). Laplace’s approximations for nonlinear mixed models. *Biometrika* **80**, 791–795.
- Wolfinger, R. and O’Connell, M. (1993). Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* **48**, 233–243.
- Wong, R. J., Lin, D. T., Schöder, H., Patel, S. G., Gönen, M., Wolden, S., Pfister, D. G., Shah, J. P., Larson, S. M., and Kraus, D. H. (2002). Diagnostic and prognostic value of [<sup>18</sup>F]fluorodeoxyglucose positron emission tomography for recurrent head and neck squamous cell carcinoma. *Journal of Clinical Oncology* **20**, 4199–4208.
- Zeger, S. L. and Karim, R. M. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association* **86**, 79–86.
- Zhang, X., Lu, N., Feng, C., Thurston, S. W., Xia, Y., Zhu, L., and Tu, X. M. (2011). On fitting generalized linear mixed-effects models for binary responses using different statistical packages. *Statistics in Medicine* DOI: 10.1002/sim.4265.

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 components <sup>g</sup>
PQL <sub>1</sub> <sup>a</sup>	Estimate (s.e.) p-value	0.02 (0.01) 0.09	1.02(0.45) 0.03	-0.60 (0.48) 0.23	0.36 (0.50) 0.48	0.377 (0.51) -
PQL <sub>2</sub> <sup>b</sup>	Estimate (s.e.) p-value	0.04 (0.02) 0.04	0.94 (0.48) 0.06	-0.98 (0.55) 0.09	0.22 (0.58) 0.71	3.321 (0.93) -
Laplace	Estimate (s.e.) p-value	0.02 (0.02) 0.12	1.08 (0.47) 0.03	-0.64 (0.51) 0.22	0.39 (0.52) 0.46	0.371 (0.76) -
Adaptive quadrature <sup>d</sup>	Estimate (s.e.) p-value	0.03 (0.02) 0.12	1.11 (0.49) 0.03	-0.67 (0.54) 0.22	0.39 (0.54) 0.47	0.646 (1.14) -
Gibbs sampling <sup>c</sup>	Estimate (95%CI)	-0.73 (-4.06, 1.31)	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 BPL <sup>e</sup>	Estimate (s.e.) p-value	0.04 (0.02) 0.04	0.94 (0.48) 0.06	-0.95 (0.54) 0.09	0.23 (0.57) 0.69	3.105 (1.05) -
Hybrid BL <sup>f</sup>	Estimate (s.e.) p-value	0.03 (0.01) 0.08	1.10 (0.48) 0.03	-0.66 (0.51) 0.21	0.39(0.53) 0.47	0.558 (0.55) -

<sup>a</sup>PQL<sub>1</sub> stands for PQL with dispersion parameter allowed to be estimated.

<sup>b</sup>PQL<sub>2</sub> stands for PQL with dispersion parameter fixed at 1.

<sup>c</sup>Median posterior estimate along with 95% credible intervals.

<sup>d</sup>Adaptive Gaussian quadrature as implemented by SAS PROC NLMIXED/PROC GLIMMIX.

<sup>e</sup>“Hybrid BPL” stands for hybrid Bayesian Pseudo-Likelihood approach.

<sup>f</sup>“Hybrid BL” stands for hybrid Bayesian Laplacian approach.

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

<sup>a</sup>Results reported in Table 1 of Breslow and Clayton (1993).

<sup>b</sup>Variance components estimated using the Breslow and Lin correction; first-order (1<sup>st</sup> CPQL) and second-order (2<sup>nd</sup> CPQL) corrections reported in Table 2 apply only to the regression coefficients.

<sup>c</sup>Results reported in Table 3 of Shun (1997).

<sup>d</sup>Results reported in Web Table 4 of Fong et al. (2010).

<sup>e</sup>The estimation method for variance components is identical under both hybrid Bayesian Pseudo-Likelihood (Hybrid BPL) and hybrid Bayesian with Laplacian (Hybrid BL) and so we report only a single set of variance components estimates here.



Table 3: Estimated fixed effects for the pooled salamander data.

Method	Intercept	$WS_f$	$WS_m$	$WS_f \times WS_m$
REML <sup>a</sup>	1.06	-3.05	-0.72	3.77
Gibbs <sup>a</sup>	1.03 (0.43)	-3.01 (0.60)	-0.69 (0.50)	3.74 (0.68)
PQL <sup>a</sup>	0.79 (0.32)	-2.29 (0.43)	-0.54 (0.39)	2.82 (0.50)
PQL( $\hat{\sigma}_{CP}^2$ ) <sup>a,b</sup>	0.82 (0.35)	-2.40 (0.46)	-0.57 (0.42)	2.95 (0.51)
1 <sup>st</sup> CPQL <sup>a,c</sup>	1.19 (0.37)	-3.39 (0.55)	-0.82 (0.43)	4.19 (0.64)
2 <sup>nd</sup> CPQL <sup>a,d</sup>	0.68 (0.37)	-2.16 (0.55)	-0.49 (0.43)	2.65 (0.64)
Laplace <sup>a</sup>	1.01 (0.39)	-2.90 (0.56)	-0.70 (0.46)	3.59 (0.64)
Hybrid BPL <sup>e</sup>	0.84 (0.37)	-2.44 (0.48)	-0.58 (0.43)	3.00 (0.52)
Hybrid BL <sup>f</sup>	1.01 (0.38)	-2.90 (0.50)	-0.70 (0.46)	3.58 (0.56)

<sup>a</sup>Results reported in Table 2 of Breslow and Clayton (1993).

<sup>b</sup>PQL( $\hat{\sigma}_{CP}^2$ ) stands for PQL estimation of regression coefficients after prespecifying the variance components at the Breslow-Lin corrected estimate.

<sup>c</sup>“1<sup>st</sup> CPQL” stands for Breslow and Lin’s first-order correction to the regression coefficients.

<sup>d</sup>“2<sup>nd</sup> CPQL” stands for Breslow and Lin’s second-order correction to the regression coefficients.

<sup>e</sup>“Hybrid BPL” stands for hybrid Bayesian Pseudo-Likelihood approach.

<sup>f</sup>“Hybrid BL” stands for hybrid Bayesian Laplacian approach.



Table 4: Mean values of parameter estimates from the salamander data simulations.

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

Method	$\sigma_f^2$	$\sigma_m^2$	$\alpha_0$	$\alpha_1$	$\alpha_2$	$\alpha_3$
True value	0.50	0.50	1.06	-3.05	-0.72	3.77
REML <sup>a,b</sup>	0.55	0.54	1.09	-3.14	-0.74	3.88
PQL <sup>a</sup>	0.33	0.32	0.94	-2.73	-0.64	3.38
PQL( $\hat{\sigma}_{CP}^2$ ) <sup>a,c</sup>	0.46	0.46	0.96	-2.78	-0.66	3.44
PQL( $\sigma^2$ ) <sup>d</sup>			0.95	-2.77	-0.64	3.42
Laplace <sup>a</sup>	0.41	0.42	1.05	-3.05	-0.71	3.76
Hybrid BPL <sup>e</sup>	0.50	0.52	0.95	-2.77	-0.66	3.44
Hybrid BL <sup>f</sup>	0.50	0.49	1.09	-3.15	-0.74	3.89

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

Method	$\sigma_f^2$	$\sigma_m^2$	$\alpha_0$	$\alpha_1$	$\alpha_2$	$\alpha_3$
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
PQL( $\hat{\sigma}_{CP}^2$ )	0.78	0.77	0.91	-2.61	-0.62	3.22
PQL( $\sigma^2$ )			0.92	-2.69	-0.63	3.33
Laplace	0.83	0.84	1.06	-3.02	-0.75	3.76
Hybrid BPL	0.87	0.88	0.9	-2.62	-0.63	3.25
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	$\sigma_f^2$	$\sigma_m^2$	$\alpha_0$	$\alpha_1$	$\alpha_2$	$\alpha_3$
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
PQL( $\hat{\sigma}_{CP}^2$ )	1.10	1.01	0.83	-2.44	-0.57	3.03
PQL( $\sigma^2$ )			0.90	-2.62	-0.63	3.23
Laplace	1.46	1.28	1.07	-3.03	-0.74	3.78
Hybrid BPL	1.30	1.17	0.85	-2.49	-0.57	3.08
Hybrid BL	1.32	1.18	1.02	-2.99	-0.71	3.73

<sup>a</sup>Results reported in Table 3 of Breslow and Clayton (1993).

<sup>b</sup>Results for REML are only reported for  $\sigma_f^2 = \sigma_m^2 = 0.50$  in Breslow and Clayton (1993).

<sup>c</sup>PQL( $\hat{\sigma}_{CP}^2$ ) stands for PQL estimation of regression coefficients after prespecifying the variance components at the Breslow-Lin corrected variance components.

<sup>d</sup>PQL( $\sigma^2$ ) stands for PQL with variance components prespecified at the true values.

<sup>e</sup>“Hybrid BPL” stands for hybrid Bayesian Pseudo-Likelihood approach.

<sup>f</sup>“Hybrid BL” stands for hybrid Bayesian Laplacian approach.

Table 5: Coverage probability of nominal 95% intervals from the salamander data simulations.

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

Method	$\alpha_0$	$\alpha_1$	$\alpha_2$	$\alpha_3$
PQL	0.92	0.86	0.94	0.86
PQL( $\hat{\sigma}_{CP}^2$ ) <sup>a</sup>	0.94	0.90	0.95	0.89
PQL( $\sigma^2$ ) <sup>b</sup>	0.95	0.91	0.96	0.91
Laplace	0.94	0.96	0.94	0.95
Hybrid BPL <sup>c</sup>	0.94	0.89	0.95	0.87
Hybrid BL <sup>d</sup>	0.94	0.93	0.95	0.92

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

Method	$\alpha_0$	$\alpha_1$	$\alpha_2$	$\alpha_3$
PQL	0.89	0.72	0.94	0.73
PQL( $\hat{\sigma}_{CP}^2$ )	0.93	0.80	0.96	0.81
PQL( $\sigma^2$ )	0.95	0.89	0.96	0.86
Laplace	0.95	0.94	0.95	0.97
Hybrid BPL	0.93	0.82	0.96	0.82
Hybrid BL	0.93	0.91	0.96	0.92

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

Method	$\alpha_0$	$\alpha_1$	$\alpha_2$	$\alpha_3$
PQL	0.87	0.60	0.94	0.56
PQL( $\hat{\sigma}_{CP}^2$ )	0.91	0.72	0.95	0.68
PQL( $\sigma^2$ )	0.97	0.90	0.97	0.85
Laplace	0.95	0.94	0.94	0.95
Hybrid BPL	0.92	0.77	0.95	0.73
Hybrid BL	0.94	0.91	0.94	0.92

<sup>a</sup>PQL( $\hat{\sigma}_{CP}^2$ ) stands for PQL estimation of regression coefficients after prespecifying the variance components at the Breslow-Lin corrected variance components.

<sup>b</sup>PQL( $\sigma^2$ ) stands for PQL with variance components prespecified at the true values.

<sup>c</sup>“Hybrid BPL” stands for hybrid Bayesian Pseudo-Likelihood approach.

<sup>d</sup>“Hybrid BL” stands for hybrid Bayesian Laplacian approach.

Table 6: Estimated fixed effects and variance components for the Guatemalan simulated datasets; a “-” indicates data not available.

Intercept	Community	Family	Child	Community	Family	Community	Family	Community	Family
$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\sigma_1^2$	$\sigma_2^2$	$\sigma_1$	$\sigma_2$	$\sigma_1$	$\sigma_2$
Method/True Values	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
PQL <sub>1</sub> <sup>a</sup>	0.831	0.806	0.808	0.633	0.204	0.792	0.443	0.792	0.443
PQL <sub>2</sub> <sup>b,c</sup>	0.990	0.945	0.978	0.892	2.063	0.939	1.431	0.939	1.431
Laplace <sup>c</sup>	0.621	0.909	0.903	0.783	0.353	0.881	0.583	0.881	0.583
AGQ3 <sup>c</sup>	1.014	0.969	0.964	-	-	0.944	0.898	0.944	0.898
AGQ5 <sup>c</sup>	1.037	0.989	0.982	-	-	0.972	0.975	0.972	0.975
AGQ7 <sup>c</sup>	1.039	0.990	0.984	-	-	0.974	0.979	0.974	0.979
Bayesian IG prior <sup>d</sup>	0.991	1.006	0.982	1.023	0.964	-	-	-	-
Bayesian Uniform prior <sup>d</sup>	1.015	1.031	1.007	1.108	1.130	-	-	-	-
Hybrid BL <sup>e</sup>	1.009	0.973	0.970	0.704	0.821	0.833	0.900	0.833	0.900

<sup>a</sup>PQL<sub>1</sub> stands for PQL with dispersion parameter allowed to be estimated.

<sup>b</sup>PQL<sub>2</sub> stands for PQL with dispersion parameter fixed at 1.

<sup>c</sup>Results reported in Table 1 of Pinheiro and Chao (2006).

<sup>d</sup>Results reported in Table 10 of Browne and Draper (2006).

<sup>e</sup>“Hybrid BL” stands for hybrid Bayesian Laplacian approach.

Table 7: Estimated coverages of 95% confidence intervals for the Guatemalan simulated datasets; a “ – ” indicates data not available.

Method/True Values	Intercept	Community	Family	Child
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
	0.665	1.000	1.000	1.000
PQL <sub>1</sub> <sup>a</sup>	0.90	0.92	0.42	0.83
PQL <sub>2</sub> <sup>b</sup>	0.98	0.97	0.87	0.92
Laplace	0.97	0.97	0.85	0.92
AGQ3	-	-	-	-
AGQ5	-	-	-	-
AGQ7	-	-	-	-
Bayesian IG prior <sup>c</sup>	0.93	0.96	0.93	0.92
Bayesian Uniform prior <sup>c</sup>	0.94	0.96	0.93	0.94
Hybrid BL <sup>d</sup>	0.97	0.97	0.94	0.93

<sup>a</sup>PQL<sub>1</sub> stands for PQL with dispersion parameter allowed to be estimated.

<sup>b</sup>PQL<sub>2</sub> stands for PQL with dispersion parameter fixed at 1.

<sup>c</sup>Results reported in Table 10 of Browne and Draper (2006).

<sup>d</sup>“Hybrid BL” stands for hybrid Bayesian Laplacian approach.

