Pseudo-Likelihood or Quadrature? What We Thought We Knew, What We Think We Know, and What We Are Still Trying to Figure Out

Walt STROUP and Elizabeth CLAASSEN

Two predominant computing methods for generalized linear mixed models (GLMMs) are linearization, e.g., pseudo-likelihood (PL), and integral approximation, e.g., Gauss–Hermite quadrature. The primary GLMM package in R, LME4, only uses integral approximation. The primary GLMM procedure in SAS®, PROC GLIMMIX, was originally developed using linearization, but integral approximation methods were added in the 2008 release. This presents a dilemma for GLMM users: Which method should one use, and why? Linearization methods are more versatile and able to handle both conditional and marginal GLMMs. Linearization can be implemented with REML-like variance component estimation, whereas quadrature is strictly maximum likelihood. However, GLMM software documentation and the literature on which it is based tend to focus on linearization's limitations. Stroup (Generalized linear mixed models: modern concepts, methods and applications, CRC Press, Boca Raton, 2013) reiterates this theme in his GLMM textbook. As a result, "conventional wisdom" has arisen that integral approximation—quadrature when possible—is always best. Meanwhile, ongoing experience with GLMMs and research about their small sample behavior suggest that "conventional wisdom" circa 2013 is often not true. Above all, it is clear there is no one-size-fits-all best method. The purpose of this paper is to provide an updated look at what we now know about quadrature and PL and to offer some general operating principles for making an informed choice between the two. A series of simulation studies investigating distributions and designs representative of research in agricultural and related disciplines provide an overview of each method with respect to estimation accuracy, type I error control, and robustness (or lack thereof) to model misspecification.
Supplementary materials accompanying this paper appear online.

Key Words: Generalized linear mixed model; Linearization; Integral approximation; Type I error control; Model misspecification; Residual maximum likelihood.

W. Stroup (✉) Department of Statistics, University of Nebraska-Lincoln, Lincoln, USA (E-mail: *wstroup1@unl.edu*). E. Claassen JMP Division, SAS Institute, Cary, USA.

1. INTRODUCTION

Over the past two decades, generalized linear mixed model (GLMM) software has been developed—e.g., PROC GLIMMIX in SAS/STAT®and R packages glmmPQL and LME4— and the GLMM has gained acceptance as a mainstream method of analyzing non-Gaussian data from studies that call for a mixed model approach. Unlike fixed-effects-only generalized linear models (GLMs) and linear mixed models (LMMs) for Gaussian data, both of which have likelihood functions from which estimating equations can be derived, GLMM likelihood functions are generally intractable and require some form of numerical approximation.

Two forms of approximation—linearization and integral approximation—are the most commonly used. Linearization methods include **penalized quasi-likelihood** (PQL) (Breslow and Clayton 1993) and **pseudo-likelihood** (PL) (Wolfinger and O'Connell 1993). Integral approximation includes Laplace approximation and quadrature—specifically adaptive Gauss–Hermite quadrature (Pinheiro and Bates 1995). Despite PL being the default, SAS software documentation emphasizes cases in which PL produces undesirable results and quadrature is preferred. Because LME4 uses integral approximation only, quadrature is implicitly the method of choice. Stroup (2013) reinforces this theme (see, e.g., pp. 300–303). By the early 2010s, de facto conventional wisdom held that one should use GLMM software's quadrature algorithms, whenever possible, for GLMM estimation and inference.

In the meantime, much experience has been gained using GLMMs. This experience suggests that the conventional wisdom circa the early 2010s is, in many cases, misleading or simply wrong. This paper presents a fresh look at the behavior of PL and quadrature as implemented by available GLMM software in a number of common research scenarios that use GLMMs. Our goal is to provide some general guidelines. When is PL preferred, and why? When is quadrature recommended, and why? When does neither method appear capable of producing useable results? How would a data analyst know?

The paper is organized as follows:

- Section 2 describes a brief review of relevant GLMM background.

- Section 3 describes history essential to understanding the dilemmas faced by GLMM users.

- Section 4 shows the description of simulation studies intended to characterize PL and quadrature in comparative experiments with respect to type I error control and accuracy of treatment mean estimation.

- Section 5 shows the simulation study results.

- Section 6 describes the summary—overarching principles to the extent that we understand them to date, and some thoughts on next steps.

To reiterate, the focus of this paper is on the performance of PL and quadrature algorithms as they are implemented in software currently available for GLMMs in agricultural research.

2. GLMM BACKGROUND

A generalized linear mixed model is defined as follows. Let $\mathbf{y}$ be a vector of observations whose distribution, conditional on $\mathbf{b}$, a vector of random model effects, can be written

$$\mathbf{y}|\mathbf{b} \sim \mathcal{D}\left(h\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}\right), \mathbf{V}_{\mu}^{1/2}\mathbf{P}\mathbf{V}_{\mu}^{1/2}\right)$$

where $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$ is the mixed model linear predictor,

$\quad$ $\mathbf{X}$ and $\mathbf{Z}$ are matrices of constants,

$\quad$ $\boldsymbol{\beta}$ is a vector of fixed effects,

$\quad$ $h\left(\bullet\right)$ is the inverse link function,

$\quad$ $\mathbf{b} \sim N\left(\mathbf{0}, \mathbf{G}\right)$

$\quad$ $\mathbf{V}_{\mu}^{1/2}$ is a diagonal matrix whose elements are $\sqrt{v\left(\mu_i\right)}$, where $v\left(\mu_i\right)$ is the variance function for the ith observation,

$\mathbf{P}$ is a matrix of scale or covariance parameters depending on the distribution and model. Let $\boldsymbol{\sigma}$ denote the vector of scale or covariance parameters on which $\mathbf{P}$ depends.

The linear predictor models the expected value of $\mathbf{y}|\mathbf{b}$, hereafter denoted $\boldsymbol{\mu}$, through the inverse link function. Alternatively, GLMMs are often specified as $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} = g\left(\boldsymbol{\mu}\right)$, where $g\left(\bullet\right)$ is called the link function.

The standard GLMM estimation approach is maximum likelihood. This requires maximizing

$$f\left(\mathbf{y}\right) = \int_{\mathbf{b}} f\left(\mathbf{y}|\mathbf{b}\right) f\left(\mathbf{b}\right) d\mathbf{b}$$

with respect to $\boldsymbol{\sigma}$, $\boldsymbol{\beta}$, and $\mathbf{b}$. In general, this integral is intractable. Estimation requires some form of approximation. The two most common approaches are linearization and integral approximation. What follows are heuristic overviews of each approach. For more in-depth discussions of these methods, refer to the references cited in Sect. 1.

Linearization

Linearization results in maximum likelihood estimating equations for the generalized linear model (GLM)

$$\mathbf{X}'\mathbf{W}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{W}\mathbf{y}^*$$

where $\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\left(\mathbf{y} - \boldsymbol{\mu}\right)$, $\mathbf{W} = \mathbf{D}\mathbf{V}_{\mu}^{1/2}\mathbf{P}\mathbf{V}_{\mu}^{1/2}\mathbf{D}$ and $\mathbf{D}$ is a diagonal matrix whose ith element is $\partial\eta_i/\partial\mu_i$. Noting that $E\left(\mathbf{y}^*\right) = \mathbf{X}\boldsymbol{\beta}$ and $Var\left(\mathbf{y}^*\right) = \mathbf{W}^{-1}$, the GLM estimating equations can be viewed as generalized least squares with the pseudo-variate $\mathbf{y}^*$.

Pseudo-likelihood, the form of linearization used by PROC GLIMMIX, extends the pseudo-variate GLS approach to the mixed model. Linearization begins by taking the first-order Taylor series expansion of $g\left(\mathbf{y}\right)$ evaluated at $\boldsymbol{\mu}$. This results in the pseudo-variate extended to the mixed model linear predictor.

$$g\left(\boldsymbol{\mu}\right) + \partial g\left(\mathbf{y}\right)/\partial\mathbf{y}|_{\boldsymbol{\mu}}\left(\mathbf{y} - \boldsymbol{\mu}\right) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{D}\left(\mathbf{y} - \boldsymbol{\mu}\right).$$

Pseudo-likelihood in effect assumes that $\mathbf{y}^*|\mathbf{b}$ is approximately distributed $N\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \mathbf{W}^{-1}\right)$ and uses $\mathbf{y}^*$ and $\mathbf{W}$ accordingly in the mixed model estimating equations. Covariance estimating equations are adapted from Harville (1977). The residual likelihood forms are used to obtain REML-like PL covariance parameter estimates. Full likelihood forms are used to obtain ML-like PL estimates. In GLIMMIX, the methods are called RSPL and MSPL, respectively, RSPL being the default.

Integral Approximation

As the name implies, integral approximation refers to the GLMM marginal likelihood

$$f\left(\mathbf{y}\right) = \int_{\mathbf{b}} f\left(\mathbf{y}|\mathbf{b}\right) f\left(\mathbf{b}\right) d\mathbf{b}.$$

Two commonly used methods are the Laplace approximation and adaptive Gauss–Hermite quadrature. This paper focuses on quadrature.

The following illustrates—very heuristically—the basic idea. Noting the GLMM random effects assumption, $\mathbf{b} \sim N\left(\mathbf{0}, \mathbf{G}\right)$, the marginal likelihood integral can be written

$$\int_{\mathbf{b}} f\left(\mathbf{y}|\mathbf{b}\right) \left(2\pi\right)^{-r/2} |\mathbf{G}|^{-1/2} \exp\left(-\mathbf{b}'\mathbf{G}\mathbf{b}/2\right) d\mathbf{b}.$$

This can be rewritten

$$\int_{\mathbf{b}} f\left(\mathbf{y}|\mathbf{z}\right) \left(\pi\right)^{-r/2} \exp\left(\mathbf{z}'\mathbf{z}\right) d\mathbf{z}$$

where $\mathbf{z}$ is the vector of z-scores obtained from the random vector $\mathbf{b}$. The Gauss–Hermite approximation of this integral is

$$\sum \mathbf{w} f\left(\mathbf{y}|\mathbf{x}\right) \left(\pi\right)^{-r/2}$$

where $\mathbf{x}$ and $\mathbf{w}$ are quadrature abscissas and weights, respectively.

In practice, the abscissa weights are centered and scaled. See Pinheiro and Bates (1995) or PROC GLIMMIX documentation for details. Maximum likelihood estimates are obtained using the log of the quadrature approximation. One-point quadrature is equivalent to the Laplace approximation.

3. SOME HISTORY

Linearization methods date to publications by Breslow and Clayton (1993) and Wolfinger and O'Connell (1993). Subsequently, publications by Breslow and Lin (1995), Lin and Breslow (1996), Murray et al. (2004), and Pinheiro and Chao (2006) document problems with linearization methods. Summarized by Bolker et al. (2009), these papers focus on

Table 1.   Average estimates of σ_b^2, true $\sigma_b^2 = 1$

Method	200 replications, $N = 4$	10 replications, $N = 200$
RSPL	0.609	0.986
MSPL	0.599	0.881
Laplace	0.835	0.902
Quadrature	0.994	0.902

Table 2.   From Claassen (2014). Recommended method. binomial(N, p) Data with S subjects

	Small S (few replications)	Large S (many replications)
Small N∗	Quadrature is least bad	Use quadrature
Large N	Use RSPL	Does not matter—use either method

∗ Binary data are the extreme case of small N

binary data, binomial distributions with Np or $N(1 - p) < 5$, and highly skewed count distributions, e.g., those with expected count < 5. Adaptive Gauss–Hermite quadrature, which appeared in a publication by Pinheiro and Bates (1995), seemed to address these problems.

Although RSPL is the GLIMMIX default, SAS documentation focuses heavily on "the down side of PL," leaving the impression that quadrature, when possible, is the method of choice for GLMMs. R's LME4 package reinforces this impression, integral approximation being the only method available with LME4.

A 2-day short course at JSM (Schabenberger and Stroup, 2008) and a subsequent GLMM textbook (Stroup, 2013) continued the "down side of PL" theme. Both use the following example to illustrate. One thousand simulated data sets are generated from a random coefficient binomial GLMM. Specifically, $y_i | b_i \sim \text{Bin}(N, p_i)$, $b_i \sim N(0, \sigma_b^2)$ and $\eta_i = \log[p_i/(1 - p_i)] = \eta + b_i$. In the example, $\sigma_b^2 = 1$, $\eta = -1$ (or, equivalently, $p \cong 0.27$) , $i = 1, 2, 3, ..., 200$ and $N = 4$. In other words, these are binomial data with a lot of replication (200 subjects or, equivalently, 200 experimental units) but small binomial N at each replication. Estimates of the random effect variance were obtained from RSPL, MSPL, Laplace, and quadrature. The left-hand column of Table 1 shows the results.

The PL estimates are biased downward by approximately 40%, whereas the quadrature estimate is essentially unbiased.

This example shows the dominant perception as of 2013. Missing in the literature was attention to scenarios, common in agricultural research, in which replication is minimal.

Claassen (2014) revisited this example, but with $N = 200$ and only ten subjects, i.e., ten replicate experimental units. The right-hand column of Table 1 shows the results.

In Claassen's version, RSPL shows negligible bias, whereas MSPL and both integral approximation methods are downward biased by approximately 10%. Out of these, two examples came recommendations for binomial data, shown in Table 2.
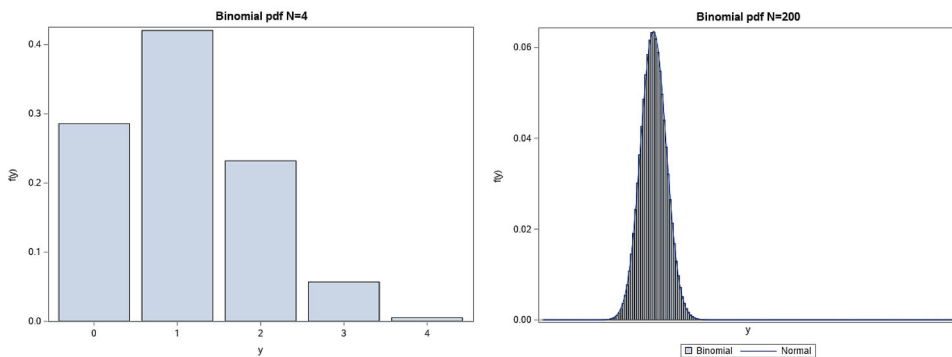
Figure 1. Plots of binomial p.d.f..

Table 2 raises the following questions. Why the difference? Are there organizing principles that inform or extend to GLMMs with other distributions and other designs?

In the time since 2013, Stroup pursued this question in his modeling class, having students compare the small sample behavior of RSPL and quadrature. Scenarios were drawn from Stroup's consulting with researchers whose experiments called for GLMM analysis. The exercises were originally intended to give students hands-on experience generating evidence supporting "quadrature is better" conventional wisdom. Instead, in an unexpectedly large number of cases, RSPL's performance was superior—often dramatically so.

These exercises led to a conjecture, based partly on pseudo-likelihood's implicit assumption that $\mathbf{y}^*|\mathbf{b}$ is approximately normally distributed, and partly on the known downward bias of maximum likelihood variance component estimates. Figure 1 shows plots of the binomial p.d.f. with $N = 4$ and $N = 200$. In both cases, $p = 0.27$.

With $N = 4$, the binomial distribution is strongly right skewed. On the other hand, the right-hand plot shows the binomial ($N = 200$, $p = 0.27$) p.d.f. with the normal distribution superimposed; the two distributions are almost indistinguishable. Pseudo-likelihood's assumption of normality for $\mathbf{y}^*|\mathbf{b}$ seems reasonable when $N = 200$ but not when $N = 4$.

Now consider variance estimation. Start with students' introduction to variance estimation: "Let $\{y_i; \ i = 1, 2, ..., S\}$ be a random sample from a population with distribution $N\left(\mu, \sigma^2\right)$. The unbiased estimate of σ^2 is $\sum_i (y_i - \bar{y})^2/(S - 1)$." The maximum likelihood estimate of σ^2 is $\sum_i (y_i - \bar{y})^2/S$. Maximizing the likelihood of the residuals, $\{(y_i - \bar{y}); \ i = 1, 2, ..., S\}$ yields the *residual* maximum likelihood (REML) estimate of σ^2, $\sum_i (y_i - \bar{y})^2/(S - 1)$. Now suppose $\sigma^2 = 1$, $\sum_i (y_i - \bar{y})^2 = 9$ and S = 10. The MLE of σ^2 is 0.9, whereas the REML estimate is 1. This is exactly what we observe in Claassen's version of the binomial random intercept example: $\mathbf{y}|\mathbf{b}$, and hence, $\mathbf{y}^*|\mathbf{b}$, satisfies PL's "$\mathbf{y}^*|\mathbf{b}$ approximately normal" assumption, and RSPL computes a REML-like estimator, which averages close to 1. The other methods compute MLEs and average close to 0.9.

When $N = 4$ but S = 200, $\mathbf{y}^*|\mathbf{b}$ fails PL's "$\mathbf{y}^*|\mathbf{b}$ approximately normal" assumption. Pseudo-likelihood performs poorly. Quadrature performs well because the difference between S and S-1 is negligible when S = 200, making ML's downward bias negligible.

Although our conjecture seems to explain the difference between the two versions of the random coefficient binomial example, does this explanation hold for other models and other distributions? In the following sections, we explore this question via simulation.

In addition, we explore one more question. An oft-cited advantage of integral approximation is that one is working with the actual likelihood, not a linear approximation. However, in many cases, this supposed advantage may actually be a liability. Thelonious Monk, the iconic jazz musician, once said, "There are no wrong notes, but some are more right than others." Recasting his quote in modeling terms, "There are no right models, but some are less wrong than others." Integral approximation depends crucially on having a "less wrong than others" likelihood, i.e., one that is consistent with the processes and probability characteristics that give rise to the data. However, what happens when the likelihood is misspecified? How robust is quadrature? How robust is PL? We include in our simulation study situations in which there are competing plausible "how the data arise" scenarios that provide a good basis for assessing robustness to model misspecification.

To summarize, the main purpose of this paper is to compare REML-like PL with quadrature in scenarios common in agricultural research, but largely ignored by the existing GLMM literature.

4. SIMULATION SCENARIOS

Our focus is on designs typically used by agricultural researchers and distributions associated with response variables common to their disciplines. An exhaustive study of all designs and distributions being impractical for a single publication, we limit simulation scenarios as follows.

Designs. Most researchers use block designs in some form, e.g., complete block designs (with or without missing data), incomplete block designs, designs with multi-level or split-plot features, etc. Even observational studies frequently have implicit retrospective blocking criteria, such as matched pairs, common locations or habitats, etc. Most studies estimate and compare treatment means or estimate a regression relationship.

The simulations in this paper use two designs, a randomized complete block (RCB) and a balanced incomplete block (BIB) design. Both designs have six treatments. In both designs, we consider two cases, one with five replications per treatment and the other with 15. The smaller design represents cases in which replication is minimal—a common occurrence in agricultural research. We include the larger design to illustrate the effect of increasing replication on the relative performance of the quadrature and REML-like PL methods.

For the RCB five or 15 replications translates to five or 15 blocks. The BIB used in this simulation study has a block size of three. The BIB with five replications per treatment has ten blocks, and the design with 15 replications has 30 blocks. Note that the BIB, the RCB with missing data and designs with split-plot features are all special cases of incomplete block design structure. Our experience is that the BIB results reported in this paper apply to incomplete block structures in general.

Distributions. Commonly measured response variables that are reasonably assumed to follow a non-Gaussian distribution are discrete proportions (binomial), discrete counts (Pois-

Table 3.   Source of variation table for count data

Source	Poisson-gamma		Poisson-normal	
	Process	Observation	Process	Observation
Block	$b_j \sim N\left(0, \sigma_b^2\right)$		$b_j \sim N\left(0, \sigma_b^2\right)$	
Treatment	τ_i		τ_i	
Exp unit (block)	$u_{ij} \sim \Gamma\left(1/\phi, \phi\right)$	$y_{ij}\vert b_j, u_{ij} \sim$ Poisson $\left(\lambda_{ij} u_{ij}\right)$	$u_{ij} \sim N\left(0, \sigma_u^2\right)$	$y_{ij}\vert b_j, u_{ij} \sim$ Poisson $\left(\lambda_{ij}\right)$

son, negative binomial), continuous proportions (beta), continuous, non-negative, right-skewed including but not limited to time-to-event data (gamma, log-normal).

For the simulation studies in this paper, we generated simulated data according to commonly used "how we think the data arise" scenarios for each type of response variable and then analyzed the data with GLMMs typically used for these data. The remainder of this section consists of three subsections: Sect. 4.1 describes how simulated data were generated; Sect. 4.2 describes GLMMs used to analyze these data; and Sect. 4.3 describes the information tabulated from each scenario and how it was used to characterize the performance of the pseudo-likelihood and quadrature algorithms.

In the interest of space, the remainder of this paper focuses on count data scenarios. These fully illustrate the issues involved, and the results for the other distributions are similar to those reported below. Descriptions and results for binomial, beta, gamma, and log-normal scenarios are available with this paper's supplementary materials online.

4.1. DATA GENERATION SCENARIOS

Following Stroup (2013), a model is "sensible" if there is a one-to-one correspondence between sources of variation and parameters in the model to account for these sources (see Chapter 2 discussion of translating design to model). In a blocked design, there are three sources: blocks, experimental units within blocks, and treatments. For each type of response variable, a plausible "how data arise" scenario must specify how variation is believed to occur for each source of variation.

Table 3 shows two common GLMM processes for count data, the Poisson-gamma and Poisson-normal. The names refer to the distribution of $\mathbf{y}\vert\mathbf{b}$ and the distribution of the experimental unit within block effect, respectively. Both processes assume Gaussian random variation among block effects and fixed treatment effects. The difference is the distribution associated with effects at the experimental unit level.

For the Poisson-gamma process, integrating out the unit-level effect, u_{ij} yields $y_{ij}\vert b_j \sim$ Negative binomial $\left(\lambda_{ij}, \phi\right)$ (see Hilbe, 2014, for in-depth discussion). For this reason, the negative binomial GLMM is commonly used for data assumed to arise via the Poisson-gamma process. We maintain this distinction in terminology throughout the rest of this paper. Poisson-gamma refers to the process giving rise to the data, and negative binomial GLMM refers to the model used to analyze the data.
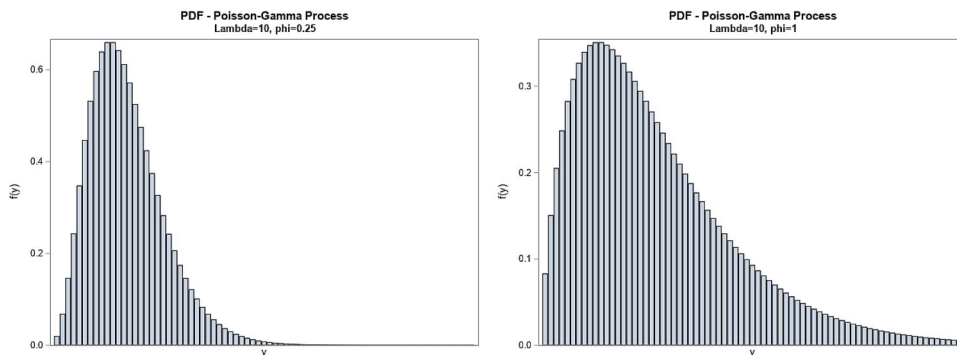
Figure 2. Plots of count data distributions used in simulation study.

Table 4. Simulation scenarios for count data

Generating distribution	Scale	Design of experiment			
		RCB		BIB	
Poisson-gamma	$\phi = 0.25$	5 blocks	15 blocks	10 blocks	30 blocks
	$\phi = 1$				
Poisson-normal	$\sigma_u^2 = 0.25$				
	$\sigma_u^2 = 1$				

Data were generated using both processes. For the Poisson-gamma (a.k.a. negative binomial) process, we used $\phi = 0.25$ to simulate data in which $y_{ij}|b_j$ has a reasonably symmetric distribution and $\phi = 1$ to simulate data in which $y_{ij}|b_j$ is strongly right-skewed. Figure 2 shows probability density function plots of the two distributions, with $\phi = 0.25$ on the left.

For the Poisson-normal process, we used $\sigma_u^2 = 0.25$ and $\sigma_u^2 = 1$, unit-level variance that mimics the two negative binomial scenarios. We generated a total of eight RCB data sets and eight BIB data sets, representing the two block sizes, two distributions and two unit-level scale or variance parameters given above. Table 4 summarizes the count data simulation scenarios.

The treatment effects, τ_i, were set to 0 for all simulated data sets in order to assess type I error rates. The true treatment mean was set to $\lambda_i = \exp(\eta + \tau_i) = 10$, where η denotes the model's intercept. In all scenarios, the block variance was set to $\sigma_b^2 = 0.5$.

4.2. GLMMS USED FOR STATISTICAL ANALYSIS

Because both REML-like PL and quadrature algorithms are both readily available in SAS PROC GLIMMIX and because it has the flexibility to implement all of the models under consideration, this simulation study uses PROC GLIMMIX. In this section, we describe the models used for each scenario. SAS statements for all models used in this paper are available in online supplemental materials.

For count data, we analyzed all data sets using both the negative binomial and Poisson-gamma GLMMs.

Negative binomial GLMM:

Assumed distributions: $y_{ij}|b_j \sim$ Negative binomial (λ_{ij}, ϕ); $b_j \sim N(0, \sigma_b^2)$.

Linear predictor: $\eta_{ij} = \log(\lambda_{ij}) = \eta + \tau_i + b_j$.

Poisson-normal GLMM:

Assumed distributions: $y_{ij}|b_j, u_{ij} \sim$ Poisson (λ_{ij}); $b_j \sim N(0, \sigma_b^2)$; $u_{ij} \sim N(0, \sigma_u^2)$.

Linear predictor: $\eta_{ij} = \log(\lambda_{ij}) = \eta + \tau_i + b_j + u_{ij}$.

In both cases, η denotes the intercept and the other terms and assumptions are specified in Table 3. An alternative form of the Poisson-normal GLMM uses the compound symmetry (CS) structure.

Linear predictor: $\log(\lambda_{ij}) = \eta + \tau_i + u_{ij}$.

Distributions:
$$\begin{bmatrix} u_{1j} \\ u_{2j} \\ u_{3j} \\ u_{4j} \\ u_{5j} \\ u_{6j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}', \begin{bmatrix} \sigma^2 + \sigma_{cs} & \sigma_{cs} & \sigma_{cs} & \sigma_{cs} & \sigma_{cs} & \sigma_{cs} \\ \sigma_{cs} & \sigma^2 + \sigma_{cs} & \sigma_{cs} & \sigma_{cs} & \sigma_{cs} & \sigma_{cs} \\ \sigma_{cs} & \sigma_{cs} & \sigma^2 + \sigma_{cs} & \sigma_{cs} & \sigma_{cs} & \sigma_{cs} \\ \sigma_{cs} & \sigma_{cs} & \sigma_{cs} & \sigma^2 + \sigma_{cs} & \sigma_{cs} & \sigma_{cs} \\ \sigma_{cs} & \sigma_{cs} & \sigma_{cs} & \sigma_{cs} & \sigma^2 + \sigma_{cs} & \sigma_{cs} \\ \sigma_{cs} & \sigma_{cs} & \sigma_{cs} & \sigma_{cs} & \sigma_{cs} & \sigma^2 + \sigma_{cs} \end{bmatrix} \right)$$

$y_{ij}|u_{ij} \sim$ Poisson (λ_{ij})

Assuming $\sigma_{cs} > 0$, the CS covariance, σ_{cs}, and the block variance, σ_b^2, are equivalent, as are the scale variance, σ^2 and the unit-level variance σ_u^2. The advantage of the CS parameterization is that it avoids the problem of type I error rate inflation when a negative solution for the block variance occurs. See Stroup and Littell (2002) for a discussion of this issue. Note that the CS covariance parameterization can be used in conjunction with the Poisson-normal GLMM but not the negative binomial GLMM. In the negative binomial GLMM, the CS scale parameter σ^2 and the negative binomial scale parameter ϕ are confounded. See Stroup (2013, pp. 107–108) for details. See also Stroup et al. (2018) for a discussion of negative variance component solutions with Poisson-normal and negative binomial GLMMs.

Note that all models used in this study are intended to estimate or test differences among $E(\mathbf{y}_i|\mathbf{b})$ the "broad inference space" ith treatment mean of $\mathbf{y}|\mathbf{b}$, i.e., λ_i for count data. This is the expected value of "a typical member of the population" as opposed to the mean of the marginal distribution, $E(\mathbf{y}_i)$. As noted in Gbur, et al. (2011) and Stroup (2013, 2014), the broad inference space $E(\mathbf{y}_i|\mathbf{b})$ is the appropriate target of inference in most academic, basic or discovery research. Some versions of PQL add a mandatory overdispersion parameter, which alters the target of inference from $E(\mathbf{y}_i|\mathbf{b})$ to $E(\mathbf{y}_i)$. This is an important subtlety that should not be overlooked.

4.3. INFORMATION USED FROM STATISTICAL ANALYSIS

For each scenario, the models shown in Sect. 4.2 were applied to the data for each simulated experiment relevant to that model. For example, data sets generated using the Poisson-gamma process were analyzed using the negative binomial and Poisson-normal GLMM, each with the PROC GLIMMIX RSPL and quadrature algorithm. In each run, output data sets were created to track the following information: the estimated treatment means on the data scale, $h(\hat{\eta} + \hat{\tau}_i)$; the estimated block variance (the one variance parameter common to all scenarios); the **rejection rate**, defined the proportion of data sets in which $H_0 : \tau_i = 0$

Table 5. Randomized complete block design. Poisson-gamma data generating process

Scale parameter	N blocks	Analysis model	Estimation method	Rejection rate	Coverage	mean $\hat{\lambda}$	Mean $\hat{\sigma}^2_{blk}$	Proportion converged
$\varphi = 0.25$	5	NB	RSPL	0.062	0.951	10.3	0.50	0.993
			Quad	0.128	0.917	10.4	0.41	0.991
		PN	RSPL	0.051	0.951	9.8	0.49	0.999
			Quad	0.144	0.911	9.6	0.41	0.996
	15	NB	RSPL	0.060	0.953	10.0	0.49	1.000
			Quad	0.068	0.945	10.1	0.46	1.000
		PN	RSPL	0.056	0.944	9.4	0.47	1.000
			Quad	0.077	0.913	9.1	0.46	1.000
$\varphi = 1$	5	NB	RSPL	0.136	0.914	10.3	0.49	0.953
			Quad	0.124	0.908	10.8	0.37	1.000
		PN	RSPL	0.054	0.909	7.8	0.32	0.996
			Quad	0.126	0.819	7.4	0.37	0.995
	15	NB	RSPL	0.122	0.933	9.9	0.49	0.930
			Quad	0.082	0.936	10.4	0.44	1.000
		PN	RSPL	0.050	0.800	7.1	0.42	0.999
			Quad	0.069	0.719	6.6	0.44	1.000

Rejection Rate: proportion of data sets in which H_0 : all λ_i is rejected at $\alpha = 0.05$

Coverage: proportion of 95% confidence intervals for λ_i that contain true mean ($\lambda_i = 10$)

for all $i = 1, 2, ..., 6$ was rejected at $\alpha = 0.05$; **coverage**, defined as the proportion of 95% confidence intervals containing the true mean (i.e., the true $h(\eta + \tau_i)$); **convergence** defined as the proportion of data sets that converged to a solution.

Average treatment mean estimates, mean block variance estimates, percent of data sets in which H_0 was rejected, and percent of data sets in which the 95% confidence interval contained the true mean and convergence proportion were tabulated for each model and each estimation algorithm. These are reported in the following section.

5. RESULTS

This section is divided into three subsections. Section 5.1 contains results for the count data scenarios using the randomized complete block design. Section 5.2 contains the balanced incomplete block count data results. The count data scenarios illustrate most of the issues that characterize the small-sample behavior of REML-like PL and quadrature. Section 5.3 contains selected results from the other distributions to complete the picture within the space constraints of this paper.

5.1. COUNT DATA: RANDOMIZED COMPLETE BLOCK DESIGN

Tables 5 and 6 show results for the RCB count data scenarios. In all tables, as well as the discussion in this section, REML-like PL is referred to by its GLIMMIX acronym RSPL.

For both processes in which the distribution of the observations conditional on the random effects is approximately symmetric ($\phi = 0.25$ for the Poisson-gamma process, $\sigma^2_u = 0.25$

Table 6. Randomized complete block design. Poisson-normal data generating process

Scale parameter	N blocks	Analysis model	Estimation method	Rejection rate	Coverage	mean $\hat{\lambda}$	Mean $\hat{\sigma}^2_{blk}$	Proportion converged
$\sigma^2_u = 0.25$	5	NB	RSPL	0.054	0.932	11.7	0.50	0.994
			Quad	0.104	0.892	11.8	0.40	0.994
		PN	RSPL	0.035	0.994	11.0	0.49	0.999
			Quad	0.115	0.907	10.8	0.41	0.997
	15	NB	RSPL	0.058	0.930	11.4	0.50	1.000
			Quad	0.085	0.908	11.5	0.48	1.000
		PN	RSPL	0.051	0.948	10.7	0.49	1.000
			Quad	0.073	0.939	10.3	0.48	1.000
$\sigma^2_u = 1$	5	NB	RSPL	0.138	0.879	16.2	0.52	0.962
			Quad	0.167	0.835	16.9	0.41	1.000
		PN	RSPL	0.039	0.950	12.0	0.48	0.997
			Quad	0.106	0.917	11.7	0.41	0.995
	15	NB	RSPL	0.127	0.765	15.5	0.51	0.939
			Quad	0.139	0.709	16.2	0.48	1.000
		PN	RSPL	0.051	0.948	11.2	0.47	1.000
			Quad	0.064	0.941	10.6	0.45	0.998

for the Poisson-normal process) RSPL consistently outperforms quadrature with respect to all criteria evaluated. With 1000 simulated experiments, type I error rate should be between 0.03 and 0.07 and confidence interval coverage should be between 0.93 and 0.97. RSPL meets these criteria in all cases—for designs with five or 15 blocks, regardless of whether the model is correct (e.g., the negative binomial GLMM is fit to Poisson-gamma data) or misspecified (e.g., the negative binomial model is fit to Poisson-normal data). For designs with five blocks, quadrature consistently shows elevated type I error rate (consistently greater than 0.10) and confidence interval coverage below 0.93. Quadrature's performance improves somewhat for designs with 15 blocks, but only if the model is correctly specified.

Variance component estimates are what one would expect. With five blocks, there are a total of 30 observations ($N = 30$) and 24 residual DF. The RSPL estimates of σ^2_B average close to 0.5 and the quadrature estimates, which are MLEs, are close to 0.4, matching the ratio of N to the residual DF. With 15 blocks, ($N = 90$) and residual DF is 84. RSPL estimates remain close of 0.5, but quadrature estimates are close to 0.47—the value one would expect given the ratio of N to residual DF. Note that the discrepancy between RSPL estimates and the ML estimates from quadrature result from the discrepancy between N and the residual DF. If there were more treatments relative to the number of observations, this discrepancy would increase and in turn further compromise quadrature's performance relative to RSPL with respect to type 1 error control and confidence interval coverage.

Convergence is not an issue for either method, regardless of the GLMM fit or the number of blocks. This changes when $\mathbf{y}|\mathbf{b}$ is right-skewed ($\phi = 1$). See below.

Using the correct or misspecified model has negligible effect on quadrature's performance for the cases with five blocks. Increasing replication to 15 blocks reduces the bias inherent in maximum likelihood variance component estimates and their resulting inferential statistics—*provided the GLMM and the data-generating process agree*. However, increas-

ing replication does not improve quadrature's type I error control or coverage when the negative binomial GLMM is fit to Poisson-normal data or the Poisson-normal GLMM is fit to Poisson-gamma data. Quadrature does not appear to be robust to model misspecification, whereas RSPL does.

For right-skewed processes ($\phi = 1$ for the Poisson-gamma, $\sigma_u^2 = 1$ for the Poisson-normal) with the five block designs, quadrature's type I error rate and confidence interval coverage are similar to what they were with the approximately symmetric $\mathbf{y}|\mathbf{b}$ distribution. RSPL's type I error rate and confidence interval coverage are similar to quadrature when negative binomial GLMMs are fit, regardless of whether the generating process is Poisson-gamma or Poisson-normal. However, RSPL's type I error control remains within the 0.03 and 0.07 limits when the Poisson-normal GLMM is used, regardless of the generating process. Confidence interval coverage for RSPL with the Poisson-normal GLMM is within the 0.93 to 0.97 limits for the Poisson-normal generating process, but coverage is 0.909 for Poisson-gamma process.

With 15 blocks, quadrature's performance improves provided the GLMM is consistent with the generating process—i.e., negative binomial GLMM fit to Poisson-gamma data or Poisson-normal model fit to Poisson-gamma data. For these cases, type I error rate is below 0.07 and confidence interval coverage is above 0.93. If the model is misspecified, quadrature performance improves relative to five blocks, but type I error rate is still greater than 0.07 and coverage is less than 0.93. With 15 blocks, RSPL's performance is within type I error limits for all cases except when fitting the negative binomial GLMM to Poisson-normal data. Coverage is consistently somewhat better than quadrature.

There is one case, fitting a negative binomial GLMM to data with a strongly right-skewed distribution of $\mathbf{y}|\mathbf{b}$, in which neither quadrature nor RSPL perform well—both have elevated type I error rates and poor coverage. Quadrature improves somewhat with 15 blocks, but its performance is still problematic.

For the processes in which the distribution of $\mathbf{y}|\mathbf{b}$ is right-skewed, convergence is a non-issue for all GLMMs fit with quadrature, regardless of the number of blocks. Convergence is only an issue for RSPL when fitting the negative binomial GLMM. With five blocks, the convergence proportion is around 0.96 and drops to around 0.94 with 15 blocks.

Notice that in two cases with 15 blocks, fitting a negative binomial GLMM to Poisson-normal data and fitting a Poisson-normal GLMM to Poisson-gamma data, there is a large discrepancy between the true λ_i, which is 10, and the average of their estimates. For the negative binomial GLMM fit to Poisson-normal data, $\hat{\lambda}_i$ averages 15.5 with RSPL and 16.2 with quadrature. For the Poisson-normal GLMM fit to Poisson-gamma data, $\hat{\lambda}_i$ averages 7.1 with RSPL and 6.6 with quadrature. Coverage suffers accordingly. The discrepancy results from the fact that the negative binomial GLMM and Poisson-normal GLMM target different characteristics of the marginal distribution of $\mathbf{y}$. As described in Stroup (2013), both generating processes produce strongly right-skewed marginal distributions of $\mathbf{y}$. Stroup's Chapter 3 (pp. 99–107) discusses the impact on the resulting inference space for GLMMs in general terms. Chapter 11 (pp. 356–60) provides additional specifics relevant to count data. The marginal distributions of $\mathbf{y}$ produced by the Poisson-normal and Poisson-gamma processes, while intractable analytically, can be characterized empirically by simulation. One can then demonstrate by simulation that Poisson-normal GLMM's broad inference space

Table 7.   Balanced incomplete block design. Poisson-gamma data generating process

Scale parameter	N blocks	Analysis model	Estimation method	Rejection rate	Coverage	Mean $\hat{\lambda}$	Mean $\hat{\sigma}^2_{blk}$	Proportion converged
$\varphi = 0.25$	10	NB	RSPL	0.042	0.947	10.2	0.49	0.988
			Quad	0.116	0.932	10.2	0.46	0.987
		PN	RSPL	0.036	0.950	9.7	0.48	0.999
			Quad	0.150	0.918	9.4	0.46	0.986
	30	NB	RSPL	0.053	0.947	10.0	0.49	0.998
			Quad	0.058	0.944	10.1	0.48	1.000
		PN	RSPL	0.041	0.937	9.4	0.47	1.000
			Quad	0.075	0.906	9.1	0.48	1.000
$\varphi = 1$	10	NB	RSPL	0.122	0.903	9.9	0.49	0.877
			Quad	0.115	0.912	10.7	0.39	0.999
		PN	RSPL	0.033	0.913	7.6	0.42	1.000
			Quad	0.128	0.848	7.2	0.44	1.000
	30	NB	RSPL	0.126	0.899	9.4	0.50	0.840
			Quad	0.084	0.935	10.3	0.44	1.000
		PN	RSPL	0.045	0.787	7.0	0.42	1.000
			Quad	0.073	0.700	6.5	0.46	1.000

estimate of λ targets the median of the marginal distribution and the negative binomial targets a value somewhere between the median and the marginal mean. Thus, fitting a Poisson-normal GLMM to Poisson-gamma data will produce $\hat{\lambda}_i$ below λ_i and fitting a negative binomial GLMM to Poisson-normal data will do the opposite. The effect is exacerbated by a large-scale parameter value (e.g., $\phi = 1$ for the Poisson-gamma or $\sigma^2_u = 1$ for the Poisson-normal) and by increasing replication.

5.2. Count Data: Balanced Incomplete Block Design

Table 7 shows results for count data with the balanced incomplete design when the data generating process is Poisson-gamma. Table 8 shows count BIB results for the Poisson-normal data generating process. Overall, the results are similar to those with the RCB reported in the previous section. The main differences are the type I error performance and coverage of quadrature and the convergence rate of RSPL fitting the negative binomial GLMM when the data generating process is strongly right-skewed ($\phi = 1$ for the Poisson-gamma data and $\sigma^2_u = 1$ for the Poisson-normal data).

As with the five-block RCB, the ten-block BIB quadrature type I error rates are consistently well above 0.07 and coverage well below 0.93. When replication is increased to 15 with the RCB, quadrature type I error rate improves to near or below 0.07 and coverage to near or above 0.93 for GLMMs that match the generating process. With the 30-block BIB, quadrature's type I error rate does not improve as much, with type I error rates remaining above 0.07 and coverage below 0.93.

For the negative binomial GLMM fit with RSPL to the highly skewed cases, convergence was between 0.94 and 0.96 for the RCB. It decreases to below 0.9 for the BIB. This suggests that the latter convergence rate would be expected for RCBs with missing data, and the

Table 8. Balanced incomplete block design. Poisson-normal data generating process

Exp unit variance	N blocks	Analysis model	Estimation method	Rejection rate	Coverage	mean $\hat{\lambda}$	Mean $\hat{\sigma}^2_{blk}$	Proportion converged
$\sigma^2_{eu} = 0.25$	10	NB	RSPL	0.062	0.939	11.4	0.50	0.984
			Quad	0.134	0.915	11.5	0.46	0.991
		PN	RSPL	0.052	0.950	10.8	0.49	1.000
			Quad	0.155	0.921	10.6	0.47	1.000
	30	NB	RSPL	0.060	0.917	11.2	0.49	0.997
			Quad	0.082	0.906	11.3	0.49	1.000
		PN	RSPL	0.043	0.942	10.6	0.47	1.000
			Quad	0.084	0.939	10.2	0.49	1.000
$\sigma^2_{eu} = 1$	10	NB	RSPL	0.160	0.863	15.5	0.54	0.852
			Quad	0.182	0.845	16.7	0.47	1.000
		PN	RSPL	0.052	0.951	11.9	0.49	1.000
			Quad	0.146	0.918	11.5	0.48	0.997
	30	NB	RSPL	0.160	0.772	14.7	0.55	0.815
			Quad	0.118	0.730	15.7	0.56	1.000
		PN	RSPL	0.057	0.941	11.1	0.47	1.000
			Quad	0.087	0.939	10.5	0.50	1.000

convergence rate for RSPL might be even lower for unbalanced incomplete block designs when the distribution of **y**|**b** is highly skewed.

To summarize, aside from these minor differences, the overall behavior of RSPL relative to quadrature is similar for complete and incomplete block scenarios.

5.3. OTHER DISTRIBUTIONS

Given the similarity of the BIB and RCB results for count data, simulations for the binomial, beta, and gamma generating processes are limited to the BIB case. Results for these scenarios can be obtained in the online supplementary materials. The main overview comment is that the results are consistent with those reported in the above sections for count data.

6. CONCLUSIONS AND FUTURE WORK

Based on these combined results, the conjecture described in Sect. 3 appears to hold. ML versus REML matters, as does the shape of the distribution of **y**|**b**. Quadrature's downward-biased variance MLEs result in inflated type I error rates and poor coverage in designs with minimal replication, but improve—provided the analysis GLMM is consistent with the data-generating process—when replication is increased relative to the number of treatments. REML-like PL (PROC GLIMMIX acronym RSPL) produces type I error rates and confidence interval coverage consistent with nominal α level and confidence coefficient, even if the model is misspecified, provided the distribution of **y**|**b** is at least approximately symmetric. RSPL appears to perform acceptably unless the distribution of **y**|**b** is very strongly skewed. Quadrature performs well in scenarios with ample replication and when the GLMM

Table 9.   Recommended method

	Small S (few e.u./treatment)	Large S (many replications)
$f(y|b)$ skewed	Quadrature is least bad—*provided the GLMM fit and the process by which the data arise are well matched*	Use quadrature—*but take care to match GLMM to data generating process to the extent possible—model misspecification matters*
$f(y|b)$ at least approximately symmetric	Use RSPL	Use RSPL

being fit is well matched to the data generating process. Quadrature struggles when replication is minimal and when the GLMM is misspecified—even slightly as the Poisson-normal and Poisson-gamma cases illustrate.

Given these results, Table 9 expands on Table 2 for our recommendations for analysis method for any GLMM.

Although quadrature's performance improves with increased replication, RSPL is still preferred when the distribution of **y**|**b** appears to be reasonably symmetric for three reasons. First, RSPL is more forgiving of model misspecification than quadrature. Second, RSPL is less computationally demanding than quadrature. Third, RSPL can be used with much more complex covariance structures, e.g., those that arise in multi-level or repeated measures designs. Admittedly, the characterizations of distribution shape and replication in Table 9 are not exact. If there is any doubt, we encourage data analysts to run simulations similar to those illustrated in this paper to decide if REML-like PL or quadrature (or Laplace) is more appropriate given the characteristics of the data to be analyzed.

These recommendations point to an obvious need for work on diagnostics to identify model misspecification. In our experience, methods used with Gaussian data such as residual plots and Q–Q plots are inadequate to identify differences in data generating processes with non-Gaussian data. See Stroup et al. (2018) for a discussion of this issue and some suggestions of how to handle model specification pending the development of more informative diagnostics.

We also emphasize that the issues discussed here, all obtained from simulations with equal treatment means, are exacerbated when treatment means are unequal. In all of the distributions presented, variances are functions of means. Therefore, when the treatment means are different, heterogeneity of variance becomes a factor and affects both confidence interval coverage and power characteristics. Space limits this paper to the equal means case. Suffice it to say, however, that if a procedure cannot control type I error, as is so often the case with quadrature estimation, it should not be used.

Analysts need also to be aware of the default fitting method in their software of choice. LME4 in R uses integral approximation—and hence ML variance estimation—only. LME4 has no REML-like option. The closest packages R users have pseudo-likelihood are glmm-PQL and HGLMMM. However, glmmPQL only implements the ML version of PQL and it imposes a scale parameter (analogous to RANDOM _ RESIDUAL_ in PROC GLIMMIX) that is often superfluous, alters the inference space and cannot be turned off. For example,

glmmPQL cannot fit either the Poisson-normal or negative binomial GLMM as defined in this paper. Molas and Lesaffre (2011) state that HGLMMM uses a REML-like objective function to estimate variance parameters. However, it too imposes a scale parameter that creates the same problems as glmmPQL. SAS/STAT PROC GLIMMIX allows users the option of RSPL (the default), MSPL, quadrature or Laplace integral approximation. All analyses described in this paper can be implemented using PROC GLIMMIX.

Readers should note that the REML-like PL versus quadrature issue is not limited to generalized linear mixed models; they apply to nonlinear mixed models (NLMMs). The primary tools for fitting NLMMs in R and SAS, respectively, are the NLME package and PROC NLMIXED. Both use adaptive quadrature. Littell et al. (2006) discuss %NLINMIX, a macro that works in conjunction with PROC MIXED to fit NLMMs using PL. This macro was widely used before the NLME package and PROC NLMIXED were introduced. Future work will address the question, "Is quadrature always the method of choice for NLMMs, and if not, when would REML-like PL be a better choice?".

Finally, the results presented in this paper may have implications for Bayesian analysis. Bayesian methods are likelihood-based—that is, estimation uses a full likelihood and a prior to obtain a posterior distribution. This applies to variance components as well as other model parameters. What happens if the likelihood used for the analysis and the process by which the data arise are not well matched? To what extent are the same issues presented in this paper in play with Bayesian inference? Although this is a topic beyond the scope of this paper, it is something that data analysts using Bayesian methods should consider.

The data analysis for this paper was generated using SAS/STAT software, version 9.4 (TS1M6), of the SAS System for Windows. Copyright ©2016 SAS Institute Inc., SAS, and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

REFERENCES

Bolker, B.M., Brooks, M.E., Clark, C.J. Geange, S.M., Poulsen, J.R., Stevens, H.H, and White, J.S. (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution.* **24**:3: 127–135.

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* 81:81–91.

Claassen, E. A. (2014). *A Reduced Bias Method of Estimating Variance Components on Generalized Linear Mixed Models*. PhD Dissertation. Lincoln, NE: University of Nebraska.

Gbur, E.E., Stroup, W.W., McCarter, K.S., Durham, S., Young, L.J., Christman, M., West, M., and Kramer, M. (2012). *Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences*. Madison, WI.: American Society of Agronomy, Soil Science Society of America, Crop Science Society of America.

Harville, D. A. (1977). "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." Journal of the American Statistical Association 72:320–338.Hilbe, J. M. (2014). *Modeling Count Data*. Cambridge: Cambridge University Press.

Molas, M. and Lesaffre, E. 2011. Hierarchical generalized linear models – the R package HGLMMM. *Journal of Statistical Software.* 39:13.

Murray, D. M., Varnell, S. P., and Blitstein, J. L. (2004). "Design and Analysis of Group-Randomized Trials: A Review of Recent Methodological Developments." *American Journal of Public Health* 94:423–432.

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.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

SAS Institute Inc. (2019) *User's Guide for SAS/STAT 14.3*. Cary, NC: SAS Institute Inc. SAS/STAT 14.3. Cary, NC: SAS Institute Inc.

Schabenberer, O. and Stroup, W.W. W. (2008). Generalized Linear Mixed Models: Theory and Applications. Short Course, 2008 Joint Statistical Meetings.

Stroup, W.W. (2013) *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. Boca Raton, FL: CRC Press.

Stroup, W.W. (2014) Rethinking the analysis of non-normal data in plant and soil science. *Agronomy Journal* 106:1–17.

Stroup, W.W. and Littell, R.C. (2002). "Impact of Variance Component Estimates on Fixed Effect Inference in Unbalanced Linear Mixed Models. *Conference on Applied Statistics in Agriculture*. https://doi.org/10.4148/2475-7772.1198

Stroup, W., Milliken, G.A., Claassen, E.A. and Wolfinger, R.D. (2018) *SAS for Mixed Models: Basic Concepts and Applications.* Cary, NC: SAS Institute Inc.

Wolfinger, R. D., and O'Connell, M. A. (1993). "Generalized Linear Mixed Models: A Pseudo-likelihood Approach." *Journal of Statistical Computation and Simulation* 48:233–243.