Aquí va el título del póster
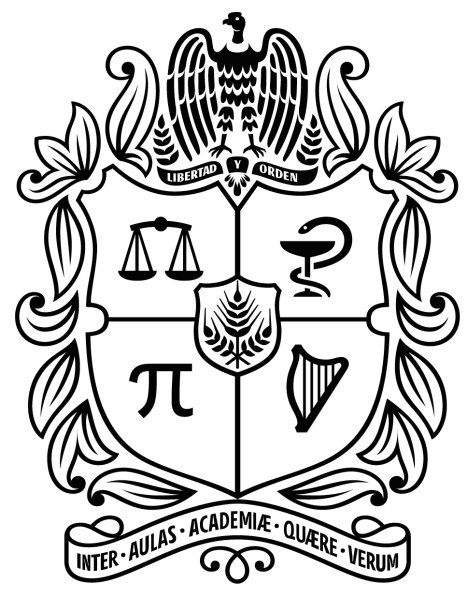
Fulanito de tal 1
Fulanito de tal 2
Fulanito de tal 3

Universidad Nacional de Colombia
Universidad Pontificia Bolivariana
Universidad del Norte

`yreod@unal.edu.co, wusjf@udea.edu.co, dyuej@uninorte.edu.co`

Resumen

Clustered data are, nowadays, available in multiple disciplines. Statistical modeling of these data is usually performed using Generalized Linear Mixed Models (GLMMs), proposed by Breslow & Clayton (1993), to account for the hierarchical structure in the data. One problem of interest is to determine whether, in the fitted model, the variance component (σ^2) of the random effects is statistically zero (that is, the random effect does not explain much of the variance and should be excluded from the model). In the literature, this feature is assessed using a variance component test. Here we present the results of a statistical simulation study assessing the performance of two variance component tests in the context of GLMMs, when a binary and Gamma response variables are considered. In particular, the likelihood ratio (LR) and the permutation tests are evaluated, and the impact of misspecifying the true distribution of the random effects is measured as a function of the number of individuals per cluster and the true value of σ^2.

1. Introduction

Clustered data are commonly collected in studies on medical, social, and behavioral sciences. This type of data arise when there is a hierarchical or clustered structure in the data such as individuals nested in institutions or organizations (i.e., students in schools, employees in firms, or patients in hospitals). Mixed models, hierarchical models or multilevel regression models provide an attractive framework to accommodate the overdispersion and dependence of this type of data (Zhu & Zhang, 2006). Many of the models used in these fields fall under the frame of Generalized Linear Mixed Models (GLMMs). A fundamental question in this models is about the heterogeneity among clusters, which is equivalent to test whether σ^2, the variance component associated to the random effects, is statistically zero. This test is known as a *variance component test*, and can be approached using the likelihood ratio (LR) and permutation tests (LRs), and have important implications in statistical modeling of clustered data. Here we use a statistical simulation approach to determine, which of the aforementioned tests is more appropriate to determine whether σ^2 is effectively zero. To do this, different statistical distributions and several prespecified values of σ^2 are utilized to generate the vector of random effects, followed by the underlying data to be modeled.

2. GLMMs

GLMMs, proposed by Breslow & Clayton (1993), have been extensively in many applications where clustered and/or longitudinal data are available. Let y_{ij} the jth response variable within the ith cluster ($i = 1, 2, \ldots, m; j = 1, 2, \ldots, n_i$). In a GLMMs with random intercept, it is assumed that, conditional to the random effects b_i, the outcome y_{ij} is independent with the following structure:

$$
\begin{aligned}
y_{ij} \mid b_i &\sim \text{independent in } F_y, \\
g(\mu_{ij}) &= X_{ij}\beta + b_i, \\
b_i &\overset{ind}{\sim} N(0, \sigma^2),
\end{aligned}
\tag{1}
$$

where F_y belongs to the exponential family, $g(\cdot)$ is a known link function, X_{ij} is the vector of covariates for the jth observation in the i cluster, and β is the parameter vector for the fixed effects.

3. Variance component test

Among the tests available in the literature for testing

$$
H_0 : \sigma^2 = 0 \text{ vs. } H_A : \sigma^2 > 0,
\tag{2}
$$

where σ^2 is the variance of random intercept, the LR test is the most commonly used test because of its theoretical properties and straightforward construction. In a GLMM with random intercept, we are interested in testing H_0 in

(2) using a type I error probability α. The statistic for the LR test is calculated as

$$
T = -2 \log\left(\frac{L_0}{L_A}\right)
\tag{3}
$$

where L_0 and L_A are the likelihood model under H_0 and H_A, respectively. Under H_0, the asymptotic null distribution of T is a GLMM with random intercept is a 50:50 mixture between a χ_0^2 and χ_1^2 distributions (Zhang & Lin, 2008).

The permutation test (PT) is a modification of LR test proposed by Fitzmaurice & Lipsitz (2007). The idea behind the PT is that, when H_0 in (2) is true, the heterogeneity between clusters is non-existing. This result implies that we could mix the clusters without changing the decision. In what follows we present the implementation of the PT given by Fitzmaurice & Lipsitz (2007).

1. Calculate the LR test statistic in the original sample and denote it by T_{obs}.
2. Permute the cluster indexes while holding fixed the number of units within a cluster, n_i, and calculate the LR test statistic T.
3. Repeat step 2, k times, to obtain $T_1, T_2, \ldots, T_k$, where T_k is the LR statistic in (3) for the kth permutation of the original clusters.
4. Determine the PT p-value as the proportion of permutation samples where $T_i \geq T_{obs}$ for $i = 1, 2, \ldots, k$.

4. Simulation study

In order to compare the two variance component tests, we consider a GLMM with random intercept and a response variable following a binary distribution. In both cases, the random intercept was sampled from four true statistical distributions (Normal, Log-Normal, Exponential and Uniform). For the fitting procedure, normality of the random intercept was assumed. The model considered in this case can be summarised as follows:

$$
\text{logit}\{P(y_{ij} = 1 \mid b_i)\} = \beta_0 + \beta_b x_1 + \beta_w x_2 + b_i,
\tag{4}
$$

where $i = 1, 2, \ldots, m$ represents the cluster and $j = 1, 2, \ldots, n_i$ represents the number of observations per cluster. The between-cluster covariate $x_1 \sim \text{Poisson}(\lambda = 2)$ and x_2 is a within-cluster covariate following a $U(0, 1)$ distribution. The true model parameters are $\beta_0 = -2.5$, $\beta_b = 2$ and $\beta_w = 1$.

The random effects b_i in model (4) were generated from normal, uniform, exponential and log-normal distributions. For exponential and log-normal case, the random values were transformed to ensure the zero mean and variance σ^2. In the next figure we show the distributions used to generate the random intercepts for the case $\sigma^2 = 1$.
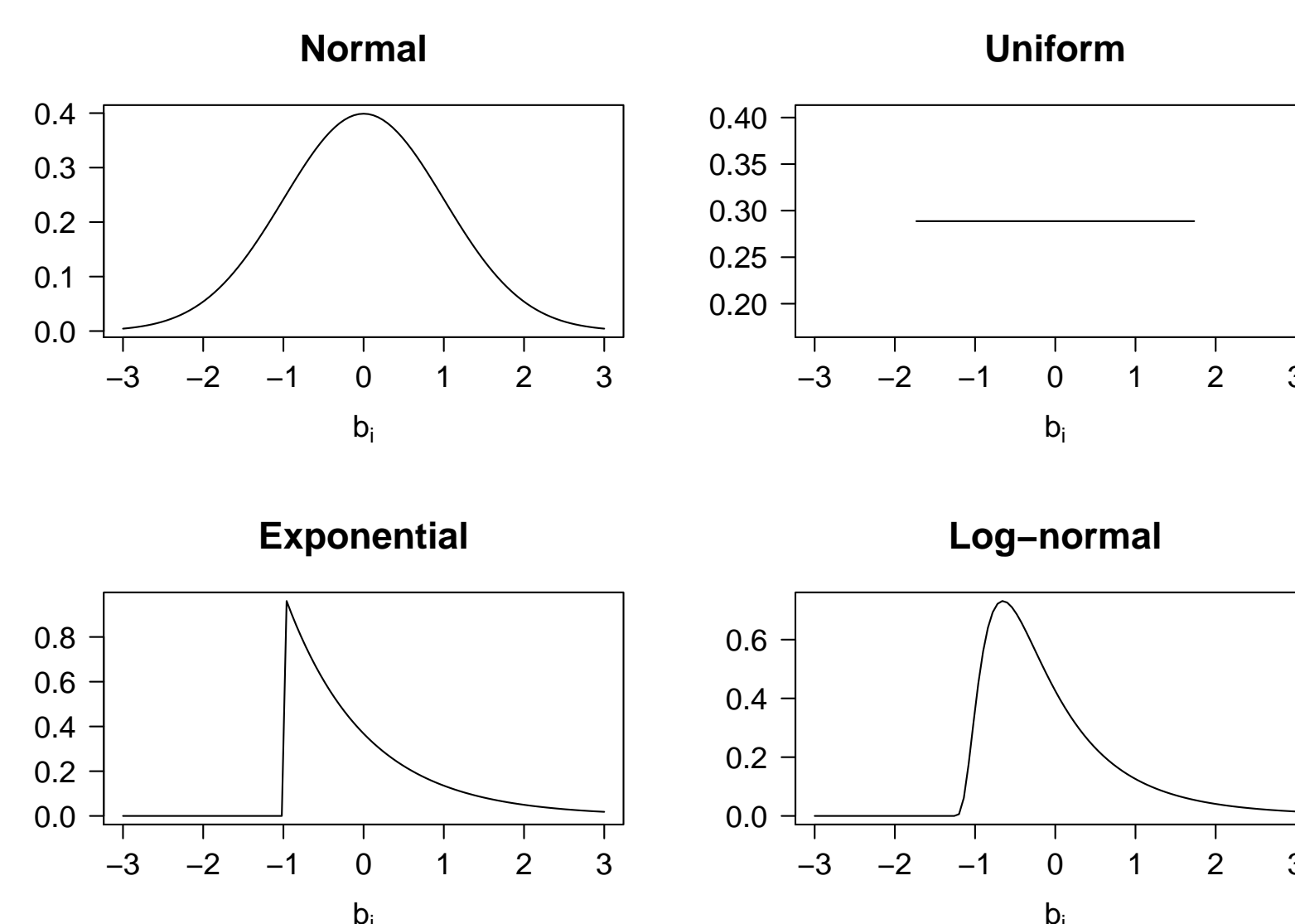

Figura 1: *Distributions for the random effect considered to generate b_i. In this figure the mean of random effects was zero with $\sigma^2 = 1$.*

5. Results

Figure 2 displays the main results for the binary GLMM outlined in expression (4). Overall, the average rejection rate (ARR) of H_0 in (2) of the LR test and the PT are affected for the number of individuals per cluster (parameter n_i in our simulation approach) and the pre-specified value of σ^2, but have a similar behaviour regardless of the true distribution of the random effects. To directly compare the ARRs between the LR test and PT, the ratio

$$
\gamma = \frac{\text{ARR}_{\text{LR}}}{\text{ARR}_{\text{PT}}}
\tag{5}
$$

was calculated. Here, values of $\gamma > 1$ indicate that the LR test outperforms the PT; values of $\gamma < 1$ indicate that the LR test outperforms the PT, and $\gamma = 1$ indicate that the LR test and PT produce equivalent rejection rates. When the number of individuals per cluster is small (that is, $n_i < 5$), the PT has a higher ARR than the LR test (see third column in Figure 2). Our results show that the LR test and PT are not significantly affected by any of the aforementioned parameters, but that the PT outperforms the LR test. This result has important implications when modelling clustered and/or longitudinal data.
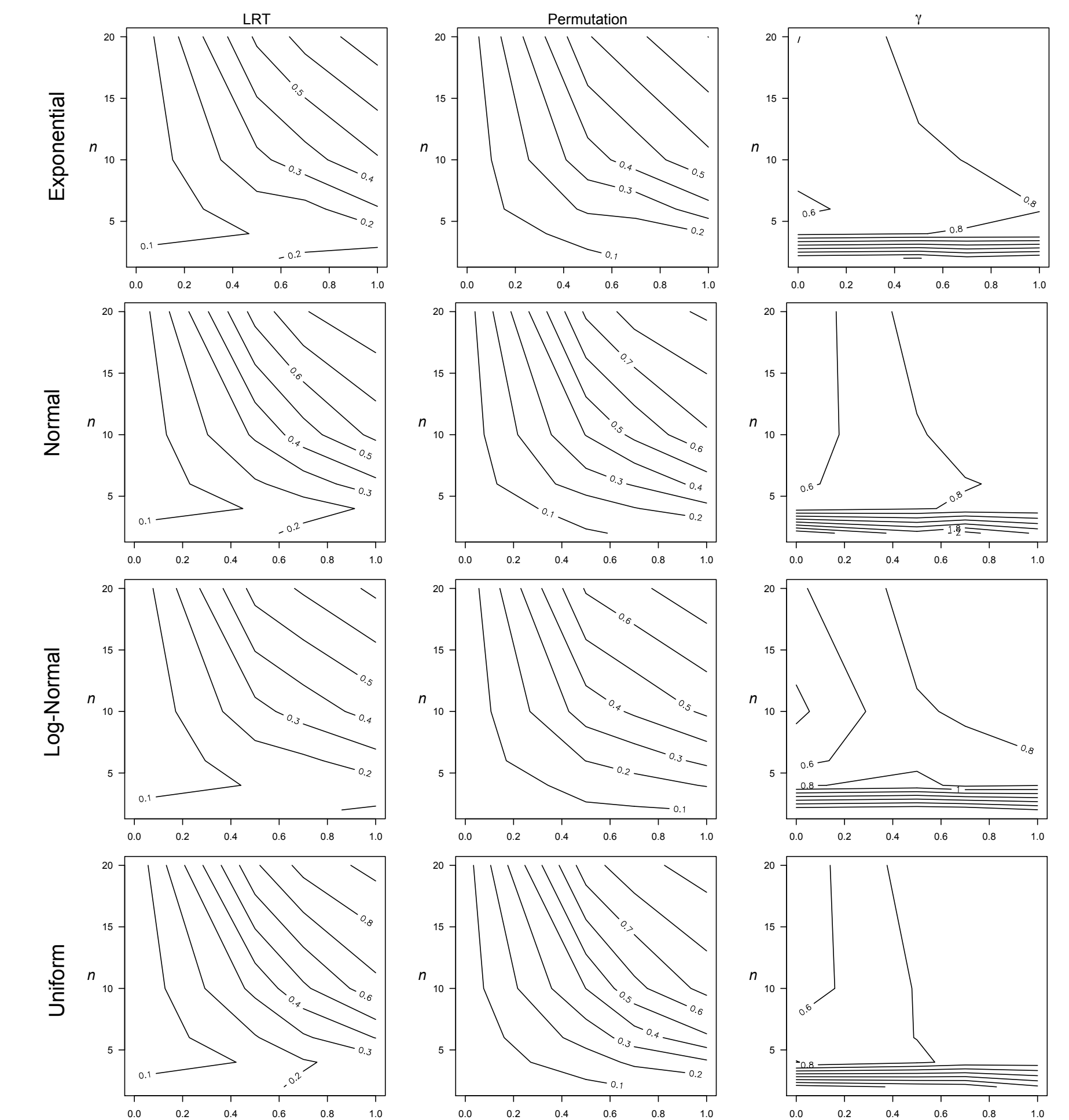

Figura 2: *ARR of $H_0 : \sigma^2 = 0$ in (2) in the binary GLMM as a function of n_i and σ^2 for each distribution of the random effects. The third column corresponds to γ.*

6. Conclusions

We found that LR and permutation tests have similar ARR patterns regardless the change in the true random effect distribution. We found that ARR increases as n and/or σ^2 increase.

References

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9−25.

Fitzmaurice, G.M. and Lipsitz, S.R. (2007). A Note on Permutation Tests for Variance Components in Multilevel Generalized Linear Mixed Models. *Biometrics*, **63**, 942−946.

Zhang, D. and Lin, X. (2008). *Variance Component Testing in Generalized Linear Mixed Models for Longitudinal/Clustered Data and other Related Topics* in Dubson, D.B. (2008) Random Effect and Latent Variable Model Selection. New York: Springer.