

Evaluation of Third-Order Method for the Tests of Variance Component in Linear Mixed Models

Yanyan Wu¹, Augustine Wong², Georges Monette², Laurent Briollais³

¹Department of Public Health Sciences, University of Hawaii at Manoa, Honolulu, HI, USA

²Department of Mathematics and Statistics, York University, Toronto, ON, Canada

³Lunenfeld-Taunenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada

Email: yywu@hawaii.edu, august@yorku.ca, georges@yorku.ca, laurent@lunenfeld.ca

Received 21 March 2015; accepted 18 May 2015; published 22 May 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Mixed models provide a wide range of applications including hierarchical modeling and longitudinal studies. The tests of variance component in mixed models have long been a methodological challenge because of its boundary conditions. It is well documented in literature that the traditional first-order methods: likelihood ratio statistic, Wald statistic and score statistic, provide an excessively conservative approximation to the null distribution. However, the magnitude of the conservativeness has not been thoroughly explored. In this paper, we propose a likelihood-based third-order method to the mixed models for testing the null hypothesis of zero and non-zero variance component. The proposed method dramatically improved the accuracy of the tests. Extensive simulations were carried out to demonstrate the accuracy of the proposed method in comparison with the standard first-order methods. The results show the conservativeness of the first order methods and the accuracy of the proposed method in approximating the p -values and confidence intervals even when the sample size is small.

Keywords

Family Data, Genetic Variant, Likelihood Ratio Test, Random Effects, Third-Order Method, Variance Component

1. Introduction

Mixed models are a powerful tool with a wide range of applications including hierarchical modeling and longitudinal studies. In a variety of applied statistical problems, there is a need for inference on variance component,

How to cite this paper: Wu, Y.Y., Wong, A., Monette, G. and Briollais, L. (2015) Evaluation of Third-Order Method for the Tests of Variance Component in Linear Mixed Models. *Open Journal of Statistics*, 5, 233-244.

<http://dx.doi.org/10.4236/ojs.2015.54025>

for example, the random effects ANOVA models, clustering [1] and homogeneity in stratified analyses [2].

To introduce our ideas, we focus on a simple but generic setting of random intercept mixed model:

$$y_{ij} = x'_{ij}\beta + b_i + \varepsilon_{ij}, \quad (1)$$

where y_{ij} is the response for member $j = 1, \dots, n_i$ of group $i = 1, \dots, m$, x_{ij} is a vector of known covariates, β is a vector of fixed effects, and $b_i \sim N(0, \sigma_b^2)$ is the group specific random effect, assumed to be independently distributed from the residual error $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. Assembling the y_{ij} into a vector Y and the rows x'_{ij} into a fixed effect regressor matrix X , the marginal distribution of response Y is:

$$Y \sim N(X\beta, \sigma_\varepsilon^2 \Delta),$$

in which $\Delta = I_N + (\sigma_b^2 / \sigma_\varepsilon^2) ZZ' = I_N + \lambda ZZ'$, N is the total sample size and Z is the random effect regressor matrix. The parameter λ can be considered as a signal-to-noise ratio since σ_b^2 determines the size of the “signal” given by $E(Y|b) = X\beta + Zb$. This λ is our parameter of interest.

The inference for variance component in mixed models has long been a methodological challenge because the variance component is restricted to nonnegative values, one typically needs one-sided tests of the null hypothesis especially when variance component is on the boundary, *i.e.*:

$$H_o : \lambda = 0 (\sigma_b^2 = 0) \quad \text{versus} \quad H_a : \lambda > 0 (\sigma_b^2 > 0).$$

Using theory originally developed by Chernoff [3], Liang and Self [4] and Stram and Lee [5] proved that the likelihood ratio test for testing a zero variance component has approximately a $0.5\chi_0^2 + 0.5\chi_1^2$ mixture distribution under the null hypothesis where χ_0^2 represents a distribution with a point mass at 0. The p-value for the above one-sided test can be obtained from the mixture Chi-square distribution or equivalently from the standard normal distribution.

The restricted or residual maximum likelihood method (REML), proposed by Patterson and Thompson [6], takes into account the loss of the degrees of freedom involved in estimating the fixed effects and reduces the downward bias of the maximum likelihood estimates of the variance component. However, simulation studies show that the likelihood ratio test based with the restricted maximum likelihood estimation is still conservative [7].

In this paper, we consider the problem of testing the null hypothesis of zero and non-zero variance component in a linear mixed model and show the conservativeness of the first-order methods: likelihood ratio test, Wald test and score test. The likelihood-based third-order method by Fraser *et al* [8] is proposed to improve the accuracy of the inference. Our simulation studies show that the proposed method provides accurate approximation of p -values and confidence intervals even when the sample size is small. The likelihood ratio test approach performs much better than the Wald test and score test. One of the key contributions of this paper is that it gives the first detailed investigation of the accuracy of the tests of variance component in linear mixed models by various methods through extensive simulations with varying number of groups and group sizes.

We describe the general setting of the likelihood-based third-order method in Section 2. In Section 3, the method is applied to the random intercept model and then extended to the random intercept and random slope mixed model. We apply the methods discussed in this paper to a real data set from Genetic Analysis Workshop (GAW 18) in Section 4. Simulation results to compare the accuracy of the methods discussed in this paper are presented in Section 5. Section 6 provides a conclusion and discussion.

2. Likelihood-Based Third-Order Method

We consider inference for a scalar parameter of interest $\psi = \psi(\theta)$ of a continuous statistical model with probability density function $f(y; \theta)$, in which θ is a vector parameter. Let $\ell(\theta) = \ell(\theta; y)$ be the observed log likelihood function. The two commonly used likelihood-based first-order methods are the Wald test and the signed log-likelihood ratio test. The Wald statistic for the parameter of interest is:

$$q = \frac{\hat{\psi} - \psi}{\hat{\sigma}_\psi} \quad (2)$$

where $\hat{\psi} = \psi(\hat{\theta})$, $\hat{\theta}$ is the maximum likelihood estimate (MLE) of θ satisfying $\partial \ell(\theta) / \partial \theta|_{\theta=\hat{\theta}} = 0$, ψ is the

hypothesized value of $\psi(\theta)$ and $\hat{\sigma}_\psi^2$ is an approximated variance of $\hat{\psi}$, which can be obtained by the Delta method:

$$\hat{\sigma}_\psi^2 = \psi_{\theta'}(\hat{\theta}) j_{\theta\theta'}^{-1}(\hat{\theta}) \psi_{\theta}(\hat{\theta})$$

where

$$\psi_{\theta}(\hat{\theta}) = \frac{\partial \psi(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}}, \quad j_{\theta\theta'}(\hat{\theta}) = -\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta=\hat{\theta}}.$$

The Wald statistic follows asymptotically a standard normal distribution with first-order accuracy, $O(n^{-1/2})$. Hence the $100(1-\alpha)\%$ confidence interval for ψ can be approximated by $(\hat{\psi} - z_{\alpha/2} \hat{\sigma}_\psi, \hat{\psi} + z_{\alpha/2} \hat{\sigma}_\psi)$ where z_{α} is the $(1-\alpha)100$ th percentile of the standard normal distribution.

The signed log-likelihood ratio statistic takes the form:

$$r = r(\psi) = \text{sign}(\hat{\psi} - \psi) \left[2 \left\{ \ell(\hat{\theta}) - \ell(\tilde{\theta}) \right\} \right]^{1/2} \quad (3)$$

where $\tilde{\theta}$ is the constrained maximum likelihood estimator of θ that maximizes $\ell(\theta)$ subject to $\psi(\theta) = \psi$. Note that r also has a standard normal distribution with first-order accuracy. Hence, the $100(1-\alpha)\%$ confidence interval for ψ is $\{\psi : |r(\psi)| \leq z_{\alpha/2}\}$.

In recent years, various adjustments to the first-order methods have been proposed. In this paper, we consider the modified signed log-likelihood ratio statistic r^* , introduced by Barndorff-Nielsen [9] [10], which takes the form

$$r^* = r^*(\psi) = r - \frac{1}{r} \log \frac{r}{Q}, \quad (4)$$

where r is the signed log-likelihood ratio statistic as defined in Equation (3), and Q is a quantity that requires the existence of an ancillary statistic. When Q exists, Barndorff-Nielsen [9] [10] showed that r^* is asymptotically distributed as the standard normal distribution with third-order accuracy, $O(n^{-3/2})$. Hence, the $100(1-\alpha)\%$ confidence interval for ψ is $\{\psi : |r^*(\psi)| \leq z_{\alpha/2}\}$. However, it is well-known that the ancillary statistic may not exist, and even if it exists, it may not be unique.

Fraser *et al.* [8] presented a systematic method to obtain Q without requiring the existence of an ancillary statistic. More specifically, Fraser *et al.* [8] show that Q is the standardized maximum likelihood estimate departure calculated in the canonical parameter space. To obtain Q , all it requires is the observed likelihood function $\ell(\theta) = \ell(\theta; y)$, and the parameter of interest $\psi(\theta)$. Let $\varphi = \varphi(\theta)$ be the canonical parameter of an exponential family model. Then Q is a modification of q in Equation (2) that is calculated in $\varphi(\theta)$ scale, and it takes the form:

$$Q = \text{sign}(\hat{\psi} - \psi) \frac{|\chi(\hat{\theta}) - \chi(\tilde{\theta})|}{\hat{\sigma}_\chi} \quad (5)$$

where $\chi(\theta)$ is a re-calibrated scalar parameter on the $\varphi(\theta)$ scale obtained from the gradient vector of $\psi(\theta)$ at the constrained maximum likelihood estimator $\tilde{\theta}$:

$$\chi(\theta) = \psi_{\theta'}(\tilde{\theta}) \varphi_{\theta}^{-1}(\tilde{\theta}) \varphi(\theta) = \frac{\partial \psi(\theta)}{\partial \theta'} \left\{ \frac{\partial \varphi(\theta)}{\partial \theta} \right\}^{-1} \Big|_{\theta=\tilde{\theta}} \varphi(\theta).$$

Note that $|\chi(\hat{\theta}) - \chi(\tilde{\theta})|$ is the maximum likelihood departure $|\hat{\psi} - \psi|$ calculated in $\varphi(\theta)$ scale. And $\hat{\sigma}_\chi^2$ is an estimated variance of $|\chi(\hat{\theta}) - \chi(\tilde{\theta})|$, which takes the form:

$$\hat{\sigma}_\chi^2 = \left\{ \psi_{\theta'}(\tilde{\theta}) j_{\theta\theta'}^{-1}(\tilde{\theta}) \psi_{\theta}(\tilde{\theta}) \right\} \frac{|j_{(\theta\theta')}(\tilde{\theta})|}{|j_{(\theta\theta')}(\hat{\theta})|}, \quad (6)$$

where

$$|j_{\theta\theta'}(\tilde{\theta})| = |j_{\theta\theta'}(\tilde{\theta})| |\varphi_{\theta}(\tilde{\theta})|^{-2} \quad \text{and} \quad |j_{\theta\theta'}(\hat{\theta})| = |j_{\theta\theta'}(\hat{\theta})| |\varphi_{\theta}(\hat{\theta})|^{-2}$$

The subscripts $(\theta\theta')$ indicate that the information has been recalibrated in terms of the new parameterization $\varphi(\theta)$.

The canonical parameter $\varphi(\theta)$ of a model may not be explicitly available; however, Fraser and Reid [11] show that it can be approximated by

$$\varphi'(\theta) = \frac{\partial}{\partial y} \ell(\theta; y) \cdot V \quad (7)$$

where V are the tangent vectors constructed using a vector of pivotal quantities $z = z(y; \theta)$:

$$V = - \left(\frac{\partial z}{\partial y'} \right)^{-1} \left(\frac{\partial z}{\partial \theta'} \right) \Big|_{\theta=\hat{\theta}}. \quad (8)$$

A simple and easy choice of V is given by the successive distribution functions $z = F(y; \theta)$, which are uniformly distributed. Thus, the quantity Q can be calculated by Equation (5) and p -value and the $100(1-\alpha)\%$ confidence interval for ψ can be obtained from Equation (4).

3. Third-Order Method For Linear Mixed Models

In this section, we derive all the key equations of the first-order and third-order methods for the random intercept mixed model and the random intercept and random slope mixed model.

Third-Order Method for Random Intercept Mixed Model

Consider Model (1), the log-likelihood function can be written as

$$\ell(\theta) = \ell(\beta, \sigma_{\varepsilon}^2, \lambda) = -\frac{N}{2} \log \sigma_{\varepsilon}^2 - \frac{1}{2} \log |\Delta| - \frac{(Y - X\beta)' \Delta^{-1} (Y - X\beta)}{2\sigma_{\varepsilon}^2}. \quad (9)$$

The maximum likelihood estimates $\hat{\theta} = (\hat{\beta}, \hat{\sigma}_{\varepsilon}^2, \hat{\lambda})$ can be obtained from maximizing the log-likelihood function (Equation (9)). For a fixed λ , the constrained maximum likelihood estimate of θ is $\tilde{\theta} = (\tilde{\beta}, \tilde{\sigma}_{\varepsilon}^2, \lambda)$. In particular, $\hat{\Delta} = I + \hat{\lambda}ZZ'$ and $\tilde{\Delta} = I + \lambda ZZ'$. Let $\delta = Y - X\beta$. The Wald statistic q can be obtained by Equation (2) where the observed information matrix evaluated at $\hat{\theta}$ is

$$j_{\theta\theta'}(\hat{\theta}) = - \frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'} \Big|_{\hat{\theta}} = \begin{pmatrix} \frac{1}{\hat{\sigma}_{\varepsilon}^2} X' \hat{\Delta}^{-1} X & \frac{1}{\hat{\sigma}_{\varepsilon}^2} X' \hat{\Delta}^{-1} \delta & -\frac{1}{\hat{\sigma}_{\varepsilon}^2} X' \hat{\Delta}_{\lambda}^{-1} \delta \\ \frac{1}{\hat{\sigma}_{\varepsilon}^4} \delta' \hat{\Delta}^{-1} X & -\frac{N}{2\hat{\sigma}_{\varepsilon}^4} + \frac{1}{\hat{\sigma}_{\varepsilon}^6} \delta' \hat{\Delta}^{-1} \delta & -\frac{1}{2\hat{\sigma}_{\varepsilon}^4} \delta' \hat{\Delta}_{\lambda}^{-1} \delta \\ -\frac{1}{\hat{\sigma}_{\varepsilon}^2} \delta' \hat{\Delta}_{\lambda}^{-1} X & -\frac{1}{2\hat{\sigma}_{\varepsilon}^4} \delta' \hat{\Delta}_{\lambda}^{-1} \delta & \frac{1}{2\hat{\sigma}_{\varepsilon}^2} \delta' \hat{\Delta}_{\lambda}^{-1} \delta - \frac{1}{2} \sum_{i=1}^m \frac{n_i^2}{(1+n_i \hat{\lambda})^2} \end{pmatrix}$$

where $\Delta_{\lambda}^{-1} = \partial \Delta^{-1} / \partial \lambda$ and $\Delta_{\lambda\lambda}^{-1} = -\partial^2 \Delta^{-1} / \partial \lambda^2$ are the first and second derivatives of Δ^{-1} .

The signed log-likelihood ratio statistic, r , defined in Equation (3) can be simplified to:

$$r = \text{sign}(\hat{\lambda} - \lambda) \left[N \log \tilde{\sigma}_{\varepsilon}^2 + \log |\tilde{\Delta}| - N \log \hat{\sigma}_{\varepsilon}^2 - \log |\hat{\Delta}| \right]^{1/2} \quad (10)$$

The constrained information matrix $j_{\theta\theta'}(\tilde{\theta})$ that will be used for the third-order method can be simply calcu-

lated by replacing $\hat{\theta} = (\hat{\beta}, \hat{\sigma}_\varepsilon^2, \hat{\lambda})$ in $j_{\theta\theta'}(\hat{\theta})$ by the constrained maximum likelihood estimate $\tilde{\theta} = (\tilde{\beta}, \tilde{\sigma}_\varepsilon^2, \lambda)$. The usual form of the score test statistic is given by

$$S = \left\{ \frac{\partial \ell(\lambda)}{\partial \lambda} \right\}^2 \left\{ -\frac{\partial^2 \ell(\lambda)}{\partial \lambda^2} \right\}^{-1},$$

under the null hypothesis. Nuisance parameters are suppressed from notation. In the special case of testing zero variance component with $n_i \equiv n$, the score test statistic is ([12]):

$$S = \frac{mn(C-1)^2}{2 \cdot 2nC-1} \quad \text{with} \quad C = \frac{1}{mn\tilde{\sigma}_\varepsilon^2} \sum_{i=1}^m \left(\sum_{j=1}^n \tilde{\delta}_{ij} \right)^2.$$

Similar to the signed likelihood ratio test statistic, the signed score test statistic is:

$$s = \text{sign}(\hat{\lambda} - \lambda) \sqrt{S}. \quad (11)$$

Patterson and Thompson [6] proposed restricted or residual likelihood method which takes into account the loss of degrees of freedom involved in estimating the fixed effects and reduces the downward bias of maximum likelihood estimates of variance parameters. The restricted log-likelihood function is defined as

$$\ell^{RE}(\theta) = -\frac{N-p}{2} \log \sigma_\varepsilon^2 - \frac{1}{2} \log |\Delta| - \frac{1}{2} \log |X' \Delta^{-1} X| - \frac{(Y - X\beta)' \Delta^{-1} (Y - X\beta)}{2\sigma_\varepsilon^2}. \quad (12)$$

The signed log-likelihood ratio statistic and Wald statistic with the restricted maximum likelihood can be obtained using the restricted log-likelihood function (Equation (12)) accordingly.

To obtain the third-order inference for λ , we need to find the canonical parameter $\varphi(\theta)$. The canonical parameter $\varphi(\theta)$ is approximated by taking the sample space gradient at the observed data point in the directions of a set of vectors V (Equation (7)):

$$\varphi'(\theta) = \frac{\partial}{\partial y} \ell(\theta; y) \cdot V.$$

where the sample space gradient is

$$\frac{\partial}{\partial y} \ell(\theta; y) = -\frac{(y - X\beta)' \Delta^{-1}}{\sigma_\varepsilon^2} = -\frac{\delta' \Delta^{-1}}{\sigma_\varepsilon^2}. \quad (13)$$

Let $\Delta^{-1} = L'L$. The tangent direction V is obtained from the pivotal quantity $z(y; \theta) = L(y - X\beta)/\sigma_\varepsilon = L\delta/\sigma_\varepsilon$ (Equation (8)):

$$V = -\left\{ \frac{\partial z}{\partial y'} \right\}^{-1} \left\{ \frac{\partial z}{\partial \theta'} \right\} \bigg|_{\theta=\hat{\theta}} = \left(-X \quad -\frac{\hat{\delta}}{2\hat{\sigma}_\varepsilon^2} \quad \left\{ L^{-1} \frac{\partial L}{\partial \lambda} \right\} \bigg|_{\lambda=\hat{\lambda}} \right) \hat{\delta}. \quad (14)$$

Let A denotes $\left\{ L^{-1} \frac{\partial L}{\partial \lambda} \right\} \bigg|_{\lambda=\hat{\lambda}}$ in Equation (14). Multiply Equations (13) and (14), we have the canonical parameter $\varphi'(\theta)$:

$$\varphi'(\theta) = \frac{\partial}{\partial y} \ell(\theta; y) \cdot V = \left(\frac{\delta' \Delta^{-1} X}{\sigma_\varepsilon^2} \quad \frac{\delta' \Delta^{-1} \hat{\delta}}{2\sigma_\varepsilon^2 \hat{\sigma}_\varepsilon^2} \quad -\frac{\delta' \Delta^{-1} A \hat{\delta}}{\sigma_\varepsilon^2} \right).$$

The Jacobian for $\varphi(\theta)$ is

$$\varphi_{\theta'}(\theta) = \frac{\partial \varphi(\theta)}{\partial \theta'} = \begin{pmatrix} -\frac{1}{\sigma_\varepsilon^2} X' \Delta^{-1} X & -\frac{1}{\sigma_\varepsilon^4} X' \Delta^{-1} \delta & \frac{1}{\sigma_\varepsilon^2} X' \Delta_\lambda^{-1} \delta \\ -\frac{1}{2\sigma_\varepsilon^2 \hat{\sigma}_\varepsilon^2} \hat{\delta}' \Delta^{-1} X & -\frac{1}{2\sigma_\varepsilon^4 \hat{\sigma}_\varepsilon^2} \hat{\delta}' \Delta^{-1} \delta & \frac{1}{2\sigma_\varepsilon^2 \hat{\sigma}_\varepsilon^2} \hat{\delta}' \Delta_\lambda^{-1} \delta \\ \frac{1}{\sigma_\varepsilon^2} \hat{\delta}' A' \Delta^{-1} X & \frac{1}{\sigma_\varepsilon^4} \hat{\delta}' A' \Delta^{-1} \delta & -\frac{1}{\sigma_\varepsilon^2} \hat{\delta}' A' \Delta_\lambda^{-1} \delta \end{pmatrix}.$$

Thus Q can be obtained by Equation (5) and we have the third-order approximations for λ by Equation (4). All the equations and the maximizing problems can be programmed into R.

4. Application: A Real Data Example

We use a data set from the Genetic Analysis Workshop (GAW 18) to illustrate the application of our proposed method. The Genetic Analysis Workshops (www.gaworkshop.org), which began in 1982, were initially motivated by the development and publication of several new algorithms for statistical genetic analysis, using different method of analysis. It is a collaborative effort among researchers worldwide to evaluate and compare statistical genetic methods, relevant to current analytical problems in genetic epidemiology and statistical genetics. The GAW 18 data set [13] was drawn from the San Antonio Family Study (SAFS) with a total of 959 participants from 20 families and it included the whole genome sequence data from all individuals. For each individual, age, sex, the use of hypertension diagnosis and antihypertensive medications (HTN) and tobacco smoking status (SMOKING) were recorded.

Our goal was to test the homogeneity of baseline systolic blood pressure (SBP) among families and if a genetic factor contributed to the variation of SBP among families. The test of homogeneity among families is a challenging task. The use of variance component of random effects in mixed models provides an efficient method to account for the heterogeneity among families.

4.1. Test of Homogeneity of SBP among Families

We first evaluated the homogeneity of SBP levels among families using a random intercept mixed model. We analysed the SBP levels with “AGE” as fixed effect in a random intercept mixed model accounting for the variabilities of SBP among families. A non-zero variance of random intercept implies there exists a significant variation of SBP levels among families and the group specific random intercepts allow us to identify the families that deviate the most from the population average SBP levels at baseline.

Let y_{ij} denote the SBP measurements for the i th individual in the j th family, the linear mixed model formula for GAW 18 data with random intercept can be written as

$$y_{ij} = \beta_0 + b_{0j} + \beta_1 AGE_{ij} + \varepsilon_{ij}, \quad (15)$$

where b_{0j} is the random intercept for the j th family.

To assess the variability of SBP, we obtain the 95% confidence interval of the variance component of random intercept λ . The 95% confidence intervals of λ obtained from third-order and first-order methods are recorded in **Table 1**. The third-order method yields a 95% confidence interval of (0.001, 0.0802) containing no zero, suggesting a significant variation of SBP levels among families oppose to the results from first-order methods. The simulation studies have shown the superior accuracy of the third-order method and the conservativeness of the first-order methods. Thus we conclude that there is a significant variation of SBP among the 20 families based on the result from the third-order method. **Figure 1(a)** is the normal quantile plot of the random intercept b_{0j} for the 20 families. The normal quantile plot also confirms the variability of SBP levels among families.

4.2. Test of Homogeneity of SBP among Families in Relation to a Genetic Factor

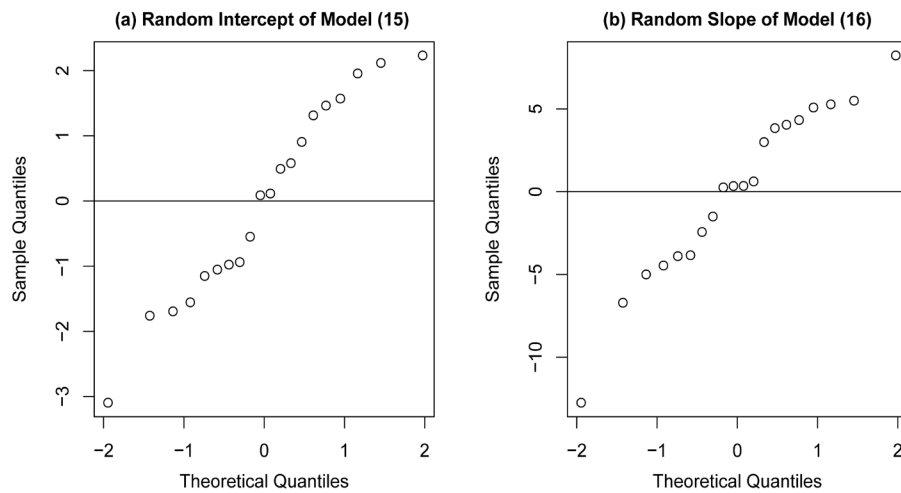
In a second step, we examine whether a combined genetic factor contributes to the variation of SBP levels among the 20 families using a random intercept and random slope mixed model. In a previous genome-wide association study (GWAS) from GAW 18 data, we showed that four single-nucleotide polymorphisms (SNPs) listed in **Table 2** from different regions were associated with systolic blood pressure levels adjusting for demographical

Table 1. The 95% confidence intervals of λ , test homogeneity of SBP among families.

Method	95% CI
Third-order	(0.001, 0.0802)
LRT-REML	(0, 0.0753)
LRT-ML	(0, 0.0675)
Wald-REML	(0, 0.0503)
Wald-ML	(0, 0.0482)
Score	(0, 0.0488)

Table 2. Four selected SNPs from GWAS study.

POS	GENE	SNP	GWAS P-value
127,074,020	PLXNA1	rs11716834	0.0000011
138,919,221	PISRT1	rs13100625	0.00000039
173,866,797	NLGN1	rs11927722	0.0000017
173,986,397	NLGN1	rs549641	0.00009

**Figure 1.** Normal quantile plots of random intercept for Model (15) and random slope for Model (16).

variables [14]. We use the average number of minor alleles from these four casual SNPs for each individual (denoted as $grSNP_{ij}$) was used as both as fixed effect and random slope in the model. To assess whether the effect of this combined genetic factor on SBP varies among families, we performed a test of variance component of the random slope in a random intercept and random slope mixed model. A non-zero variation of random slope implies significant variation of SBP levels among families due to the genetic factor. Backward elimination technique was used for model selection and led to the following random intercept and random slope mixed model:

$$y_{ij} = \beta_0 + b_{0j} + (\beta_1 + b_{1j}) grSNP_{ij} + \beta_2 AGE_{ij} + (\text{SEX}_{ij} * \text{SMOKING}_{ij} * \text{HTN}_{ij}) \beta_{\text{SEX*SMOKING*HTN}} + \varepsilon_{ij} \quad (16)$$

where b_{0j} is the random intercept and b_{1j} is the random slope of the average number of minor alleles for the j th family and $(\text{SEX}_{ij} * \text{SMOKING}_{ij} * \text{HTN}_{ij})$ denotes the three-way interaction term including the lower order variables. The normal quantile plot of the random slope b_{1j} is shown in **Figure 1(b)** and displays substantial variability among families. The 95% confidence intervals of λ_{11} obtained from third-order method, LRT-REML

and LRT-ML are (0.1035, 1.1008), (0.0554, 1.0445) and (0.0457, 0.9658) respectively. Not surprisingly, result from the likelihood ratio test methods are more conservative than with the third-order method.

Our two sets of analysis therefore demonstrated the interest of the third-order approach, the only method to find significant heterogeneity among families in terms of baseline SNP and change of SPB due to a combined genetic factor.

5. Simulation Studies

In this section, we present results from simulation studies for the random intercept mixed model and random intercept and random slope mixed model to assess the improvement of the third-order method and examine the accuracy of the proposed third-order method compared to the following first-order methods:

- 1) Third-order: the third-order method based on maximum likelihood;
- 2) LRT-REML: the signed log-likelihood ratio statistic method with restricted maximum likelihood estimation;
- 3) LRT-ML: the signed log-likelihood ratio statistic method with maximum likelihood estimation;
- 4) Wald-REML: Wald statistic with restricted maximum likelihood estimation.
- 5) Wald-ML: Wald statistic with maximum likelihood estimation;
- 6) Score: Score test with maximum likelihood estimation.

The statistical software R (“nlme” package) was used to carry out all our analyses.

5.1. Simulation Results for the Random Intercept Mixed Model

We present simulation results for the null distribution of the zero variance component λ : $H_0: \lambda = 0$ ($\sigma_b^2 = 0$) versus $H_a: \lambda > 0$ ($\sigma_b^2 > 0$) and for non-zero variance component λ : $H_0: \lambda = 4$ versus $H_a: \lambda > 4$ in a random intercept mixed model.

5.1.1. $H_0: \lambda = 0$ versus $H_a: \lambda > 0$

Balanced data sets with equal group size $n_i = n = 10$ were generated with fixed effect $\beta = (\beta_0, \beta_1) = (1, 2)$, $\sigma_\epsilon^2 = 1$ for the test of zero variance component $\lambda = 0$. For each simulation configuration, 10,000 iterations were used for a number of groups $m = 5, 10, 20, 50$.

Probability-probability plots (pp-plot), similar to a quantile-quantile (qq-plot) but on the p -value scale rather than on the scale of the test statistic [6], are used to examine the accuracy of the p -values obtained from the six methods. For a perfect approximation, the p -values line up on the 45° diagonal line. The nominal p -values calculated from the six methods, are plotted versus the empirical p -values in Figure 2(a). Figure 2(b) shows only the 0 to 15th percentile of Figure 2(a), that allows us have a clearer look at the differences. Figure 2 shows that when group size $n = 10$, the third-order method is surprisingly accurate and consistent regardless of the number of groups m . The likelihood ratio test is the most accurate among the first-order methods while the likelihood ratio test with restricted maximum likelihood estimation reduces the bias of maximum likelihood estimation. The approximations from Wald test and score test are poor, thus omitted for the rest of this paper. The figures also show that the proportion of point mass at 0 converges to 50% when sample size increases.

To further examine the accuracy of the third-order and first-order approximations, a 3-dimensional plot is used to visualize the bias for a number of groups m varying from 3 to 40 and group size n from 5 to 40 (Figure 3). The bias is calculated by subtracting the nominal p -value at the 5th percentile from its empirical (or true) p -value 0.05. A bias of zero implies perfect approximation and a positive bias implies conservativeness (greater than the true p -value). Figure 3 shows the same pattern as Figure 2: the third-order method is accurate and consistent. The likelihood ratio test with restricted maximum likelihood estimation is more accurate than that of maximum likelihood estimation. We observe an interesting pattern that is for a given number of groups m , the bias is small when the group size n is small, it increases with group size n and then flattens down and decreases slowly. For a given number of group size n , the bias decreases with m . The phenomenon suggests that the accuracy of first-order methods depend on both number of groups m and the group size n .

5.1.2. $H_0: \lambda = 4$ versus $H_a: \lambda > 4$

In this section, we show simulation results for a non-zero variance component $H_0: \lambda = 4$ versus $H_a: \lambda > 4$. Balanced data sets with equal group size $n_i = n = 10$ were generated with fixed effect $\beta = (\beta_0, \beta_1) = (1, 2)$,

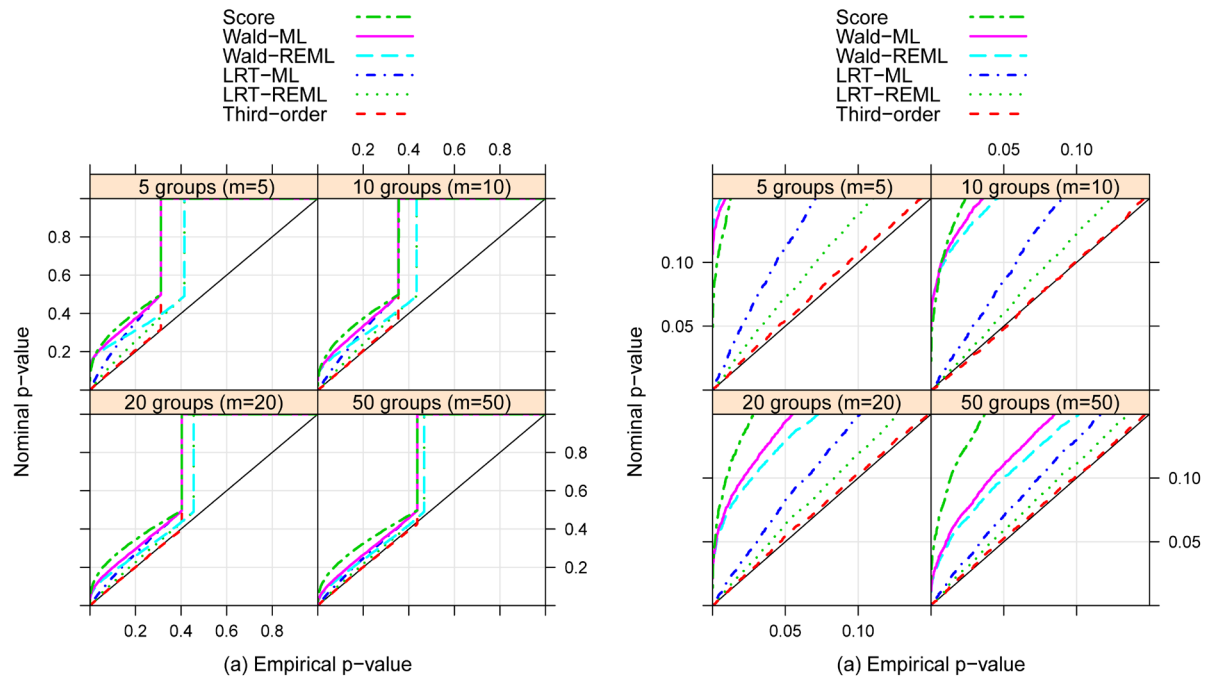


Figure 2. Probability-probability plot for testing $\lambda = 0$ vs. $\lambda > 0$ for number of groups $m = 5, 10, 20, 50$ with equal group size $n = 10$. For a perfect approximation, the p -values line up on the 45° diagonal line. This plot shows the conservativeness of the first-order methods.

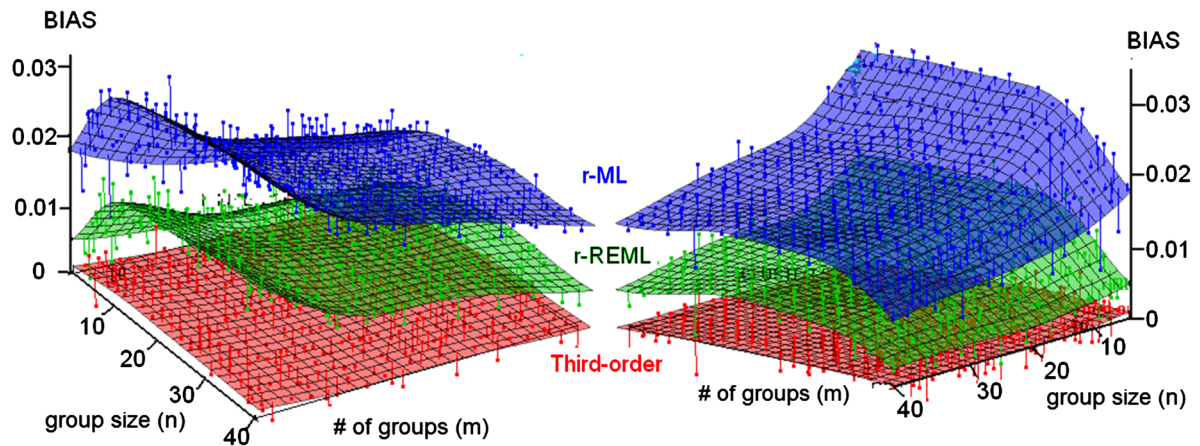


Figure 3. 3-D bias plots testing $\lambda = 0$ vs. $\lambda > 0$ for various number of groups m and group size n . The bias is calculated subtracting the nominal p -value from the five methods at the 5th percentile from its empirical (or true) p -value 0.05. A bias of zero implies perfect approximation and a positive bias implies that the approximation is conservative (greater than the true p -value).

$\sigma_\epsilon^2 = 1$ and $\lambda = 4$. For each simulation configuration, 10,000 iterations were used for a number of groups $m = 5, 10, 20, 50$. Again, the pp-plot in [Figure 4](#) confirms the superior accuracy of the third-order method.

In addition, we examined the lower and upper error probabilities and the central coverage probabilities of third-order method and LRT methods. We generated data with $m = 5, 10, 50$ and $n = 5, 10, 50$ with 10,000 iterations for each configuration and recorded the coverage probabilities in [Table 3](#). The performance of a method is assessed by the following values: the central coverage probability is the proportion of intervals that contains the true λ (CCP), the lower error probability is the proportion of true λ that falls outside the lower bound of the confidence interval (L) and the upper error probability which is the proportion of true λ that falls outside the upper bound of the confidence interval (U). The true CCP, L, and U are 0.95, 0.025 and 0.025 with standard errors

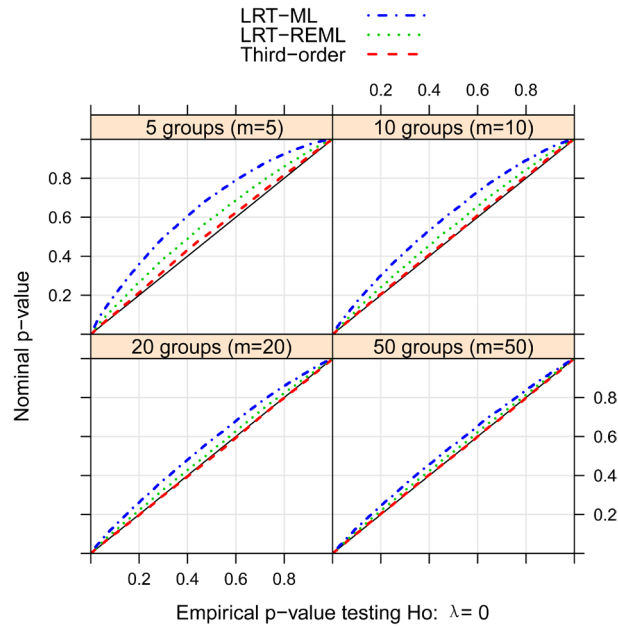


Figure 4. Probability-probability plot for testing $\lambda = 4$ vs. $H_a: \lambda > 4$ for number of groups $m = 5, 10, 20, 50$ with equal group size $n = 10$.

Table 3. Central coverage probabilities and error probabilities for testing $H_0: \lambda = 4$ versus $H_a: \lambda > 4$. The nominal values are 0.95 0.025, 0.025 with standard errors 0.0022, 0.0016 and 0.0016 for CCP, L and U respectively.

m	n	Coverage	Third-Order	LRT-REML	LRT-ML
5	5	CCP	0.9465	0.9373	0.9026
		L	0.0245	0.0195	0.0148
		U	0.0290	0.0432	0.0826
5	10	CCP	0.9486	0.9393	0.9002
		L	0.0230	0.0173	0.0107
		U	0.0284	0.0434	0.0891
5	50	CCP	0.9505	0.9425	0.9030
		L	0.0239	0.0165	0.0071
		U	0.0256	0.0410	0.0899
10	5	CCP	0.9553	0.9509	0.9353
		L	0.0217	0.0188	0.0154
		U	0.0230	0.0303	0.0493
10	10	CCP	0.9505	0.9456	0.9312
		L	0.0246	0.0199	0.0134
		U	0.0249	0.0345	0.0554
10	50	CCP	0.9503	0.9471	0.9255
		L	0.0233	0.0166	0.0104
		U	0.0264	0.0363	0.0641
50	5	CCP	0.9517	0.9524	0.9499
		L	0.0237	0.0207	0.0178
		U	0.0246	0.0269	0.0323
50	10	CCP	0.9518	0.9492	0.9468
		L	0.0220	0.0206	0.0165
		U	0.0262	0.0302	0.0367
50	50	CCP	0.9497	0.9515	0.9472
		L	0.0250	0.0200	0.0154
		U	0.0253	0.0285	0.0374

0.0022, 0.0016 and 0.0016, respectively. The table shows that the third-order method has the best coverage probabilities. It is always accurate with symmetric lower and upper error coverage probabilities. The LRT-REML has good central coverage probabilities in some cases however the lower and upper coverage probabilities are asymmetric with the lower error probabilities tends to be less than the true value and the upper probabilities tend to be larger.

5.2. Simulation Results for Random Intercept and Random Slope Mixed Model

Testing zero variance component of random slope in a random intercept and random slope mixed model implies that we are testing both $\lambda_{01} = 0$ and $\lambda_{11} = 0$. The third-order method is currently limited to the test of a scalar parameter. Theoretical development for vector parameter using third-order approach has not been proposed yet. Instead of testing $\lambda_{11} = 0$, we could obtain the 95% confidence interval of λ_{11} . Thus we show the simulation results for the test of a non-zero variance component of random slope $H_0 : \lambda_{11} = 4$ versus $H_a : \lambda_{11} > 4$. We generated balanced data sets with equal group size $n_i = n = 10$ with fixed effect $\beta = (\beta_0, \beta_1) = (1, 2)$, $\sigma_e^2 = 1$, $\lambda_{00} = 1$, $\lambda_{01} = 1$ and $\lambda_{11} = 4$, and a number of groups $m = 5, 10, 20, 50$. For each simulation configuration, 10,000 iterations were used. **Figure 5** is the pp-plot for $H_0 : \lambda_{11} = 4$ versus $H_a : \lambda_{11} > 4$. The third-order method has shown its superior accuracy and symmetric tail probabilities for the test of variance component of random slope in a random intercept and random slope mixed model. The LRT with REML is less conservative than LRT with ML.

In this section, we presented simulation results for the tests of variance component in linear mixed models. The simulation results demonstrated that the third-order method is accurate and consistent. The LRT method perform the best among all first-order methods and the REML reduce the bias comparing the ML. The computation time for third-order p -values is fast and depends on sample size. It takes only a few seconds for a sample of 40 group with group size of 40 on a personal computer with a Quad CPU @3 GHz and 8 GB RAM.

6. Discussion

A third-order likelihood-based method is proposed for the tests of zero and non-zero variance component in the random intercept mixed model and the random intercept and random slope mixed model. The third-order method requires only the observed likelihood function and derivation of the pivotal quantity and ancillary direction. The computation of third-order method is efficient. Our simulation studies show that the proposed third-order

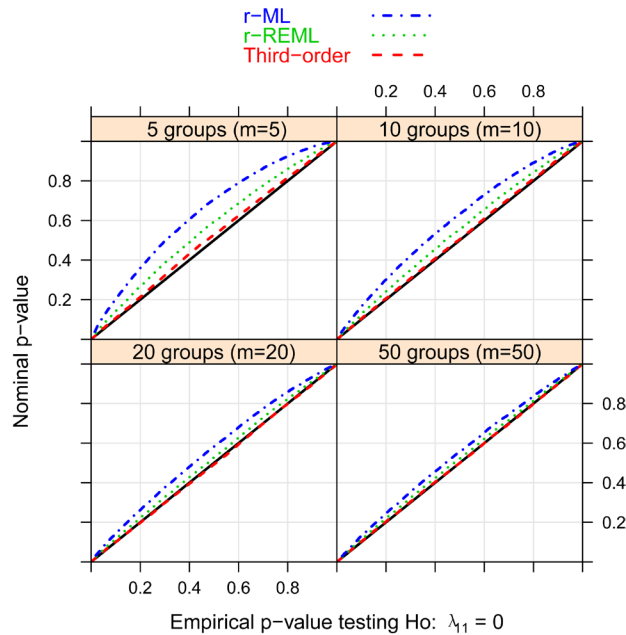


Figure 5. Probability-probability plot for testing $\lambda_{11} = 4$ vs. $\lambda_{11} > 4$ for number of groups $m = 5, 10, 20, 50$ with equal group size $n = 10$.

likelihood-based method provides substantial improvement over traditional first-order methods. It provides extremely accurate and consistent approximation for the tests of variance component in mixed models even when the number of groups and group sizes are as small as five and three respectively. The likelihood ratio test performs the best among the first-order methods. The likelihood ratio test with restricted maximum likelihood estimation is less conservative than the maximum likelihood estimation. The likelihood ratio test with restricted maximum likelihood estimation is recommended for the test of variance component in mixed models whenever the first-order method is used. The application to the GAW 18 data illustrated the advantages of the third-order methods for the test of homogeneity among families in relation to a genetic factor. The simulation results also yielded a comprehensive evaluation of the accuracy of the first-order methods. The third-order method provides also opportunities for further applications in the context of longitudinal studies with repeated measurements and genetic studies. For example, there has been increased interest in genetic analysis to test the combined effect of several rare variants using random effects in the mixed model framework [14]. Since some popular approaches for this type of analysis are based on the score test of random effect, we think that the third-order approach could provide a more accurate alternative in this context, that we are planning to explore in future work.

References

- [1] Britton, T. (1997) Tests to Detect Clustering of Infected Individuals within Families. *Biometrics*, **53**, 98-109. <http://dx.doi.org/10.2307/2533100>
- [2] Liang, K.Y. (1987) A Locally Most Powerful Test for Homogeneity with Many Strata. *Biometrics*, **51**, 259-264. <http://dx.doi.org/10.1093/biomet/74.2.259>
- [3] Chernoff, H. (1954) On the Distribution of the Likelihood Ratio. *Annals of Mathematical Statistics*, **25**, 573-578. <http://dx.doi.org/10.1214/aoms/1177728725>
- [4] Self, S.F. and Liang, K.Y. (1987) Asymptotic Properties of Maximum Likelihood Estimator and Likelihood Ratio Tests under Nonstandard Conditions. *Journal of the American Statistical Association*, **82**, 605-610. <http://dx.doi.org/10.1080/01621459.1987.10478472>
- [5] Stram, D.O. and Lee, J.W. (1994) Variance Components Testing in the Longitudinal Mixed Effects Model. *Biometrics*, **50**, 1117-1177. <http://dx.doi.org/10.2307/2533455>
- [6] Patterson, H.D. and Thompson, R. (1971) Recovery of Interblock Information When Block Sizes Are Unequal. *Biometrika*, **58**, 545-554. <http://dx.doi.org/10.1093/biomet/58.3.545>
- [7] Crainiceanu, D.M. and Ruppert, D. (2004) Likelihood Ratio Tests in Linear Mixed Models with One Variance Component. *Journal of the Royal Statistical Society: Series B*, **66**, 165-185. <http://dx.doi.org/10.1111/j.1467-9868.2004.00438.x>
- [8] Fraser, D.A.S., Reid, N. and Wu, J. (1999) A Simple General Formula for Tail Probabilities for Frequentist and Bayesian Inference. *Biometrika*, **86**, 249-264. <http://dx.doi.org/10.1093/biomet/86.2.249>
- [9] Barndorff-Nielsen, O.E. (1986) Inference on Full and Partial Parameters, Based on the Standardized Signed Log-Likelihood Ratio. *Biometrika*, **73**, 307-322. <http://dx.doi.org/10.2307/2336207>
- [10] Barndorff-Nielsen, O.E. (1991) Modified Signed Log-Likelihood Ratio Statistic. *Biometrika*, **78**, 557-563. <http://dx.doi.org/10.1093/biomet/78.3.557>
- [11] Fraser, D.A.S. and Reid, M. (1995) Ancillaries and Third-Order Significance. *Utilitas Mathematica*, **47**, 33-53.
- [12] Verbeke, G. and Molenberghs, G. (2003) The Use of Score Tests for Inference on Variance Components. *Biometrics*, **59**, 254-262. <http://dx.doi.org/10.1111/1541-0420.00032>
- [13] Almasy, L., Dyer, T.D., Peralta, J.M., Jun, G., Fuchsberger, C., Almeida, M.A., Kent, J.W., Fowler, S., Duggirala, R. and Blangero, J. (2014) Data for Genetic Analysis Workshop 18: Human Whole Genome Sequence, Blood Pressure, and Simulated Phenotypes in Extended Pedigrees. *BMC Proceedings*, **8**, S2.
- [14] Wu, Y.Y. and Briollais, L. (2014) Mixed-Effects Models for Joint Modeling of Sequence Data in Longitudinal Studies. *BMC Proceedings*, **8**, S92.