



**Manchester  
Metropolitan  
University**

---

Rao, Kaidi, Drikvandi, Reza ORCID logoORCID: <https://orcid.org/0000-0002-7245-9713> and Saville, Benjamin (2019) Permutation and Bayesian tests for testing random effects in linear mixed-effects models. *Statistics in Medicine*. ISSN 0277-6715

---

**Downloaded from:** <https://e-space.mmu.ac.uk/623734/>

**Version:** Accepted Version

**Publisher:** Wiley

**DOI:** <https://doi.org/10.1002/sim.8350>

Please cite the published version

<https://e-space.mmu.ac.uk>

# Permutation and Bayesian tests for testing random effects in linear mixed-effects models

Kaidi Rao<sup>1</sup>, Reza Drikvandi<sup>2,\*</sup>, and Benjamin Saville<sup>3,4</sup>

<sup>1</sup> Statistics Section, Department of Mathematics, Imperial College London, UK

<sup>2</sup> Department of Computing and Mathematics, Manchester Metropolitan University, UK

<sup>3</sup> Berry Consultants, Austin, TX, 78746, USA

<sup>4</sup> Adjunct Faculty, Department of Biostatistics, Vanderbilt University, USA

\* Corresponding author. Email: r.drikvandi@mmu.ac.uk

## Abstract

In many applications of linear mixed-effects models to longitudinal and multilevel data especially from medical studies, it is of interest to test for the need of random effects in the model. It is known that classical tests such as the likelihood ratio, Wald and score tests are not suitable for testing random effects because they suffer from testing on the boundary of the parameter space. Instead, permutation and bootstrap tests as well as Bayesian tests, which do not rely on the asymptotic distributions, avoid issues with the boundary of the parameter space. In this paper, we firstly develop a permutation test based on the likelihood ratio test statistic, which can be easily used for testing multiple random effects and any subset of them in linear mixed-effects models. The proposed permutation test would be an extension to both the test of Fitzmaurice et al.<sup>1</sup> (Biometrics, 2007) and the test of Drikvandi et al.<sup>2</sup> (Biostatistics, 2013). We secondly aim to compare permutation tests and Bayesian tests for random effects to find out which test is more powerful under which situation. Nothing is known about this in the literature, whilst this is an important practical problem due to the usefulness of both methods in tackling the challenges with testing random effects. For this, we consider the Bayesian test of Saville and Herring<sup>3</sup> (Biometrics, 2009), where we also propose a new alternative computation for this Bayesian test to avoid some computational issue it encounters in testing multiple random effects. Extensive simulations and a real data analysis are used for evaluation of the proposed permutation test and its comparison with the Bayesian test. We find that both tests perform well, albeit the permutation test with the likelihood ratio statistic tends to provide a relatively higher power when testing multiple random effects.

**Keywords:** *Bayesian test; Longitudinal data; Linear mixed-effects models; Permutation test; Random effects; Variance components.*

## 1. Introduction

This paper firstly aims to develop a permutation test for random effects based on the likelihood ratio test statistic, to be applicable for testing multiple random effects and any subset of them in linear mixed-effects models. The proposed permutation test would be an extension to both tests of Fitzmaurice et al.<sup>1</sup> and Drikvandi et al.<sup>2</sup>. We secondly aim to compare permutation tests and Bayesian tests for random effects to find out which test is more powerful under which situation. The importance and background of these two objectives are given in the following.

Longitudinal studies arise in many different domains, notably in medical studies, where they help collect repeated measurements on individuals or subjects. Longitudinal data are therefore useful for studying and understanding changes over time. For example, a study on how alcohol consumption will affect the blood pressure can be conducted by comparing the repeated measurements of blood pressure, taken over a certain period of time, between an alcohol-consuming group and a control group.

Linear mixed-effects models<sup>4,5</sup> are well suited for the analysis of longitudinal data, where random effects are used to capture the between-subject variability due to unmeasured covariates or unknown biological differences. Random effects are unobserved random variables that vary across subjects. Testing for the need of random effects in linear mixed-effects models is important both theoretically and practically. Several practical examples on testing random effects are given in<sup>2,6</sup>, and moreover theoretical and computational investigations suggest to include only necessary random effects in the model on the one hand, and on the other hand to not ignore any important random effects<sup>7,8</sup>.

From a statistical perspective, to test for the need of random effects is equivalent to testing whether the variance components of random effects equal zero. Denoting the covariance matrix of random effects by  $\Sigma$ , we can express this test as follows

$$\begin{cases} H_0 : \Sigma = \mathbf{0} \\ H_1 : \Sigma \text{ is a non-zero non-negative definite matrix.} \end{cases} \quad (1)$$

This is a non-standard testing problem because the null hypothesis places the variances of random effects on the boundary of the parameter space. Classical tests such as the likelihood ratio,

Wald and score tests suffer from testing on the boundary of the parameter space, because the regularity conditions do not hold under such situations. As a consequence, the usual asymptotic chi-squared distribution of the likelihood ratio or score statistic is not valid. It is shown, under some additional conditions, that the correct asymptotic distribution is a mixture of chi-squared distributions under the null hypothesis<sup>6,9-11</sup>. However, the weights of the asymptotic mixture distribution are unknown, except for some special cases. For instance, for testing a single random effect, the correct asymptotic distribution is a 50 : 50 mixture of  $\chi_q^2$  and  $\chi_{q-1}^2$  where  $q$  is the total number of random effects in the model under alternative hypothesis. Also, for testing  $k$  uncorrelated random effects (i.e.,  $\Sigma$  is diagonal), the correct asymptotic distribution is given by  $\sum_{m=0}^k 2^{-k} \binom{k}{m} \chi_m^2$ . Determining the mixture's weights is very complex for a broad number of cases, as shown by Shapiro<sup>10</sup>. In addition to this limitation, the tests based on the asymptotic mixture distribution are shown to have incorrect Type I error rate and low power in small samples<sup>1,2,12</sup>.

Tests for random effects that avoid the boundary issue are therefore of great importance. Permutation and bootstrap tests as well as Bayesian tests do not rely on asymptotic distributions and hence are more suitable for testing random effects in mixed-effects models. Sinha<sup>13</sup> suggested a bootstrap test based on the score statistic; however, his bootstrap procedure is difficult to apply for testing multiple random effects or a subset of them. Fitzmaurice et al.<sup>1</sup> developed a permutation test, based on the likelihood ratio statistic, for testing a single random effect, and Lee and Braun<sup>14</sup> extended their permutation test for testing all random effects. Drikvandi et al.<sup>2</sup> proposed a more general permutation test for testing all random effects and any subset of them, and with a different test statistic. In this paper, we will use the general permutation procedure of Drikvandi et al.<sup>2</sup> to develop a permutation test based on the likelihood ratio test statistic, which can be easily applied for testing all random effects and any subset of them in linear mixed-effects models. It should be mentioned that the permutation procedure of Drikvandi et al.<sup>2</sup> coincides with that of Lee and Braun<sup>14</sup> for testing all random effects; however, the permutation procedure of Drikvandi et al.<sup>2</sup> is more general and can be easily used for testing any subset of random effects. Lee and Braun<sup>14</sup> pointed out that their permutation test is able to test a subset of random effects; however, they have never shown or investigated this in their paper.

It is also of interest to understand how permutation tests compare to the Bayesian tests, due to the usefulness of both methods in testing random effects. Nothing is known about this in the literature, and we investigate this problem to find out which test is more powerful. For this

purpose, we will compare the proposed permutation test with the Bayesian test of Saville and Herring<sup>3</sup> which was developed based on Bayes factors and a Laplace approximation to overcome the boundary challenge. It should be emphasised that we propose a new alternative computation for this Bayesian test to avoid some computational issue it encounters in testing multiple random effects. Extensive simulations and a real data analysis are used for evaluation of the proposed permutation test and its comparison with the Bayesian test of Saville and Herring<sup>3</sup>.

## 2. Linear mixed-effects model

Linear mixed-effects (LME) models incorporate random effects to capture the between-subject variability due to unmeasured covariates or unknown biological differences between subjects. This is an effective way to allow the parameters (e.g., intercepts and slopes) to vary across subjects. As an example, in the study of changes in blood pressure, a random intercept model can be used to allow the baseline blood pressure to vary across subjects in the study groups<sup>12</sup>.

Consider a longitudinal study with  $N$  subjects. Let  $Y_{ij}$  be the response variable for subject  $i$  taken at measurement time  $j$ ,  $j = 1, \dots, n_i$ . Denoting  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ , we can write the LME model as follows<sup>4</sup>:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad (2)$$

where  $\mathbf{X}_i$  and  $\mathbf{Z}_i$  are, respectively,  $n_i \times p$  and  $n_i \times q$  design matrices associated with fixed effects and random effects for subject  $i$ ,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})^T$  is a  $p \times 1$  vector of fixed-effects parameters,  $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})^T$  is a  $q \times 1$  vector of random effects for subject  $i$ , and  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})^T$  is an  $n_i \times 1$  vector of measurement errors for subject  $i$ . It is assumed that  $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$  and  $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_{n_i})$ , and further  $\mathbf{b}_i$  and  $\boldsymbol{\varepsilon}_i$  are independent.

It is more convenient to stack the  $N$  response vectors into a single response vector  $\mathbf{Y} = (\mathbf{Y}_1^T, \mathbf{Y}_2^T, \dots, \mathbf{Y}_N^T)^T$ , and also the fixed-effects and random effects design matrices into  $\mathbf{X} = [\mathbf{X}_1^T, \dots, \mathbf{X}_N^T]^T$  and  $\mathbf{Z} = \text{Diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_N)$ , respectively. Similarly, we write  $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_N^T)^T$  and  $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^T, \dots, \boldsymbol{\varepsilon}_N^T)^T$ . Then, the LME model (2) can be rewritten as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}.$$

Note that the covariance matrix of the random components  $\mathbf{b}$  and  $\boldsymbol{\varepsilon}$  is given by  $\text{Cov}(\mathbf{b}, \boldsymbol{\varepsilon}) = \begin{bmatrix} \boldsymbol{\Sigma} \otimes \mathbf{I}_{N_T} & \mathbf{0} \\ \mathbf{0} & \sigma_\varepsilon^2 \mathbf{I}_{N_T} \end{bmatrix}$ , where  $\otimes$  is the Kronecker product and  $\mathbf{I}_{N_T}$  is the identity matrix of order  $N_T$ , with  $N_T = \sum_{i=1}^N n_i$ .

The marginal distribution of  $\mathbf{Y}$ , after integrating out the random effects  $\mathbf{b}$  over their assumed distribution, would be  $\mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$  where  $\mathbf{V} = \mathbf{Z}(\boldsymbol{\Sigma} \otimes \mathbf{I}_{N_T})\mathbf{Z}^T + \sigma_\varepsilon^2 \mathbf{I}_{N_T}$ <sup>5</sup>. Therefore, we can write the marginal log-likelihood function of the model as follows

$$l(\boldsymbol{\theta}) = -\frac{N_T}{2} \log(2\pi) - \frac{1}{2} \log(|\mathbf{V}|) - \frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \quad (3)$$

where  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_\varepsilon^2, \boldsymbol{\Sigma})$  represents all the unknown parameters. The maximum likelihood estimates of parameters  $\boldsymbol{\theta}$  are obtained by maximising the log-likelihood function (3). This can be done using standard software such as the **lmer** function in R or the **PROC MIXED** in SAS.

### 3. The permutation test

To test for the need of a single random effect is equivalent to testing whether or not its variance is 0. When we test multiple random effects we need to test if their corresponding covariance matrix is  $\mathbf{0}$ . For instance, the null hypothesis for testing all random effects is  $H_0: \boldsymbol{\Sigma} = \mathbf{0}$ , while the alternative hypothesis says  $\boldsymbol{\Sigma}$  is a non-zero non-negative definite matrix. As discussed in the introduction, such tests for random effects require testing on the boundary of the parameter space which causes difficulties in applying the classical tests.

In this section, we develop a permutation test for testing multiple random effects, which is an extension to both the test of Fitzmaurice et al.<sup>1</sup> and the test of Drikvandi et al.<sup>2</sup>. It should be clarified that the permutation test of Fitzmaurice et al.<sup>1</sup> was developed for testing only a single random effect, and the permutation test of Drikvandi et al.<sup>2</sup> was constructed based on an ad hoc non-parametric test statistic. As discussed in the introduction, the permutation procedure of Drikvandi et al.<sup>2</sup> is more general than the permutation procedure of Lee and Braun<sup>14</sup>.

We consider the likelihood ratio statistic as our test statistic which leads to a powerful permutation test as shown in Section 5. The test statistic is defined as

$$R = 2l(\hat{\boldsymbol{\theta}}) - 2l(\hat{\boldsymbol{\theta}}_0),$$

where  $\hat{\boldsymbol{\theta}}_0$  is the ML estimates of parameters under the null hypothesis and  $\hat{\boldsymbol{\theta}}$  is the ML estimates under the entire parameter space. As already discussed, the correct asymptotic distribution of the likelihood ratio statistic is generally not available, so we use the general permutation procedure of Drikvandi et al.<sup>2</sup> to approximate its finite-sample distribution for testing all random effects and any subset of them.

### 3.1. Likelihood-based permutation test for testing all random effects

We first consider the case of testing all random effects, with the null and alternative hypotheses formulated in (1). It is straightforward to show that the likelihood ratio statistic for testing all random effects is as follows

$$R = N_T \log(\hat{\sigma}_{0\epsilon}^2) + \frac{1}{\hat{\sigma}_{0\epsilon}^2} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{OLS})^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{OLS}) - \log(|\hat{\mathbf{V}}|) - (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS})^T \hat{\mathbf{V}}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS}), \quad (4)$$

where  $\hat{\sigma}_{0\epsilon}^2$  is the estimate of  $\sigma_\epsilon^2$  under the null hypothesis,  $\hat{\mathbf{V}}$  is the estimate of  $\mathbf{V}$  in the presence of all random effects (i.e., under the alternative hypothesis), and

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{GLS} &= (\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{Y} \\ \hat{\boldsymbol{\beta}}_{OLS} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \end{aligned} \quad (5)$$

are, respectively, the generalised least squares (GLS) and ordinary least squares (OLS) estimates of  $\boldsymbol{\beta}$ .

To approximate the null distribution of the likelihood ratio test statistic, we use the permutation procedure of Drikvandi et al.<sup>2</sup> which they applied to a non-parametric test statistic. If we define  $\mathbf{Y}^* = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$  then  $\mathbf{Y}^* = \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}$ . Since under  $H_0 : \boldsymbol{\Sigma} = \mathbf{0}$  all the random effects  $\mathbf{b}$  disappear from the model, we get  $\mathbf{Y}^* = \boldsymbol{\epsilon}$  under the null. Thus, all elements of  $\mathbf{Y}^*$  are i.i.d., and hence exchangeable, under the null hypothesis. The exchangeability allows us to perform a permutation test using the adjusted data  $\mathbf{Y}^*$ . However,  $\mathbf{Y}^*$  depends on the unknown parameters  $\boldsymbol{\beta}$  which we need to estimate. We replace  $\boldsymbol{\beta}$  with  $\hat{\boldsymbol{\beta}}_{GLS}$  in (5) to get  $\hat{\mathbf{Y}}^* = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS}$ , which are actually the residuals. The elements of  $\hat{\mathbf{Y}}^*$  are not i.i.d. anymore; however, they are exchangeable under the null hypothesis<sup>2</sup>. Following Drikvandi et al.<sup>2</sup>, we permute the residuals  $\hat{Y}_{ij}^*$  among subjects for each measurement time  $j$ .

Using the likelihood ratio test statistic (4) and the above permutation procedure of Drikvandi et al.<sup>2</sup>, we set up the permutation test for testing all random effects as follows:

1. Compute the likelihood ratio test statistic (4) for the original sample and denote it by  $R_{obs}$ .
2. Obtain a permutation sample by randomly permuting the residuals  $\hat{Y}_{ij}^*$  among subjects for each  $j$ , and compute the test statistic (4) for the permutation sample.
3. Repeat the above step  $B$  times, to get  $B$  test statistics, say,  $R^{(b)}$ ,  $b = 1, \dots, B$ .
4. Calculate the empirical  $p$ -value being the proportion of  $R^{(b)}$  exceeding  $R_{obs}$ .
5. Given the significant level  $\alpha$ , reject  $H_0$  if the empirical  $p$ -value is smaller than  $\alpha$ .

### 3.2. Likelihood-based permutation test for testing a subset of random effects

It is often of interest to test whether or not a subset of random effects are needed in the model. This is a more challenging testing problem because the boundary issue imposes additional constraints and conditions for obtaining the weights of the asymptotic mixture distribution<sup>11</sup>.

Let us rewrite the LME model (2) as follows

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i^{(1)}\mathbf{b}_i^{(1)} + \mathbf{Z}_i^{(2)}\mathbf{b}_i^{(2)} + \boldsymbol{\varepsilon}_i, \quad (6)$$

where  $\mathbf{b}_i^{(1)}$  and  $\mathbf{b}_i^{(2)}$  are, respectively,  $r \times 1$  and  $(q-r) \times 1$  vectors of random effects, and  $\mathbf{Z}_i^{(1)}$  and  $\mathbf{Z}_i^{(2)}$  are  $n_i \times r$  and  $n_i \times (q-r)$  random-effects design matrices respectively. Also, let

$$Cov(\mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)}) = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

To test whether the random effects  $\mathbf{b}_i^{(2)}$  can be left out from the model, whilst retaining the random effects  $\mathbf{b}_i^{(1)}$ , we need to perform the following test

$$\left\{ \begin{array}{l} H_0 : \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \\ H_1 : \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{bmatrix} \end{array} \right.$$

We again use the likelihood ratio statistic as our test statistic; however, we need to change the definition of  $\mathbf{Y}^*$  in order to make the permutation procedure in the previous section applicable to



the above test for testing a subset of random effects. For this, we define  $\mathbf{Y}^* = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}^{(1)}\mathbf{b}^{(1)}$ , where  $\mathbf{Z}^{(1)} = \text{Diag}(\mathbf{Z}_1^{(1)}, \dots, \mathbf{Z}_N^{(1)})$  and  $\mathbf{b}^{(1)} = (\mathbf{b}_1^{(1)T}, \dots, \mathbf{b}_N^{(1)T})^T$ . Equivalently, this can be written as  $\mathbf{Y}^* = \mathbf{Z}^{(2)}\mathbf{b}^{(2)} + \boldsymbol{\varepsilon}$ , where  $\mathbf{Z}^{(2)} = \text{Diag}(\mathbf{Z}_1^{(2)}, \dots, \mathbf{Z}_N^{(2)})$  and  $\mathbf{b}^{(2)} = (\mathbf{b}_1^{(2)T}, \dots, \mathbf{b}_N^{(2)T})^T$ . Since the random effects  $\mathbf{b}^{(2)}$  vanish under  $H_0 : \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ , we get  $\mathbf{Y}^* = \boldsymbol{\varepsilon}$  under the null hypothesis. Thus, all elements of the new  $\mathbf{Y}^*$  are i.i.d., and hence exchangeable, under the above null hypothesis. Because  $\mathbf{Y}^*$  involves  $\boldsymbol{\beta}$  and  $\mathbf{b}^{(1)}$  which are unknown, we replace the fixed-effects parameters  $\boldsymbol{\beta}$  by  $\hat{\boldsymbol{\beta}}_{GLS}$  and the random effects  $\mathbf{b}^{(1)}$  by their empirical Bayes estimates  $\hat{\mathbf{b}}^{(1)}$ <sup>5</sup> to obtain  $\hat{\mathbf{Y}}^* = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS} - \mathbf{Z}^{(1)}\hat{\mathbf{b}}^{(1)}$ , which are also exchangeable under the null hypothesis<sup>2</sup>. We then use the new  $\hat{\mathbf{Y}}^*$  to perform the permutation test for the subset of random effects  $\mathbf{b}_i^{(2)}$ .

## 4. The Bayesian test

Saville and Herring<sup>3</sup> introduced a Bayesian test for testing random effects using Bayes factors with Laplace approximation and suggested a reparameterisation to overcome the boundary issue. They used the Nelder-Mead simplex algorithm<sup>15</sup> to estimate the posterior mode for Laplace approximation. This algorithm sometimes results in a saddle point instead of optimum point in high dimensional approximations, which causes problems in testing multiple random effects. To overcome this difficulty, we propose to use the bound constrained optimisation method of Byrd et al.<sup>16</sup> which effectively avoids such problem in estimating the posterior mode. This is very helpful for testing multiple random effects.

### 4.1. Reparameterisation of LME model for the Bayesian approach

Recall the LME model (2), where we again assume that the random effects  $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})^T$  follow the multivariate normal distribution  $\mathcal{N}(0, \boldsymbol{\Sigma})$ . The following reparameterisation, which involves a factorisation of the covariance matrix  $\boldsymbol{\Sigma}$ , was suggested by Chen and Dunson<sup>17</sup> and Saville and Herring<sup>3</sup>

$$\boldsymbol{\Sigma} = \sigma_\varepsilon^2 \mathbf{D} = \begin{bmatrix} \lambda_1^2 \sigma_\varepsilon^2 & \gamma_{12} \sigma_\varepsilon^2 \sqrt{\lambda_1 \lambda_2} & \dots \\ \vdots & \ddots & \\ \gamma_{1q} \sigma_\varepsilon^2 \sqrt{\lambda_1 \lambda_q} & & \lambda_q^2 \sigma_\varepsilon^2 \end{bmatrix},$$

where  $\gamma_{jk}$  represents the correlation between the  $j$ -th and  $k$ -th random effects, and the scalar  $\lambda_h$  controls the contribution of the  $h$ -th random effect.

The LME model (2) can then be decomposed and reexpressed as follows<sup>17</sup>:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\Lambda}\boldsymbol{\Gamma}\mathbf{b}_i + \varepsilon_i, \quad (7)$$

where  $\boldsymbol{\Lambda} = \text{Diag}(\lambda_1, \dots, \lambda_q)$  and  $\boldsymbol{\Gamma}$  is a  $q \times q$  lower triangular matrix with 1 in the diagonal entries and  $\gamma_{jk}$  in the lower off-diagonal entries. Note that  $\boldsymbol{\Lambda}\boldsymbol{\Gamma}$  is the lower triangular Cholesky decomposition of  $\mathbf{D}$  such that  $\mathbf{D} = \boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\Lambda}^T\boldsymbol{\Gamma}^T$ .

## 4.2. Laplace approximation for calculation of the Bayes factor

Let  $\boldsymbol{\theta} = (\boldsymbol{\zeta}, \mathbf{b}, \sigma_\varepsilon^2)$  represent all the model parameters, where  $\boldsymbol{\zeta} = (\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma})$ , in which  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)$  and  $\boldsymbol{\gamma} = (\gamma_{12}, \dots, \gamma_{q-1,q})$ , contains all parameters other than the random effects  $\mathbf{b}$  and error variance  $\sigma_\varepsilon^2$ . Centring and scaling each of the covariates  $x_i$  by two times the standard deviation of  $x_i$ , Saville and Herring<sup>3</sup> suggested the following prior distributions:

$$\begin{aligned} \boldsymbol{\beta} &\sim \mathcal{N}(\mathbf{0}, 10 \times \mathbf{I}), \\ \sigma_\varepsilon^2 &\sim \text{InvGamma}(v, w), \\ \boldsymbol{\lambda} &\sim \log\mathcal{N}(\log(0.3) \times \mathbf{1}, 2 \times \mathbf{I}), \\ \boldsymbol{\gamma} &\sim \mathcal{N}(\mathbf{0}, \mathbf{1}). \end{aligned} \quad (8)$$

Suppose that  $M_k$ ,  $k = 0, 1$ , represents the model under the null and alternative hypotheses, respectively. Marginalising out  $\mathbf{b}_k$  and  $\sigma_\varepsilon^2$ , we get the marginal likelihood as follows

$$\mathcal{P}(\mathbf{Y}|M_k) = \int \mathcal{P}(\mathbf{Y}|\boldsymbol{\zeta}_k, M_k)\pi(\boldsymbol{\zeta}_k|M_k)d\boldsymbol{\zeta}_k. \quad (9)$$

The marginal likelihood (9), needed for calculation of the Bayes factor, is generally not available in closed form for the LME model (7). Saville and Herring<sup>3</sup> used the following Laplace approximation

$$\hat{\mathcal{P}}(\mathbf{Y}|M_k) = (2\pi)^{\frac{d_k}{2}} |\tilde{\mathbf{H}}_k|^{-\frac{1}{2}} \mathcal{P}(\mathbf{Y}|\tilde{\boldsymbol{\zeta}}_k, M_k)\pi(\tilde{\boldsymbol{\zeta}}_k|M_k), \quad (10)$$

where  $d_k$  is the dimension of integral in the marginal likelihood (9), and  $\tilde{\mathbf{H}}_k$  is the inverse negative Hessian matrix of  $\log\left(\mathcal{P}(\mathbf{Y}|M_k, \boldsymbol{\zeta}_k)\pi(\boldsymbol{\zeta}_k|M_k)\right)$  evaluated at the posterior mode  $\tilde{\boldsymbol{\zeta}}_k$ . They applied the Nelder-Mead simplex algorithm<sup>15</sup> to calculate the posterior mode  $\tilde{\boldsymbol{\zeta}}_k$ , but this algorithm sometimes results in a saddle point instead of optimum point in high dimensional approximations. As a consequence, the Laplace approximation (10) is undefined since the determinant of  $\tilde{\mathbf{H}}_k$  is negative. We propose to use the bound constrained optimisation method of

Byrd et al.<sup>16</sup> which effectively avoids such problem in estimating the posterior mode for testing multiple random effects.

The Laplace approximation also fails when the posterior mode  $\tilde{\zeta}_k$  lies on the boundary<sup>18,19</sup>. This is the case here because  $\lambda_h^2 = 0$  puts the true variance value on the boundary of the parameter space. To overcome this issue, Saville and Herring<sup>3</sup> suggested to use the reparameterisation  $\lambda_h = e^{\phi_h}$ . The diagonal matrix  $\mathbf{\Lambda}$  then needs to be reconstructed as  $\mathbf{\Lambda} = \text{Diag}(e^{\phi_1}, \dots, e^{\phi_q})$ . Also, the prior for  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_q)$  would be  $\mathcal{N}(\log(0.3) \times \mathbf{1}, 2 \times \mathbf{I})$ .

The approximated Bayes factor using the Laplace approximation would then be as follows

$$\hat{BF} = \frac{\hat{\mathcal{P}}(\mathbf{Y}|M_1)}{\hat{\mathcal{P}}(\mathbf{Y}|M_0)},$$

and we interpret it using the scale suggested by Jeffreys<sup>20</sup> and Wasserman<sup>21</sup> to conduct the Bayesian test for random effects.

## 5. Comparison of the permutation test and the Bayesian test

In this section, we conducted simulations to evaluate the performance of the proposed permutation test and compare it with the Bayesian test of Saville and Herring<sup>3</sup> amended by the bound constrained optimisation method of Byrd et al.<sup>16</sup>. Beneficially, our simulations concern both testing all random effects and testing a subset of random effects in LME models. In the simulations we considered 4 cases of tests regarding random effects. We used the sample sizes of  $N = 25, 50, 100, 200$  and the number of repeated measurements  $n_i = 5$ , and set the error variance  $\sigma_\varepsilon^2$  to 1. We generated 1000 data sets, from the LME models considered in the following, to compute the empirical power of the two tests. The Type I error of the two tests were evaluated based on 2000 data sets.

### Case 1: Testing a single random intercept

For this case, we considered the mixed ANOVA model:

$$Y_{ij} = \beta + \lambda b_i + \varepsilon_{ij},$$

where  $\beta$  is the population level fixed-effect intercept and  $b_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$  is a subject-specific random intercept with  $\lambda \geq 0$  controlling the magnitude of  $b_i$  so that the variance of the random effect part is  $\lambda^2 \sigma_\varepsilon^2$ . The hypothesis test for testing the random intercept here would be equivalent to

$$\begin{cases} H_0 : \lambda = 0 \\ H_1 : \lambda > 0. \end{cases}$$

We fixed  $\beta = 0.5$  and considered several different values of  $\lambda = 0, 0.15, 0.3, 0.45, 0.6$ . We used 200 permutation samples in the simulations, as suggested by Fitzmaurice et al.<sup>1</sup> and Biard et al.<sup>22</sup>, to conduct the permutation test at the significance level of  $\alpha = 0.05$ . For the Bayesian test, we chose  $\beta \sim \mathcal{N}(0, 10)$ ,  $\sigma_\varepsilon^2 \sim \text{InvGam}(1, 1)$  and  $\phi \sim \mathcal{N}(\log(0.3), 2)$  in accordance with the priors in (8) used for the reparameterised model. In our initial simulations (not reported here), the Bayesian test with these priors showed a very low Type I error rate compared to the permutation test, which makes the power comparison unfair. As suggested by a referee, we modified the Bayesian test to have a more accurate Type I error rate. As shown by Saville and Herring<sup>3</sup>, one can get a higher Type I error rate for the Bayesian test by choosing a smaller mean for the prior distribution of  $\phi$ . We found that the prior  $\phi \sim \mathcal{N}(\log(0.15), 2)$  works well here. The simulation results, presented in Table 1, suggest that the Type I error rate of both the permutation and Bayesian tests is at the nominal level 0.05, and the two tests have a very similar power in this case. Both tests show higher power when the sample size  $N$  or the scale  $\lambda$  increases.

## Case 2: Testing a random intercept and a random slope simultaneously

For this case, we considered the following LME model with two random effects:

$$Y_{ij} = \beta_1 + b_{i1} + (\beta_2 + b_{i2})t_{ij} + \varepsilon_{ij} \quad (11)$$

where we assume the random intercept and random slope satisfy  $(b_{i1}, b_{i2}) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ , in which  $\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \gamma_{12} \\ \gamma_{12} & \sigma_2^2 \end{bmatrix}$ . The hypothesis test for testing both the random intercept and the random slope simultaneously is follows

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 = 0 \\ H_1 : \sigma_1^2 \text{ and } \sigma_2^2 \text{ are non negative and at least one of them is strictly positive.} \end{cases}$$

Table 1: Type I error rate and power of the permutation and Bayesian tests in testing a single random intercept.

$\lambda$	N = 25		N = 50	
	Bayes test	Permutation test	Bayes test	Permutation test
$\lambda = 0$	4.6	4.8	4.5	4.8
$\lambda = 0.15$	8	8	9	8
$\lambda = 0.3$	10	9	11	9
$\lambda = 0.45$	23	21	23	23
$\lambda = 0.6$	49	49	71	72
$\lambda$	N = 100		N = 200	
	Bayes test	Permutation test	Bayes test	Permutation test
$\lambda = 0$	4.6	4.7	4.7	5.1
$\lambda = 0.15$	8	6	9	10
$\lambda = 0.3$	12	11	18	20
$\lambda = 0.45$	33	34	47	51
$\lambda = 0.6$	92	92	98	99

In the simulations, we fixed  $\beta_1 = 1, \beta_2 = 2, t_{ij} = j$  and set the covariance matrix to  $\Sigma = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$  (to evaluate the Type I error rate),  $\begin{bmatrix} 0.05 & 0.02 \\ 0.02 & 0.05 \end{bmatrix}$ ,  $\begin{bmatrix} 0.08 & 0.02 \\ 0.02 & 0.08 \end{bmatrix}$ ,  $\begin{bmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{bmatrix}$ ,  $\begin{bmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{bmatrix}$  respectively to investigate the power of the permutation and Bayesian tests. We recall that, following Saville and Herring<sup>3</sup>, for the Bayesian test we centred and scaled the repeated measurement time  $t_{ij}$  by two times its standard deviation. Also, similar to Case 1, we considered the prior  $\phi \sim \mathcal{N}(\log(0.15), 2)$  to have a more accurate Type I error rate for the Bayesian test. The simulation results, shown in Table 2, indicate that the two tests have the correct Type I error rate; however, the permutation test seems to be more powerful than the Bayesian test in testing the two random effects here. The power of both tests increases when the sample size or the variance parameters get larger.

Table 2: Type I error rate and power of the permutation and Bayesian tests in testing a random intercept and a random slope simultaneously.

$\Sigma$	N = 25		N = 50	
	Bayes test	Permutation test	Bayes test	Permutation test
$\Sigma = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	4.5	5.2	4.6	5.1
$\Sigma = \begin{bmatrix} 0.05 & 0.02 \\ 0.02 & 0.05 \end{bmatrix}$	14	15	25	30
$\Sigma = \begin{bmatrix} 0.08 & 0.02 \\ 0.02 & 0.08 \end{bmatrix}$	23	22	35	38
$\Sigma = \begin{bmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{bmatrix}$	24	25	37	41
$\Sigma = \begin{bmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{bmatrix}$	35	38	47	50
$\Sigma$	N = 100		N = 200	
	Bayes test	Permutation test	Bayes test	Permutation test
$\Sigma = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	4.7	5.0	4.8	4.9
$\Sigma = \begin{bmatrix} 0.05 & 0.02 \\ 0.02 & 0.05 \end{bmatrix}$	27	40	34	40
$\Sigma = \begin{bmatrix} 0.08 & 0.02 \\ 0.02 & 0.08 \end{bmatrix}$	51	59	52	58
$\Sigma = \begin{bmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{bmatrix}$	59	65	60	68
$\Sigma = \begin{bmatrix} 0.1 & 0.09 \\ 0.09 & 0.1 \end{bmatrix}$	70	74	72	77

### Case 3: Testing a random slope in the presence of a random intercept

For this case, we again considered the LME model (11) which contains two random effects. The hypothesis test for testing the random slope while the random intercept is present in the model would be as follows

$$\begin{cases} H_0 : \sigma_2^2 = 0 \\ H_1 : \sigma_2^2 > 0, \end{cases}$$

where we set  $\sigma_1^2 = 1$  in the simulations. For this test, we set the covariance matrix to  $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$  (to evaluate the Type I error rate),  $\begin{bmatrix} 1 & -0.045 \\ -0.045 & 0.0225 \end{bmatrix}$ ,  $\begin{bmatrix} 1 & -0.09 \\ -0.09 & 0.09 \end{bmatrix}$ ,  $\begin{bmatrix} 1 & -0.135 \\ -0.135 & 0.2025 \end{bmatrix}$ ,

$\begin{bmatrix} 1 & -0.18 \\ -0.18 & 0.36 \end{bmatrix}$ . Unlike Cases 1 and 2, we here used the prior  $\phi \sim \mathcal{N}(\log(0.3), 2)$  for the Bayesian test. The simulation results, reported in Table 3, show that the two test have the correct Type I error rate, and the permutation test is relatively more powerful than the Bayesian test.

Table 3: Type I error rate and power of the permutation and Bayesian tests in testing a random slope in the presence of a random intercept.

$\Sigma$	N = 25		N = 50	
	Bayes test	Permutation test	Bayes test	Permutation test
$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	4.4	5.3	4.7	5.1
$\Sigma = \begin{bmatrix} 1 & -0.045 \\ -0.045 & 0.0225 \end{bmatrix}$	7	8	7	6
$\Sigma = \begin{bmatrix} 1 & -0.09 \\ -0.09 & 0.09 \end{bmatrix}$	10	16	11	11
$\Sigma = \begin{bmatrix} 1 & -0.135 \\ -0.135 & 0.2025 \end{bmatrix}$	10	17	21	23
$\Sigma = \begin{bmatrix} 1 & -0.18 \\ -0.18 & 0.36 \end{bmatrix}$	15	21	25	25
$\Sigma$	N = 100		N = 200	
	Bayes test	Permutation test	Bayes test	Permutation test
$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	4.5	4.9	4.5	5.0
$\Sigma = \begin{bmatrix} 1 & -0.045 \\ -0.045 & 0.0225 \end{bmatrix}$	9	12	6	11
$\Sigma = \begin{bmatrix} 1 & -0.09 \\ -0.09 & 0.09 \end{bmatrix}$	12	16	22	29
$\Sigma = \begin{bmatrix} 1 & -0.135 \\ -0.135 & 0.2025 \end{bmatrix}$	30	34	48	62
$\Sigma = \begin{bmatrix} 1 & -0.18 \\ -0.18 & 0.36 \end{bmatrix}$	51	56	71	83

#### Case 4: Testing two random slopes in the presence of a random intercept

For this case, we considered the following LME model with three random effects:

$$Y_{ij} = \beta_1 + \beta_2 x_{ij} + b_{i1} + b_{i2} z_{ij1} + b_{i3} z_{ij2} + \varepsilon_{ij},$$

where we assume the three random effects follow  $(b_{i1}, b_{i2}, b_{i3}) \sim \mathcal{N}(\mathbf{0}, \Sigma)$ , in which  $\Sigma =$

$$\begin{bmatrix} \sigma_1^2 & \gamma_{12} & \gamma_{13} \\ \gamma_{12} & \sigma_2^2 & \gamma_{23} \\ \gamma_{13} & \gamma_{23} & \sigma_3^2 \end{bmatrix}. \text{ The hypothesis test for testing the two random slopes while the random}$$

intercept is present in the model would be as follows

$$\begin{cases} H_0 : \sigma_2^2 = 0, \sigma_3^2 = 0 \\ H_1 : \sigma_2^2 \text{ and } \sigma_3^2 \text{ are non negative and at least one of them is strictly positive,} \end{cases}$$

where we set  $\sigma_1^2 = 1$  in the simulations. For this test, we generated all the covariates  $x_{ij}, z_{ij1}$  and  $z_{ij2}$  from the uniform distribution  $\mathcal{U}(0, 1)$ . For simplicity in the simulations, we assumed that the random intercept  $b_{i1}$  is uncorrelated with the random slopes  $b_{i2}$  and  $b_{i3}$  (i.e.,  $\gamma_{12} =$

$\gamma_{13} = 0$ ). We set the covariance matrix to  $\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$  (to evaluate the Type I error rate),

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.02 & 0.01 \\ 0 & 0.01 & 0.02 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0.1 \\ 0 & 0.1 & 0.5 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.2 \\ 0 & 0.2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{bmatrix}.$$

Similar to Case 3, we used the prior  $\phi \sim \mathcal{N}(\log(0.3), 2)$  for the Bayesian test. The simulation results, shown in Table 4, suggest that the two tests have the correct Type I error rate. Also, the permutation and Bayesian tests show a very similar power in this case. As expected, the power of both tests increases when the sample size or the variance parameters become larger. A referee pointed out that the power values are larger in Table 4 compared to Table 3, and a possible reason might be that the random intercept was assumed to be independent from the two random slopes. We note that the variances are larger in Table 4, and that we test two random slopes in Case 4 compared to Case 3 where only one random slope is examined. In other words, in Case 4 we test for the need of a  $2 \times 2$  block of variance components, so when that is not  $\mathbf{0}$  the two tests are more powerful to detect it.



Table 4: Type I error rate and power of the permutation and Bayesian tests in testing two random slopes in the presence of a random intercept.

$\Sigma$	N = 25		N = 50	
	Bayes test	Permutation test	Bayes test	Permutation test
$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	5.4	4.8	5.3	5.2
$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.02 & 0.01 \\ 0 & 0.01 & 0.02 \end{bmatrix}$	12	11	16	16
$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0.1 \\ 0 & 0.1 & 0.5 \end{bmatrix}$	42	35	47	44
$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.2 \\ 0 & 0.2 & 1 \end{bmatrix}$	63	63	83	79
$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{bmatrix}$	73	69	82	86
$\Sigma$	N = 100		N = 200	
	Bayes test	Permutation test	Bayes test	Permutation test
$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	4.8	4.8	5.2	5.1
$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.02 & 0.01 \\ 0 & 0.01 & 0.02 \end{bmatrix}$	23	23	24	25
$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0.1 \\ 0 & 0.1 & 0.5 \end{bmatrix}$	62	60	85	87
$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.2 \\ 0 & 0.2 & 1 \end{bmatrix}$	93	94	100	100
$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{bmatrix}$	98	100	100	100

## 6. Real data application

We apply both the permutation test and the Bayesian test to a real data set, collected from a study of hyperglycemia and relative hyperinsulinemia conducted in the Pediatric Clinical Research ward of the University of Colorado Medical Centre<sup>23,24</sup>. During the study, standard

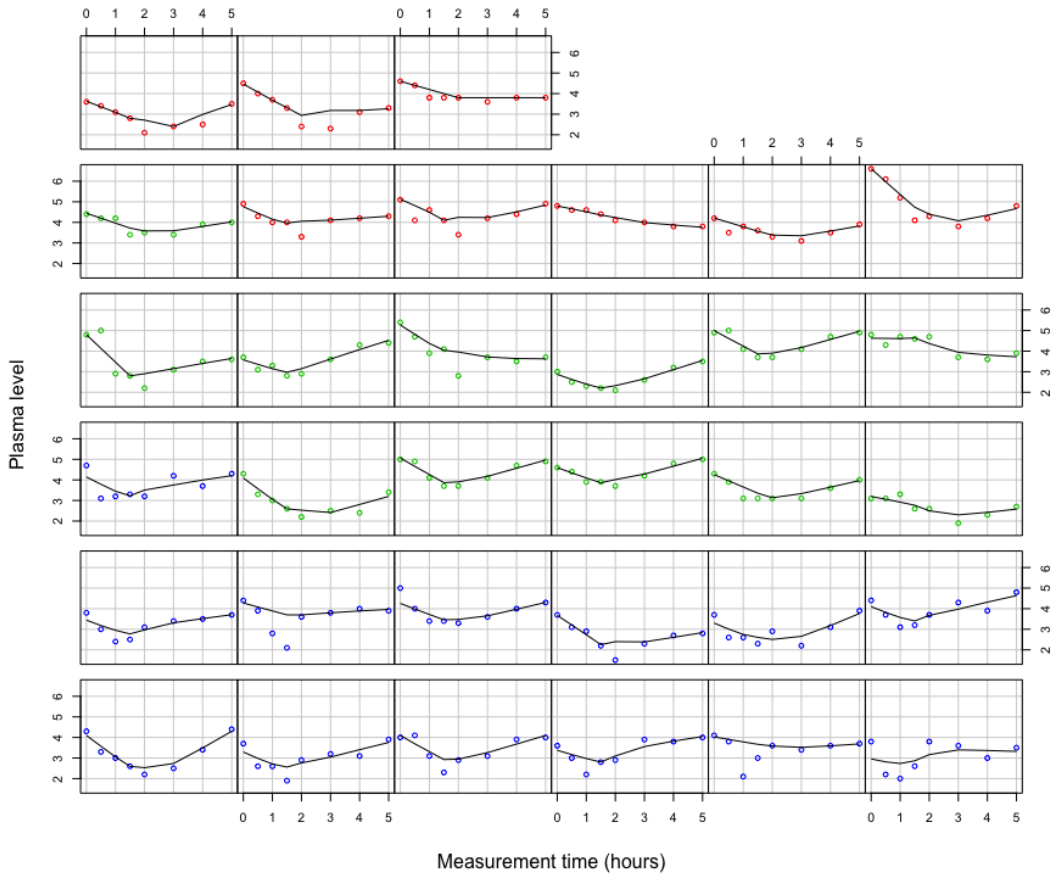


Figure 1: Individual profiles for the control patients (blue), non-hyperinsulinemic obese patients (green) and hyperinsulinemic obese patients (red) in the plasma data.

glucose tolerance tests were given to three groups of patients: 13 control patients, 12 non-hyperinsulinemic obese patients and 8 hyperinsulinemic obese patients. For each patient, the plasma level (plasma inorganic phosphate) was measured repeatedly from blood samples taken at 0, 0.5, 1, 1.5, 2, 3, 4, 5 hour after the glucose test. The main purpose was to study the changes of plasma level over time and see if they are treatment-dependent. The data are presented in Figure 1. The individual profiles show that the patients have different plasma levels at baseline (before the glucose test) as well as different plasma levels over time (after the glucose test). This suggests the use of a mixed-effects model with random intercepts and random effects to account for such heterogeneity among patients.

As suggested by Drikvandi et al.<sup>2</sup>, since the plasma level shows as a quadratic function of

time (see Figure 1), we consider the following mixed-effects model for the data:

$$Y_{ij} = \begin{cases} (\beta_1 + b_{i1}) + (\beta_2 + b_{i2})t_{ij} + (\beta_3 + b_{i3})t_{ij}^2 + \varepsilon_{ij}, & \text{if control,} \\ (\beta_4 + b_{i1}) + (\beta_5 + b_{i2})t_{ij} + (\beta_6 + b_{i3})t_{ij}^2 + \varepsilon_{ij}, & \text{if non-hyperinsulinemic obese,} \\ (\beta_7 + b_{i1}) + (\beta_8 + b_{i2})t_{ij} + (\beta_9 + b_{i3})t_{ij}^2 + \varepsilon_{ij}, & \text{if hyperinsulinemic obese,} \end{cases}$$

where  $Y_{ij}$  is the plasma level for patient  $i$  measured at time  $t_{ij}$ ,  $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9)^T$  is the vector of fixed-effects parameters,  $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$  is the error term,  $b_{i1}$  is a random intercept, and  $b_{i2}$  and  $b_{i3}$  are random slopes for linear and quadratic time effects.

Assuming the multivariate normal distribution for the random effects  $(b_{i1}, b_{i2}, b_{i3})$ , we obtain the maximum likelihood estimate of the random-effects covariance matrix as follows

$$\hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\gamma}_{12} & \hat{\gamma}_{13} \\ \hat{\gamma}_{12} & \hat{\sigma}_2^2 & \hat{\gamma}_{23} \\ \hat{\gamma}_{13} & \hat{\gamma}_{23} & \hat{\sigma}_3^2 \end{bmatrix} = \begin{bmatrix} 0.334 & -0.064 & 0.007 \\ -0.064 & 0.063 & -0.009 \\ 0.007 & -0.009 & 0.001 \end{bmatrix},$$

where the random intercept  $b_{i1}$  has a larger estimated variance compared to the two random slopes  $b_{i2}$  and  $b_{i3}$ .

First, we test if random effects are needed in the model or a model without any random effects is more suitable for analysing the plasma data. The null and alternative hypotheses for this test are

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 0 \\ H_1 : \text{All of } \{\sigma_1^2, \sigma_2^2, \sigma_3^2\} \text{ are all non negative and at least one of them is strictly positive.} \end{cases}$$

Unlike the simulations, we here use 1000 permutations for the proposed permutation test. The permutation test, with the likelihood ratio test statistic of 144.469, gives a  $p$ -value less than 0.001 based on 1000 permutations. So the permutation test rejects the null hypothesis at the significance level of 0.05. The Bayesian test of Saville and Herring<sup>3</sup>, with the priors specified in (8) and with  $\sigma_\varepsilon^2 \sim \text{InvGam}(0.01, 0.01)$ , produces an approximated Bayes factor of  $6.452265 \times 10^{28}$ , which implies that there is very strong evidence in favour of the alternative hypothesis. Therefore, both the permutation and Bayesian tests conclude the suitability of a model with random effects for the plasma data.

Next, because the estimates of variances of the random slopes  $b_{i2}$  and  $b_{i3}$  are small, we test if a model with only random intercepts  $b_{i1}$  is adequate to explain the between-patient

heterogeneity. For this, we test

$$\begin{cases} H_0 : \sigma_2^2 = \sigma_3^2 = 0 \\ H_1 : \sigma_2^2 \text{ and } \sigma_3^2 \text{ are non negative and at least one of them is strictly positive.} \end{cases}$$

For this test, the permutation test, with the likelihood ratio test statistic of 10.6359, produces a  $p$ -value of 0.019 based on 1000 permutations. Hence, the test rejects the null hypothesis at the 5% significance level. The Bayesian test of Saville and Herring<sup>3</sup> gives an approximated Bayes factor of 14.3, which provides a strong evidence in favour of the alternative hypothesis. Therefore, both the permutation and Bayesian tests confirm that there is also significant heterogeneity among patients over time.

Finally, since the estimate of  $\sigma_3^2$  is much smaller than the estimate of  $\sigma_2^2$ , we are interested in testing whether or not the random slope  $b_{i3}$  for quadratic time effects can be removed from the model. The null and alternative hypotheses for this test are as follows

$$\begin{cases} H_0 : \sigma_3^2 = 0 \\ H_1 : \sigma_3^2 > 0. \end{cases}$$

The permutation test, with the likelihood ratio test statistic of 3.77, gives a  $p$ -value of 0.044 based on 1000 permutations. So it rejects the null hypothesis at the significance level of 0.05. The Bayesian test of Saville and Herring<sup>3</sup> produces an approximated Bayes factor of 9.34, which provides some evidence in favour of the alternative hypothesis; however, it is not very strong. We also notice that the likelihood ratio test can be applied to this specific hypothesis for which the asymptotic distribution of the likelihood ratio statistic for testing 2 random effects versus 3 random effects is a 50 : 50 mixture of  $\chi_2^2$  and  $\chi_3^2$ <sup>9</sup>. It produces a  $p$ -value of 0.22 suggesting that the null hypothesis may not be rejected.

We emphasise that the likelihood ratio test is not easily applied to the first two tests considered above because the weights of the asymptotic mixture distribution are unknown. Note also that the permutation tests of Fitzmaurice et al.<sup>1</sup> and Lee and Braun<sup>14</sup> cannot be applied to the above tests on testing subsets of random effects.

The above analysis using the permutation and Bayesian tests suggests the following LME

model for the plasma data

$$Y_{ij} = \begin{cases} (\beta_1 + b_{i1}) + (\beta_2 + b_{i2})t_{ij} + \beta_3 t_{ij}^2 + \varepsilon_{ij}, & \text{if control,} \\ (\beta_4 + b_{i1}) + (\beta_5 + b_{i2})t_{ij} + \beta_6 t_{ij}^2 + \varepsilon_{ij}, & \text{if non-hyperinsulinemic obese,} \\ (\beta_7 + b_{i1}) + (\beta_8 + b_{i2})t_{ij} + \beta_9 t_{ij}^2 + \varepsilon_{ij}, & \text{if hyperinsulinemic obese.} \end{cases}$$

## 7. Conclusions and discussion

The last two paragraphs of the introduction clarify the importance and novelty of our proposed permutation test based on the likelihood ratio statistic, as well as the need for its comparison with the Bayesian test of Saville and Herring<sup>3</sup> that we amended using the bound constrained optimisation method of Byrd et al.<sup>16</sup> We have found both permutation and Bayesian tests effective and powerful for testing random effects in linear mixed-effects models, especially with multiple random effects. Both tests effectively avoid the issues with testing on the boundary of the parameter space. The permutation test avoids the boundary issue by approximating the finite-sample distribution of the likelihood ratio statistic via permuting the residuals under null hypothesis. The Bayesian test bypasses the boundary problem by reparameterising both the LME model and the random-effects covariance matrix and leads to a low dimensional approximation of the Bayes factor using Laplace approximation.

The findings and results of our simulations are summarised below. We should emphasise that these results essentially apply to the range of scenarios considered.

- **Testing a single random effect:** Both permutation and Bayesian tests perform equally well and produce reasonably high powers in detecting a significant random effect.
- **Testing all random effects:** While the two tests show high power in detecting significant random effects, the permutation test is relatively more powerful than the Bayesian test for testing all random effects.
- **Testing a single random effect in the presence of other random effects:** The permutation test is relatively more powerful than the Bayesian test, though both tests perform well in finding a significant random effect in the presence of other random effect.
- **Testing multiple random effects in the presence of other random effects:** Both permutation and Bayesian tests perform equally well and produce reasonably high powers in detecting significant random effects.

Overall, the permutation test showed at least as good and sometimes superior power to the Bayesian test, at least within the range of scenarios considered. We note that it is essential to calibrate the Type I error rate of the Bayesian test (via prior distributions) to be close to the nominal level. This complication may provide an argument for using the permutation test in practice, as the permutation test does not require a simulation study in order to calibrate the Type I error rate.

It should be pointed out that both tests require distributional assumptions which should be carefully checked when applying these tests. The permutation procedure itself does not need any distributions, however calculating the likelihood ratio statistic requires the normality assumption. For the permutation test, the appropriateness of the normality assumption on random effects can be checked by the diagnostic tools recently developed by Drikvandi et al.<sup>8</sup> and Efendi et al.<sup>25</sup>. For the Bayesian test, the sensitivities of the test result to the assumed priors can be assessed using a sensitivity analysis, as explained in Saville and Herring<sup>3</sup>.

While the Bayesian test is already extended to multilevel models by Saville et al.<sup>26</sup>, it would be of interest to generalise the permutation test to multilevel models. In the paper, we mainly studied the two-level longitudinal data, however the permutation procedure can be extended to situations when we have longitudinal data with three or more levels, as also suggested by Fitzmaurice et al.<sup>1</sup>. The idea would be, when testing the variability within each level, one simply permutes the indices corresponding to that level. A referee pointed to grouped or clustered data, such as data collected from a multicenter trial, where there is no measurement time  $j$ . For such situations, the interest is usually in testing the between-cluster variability, so one can implement the permutation test by randomly permuting the individuals between clusters while keeping the number of individuals within a cluster the same as in the original sample. Note that the extension of the permutation procedure may not be transparent to more complex situations such as nested data structures with crossed random effects models<sup>1</sup>.

Finally, we have implemented both the permutation test and the Bayesian test in R and produced efficient codes for users. The R codes for all our simulations and real data analysis are available at “<https://github.com/spaghettidog/thesis>”.

## References

1. Fitzmaurice, G. M., Lipsitz, S. R., and Ibrahim, J. G. A note on permutation tests for variance components in multilevel generalized linear mixed models. *Biometrics*, 63:942–946, 2007.

2. Drikvandi, R., Verbeke, G., Khodadadi, A., and Partovi Nia, V. Testing multiple variance components in linear mixed-effects models. *Biostatistics*, 14:144–159, 2013.
3. Saville, B. R. and Herring, A. H. Testing random effects in the linear mixed model using approximate Bayes factors. *Biometrics*, 65:369–376, 2009.
4. Laird, N. M. and Ware, J. H. Random-effects models for longitudinal data. *Biometrics*, 38: 963–974, 1982.
5. Verbeke, G. and Molenberghs, G. *Linear mixed models for longitudinal data*. Springer Series in Statistics. New York: Springer, 2000.
6. Crainiceanu, C. M. and Ruppert, D. Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society Series B*, 66:165–185, 2004.
7. Heagerty, P. J. and Kurland, B. F. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika*, 88:973–985, 2001.
8. Drikvandi, R., Verbeke, G., and Molenberghs, G. Diagnosing misspecification of the random-effects distribution in mixed models. *Biometrics*, 73:63–71, 2017.
9. Self, S. G. and Liang, K.-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82:605–610, 1987.
10. Shapiro, A. Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review*, 56:49–62, 1988.
11. Stram, D. O. and Lee, J. W. Variance components testing in the longitudinal mixed effects model. *Biometrics*, 50:1171–1177, 1994.
12. Drikvandi, R., Khodadadi, A., and Verbeke, G. Testing variance components in balanced linear growth curve models. *Journal of Applied Statistics*, 39:563–572, 2012.
13. Sinha, S. Bootstrap tests for variance components in generalized linear mixed models. *The Canadian Journal of Statistics*, 37:219–234, 2009.
14. Lee, O. E. and Braun, T. M. Permutation tests for random effects in linear mixed models. *Biometrics*, 68:486–493, 2012.

15. Nelder, J. A. and Mead, R. A simplex method for function minimization. *The computer journal*, 7:308–313, 1965.
16. Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16:1190–1208, 1995.
17. Chen, Z. and Dunson, D. B. Random effects selection in linear mixed models. *Biometrics*, 59:762–769, 2003.
18. Erkanli, A. Laplace approximations for posterior expectations when the mode occurs at the boundary of the parameter space. *Journal of the American Statistical Association*, 89: 250–258, 1994.
19. Pauler, D. K., Wakefield, J. C., and Kass, R. E. Bayes factors and approximations for variance component models. *Journal of the American Statistical Association*, 94:1242–1253, 1999.
20. Jeffreys, H. *The theory of probability*. Oxford University Press, 1961.
21. Wasserman, L. Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, 44:92–107, 2000.
22. Biard, L., Porcher, R., and Resche-Rigon, M. Permutation tests for centre effect on survival endpoints with application in an acute myeloid leukaemia multicentre study. *Statistics in Medicine*, 33:3047–3057, 2014.
23. Zerbe, G. O. Randomization analysis of the completely randomized design extended to growth and response curves. *Journal of the American Statistical Association*, 74:215–221, 1979.
24. Zerbe, G. O. and Murphy, J. R. On multiple comparisons in the randomization analysis of growth and response curves. *Biometrics*, 42:795–804, 1986.
25. Efendi, A., Drikvandi, R., Verbeke, G., and Molenberghs, G. A goodness-of-fit test for the random-effects distribution in mixed models. *Statistical Methods in Medical Research*, 26: 970–983, 2017.
26. Saville, B. R., Herring, A. H., and Kaufman, J. S. Assessing variance components in multilevel linear models using approximate Bayes factors: a case-study of ethnic disparities in birth weight. *Journal of the Royal Statistical Society Series A*, 174:785–804, 2011.