A closed-form estimator for meta-analysis and surrogate markers evaluation

Peer-reviewed author version

FLOREZ POVEDA, Alvaro; MOLENBERGHS, Geert; VERBEKE, Geert & Abad, Ariel Alonso (2019) A closed-form estimator for meta-analysis and surrogate markers evaluation. In: Journal of Biopharmaceutical Statistics, 29 (2), p. 318-332.

DOI: 10.1080/10543406.2018.1535504
Handle: http://hdl.handle.net/1942/30008

A closed-form estimator for meta-analysis and surrogate markers evaluation

Alvaro J. Flórez[1], Geert Molenberghs[1,2], Geert Verbeke[2,1] and Ariel Alonso Abad[2]

[1] *I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium.*

[2] *I-BioStat, KU Leuven, B-3000 Leuven, Belgium.*

Abstract

Estimating complex linear mixed models using an iterative full maximum likelihood estimator can be cumbersome in some cases. With small and unbalanced datasets, convergence problems are common. Also, for large datasets, iterative procedures can be computationally prohibitive. To overcome these computational issues, an unbiased two-stage closed-form estimator for the multivariate linear mixed model is proposed. It is rooted in pseudo-likelihood-based split-sample methodology, and useful, for example, when evaluating normally distributed endpoints in a meta-analytic context. However, applications go well beyond this framework. Its statistical and computational performance is assessed via simulation. The method is applied to a study in schizophrenia.

Keywords: Hierarchical data; Linear mixed model; Unequal cluster size; Surrogacy evaluation; Weighting.

1 Introduction

One of the most often used techniques to analyze clustered or hierarchical continuous data is the linear mixed model (LMM; Laird and Ware 1983; Verbeke and Molenberghs 2000). Broadly speaking, clustered data refers to a set of measurements collected from structured units, and embraces different settings, e.g., longitudinal, multilevel, and spatial data. Iterative techniques, such as Newton-Raphson, are commonly used to fit the linear mixed model (Lindstrom and Bates, 1988). Based on an initial guess for the parameters, such algorithms iteratively update these values until a convergence criterion is reached. They often perform well. However, in some cases convergence is either not achieved, or is extremely time-consuming. This more frequently occurs in complex models with many variance components, in very small, or extremely large datasets.

One of the typical cases where a combination of a complex LMM and a small dataset is commonly encountered is in the meta-analytic evaluation of surrogate endpoints proposed by Buyse et al. (2000) and studied in subsequent papers. Computational issues were examined for the first time in Tibaldi et al. (2003). In this approach, a surrogate endpoint is evaluated to asses whether it can act as a replacement outcome for a true endpoint (the most credible indicator of drug response) using multiple trials. A surrogate is convenient when it can be measured earlier, more frequently, or more cheaply than the true endpoint. The LMM framework allows assessing surrogacy at two different levels: trial and individual. Surrogacy at the individual level is defined as the association between the surrogate and the true endpoint after adjustment for the treatment and trial effect; surrogacy at the trial level is the association between the treatment effect on the surrogate and the treatment effect on the true endpoint. Extensive overviews on the meta-analytic evaluation of surrogate endpoints can be found in two books on the topic (Burzykowski et al., 2005, and Alonso Abad et al., 2016).

After fitting a LMM, surrogacy is often quantified using two metrics that are based on the variance-covariance matrix of random effects (D) and residuals (Σ), meaning that reliable convergence is needed, as well as positive-definiteness of D and Σ. Based on simulation

studies, convergence problems occur more often with few and highly unbalanced trials, and when the between-cluster variability is relatively small when compared to the residual variability (Burzykowski et al., 2005; Van der Elst et al., 2016).

There are computational issues, though, that need careful addressing. The linear mixed model is routinely fitted (iteratively) by maximizing the marginal likelihood, using Newton-Raphson-based techniques for example. These algorithms require substantial computing resources to calculate the log-likelihood function and its derivatives (among other quantities) in each iteration. Of course, it is tractable with medium to relatively large data. However, with a very large number of clusters and cluster-sizes, these procedures can be too time-consuming or computationally infeasible. Another problem of maximization procedures, is that they may fail to converge or converge to an spurious solution, e.g., to a local maximum or to a solution outside the parameter space. This is common when we are analyzing few and very unbalanced clusters.

To remove or at least alleviate these computational difficulties, we propose a non-iterative unbiased estimator for the multivariate linear mixed model. It requires less computing resources, takes only one step to find the solution, making it fast. Secondly, given that it is non-iterative, it does not suffer from convergence issues.

While useful for surrogacy evaluation, it is by no means restricted to that setting. Our estimator is based on so-called split-sample methodology and pseudo-likelihood (Molenberghs et al., 2011, 2014, 2018). Here, the sample is conveniently divided into subsamples, and the parameters are estimated in each one. Thereafter, the resulting estimates are averaged, using some weights, to obtain an overall estimator. In our case, each subsample contains one single trial, leading to the so-called trial-by-trial or cluster-by-cluster estimator (Molenberghs et al., 2018). This method has been shown to exhibit good statistical performance and computational efficiency to analyze data with different clustering structures, such as autoregressive (AR) (Hermans et al., 2017) and compound-symmetry (Molenberghs et al., 2018). For the latter, the cluster-by-cluster estimator is consistent when the number of replicates per cluster increases more rapidly than the number of clusters. In these models,

all parameters can be estimated in each cluster and, therefore, the split-sample method is applied directly. In the general linear mixed model, the variance-covariance matrix of the random effects captures the between-cluster variability, and it cannot be estimated using a single cluster. Consequently, we propose an unbiased (method-of-moments) estimator.

To evaluate the statistical performance of our estimator, we performed a simulation study to asses efficiency. As a reference, our proposal is compared to classical restricted maximum likelihood (REML) iterative estimator and others alternatives. Furthermore, we assessed the computation time consumed to fit the model using both techniques. The latter is a factor to take into account when the dataset is very large.

The remainder of the paper is organized as follows. In Section 2, the multivariate linear mixed model, along with the model usually used to assess surrogacy, is introduced and model fitting alternatives are presented. Section 3 is dedicated to the trial-by-trial estimator. Section 4 presents the simulation settings used to evaluate our estimator's (statistical and computational) performance. Results of the simulation study are shown in Section 5. A case study in schizophrenia is analyzed in Section 6. Finally, Section 7 is reserved for conclusions and discussion.

2 Multivariate linear mixed model

Defining $Y_i = (\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{im})$ as the $(n_i \times m)$ matrix of multivariate outcomes for cluster i, $i = 1, \ldots, N$, where $\boldsymbol{y}_{ik}$ is the n_i-dimensional vector of the kth outcome, the multivariate linear mixed model (LMMM) is expressed as follows (Shah et al., 1997):

$$Y_i = X_i B + Z_i E_{\boldsymbol{b}_i} + E_{\boldsymbol{\varepsilon}_i}, \tag{1}$$

where X_i $(n_i \times p)$ and Z_i $(n_i \times q)$ are matrices of known covariates; vec$B = \boldsymbol{\beta}$ is the pm-dimensional vector of fixed effects; vec $E_{\boldsymbol{b}_i} = \boldsymbol{b}_i$ the corresponding qm-dimensional vector of random effects; and vec $E_{\boldsymbol{\varepsilon}_i} = \boldsymbol{\varepsilon}_i$ the residual term. $\otimes$ signifies the Kronecker product and vec the vec-operator, which transforms a matrix into a vector by stacking

the columns of the matrix one underneath the other (more details on this operator can be found in Appendix A in the Supplementary Materials). Generally, we assume that $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \Sigma \otimes I_{n_i})$ and $\boldsymbol{b}_i \sim N(\mathbf{0}, D)$, where Σ and D are $(m \times m)$ and $(q \times q)$ unstructured variance-covariance matrices, respectively.

From (1) we see that, conditionally on the random effects, $\underset{\sim}{\boldsymbol{Y}}_i = \text{vec } Y_i$ is normally distributed with mean vector $(I_m \otimes X_i)\boldsymbol{\beta} + (I_m \otimes Z_i)\boldsymbol{b}_i$ and with variance-covariance matrix $\Sigma \otimes I_{n_i}$. Further, the marginal distribution of $\underset{\sim}{\boldsymbol{Y}}_i$ is,

$$\underset{\sim}{\boldsymbol{Y}}_i \sim N\left[(I_m \otimes X_i)\,\boldsymbol{\beta}, V_{\underset{\sim}{\boldsymbol{Y}}_i} = (I_m \otimes Z_i)\,D\,(I_m \otimes Z_i)' + \Sigma \otimes I_{n_i}\right]. \tag{2}$$

Note that $V_{\underset{\sim}{\boldsymbol{Y}}_i}$ is a function of the variance-covariance matrix of the random effects and residuals, allowing for the analysis of between- and within-cluster variation through D and Σ, respectively.

Given the hierarchical structure of the LMMM, it can be expressed in two stages (Verbeke and Molenberghs, 2000). The first stage corresponds to the "within-cluster" model and is defined as:

$$Y_i = T_i B_i + E_{\boldsymbol{\varepsilon}_i}, \tag{3}$$

where $T = (Z_i \; Z_{ci})$, Z_{ci} is a $(n_i \times r)$ matrix of within-cluster covariates not associated with random effect, B_i is a $[(q + r) \times m]$ matrix of unknown cluster-specific coefficients. The second stage model, called "between-cluster" model, is:

$$B_i = K_i B + (I_q \; 0_{q \times r})' E_{\boldsymbol{b}_i},$$

where K_i is a $[(q+r) \times p]$ matrix of known cluster-specific covariates satisfying $X_i = T_i K_i$.

2.1 Fitting alternatives

Routinely, the full REML/ML estimator of a linear mixed model is obtained using iterative techniques, usually Newton-Raphson-based (NR), Expectation-Maximization (EM) algorithm, or variations thereof. These algorithms maximize the marginal log-(restricted)likelihood

function over the parameter space $\boldsymbol{\Theta}$, i.e., all vectors $\boldsymbol{\theta} = (\boldsymbol{\beta}, \text{vech } \Sigma, \text{vech } D)'$ that lead to positive (semi-)definite D and Σ matrices (Verbeke and Molenberghs, 2000). The advantage of the NR algorithm, over EM, is its fast convergence rate (Lindstrom and Bates, 1988). However, these methods may diverge or converge to values in or outside the parameter space, to a non-positive-definite estimate for D for instance. The latter is not an issue if primary interest is with inferences on the fixed effects, where it is only required that the marginal covariance matrix ($V_{\underset{\sim}{\boldsymbol{Y}}_i}$) be positive-definite for all clusters. But, when interest lies in the variance components, such as in the surrogacy evaluation case, this is needed. As we mentioned before, these computational problems occur more often with few and highly unbalanced clusters, and when D is relatively small compared to Σ (Burzykowski et al., 2005; Van der Elst et al., 2016). To increase the convergence consistency of the NR algorithm, Lindstrom and Bates (1988) suggest the Cholesky root parameterization of D in the maximization process. it guarantees that D is always positive (semi)-definite during the estimation process (West et al., 2014).

To overcome these computational issues, Van der Elst et al. (2016) proposed multiple imputation (MI) to introduce balance in the dataset prior fitting the model along with the Cholesky root parameterization. The former is justified by the fact that proper convergence is more likely with balanced datasets. Moreover, different parametrizations of D simplify the numerical optimization. Based on a simulation study, the authors concluded that this alternative provides good convergence and statistical properties. Nevertheless, there are some limitations. The MI-based estimator can be computationally intensive, especially with a large number of clusters, and it may depend strongly on the imputation model.

3 Trial-by-trial estimator

Our estimator is motivated by meta-analysis, where the clustering variable is trial. Furthermore, it is based on the split-sample method (Molenberghs et al., 2011, 2014, 2018), with a single trial per stratum. Of course, with enough individuals per trial, this splitting

allows estimating $\boldsymbol{\beta}$ and Σ. However, D measures the between-trial variability, and it requires information on more than one trial for it to be estimated. Therefore, its estimation is based on the "trial-specific" estimates of $\boldsymbol{\beta}_i$ and the overall estimate of $\boldsymbol{\beta}$ at the second stage. Reinsel (1985) proposed a (method-of-moments) estimator of D in the univariate context. We extend the estimator to the multivariate model (1) and for a general weighting scheme for the estimator of $\boldsymbol{\beta}$.

The estimator is divided into two stages. At the fist one, we estimate the "trial-specific" parameters ($\boldsymbol{\beta}_i$ and Σ_i) in each trial. In the second stage, these estimates are combined using a weighting scheme to get overall estimates for both parameters, and an unbiased estimator for D is proposed.

The estimator is introduced in the following subsections, assuming that $T_i = Z_i$. Nevertheless, the extension to a general case requires some further but straightforward algebra. More technical details are given in Appendix A of the Supplementary Materials.

3.1 First stage

In first stage, we fit model (3) separately in each trial. Given that, conditionally on $\boldsymbol{b}_i$, the set of outcomes of the same cluster (Y_i) are independent, we use the ordinary least squared (OLS) estimator for $\boldsymbol{\beta}_i$ and Σ_i:

$$\widehat{B}_i = \left(Z_i'Z_i\right)^{-1} Z_i'Y_i \quad \text{and} \quad \widehat{\Sigma}_i = \frac{1}{n_i - q} Y_i' \left[I_{n_i} - Z\left(Z_i'Z_i\right)^{-1} Z_i'\right] Y_i.$$

Defining vec $\widehat{B}_i = \widehat{\boldsymbol{\beta}}_i$, the corresponding sampling variances are:

$$V\left(\widehat{\boldsymbol{\beta}}_i\right) = D + \Sigma \otimes \left(Z_i'Z_i\right)^{-1} \tag{4}$$

and

$$V\left(\text{vech } \widehat{\Sigma}_i\right) = \frac{2}{n_i - q} L_m \left(\Sigma \otimes \Sigma\right) L_m',$$

where vech A is the half-vec-operator applied to a symmetric matrix A. That is, stacking the columns of A into a vector, excluding the duplicate elements. L_m is the elimination

matrix of order m. That is, the matrix which, for any $(m \times m)$ symmetric matrix A, transforms vec A into vech A. More details on the vech operator and L_m can be found in Appendix A of the Supplementary Materials.

3.2 Second stage

At the second stage, to obtain overall estimators for $\boldsymbol{\beta}$ and Σ, the estimates of each trial are averaged as follows:

$$\widetilde{\boldsymbol{\beta}} = \left(\sum_{j=1}^{N} K'_{mj} W_{1j} K_{mj} \right)^{-1} \sum_{i=1}^{N} K'_{mi} W_{1i} \widehat{\boldsymbol{\beta}}_i \quad \text{and} \quad \text{vech } \widetilde{\Sigma} = \sum_{i=1}^{N} W_{2i} \text{vech } \widehat{\Sigma}_i,$$

where W_{1i} and W_{2i} are weighting matrices, and $K_{mi} = (I_m \otimes K_i)$. We considered two different weighting schemes for these parameters because the estimators of $\boldsymbol{\beta}_i$ and Σ_i are independent. Furthermore, the variances of $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\Sigma}$ are:

$$V\left(\widetilde{\boldsymbol{\beta}}\right) = \left(\sum_{j=1}^{N} K'_{mj} W_{1j} K_{mj} \right)^{-1} \left[\sum_{i=1}^{N} K'_{mi} W_{1i} V\left(\widehat{\boldsymbol{\beta}}_i\right) W'_{1i} K_{mi} \right] \left(\sum_{j=1}^{N} K'_{mj} W_{1j} K_{mj} \right)^{-1},$$

and,

$$V\left(\text{vech } \widetilde{\Sigma}\right) = \sum_{i=1}^{N} W_{2i} V\left(\text{vech } \widehat{\Sigma}_i\right) W'_{2i}.$$

There are several weighting schemes possible, where the most common ones are constant weights, $W_i = \frac{1}{N} I$, and proportional weights, $W_i = \frac{n_i}{n_\cdot} I$, where $n_\cdot = \sum_{i=1}^{N} n_i$. To ensure unbiasedness, the $\sum_{i=1}^{N} W_i = I$ constraint is required (Molenberghs et al., 2018).

Using the trial-specific estimates of $\boldsymbol{\beta}_i$, the estimator of D is based on the following matrix:

$$S_b = \sum_{i=1}^{N} (\widehat{\boldsymbol{\beta}}_i - K_{mi}\widetilde{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}_i - K_{mi}\widetilde{\boldsymbol{\beta}})' = \sum_{i=1}^{N} \widetilde{\boldsymbol{b}}_i \widetilde{\boldsymbol{b}}'_i. \tag{5}$$

By equating S_b with its expected value and solving for D, an unbiased (method-of-moments) estimator of D is found (Reinsel, 1985; Vonesh and Carter, 1987). Given that,

$\widetilde{\boldsymbol{b}}_i \sim N\left[\mathbf{0}, V\left(\widetilde{\boldsymbol{b}}_i\right)\right]$, we have that $E\left(S_b\right) = \sum_{i=1}^{N} V\left(\widetilde{\boldsymbol{b}}_i\right)$, where

$$V\left(\widetilde{\boldsymbol{b}}_i\right) = (I - H_{ii}) V\left(\widehat{\boldsymbol{\beta}}_i\right) (I - H_{ii})' + \sum_{k \neq i} H_{ik} V\left(\widehat{\boldsymbol{\beta}}_k\right) H_{ik}', \tag{6}$$

and $H_{ik} = K_{mi}\left(\sum_{j=1}^{N} K_{mj}' W_{1j} K_{mj}\right)^{-1} K_{mk}' W_{1k}$. Then,

$$\begin{aligned} E\left(S_{\boldsymbol{b}}\right) = & \sum_{i=1}^{N} (I - H_{ii}) \left[D + \Sigma \otimes \left(Z_i' Z_i\right)^{-1}\right] (I - H_{ii})' + \\ & \sum_{k \neq i} H_{ik} \left[D + \Sigma \otimes \left(Z_i' Z_i\right)^{-1}\right] H_{ik}'. \end{aligned} \tag{7}$$

Before solving (7) for D, we apply the vec-operator at both sides of the equation:

$$\text{vec } E\left(S_{\boldsymbol{b}}\right) = \sum_{i=1}^{N} \left[(I - H_{ii})^{\otimes 2} + \sum_{k \neq i} H_{ik}{}^{\otimes 2}\right] \text{vec } \left[D + \Sigma \otimes \left(Z_i' Z_i\right)^{-1}\right], \tag{8}$$

where $A^{\otimes 2} = A \otimes A$. Then, solving (8) for D, the following unbiased estimator is found:

$$\text{vec } \widetilde{D} = \left[\sum_{i=1}^{N} (I - H_{ii})^{\otimes 2} + \sum_{k \neq i} H_{ik}{}^{\otimes 2}\right]^{-1} (\text{vec } S_b - \boldsymbol{c}), \tag{9}$$

where

$$\boldsymbol{c} = \sum_{i=1}^{N} \left[(I - H_{ii})^{\otimes 2} + \sum_{k \neq i} H_{ik}{}^{\otimes 2}\right] \text{vec} \left[\widetilde{\Sigma} \otimes \left(Z_i' Z_i\right)^{-1}\right].$$

The variance of vech $\widetilde{D}$ is:

$$\begin{aligned} V\left(\text{vech } \widetilde{D}\right) = & L_{qm} \left[\sum_{i=1}^{N} (I - H_{ii})^{\otimes 2} + \sum_{k \neq i} H_{ik}{}^{\otimes 2}\right]^{-1} \left[V\left(\text{vec } S_b\right) + V\left(\boldsymbol{c}\right)\right] \times \\ & \times \left[\sum_{i=1}^{N} (I - H_{ii})^{\otimes 2} + \sum_{k \neq i} H_{ik}{}^{\otimes 2}\right]^{-1} L_{qm}', \end{aligned} \tag{10}$$

where $V\left(\text{vec } S_b\right)$ and $V\left(\boldsymbol{c}\right)$ are (S.1) and (S.2) in the Supplementary Materials, respectively.

In the random-effects multivariate meta-regression framework, several non-iterative estimators of the between-study covariance have been proposed (Chen et al., 2012; Jackson et al., 2013). Particularly, Jackson et al. (2013) proposed a (method-of-moments) estimator of D similar to (9). However, it assumes that the within-study variability is known, and not estimated as is commonly the case. Therefore, a large number of studies is necessary to justify this approximation. A comparison of our method with the Jackson et al. (2013) estimator via simulation is presented in Section C.1 of the Supplementary Materials.

3.3 Adjustment for non-positive-definite $\widetilde{D}$

The estimator (9) might lead to a non-positive-definite estimate of D. Laird and Lange (1987) and Rousseeuw and Molenberghs (1993) proposed various methods to find the nearest positive-definite matrix, the latter in the context of correlation matrices. The adjustments' purpose is to find the closest positive-definite matrix of the one that is found at first. The eigenvalue method corrects non-positive-definiteness of $\widetilde{D}$ as follows:

$$\widetilde{D}_+ = L\widetilde{E}L',$$

where L is the set of orthonormal eigenvectors of $\widetilde{D}$ and $\widetilde{E}$ is the diagonal matrix of eigenvalues of $\widetilde{D}$, except that all negative values are replaced by a small value $\delta > 0$. When applying such an adjustment, the statistical properties of the estimator of D are affected (unbiasedness and sampling variance). However, based on our simulation study (see Appendix B.2 of the Supplementary Materials), the sampling variance is very close to (10) and there is a, however small, negative bias.

3.4 About weighting scheme

There are several weighting scheme alternatives to estimate $\boldsymbol{\beta}$ and Σ. Some weights have the advantage of being parameter-free. The proportional weights are preferred when the clusters differ largely in size. However, we can also consider so-called optimal weights

(Molenberghs et al., 2018). For estimation of a parameter $\boldsymbol{\theta}$, these take the form:

$$W_i^{opt} = \left(\sum_{k=1}^{N} V_k^{-1} \right)^{-1} V_i^{-1},$$

where V_k is the variance of $\widehat{\boldsymbol{\theta}}_k$. Particularly for $\boldsymbol{\beta}$, the optimal weights are as follows:

$$W_{1i}^{opt} = \left\{ \sum_{k=1}^{N} \left[D + \Sigma \otimes (Z_k' Z_k)^{-1} \right]^{-1} \right\}^{-1} \left[D + \Sigma \otimes (Z_i' Z_i)^{-1} \right]^{-1}. \tag{11}$$

As we can observe in (11), the weighting matrices depend on unknown parameters, and therefore its application is not straightforward. One alternative is to apply an approximation of the optimal weights (Molenberghs et al., 2018). Here, we estimate D and Σ using a simple scheme, such as proportional, and later, these parameters are replaced by their estimates in (11). On the other hand, the optimal weights for Σ are scalar and parameter-free:

$$W_{2i}^{opt} = \frac{n_i - q}{\sum_{k=1}^{N} (n_k - q)} I.$$

For the following simulations, we used proportional and optimal weights for $\boldsymbol{\beta}$ and Σ, respectively. Thereafter, we updated the estimate of $\boldsymbol{\beta}$ using the approximate optimal weights. The latter is suggested if interest lies on the fixed effects.

To improve the estimation of D, we can consider iterated optimal weighting (Molenberghs et al., 2018). To do so, we cycle through estimating $\boldsymbol{\beta}$ using approximate optimal weights and estimating D, until some convergence criterion is reached. In Appendix B.3 of the Supplementary Materials, we show the MSE reduction due to the use of iterated optimal weights for D in each simulated scenario.

One interesting result of the trial-by-trial estimator is that, in the case of balanced data (same size and treatment allocation trials), it is equal to the REML estimator for a linear mixed model. Therefore, the difference of the estimators depends mostly on imbalance.

4 Simulation study

4.1 Model

We consider a surrogate meta-analytic endpoint evaluation setting. Assuming that the surrogate and true endpoints follow a bivariate normal distribution, Buyse et al. (2000) proposed the following model,

$$\begin{cases} S_{ij} = \mu_S + m_{Si} + (\alpha + a_i)Z_{ij} + \varepsilon_{Sij}, \\ T_{ij} = \mu_T + m_{Ti} + (\beta + b_i)Z_{ij} + \varepsilon_{Tij}, \end{cases} \tag{12}$$

where S_{ij} and T_{ij} represent the surrogate and true endpoints for patient j at trial i ($i = 1, \ldots, N,\ j = 1, \ldots, n_i$), respectively; Z_{ij} denotes the indicator variable for the treatment; μ_S, α are the fixed intercept and treatment effect for S; m_{Si}, a_i are the corresponding random intercepts and treatment effects; μ_T, β are the fixed intercept and treatment effect for T; m_{Ti}, b_i are the corresponding random intercepts and treatment effects; and the residuals for S and T are represented by ε_{Sij} and ε_{Tij}, respectively. This model assumes that $(m_{Si}, a_i, m_{Ti}, b_i) \sim N(\mathbf{0}, D)$ and $(\varepsilon_{Sij}, \varepsilon_{Tij}) \sim N(\mathbf{0}, \Sigma)$, where

$$D = \begin{pmatrix} d_{SS} & d_{Sa} & d_{ST} & d_{Sb} \\ & d_{aa} & d_{aT} & d_{ab} \\ & & d_{TT} & d_{Tb} \\ & & & d_{bb} \end{pmatrix} \quad \text{and } \Sigma = \begin{pmatrix} \sigma_{SS} & \sigma_{ST} \\ & \sigma_{TT} \end{pmatrix}$$

are unstructured variance-covariance matrices. Particularly, surrogacy evaluation in this framework is done by two metrics: the trial- and individual-level coefficients of determination, represented by R^2_{Trial} and R^2_{Ind}, computed as:

$$R^2_{\text{Trial}} = \frac{\begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}' \begin{pmatrix} d_{SS} & d_{Sa} \\ d_{Sa} & d_{aa} \end{pmatrix}^{-1} \begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}}{d_{bb}} \quad \text{and } R^2_{\text{Ind}} = \frac{\sigma^2_{ST}}{\sigma_{SS}\sigma_{TT}}.$$

R^2_{Trial} and R^2_{Ind} quantify the strength of the association between the treatment effects on S and T, and between S and T, respectively (Alonso Abad et al., 2016). R^2_{Trial} ranges over the unit interval if the corresponding D matrix is positive-definite (Burzykowski et al., 2005).

As mentioned before, fitting model (12) is often surrounded by computational issues (Burzykowski et al., 2005; Van der Elst et al., 2016). To address this problem, Tibaldi et al. (2003) proposed different simplifications, such as removing the random effects for the intercept or a simpler fixed-effects model.

The meta-analytic approach to evaluate surrogacy under normallity, as is shown here, can be seen as a special case of the information-theoretic framework at both levels (Van der Elst et al., 2017). Alonso and Molenberghs (2007) proposed to assess surrogacy in terms of uncertainty reduction. Here, S is a good surrogate for T at the individual (trial) level, if the uncertainty about T (the treatment effect on T) is reduced substantially when S (the treatment effect on S) is known (Molenberghs et al., 2010). One of its advantages is that a hierarchical joint model, such as (12), is not needed. Therefore, it generally requires less complicated models to fit. For details, see Alonso and Molenberghs (2007) and Van der Elst et al. (2017) [Chapters 9 and 10].

4.2 Settings

The parameter values are the same as used by Van der Elst et al. (2016) in their simulation study. For the fixed effects we used $\boldsymbol{\beta} = (\mu_S, \alpha, \mu_T, \beta)' = (450, 300, 500, 500)'$ and for the variance-covariance matrices for the random effects and error term,

$$D = \begin{pmatrix} 100 & 0 & 40 & 0 \\ & 100 & 0 & 70.7107 \\ & & 100 & 0 \\ & & & 100 \end{pmatrix} \text{ and, } \Sigma = \begin{pmatrix} 300 & 212.132 \\ & 300 \end{pmatrix},$$

respectively. This setting leads to $R^2_{\text{Trial}} = R^2_{\text{Ind}} = 0.5$. We varied the number of trials (N), the mean number of patients per trial (μ_n), at first fixing $N = 10$ and varying $\mu_n = \{10, 25, 50, 100, 150, 200,250\}$, and later, fixing $\mu_n = 10$ and varying $N = \{10, 25, 50, 100, 150, 200, 250\}$. The standard deviation of the number of patients per trial was determined as a fraction of the mean, that is $\sigma_n = \gamma\mu_n$. We considered two cases: high imbalance ($\gamma = 0.25$) and extremely high imbalance ($\gamma = 0.5$). These settings lead to 28 different scenarios.

For each scenario, we simulated the data as follows:

1. Number of patients per trial using a normal distribution, $n_i \sim N\left[\mu_n, (\gamma\mu_n)^2\right]$ (rounding n_i to the nearest integer); with the minimum number of patients per trial equals to five.

2. Treatment allocation for each trial using a binomial distribution, $n_{1i} \sim \text{Binomial}(n_i, 0.5)$; with a minimum number of two patients per treatment arm; that is, $\min(n_{1i}) = 2$ and $\max(n_{1i}) = n_i - 2$.

3. Outcomes for S_{ij} and T_{ij} following model (12).

A total of 1,000 datasets were generated for each scenario. Then, the simulated datasets were analyzed using different methods:

- the trial-by-trial estimator, proposed in Section 3, using approximate optimal and optimal weights to estimate $\boldsymbol{\beta}$ and Σ, respectively;

- the iterative REML estimator using the Cholesky decomposition of D suggested by Van der Elst et al. (2016), to enhance the rate of proper convergence;

- Particularly for surrogacy evaluation metrics, we used the information-theoretic framework based on a fixed-effects model (Alonso and Molenberghs, 2007).

To evaluate the computation time we increased the values of N and μ_n as follows: fixing $\mu_n = 250$ and varying $N = \{500, 1000, 2{,}500, 5{,}000\}$ and fixing $N = 100$ and varying $\mu_n = \{500, 1{,}000, 1{,}500, 2{,}500\}$. For this, we simulated 25 datasets per scenario.

During the simulation process we evaluated three different aspects:

- The percentage of positive-definite estimates of D;
- The asymptotic relative efficiency (ARE), defined as the MSE ratio of the alternative estimator (trial-by-trial or information-theoretic method) over the iterative REML estimator, for all parameters of model (12) and its derived quantities. For $\boldsymbol{\beta}$, Σ and D, we computed an overall ARE;
- the computation time, in seconds, needed to obtain the solution for the trial-by-trial and iterative REML estimator.

5 Simulation results

5.1 Positive-definite estimates of D

Table 1 shows that the percentage of positive-definite (pd) solutions of D increases when the number of trials and/or patients per trial gets larger. Furthermore, it is lower when the imbalance is higher. These results agree with the ones obtained by Van der Elst et al. (2016), showing that convergence problems are encountered mostly in unbalanced small data cases. Comparing both estimators, the proportion of pd estimates of D is slightly larger for the trial-by-trial than for the iterative REML estimator, in most of the settings. Furthermore, the former allows a direct use of the adjustments for non-positive-definiteness presented in Section 3.2.

5.2 Asymptotic relative efficiency

Table 2(a) displays the asymptotic relative efficiency (ARE) of the trial-by-trial estimator and the information-theoretic method in the cases where the number of trials (N) is fixed at 10 and the mean of patients per trial (μ_n) increases, for both high and extremely high imbalance. In the same way, Table 2(b) exhibits the ARE when μ_n is bounded at 10 and N varies. For the iterative REML, only the pd $\widehat{D}$ were taken into account to compute the

MSE. Meanwhile, for the trial-by-trial estimator, all cases were used, including when the adjustment for non-positive-definiteness was required.

As shown in Table 2, the trial-by-trial estimator performs differently depending on the parameter. Regarding $\boldsymbol{\beta}$, it is as efficient as the iterative REML estimator, its ARE is stable around one, even with small values of N and μ_n, and extremely large imbalance. For Σ, there is efficiency loss. Nevertheless, it decreases rapidly as N and/or μ_n increase. In the smallest setting, its MSE is around 12% larger than the iterative REML estimator. Moreover, the efficiency is slightly affected by imbalance. Concerning D, it is asymptotically efficient when μ_n increases faster than N. On the contrary, when N increases faster than μ_n, it does not seem to be asymptotically as efficient as its iterative counterpart. The ARE converges to a value above one. However, the efficiency loss is not greater than 10% even in the extremely large imbalance settings.

For R^2_{Ind}, the ARE of the trial-by-trial estimator exhibits the same behavior than the one observed for Σ. Likewise, its ARE of R^2_{Trial} behaves similarly than the one observed for D. Nonetheless, the efficiency loss is larger, especially with the smallest N and μ_n. Here, its MSE is around 1.4 and 1.6 larger than the one observed for the iterative REML estimator under high and extremely high imbalance, respectively. The large inefficiency can be explained by the use of the adjustments for non-pd estimates of D in more of the half of the simulated datasets. Comparing only the cases where both estimators led to a pd estimate for D, the ARE decreases to roughly 1.10 and 1.15 (see Figure S.3 of the Supplementary Materials) with high and extremely high imbalance, respectively.

In contrast to the trial-by-trial estimator, the information-theoretic approach provides more accurate estimates of R^2_{Trial}. Its ARE is lower than one in all scenarios. In the $N = \mu_n = 10$ case, the MSE of the iterative REML estimator is around 40% and 20% larger than the one observed by this method under high and extremely high imbalance, respectively. However, its ARE converges to one when μ_n increases faster than N. On the other hand, the estimator of R^2_{Ind} does not perform well when μ_n is small. its ARE diverges when μ_n is fixed and N increases. Nevertheless, it is efficient when μ_n increases faster

than N. More results on R^2_{Ind} and R^2_{Trial} can be found in Section B.4 of the Supplementary Materials.

5.3 Computation time

The median time spent to fit model (12) using the trial-by-trial and iterative REML estimator, in the large data settings, are displayed in Table 3. Naturally, the trial-by-trial estimator is faster than iterative REML in all scenarios. However, the difference in time depends on how quick N and μ_n increase. When N is fixed at 100 (Table 3a), the iterative REML estimator shows a steeper increasing time than the trial-by-trial estimator. Additionally, the latter is slightly affected by increasing the number of patients per trial, showing a maximum median time of 1.18 seconds when μ_n is equal to 2,500. Meanwhile, the iterative REML estimator spent roughly 138 seconds, which is 116 times longer. As expected, imbalance does not affect the trial-by-trial estimator. On the contrary, the iterative REML estimator took slightly more time in the extremely high imbalanced settings. When μ_n is bounded at 250 (Table 3b), the same behavior is observed, a faster increasing median time for the iterative REML estimator. However, the ratio does not rise as quickly as before. In these cases, imbalance does not seem to affect the computation time of the iterative estimator all that much.

6 Meta-analysis of clinical trials in schizophrenia

Next, the trial-by-trial estimator is implemented in data from a study in schizophrenia. The dataset combines five double-blind randomized clinical trials, and is available in the R library Surrogate (Van der Elst et al., 2017). The main aim was to examine the efficacy of risperidone to treat schizophrenia. In each trial, patients received risperidone or an active control for four to eight weeks. Three different instruments were used to assess patients' schizophrenic symptoms: The clinical Global Impression (CGI), the Brief Psychiatric Rating Scale (BPRS) test and the Positive and Negative Syndrome Scale (PANSS). The outcome of interest was the change in the score measure by each instrument. Surrogate

evaluation of the schizophrenia data was performed by Alonso Abad et al. (2016) using different combinations of endpoints. This is a common situation where the true endpoint is ambiguous, and any of the scales can be considered as the true endpoint with the others as possible surrogates (Molenberghs et al., 2010).

We illustrate our estimator in two different analyses. Firstly, we fitted a multivariate mixed-effects model using the three outcomes. Later, a meta-analytic surrogate endpoint evaluation was performed considering BPRS as a possible surrogate of PANSS.

The complete dataset contains information of 2,128 patients treated by 198 psychiatrists (clustering variable). Since the trial-by-trial estimator requires enough observations per cluster for estimating the cluster-specific parameters, we considered only the information of psychiatrists who examined more than two patients per treatment arm. Then, the remaining data include 1,392 patients treated by 64 psychiatrists. Although the dataset is not small, it is highly unbalanced. The mean (standard deviation) number of patients per psychiatrist is 21.8 (9.9), with a minimum (maximum) of 6 (52) patients. Even though these scales are discrete variables, their change in scores can be considered as (semi)continuous and can be approximately normally distributed (Alonso Abad et al., 2016). For comparative purposes, we analyzed the data using the REML estimator, based on the Cholesky decomposition, using the complete dataset.

6.1 Multivariate mixed model

We fitted a multivariate mixed model with BPRS, CGI, and PANSS as response variables, using treatment as covariate. Furthermore, the model included random effects for intercept and treatment effect. The trial-by-trial estimator reported non-pd estimates of D, but we used the eigenvalue method to correct non-positive-definiteness. The fitted variance matrix of residuals and random effects (after adjustment) are:

$$\widetilde{\Sigma} = \begin{pmatrix} 171.30 & 13.61 & 283.81 \\ 13.61 & 2.11 & 24.19 \\ 283.81 & 24.19 & 512.92 \end{pmatrix},$$

and,

$$\widetilde{D}_{+} = \begin{pmatrix} 36.60 & 1.42 & 3.38 & -0.08 & 64.96 & 2.57 \\ 1.42 & 0.11 & 0.19 & -0.00 & 2.74 & 0.04 \\ 3.38 & 0.19 & 0.47 & -0.03 & 6.23 & 0.17 \\ -0.08 & -0.00 & -0.03 & 0.01 & -0.09 & -0.01 \\ 64.96 & 2.74 & 6.23 & -0.09 & 117.01 & 4.31 \\ 2.57 & 0.04 & 0.17 & -0.01 & 4.31 & 0.24 \end{pmatrix},$$

respectively. Both, $\widetilde{\Sigma}$ and $\widetilde{D}_{+}$ show a strong association between the outcomes at both levels, individual and trial, with a larger correlation between BPRS and PANSS. The REML estimator, using the Cholesky decomposition, reported convergence, but to a non-pd $\widehat{D}$. However, the results on the fixed effects are still valid. The marginal variance matrix $V_{\boldsymbol{y}_i}$ is pd for every trial. Table 4 exhibits the estimates and standard errors of $\boldsymbol{\beta}$ using the trial-by-trial and REML estimators.

As Table 4 reveals, both methods provide similar estimates, but the trial-by-trial estimator reports a higher standard error. The efficiency loss is due to excluding several small clusters from the analysis. Nevertheless, there still is a significant negative treatment effect in the three outcomes, indicating that risperidone reduces schizophrenic symptoms.

6.2 Surrogacy evaluation

To evaluate surrogacy, we fitted model (12) using BPRS as a surrogate candidate of PANSS. In this case, a positive-definite estimate of D is vital during the estimation process. As before, the REML estimator, reached convergence, but to a non-pd estimate of D. Likewise, the fitted trial-by-trial estimator of D is non-pd. However, we proceeded to apply the eigenvalue method to find the nearest pd matrix. Both, the first estimate and

the posterior adjustment of $\widetilde{D}$ are:

$$\widetilde{D} = \begin{pmatrix} 37.2 & 2.0 & 66.2 & 3.4 \\ 2.0 & -1.9 & 2.3 & -2.5 \\ 66.2 & 2.3 & 118.8 & 3.8 \\ 3.4 & -2.5 & 3.8 & -2.6 \end{pmatrix} \text{ and } \widetilde{D}_+ = \begin{pmatrix} 37.4 & 1.3 & 66.1 & 2.6 \\ 1.3 & 0.1 & 2.5 & 0.0 \\ 66.1 & 2.5 & 118.9 & 4.1 \\ 2.6 & 0.0 & 4.1 & 0.3 \end{pmatrix},$$

respectively. This leads to $\widetilde{R}^2_{\text{Trial}} = 0.955$, indicating that there is a strong association between the treatment effect on the change in score measured by BPRS test and the change in score measure by PANSS test. Furthermore, the estimate of Σ is,

$$\widetilde{\Sigma} = \begin{pmatrix} 161.65 & 267.9 \\ 267.9 & 484.34 \end{pmatrix},$$

leading to $\widetilde{R}^2_{\text{Ind}} = 0.917$, showing a strong association between both tests at the patient level as well. From $\widetilde{\Sigma}$ and $\widetilde{D}$, we observed that the estimated residual variability is relatively larger than the estimated between-cluster variability. This fact, added to the highly unbalanced nature of the data, might be the cause of the non-pd estimates of D for both estimators.

As a reference and to make a comparison with our results, we evaluated surrogacy under the information-theoretic approach. Here, the point estimates of R^2_{Trial} and R^2_{Ind} are 0.903 and 0.924, respectively. These results are similar to the ones obtained by the trial-by-trial estimator and indicate a strong association between BPRS and PANSS tests at both, individual and trial, levels.

7 Discussion

We have proposed a closed-form unbiased estimator for the multivariate linear mixed-effects model. On theoretical grounds and based on a simulation study, it shows good statistical properties. Moreover, it is computationally highly efficient. Therefore, it is a good alternative to the standard iterative ML/REML estimator. Mainly, we suggest to

implement it in cases where the latter does not converge correctly, or when it is computationally too intensive or prohibitive. The first situation is commonly encountered with highly unbalanced small data, and the second with an extremely large number of clusters and/or units per cluster.

The simulation study has shown that the trial-by-trial estimator shares the same issues as its iterative counterpart in the former situation. It led to non-positive-definite estimates of D more frequently. Nevertheless, adjustments for non-positive-definiteness can easily be applied. In simulations, the eigenvalue method successfully found a positive-definite matrix near the values of the initially estimated matrix with a small cost in efficiency on the estimation of D. Of course, more alternatives can be considered for this task (Rousseeuw and Molenberghs, 1993; Higham, 2002).

Regarding statistical properties, the trial-by-trial estimator is asymptotically as efficient as the iterative REML estimator for all parameters of the multivariate linear mixed model when the number of units per cluster increases faster than the number of clusters. However, it behaves differently depending on the parameter. For $\boldsymbol{\beta}$ and Σ, the trial-by-trial estimator is as efficient as the iterative REML. For D, there is a small efficiency loss when the number of clusters and units per cluster are small, and it is affected by imbalance.

In the particular case of surrogacy evaluation, the efficiency of the estimators of R^2_{Ind} and R^2_{Trial} behaves similarly to the efficiency of estimators of Σ and D, respectively. However, the estimator of R^2_{Trial} is more affected by imbalance and the use of adjustments for non-positive-definiteness. Although it increases the MSE, the corrections allow us to always estimate R^2_{Trial}. On the other hand, there is no estimate of R^2_{Trial} with the iterative REML estimator in those cases. The implementation of these adjustments in the iterative procedure is not straightforward. An evaluation of whether or not the D matrix is pd should be undertaken in each iteration and then, if necessary, a correction should be applied. Furthermore, it is unclear how this change in the algorithm may affect the convergence and performance of the procedure. Specially for R^2_{Trial}, the information-theoretic approach provides more efficient estimates than the meta-analytic methodology

in settings with a small number of trials and patients per trial. Although it performed better, we cannot conclude that this approach is superior. The surrogacy definition and quantification are different (Alonso and Molenberghs, 2007). That said, with normally distributed endpoints, both estimators can be compared.

About the weighting scheme, approximate optimal and optimal weights for estimating $\boldsymbol{\beta}$ and Σ, respectively, are recommended. Optimal weighting is possible for the latter because it is parameter-free. For $\boldsymbol{\beta}$, on the other hand, it depends on unknown parameters. Therefore, we opt for approximate or iterated optimal weights. Based on the simulation study, approximate optimal weights for estimating $\boldsymbol{\beta}$ are highly recommended if the interest relies on fixed effects. The use of iterated optimal weights reduces the MSE of $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{D}$. However, this reduction is too small to be worth too much consideration. Moreover, it has a computational cost, especially if the number of clusters is large.

In large data settings, both estimators are statistically equivalent (practically the same estimates for all parameters). Nevertheless, the iterative method is more computationally intensive. The simulations showed that computation time of the trial-by-trial estimator can be more than 100 times faster than the standard iterative REML estimator. Given that a closed-form solution is performed separately in each cluster, these sub-processes could execute in parallel, reducing the computation time considerably, especially in cases with an extremely large number of cluster.

Given its computational advantages, we considered that our proposal is convenient not only when the dataset is large, but also when the number of outcomes becomes too large. Nevertheless, more simulation studies are needed to evaluate its statistical performance. Furthermore, the trial-by-trial estimator can be extended to non-Gaussian data. Of course, a closed-form estimator is not possible, but computational advantages can still be gained.

Acknowledgements

Financial support from the IAP research network $\#P7 = 06$ of the Belgian Government (Belgian Science Policy) is gratefully acknowledged. Alvaro J. Flórez acknowledges funding from the European Seventh Framework programme $FP7$ $2007 - 2013$ under grant agreement Nr. 602552.

References

Alonso, A. and Molenberghs, G. (2007). Surrogate marker evaluation from an information theory perspective. *Biometrics*, 63(1):180–186.

Alonso Abad, A., Bigirumurame, T., Burzykowski, T., Buyse, M., Molenberghs, G., Muchene, L., Perualila, N., Shkedy, Z., and Van der Elst, W. (2016). *Applied Surrogate Endpoint Evaluation with SAS and R.* Chapman & Hall/CRC, Boca Ratón.

Burzykowski, T., Molenberghs, G., and Buyse, M. (2005). *The Evaluation of Surrogate Endpoints.* Springer-Verlag GMBH.

Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D., and Geys, H. (2000). The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*, 1(1):49–67.

Chen, H., Manning, A., and Dupuis, J. (2012). A method of moments estimator for random effect multivariate meta-analysis. *Biometrics*, 68(4):1278–1284.

Hermans, L., Nassiri, V., Molenberghs, G., Kenward, M., Van der Elst, W., Aerts, M., and Verbeke, G. (2017). Fast, closed-form, and efficient estimators for hierarchical models with AR(1) covariance and unequal cluster sizes. *Communications in Statistics - Simulation and Computation.*

Higham, N. (2002). Computing the nearest correlation matrix - a problem from finance. *IMA Journal of Numerical Analysis*, 22:329–343.

Jackson, D., White, I., and Riley, R. (2013). A matrix-based method of moments for fitting the multivariate random effects model for meta-analysis and meta-regression. *Biometrical Journal*, 55(2):231–245.

Laird, N. and Lange, N. (1987). [prediction of future observations in growth curve models]: Comment. *Statistical Science*, 2(4):451–454.

Laird, N. and Ware, J. (1983). Random-effects models for longitudinal data. biometrics 38:963-974. *Biometrics*, 38(4):963–74.

Lindstrom, M. and Bates, D. (1988). Newton-raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83(404):1014–1022.

Molenberghs, G., Burzykowski, T., Alonso, A., Assam, P., Tilahun, A., and Buyse, M. (2010). A unified framework for the evaluation of surrogate endpoints in mental-health clinical trials. *Statistical Methods in Medical Research*, 19(3):205–236.

Molenberghs, G., Hermans, L., Nassiri, V., Kenward, M., Van der Elst, W., Aerts, M., and Verbeke, G. (2018). Clusters with random size: maximum likelihood versus weighted estimation. *Statistica Sinica*.

Molenberghs, G., Kenward, M., Aerts, M., Verbeke, G., Tsiatis, A., Davidian, M., and Rizopoulos, D. (2014). On random sample size, ignorability, ancillarity, completeness, separability, and degeneracy: sequential trials, random sample sizes, and missing data. *Statistical Methods in Medical Research*, 23(1):11–41.

Molenberghs, G., Verbeke, G., and Iddi, S. (2011). Pseudo-likelihood methodology for partitioned large and complex samples. *Statistics & Probability Letters*, 81(7):892 – 901.

Reinsel, G. (1985). Mean squared error properties of empirical bayes estimators in a multivariate random effects general linear model. *Journal of the American Statistical Association*, 80(391):642–650.

Rousseeuw, P. J. and Molenberghs, G. (1993). Transformation of non positive semidefinite correlation matrices. *Communications in Statistics, Theory and Methods*, 22(4):965–984.

Shah, A., Laird, N., and Schoenfeld, D. (1997). A random-effects model for multiple characteristics with possibly missing data. *Journal of the American Statistical Association*, 92(438):775–779.

Tibaldi, F., Cortiñas Abrahantes, J., Molenberghs, G., Renard, D., Burzykowski, T., Buyse, M., Parmar, M., Stijnen, T., and Wolfinger, R. (2003). Simplified hierarchical linear models for the evaluation of surrogate endpoints. *Journal of Statistical Computation and Simulation*, 73(9):643–658.

Van der Elst, W., Hermans, L., Verbeke, G., Kenward, M., Nassiri, V., and Molenberghs, G. (2016). Unbalanced cluster sizes and rates of convergence in mixed-effects models for clustered data. *Journal of Statistical Computation and Simulation*, 86(11):2123–2139.

Van der Elst, W., Meyvisch, P., Alonso, A., Hannah, M. E., Christopher, J. W., and Molenberghs, G. (2017). Surrogate: Evaluation of surrogate endpoints in clinical trials. R package version 0.2.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer, New York.

Vonesh, E. and Carter, R. (1987). Efficient inference for random-coefficient growth curve models with unbalanced data. *Biometrics*, 43(3):617–628.

West, B., Welch, K., and Galecki, A. (2014). *Linear Mixed Models: A Practical Guide Using Statistical Software*. Taylor & Francis Inc, 2 edition.

Table 1: *Proportion of cases where the estimators lead to a positive-definite estimate for D, (a) fixing $N = 10$ and varying μ_n, and (b) fixing $\mu_n = 10$ and varying N.*

(a)	high imbalance							extremely high imbalance						
	μ_{n_i}							μ_{n_i}						
Est.	10	25	50	100	150	200	250	10	25	50	100	150	200	250
TbT	0.46	0.88	0.97	0.99	1.00	1.00	1.00	0.43	0.78	0.92	0.96	0.98	0.98	0.98
REML	0.44	0.85	0.95	0.98	0.99	1.00	1.00	0.38	0.77	0.92	0.97	0.99	0.99	0.99
(b)	high imbalance							extremely high imbalance						
	N							N						
Est.	10	25	50	100	150	200	250	10	25	50	100	150	200	250
TbT	0.46	0.97	1.00	1.00	1.00	1.00	1.00	0.43	0.78	0.92	0.96	0.98	0.98	0.98
REML	0.43	0.78	0.92	0.96	0.98	0.98	0.98	0.38	0.77	0.92	0.97	0.99	0.99	0.99

TbT: Trial-by-trial estimator; **REML:** Iterative REML estimator using Cholesky decomposition.

Table 2: *ARE of the estimators of the linear mixed model and its derived quantities, (a) fixing $N = 10$ and varying μ_n, and (b) fixing $\mu_n = 10$ and varying N.*

(a)		high imbalance							extremely high imbalance						
		μ_{n_i}							μ_{n_i}						
Est.	parm.	10	25	50	100	150	200	250	10	25	50	100	150	200	250
TbT	$\boldsymbol{\beta}$	1.01	1.00	1.01	1.00	1.01	1.00	1.00	0.97	1.02	1.00	0.99	1.00	1.00	1.00
	Σ	1.12	1.01	1.01	1.00	1.00	1.01	1.00	1.07	1.03	1.01	1.00	1.00	1.00	1.00
	D	1.03	1.03	1.01	1.00	1.00	1.00	1.00	0.97	1.09	1.06	1.02	1.03	1.03	1.02
	R^2_{Ind}	1.10	1.04	0.99	1.00	1.00	1.00	1.00	1.00	1.02	1.02	0.99	0.99	1.00	1.00
	R^2_{Trial}	1.45	1.17	1.06	1.02	1.01	1.01	1.00	1.64	1.23	1.10	1.04	1.05	1.06	1.05
IT	R^2_{Ind}	1.56	1.18	1.05	1.03	1.01	1.02	1.00	1.30	1.25	1.07	1.04	1.02	1.03	1.01
	R^2_{Trial}	0.71	0.84	0.89	0.96	0.95	0.98	0.98	0.83	0.86	0.90	0.94	0.97	1.01	1.02
(b)		high imbalance							extremely high imbalance						
		N							N						
Parm.	Est.	10	25	50	100	150	200	250	10	25	50	100	150	200	250
TbT	$\boldsymbol{\beta}$	1.01	1.00	1.00	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00
	Σ	1.12	1.01	1.00	0.99	1.00	1.00	1.00	1.07	1.00	1.00	1.00	1.00	1.00	1.01
	D	1.03	1.04	1.04	1.03	1.04	1.04	1.03	0.97	1.07	1.06	1.06	1.07	1.08	1.06
	R^2_{Ind}	1.10	0.99	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	R^2_{Trial}	1.45	1.10	1.05	1.04	1.03	1.03	1.03	1.64	1.07	1.07	1.06	1.04	1.06	1.05
IT	R^2_{Ind}	1.56	2.40	3.94	7.00	9.44	12.78	16.22	1.30	2.43	3.47	6.86	10.21	13.10	16.52
	R^2_{Trial}	0.71	0.58	0.59	0.60	0.60	0.61	0.59	0.83	0.60	0.65	0.70	0.73	0.77	0.80

TbT: Trial-by-trial estimator; **IT:** Information-theoretic approach based on a fixed-effects model.

Table 3: *Median computation time, and ratio, of the trial-by-trial and iterative REML estimator, (a) fixing $N = 100$ and varying μ_n, and (b) fixing $\mu_n = 250$ and varying N.*

(a)	high imbalance				extremely high imbalance			
	μ_n				μ_n			
Est.	500	1,000	1,500	2,500	500	1,000	1,500	2,500
TbT	0.29	0.83	0.72	1.08	0.29	0.82	0.72	1.18
REML	6.44	19.88	44.78	112.81	7.52	37.00	42.50	137.47
Ratio	22.27	23.89	62.37	104.35	25.93	44.85	59.11	116.01
(b)	high imbalance				extremely high imbalance			
	N				N			
Est.	500	1,000	2,500	5,000	500	1,000	2,500	5,000
TbT	2.62	9.75	93.98	237.15	2.71	9.58	59.25	242.03
REML	34.16	128.26	1002.68	3723.29	35.00	128.12	825.78	3861.60
Ratio	13.03	13.16	10.67	15.70	12.93	13.37	13.94	15.96

TbT: Trial-by-trial estimator; **REML:** Iterative REML estimator using Cholesky decomposition.

Table 4: *Schizophrenia data. Estimates of the multivariate linear mixed model using the trial-by-trial and REML estimator*

	Trial-by-trial		REML	
Parm.	Estimate	Std. error	Estimate	Std. error
$\beta_{0,\text{BPRS}}$	-8.15	0.863	-7.85	0.519
$\beta_{1,\text{BPRS}}$	-1.49	0.408	-1.26	0.332
$\beta_{0,\text{CGI}}$	3.28	0.097	3.32	0.054
$\beta_{1,\text{CGI}}$	-0.16	0.046	-0.12	0.038
$\beta_{0,\text{PANSS}}$	-14.59	1.53	-13.87	0.911
$\beta_{1,\text{PANSS}}$	-2.74	0.707	-2.41	0.582