Published in final edited form as:

Stat Med. 2011 June 30; 30(14): 1735–1750. doi:10.1002/sim.4169.

Constrained S-estimators for linear mixed effects models with covariance components

Inna Chervoneva^{1,*} and Mark Vishnyakov¹

¹ Division of Biostatistics, Department of Pharmacology and Experimental Therapeutics, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, U.S.A

Summary

Linear mixed effects (LME) models are increasingly used for analyses of biological and biomedical data. When the multivariate normal assumption is not adequate for an LME model, then a robust estimation approach is preferable to the maximum likelihood one. M-estimators were considered before for robust estimation of the LME models, and recently a constrained S-estimator was proposed. This S-estimator can not be applied directly to LME models with correlated error terms and vector random effects with correlated dimensions. Therefore, a modification is proposed, which extends application of the constrained S-estimator to the LME models for multivariate responses with correlated dimensions and to longitudinal data. Also a new computational algorithm is developed for computing constrained S-estimators. Performance of the S-estimators based on the original Tukey's biweight and translated biweight is evaluated in a small simulation study with repeated multivariate responses with correlated dimensions. Proposed methodology is applied to jointly analyze repeated measures on three cholesterol components, HDL, LDL, and triglycerides.

Keywords

Multivariate linear mixed effects models; robust estimation; CTBS estimator for LME model; M-estimator

1. INTRODUCTION

The data that motivates this work come from the study of antiepileptic drugs in terms of their effect on serum lipid fractions, HDL, LDL, and triglycerides (TRIG) [1]. The widely-prescribed anticonvulsants phenytoin (PHT) and carbamazepine (CBZ) are potent inducers of cytochrome P450 enzymes, which are involved in cholesterol synthesis. The small study was performed to determine if these drugs may have an effect on cholesterol and other serologic markers of vascular risk. Thirty four epilepsy patients taking CBZ or PHT in monotherapy were switched to one of the non-inducing anticonvulsants, lamotrigine (LTG) or levetiracetam (LEV). Fasting blood samples were obtained both before and 6 weeks after the switch. Figure 1 shows the scatter plots of the lipid fractions and line plots of subject-specific changes in HDL, LDL, and TRIG measures. The goal of analyses was evaluation of the study measures before and after the switch and of the change resulting from switching the drug. It was expected that drug-related difference in lipids are correlated and more likely before the switch than after. To address this hypothesis, it is necessary to model these data as repeated multi-dimensional correlated responses.

*Correspondence to: Inna Chervoneva, Division of Biostatistics, Department of Pharmacology and Experimental Therapeutics, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, U.S.A. i_chervoneva@mail.jci.tju.edu.

When normal distribution assumption for both random effects and error terms is appropriate, a multivariate linear mixed effects (MLME) model is used for analysis of clustered and correlated data such as our example. However, the standard (restricted) maximum likelihood estimation for MLME model may not be appropriate for these data because the normal distribution assumption is likely violated due to potential outliers (Figure 1). Therefore, a robust estimation approach is needed to minimize potential bias due to outliers. Heritier et al [2] provide a comprehensive up to date review of robust methods available for the LME models. In the earlier work, weighted robustified log-likelihood was considered [3–4] for robust estimation of the LME models. Further, Richardson [5] and Welsh and Richardson [6], proposed a general class of robust bounded influence estimators for the LME models. These estimators are generalized M-estimators defined as a solution to the bounded influence estimating equations (BIEE). While local robustness and asymptotic properties of M-estimators follow from the general theory [7], these estimators do not necessarily have good break-down properties in high dimensions. For a general M-estimator of p -variate location [8], the breakdown point is at most $1/(1 + p)$. Therefore, more recent work in robust estimation of multivariate regression and LME models have shifted towards S-estimators [9–11], which may be viewed as a subclass of M-estimators tuned to have a high breakdown point [12].

Copt and Victoria-Feser [11] proposed a constrained translated biweight S-estimator (CTBS) for the LME models. This estimator is an adaptation of the S-estimator of multivariate location and scale [12,13] to the LME models with multivariate formulation and covariance structures with uncorrelated residual error terms. The CTBS estimator may be used only for a subclass of linear mixed effects models for fully balanced data, while BIEE estimator of Welsh and Richardson [6] is applicable to more general LME models. However, no easy numerical implementation of the BIEE estimator has been proposed, and advantage of the CTBS estimator is a possibility to compute this S-estimator through an iterative algorithm. Yet, without modification of the part that is related to the covariance matrix, the estimating equations and iterative algorithm in [11] are not directly applicable to MLME models for multivariate responses with correlated dimensions. Meanwhile, in applications such as our data example, it is desirable to model repeated multi-dimensional responses with correlated dimensions in error term vectors and/or vector random effects.

In this work, we propose a modification of the constrained S-estimator of Copt and Victoria-Feser [11] for the LME models with correlated covariance components, which may include vector error terms and vector random effects with correlated dimensions. Thus, our work extends application of the constrained S-estimator [11] to the LME models for multivariate responses with correlated dimensions and to longitudinal data. Still, all constrained S-estimators are applicable only to LME models that may be written in terms of independent multivariate normal vectors of the same dimensions with the same structured covariance matrix and without missing observations.

The constrained S-estimator minimizes the determinant of the structured covariance matrix of independent vector observations subject to constraint, which ensures consistency under uncontaminated normal model and would guarantee the desired breakdown point (up to $\frac{1}{2}$) in the multivariate location and scale problem. However, for fixed designs corresponding to multiple groups, the breakdown point of S-estimators may be lower [15,16]. The constrained S-estimator may be based on any redescending ρ -function that satisfies regularity conditions in [12]. In our applications, we consider the original Tukey's biweight [17] as well as translated biweight [18], while original work [11] was focused primarily on translated biweight [18], giving the rise to the term CTBS estimator.

For computing constrained S-estimators, we have developed a new computational algorithm, which ensures that constraint is satisfied and increases the chance to attain the global minimum of the determinant of the structured covariance matrix. Similar to [11], the constrained S-estimators are computed using an iterative algorithm derived from the estimating equations. However, our estimating equations incorporate the constraint directly to make equations linearly independent and to ensure that their solutions include the ones that satisfy the constraint. In case of multiple local minima of the objective function, an iterative algorithm is naturally sensitive to the starting values. Therefore, the subsampling algorithm, similar to the one described in [14], is used to compute random starting values. This approach allows bypassing the need to compute robust starting values and provides the means to increase the chance to identify the global minimum by increasing the number of random starts.

Performance of the proposed S-estimators was evaluated in a small simulation study. The simulated data were similar in the structure and parameters to the motivating data example. The contamination patterns were designed to include contaminations in two positively or negatively correlated dimensions of the three-dimensional outcome to investigate the impact of correlation on the performance of the proposed S-estimators.

The rest of this article is organized as follows. In Section 2, we introduce the constrained S-estimators for the multivariate LME models. An algorithm for computing constrained S-estimators is described in section 3. Simulation study results are summarized in section 4. In Section 5, we compare the analyses results for our data example using the robust and maximum likelihood methods. Section 6 concludes with a discussion.

2. CONSTRAINED S-ESTIMATORS FOR MULTIVARIATE LME MODELS

Let $\mathbf{y}_{ik}$ be the k^{th} , $k = 1, \dots, K_j$, J -dimensional response vector nested within the highest level cluster i , $i = 1, \dots, M$, so that the highest level clusters are independent. Any possible intermediate levels of clustering are considered to be reflected in the design matrices of

random effects. Denote by $\mathbf{y}_i = [\mathbf{y}'_{i1}, \mathbf{y}'_{i2}, \dots, \mathbf{y}'_{iK_j}]'$ the combined vector of all observations for the highest level cluster i . A multivariate LME model may be written as

$$\mathbf{y}_i = \mathbf{A}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \quad (1)$$

where $\boldsymbol{\beta}$ is a $q \times 1$ vector of fixed population parameters, $\boldsymbol{\gamma}_i$ is a $r \times 1$ vector of all random effects associated with the highest level cluster i , $\mathbf{A}_i$ and $\mathbf{Z}_i$ are design matrices for the fixed and random effects, respectively, and $\boldsymbol{\varepsilon}_i = [\boldsymbol{\varepsilon}'_{i1}, \boldsymbol{\varepsilon}'_{i2}, \dots, \boldsymbol{\varepsilon}'_{iK_j}]'$, where $\boldsymbol{\varepsilon}_{ik}$ is J -dimensional error terms vector for k^{th} response vector in cluster i . It is further assumed that $\forall i, K_j = K$, $\boldsymbol{\gamma}_i \sim \mathbf{N}_r(\mathbf{0}, \mathbf{G})$, $\boldsymbol{\varepsilon}_i \sim \mathbf{N}_p(\mathbf{0}, \mathbf{R})$, where $p = JK$, and $\mathbf{G}$ and $\mathbf{R}$ are the covariance matrices of the random effects and of the error terms, respectively. It is assumed that $\boldsymbol{\gamma}_i$ is independent of $\boldsymbol{\varepsilon}_i$ any i and j . Let $\mathbf{b}_i = [\boldsymbol{\gamma}'_i, \boldsymbol{\varepsilon}'_i]'$, then $\mathbf{b}_i \sim \mathbf{N}_{p+r}(\mathbf{0}, \mathbf{D})$ with $\mathbf{D} = \text{Diag}(\mathbf{G}, \mathbf{R})$ and one can re-write (1) as

$$\mathbf{y}_i = \mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i, \quad (2)$$

where $\mathbf{B}_i = [\mathbf{Z}_i \mathbf{I}_p]$ such that $\mathbf{B}_i = \mathbf{B} \forall i$ and $\mathbf{I}_p$ is $p \times p$ identity matrix.

Thus, for independent vectors $\mathbf{y}_i$, we have the following multivariate formulation

$$\mathbf{y}_i \sim N_p(\mathbf{A}_i\boldsymbol{\beta}, \mathbf{V}), \quad \mathbf{V} = \mathbf{B}\mathbf{D}\mathbf{B}' \quad (3)$$

Let $\boldsymbol{\theta} = [\theta_1, \dots, \theta_l]'$ be the $l \times 1$ vector of all unknown parameters in $\mathbf{D}$. We assume that $\mathbf{D}$

may be written as a linear combination $\mathbf{D} = \sum_{g=1}^l \theta_g \mathbf{L}_g$, where $\mathbf{L}_g$ are completely known and

usually singular matrices. Then, $\mathbf{V} = \sum_{g=1}^l \theta_g \mathbf{B}\mathbf{L}_g\mathbf{B}'$. Model (3) is similar but not the same as a multivariate regression or MANCOVA model, because covariance matrix $\mathbf{V}$ is structured, and instead of estimating elements of $\mathbf{V}$, we have to estimate unknown covariance parameters in $\boldsymbol{\theta}$. These parameters uniquely define $\mathbf{V}$, but inverse is not generally true in LME models, and the same matrix $\mathbf{V}$ may correspond to more than one way of splitting variability between random effects and error terms. That is why the theory of S-estimators in multivariate regression is not directly applicable. On the other hand, constrained S-estimators developed in [11] do not accommodate covariance components represented by general positive definite matrices as in (2). Indeed, in [11], $\mathbf{V}$ is represented by a linear combination of unknown variance components ($\mathbf{V} = \sum \sigma_j^2 \mathbf{Z}_j \mathbf{Z}_j'$) and components of the error term vectors $\boldsymbol{\varepsilon}_j$ are independent, so that $\mathbf{R} = \sigma_0^2 \mathbf{I}_p$. So, our covariance matrix

$\mathbf{V} = \sum_{g=1}^l \theta_g \mathbf{B}\mathbf{L}_g\mathbf{B}'$ is more general than in [11] because covariance parameters in $\boldsymbol{\theta}$ may be negative (but their sign is not fixed by parameterization) if they correspond to off-diagonal elements of matrix $\mathbf{D}$. In particular, covariance matrix $\mathbf{V}$ for our data example and simulation study includes negative correlation among dimensions of multivariate responses and may not be written in the form $\mathbf{V} = \sum \sigma_j^2 \mathbf{Z}_j \mathbf{Z}_j'$.

Model (3) may also be applied to longitudinal data with the same number of responses per independent unit. While it is not possible to express the standard autoregressive correlation

structure of $\mathbf{R}$ as $\sum_{g=1}^l \theta_g \mathbf{L}_g$, it is possible to write in this form a slightly more general symmetric Toeplitz (diagonal-constant) matrix $\mathbf{R}$.

The constrained S-estimator minimizes the determinant of $\mathbf{V}$, $\det(\mathbf{V})$, subject to constraint

$$M^{-1} \sum_{i=1}^M \rho(d_i) = b_0, \quad (4)$$

where d_i is the Mahalanobis distance for vector $\mathbf{y}_i$,

$$d_i = [(\mathbf{y}_i - \mathbf{A}_i\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{y}_i - \mathbf{A}_i\boldsymbol{\beta})]^{1/2} \quad (5)$$

parameter b_0 satisfies condition $b_0 = E_{d^2 \sim \chi_p^2}[\rho(d)]$ that ensures consistency under the normal model, and ρ -function satisfies conditions given in Appendix A. The breakdown parameter $\varepsilon^* = 0.5$ is given by the ratio

$$\varepsilon^* = \frac{E_{d^2 \sim \chi_p^2}[\rho(d)]}{\max_d \rho(d)}. \quad (6)$$

For the multivariate location or multivariate regression setting, which are particular subclasses of the LME model, ε^* is the asymptotic breakdown point, which may be controlled (up to $\frac{1}{2}$ as a maximum). However, for fixed designs corresponding to multiple groups, the breakdown point of S-estimators may be lower [15–16].

In this work we use two ρ -functions, the original Tukey's biweight [17] and translated Tukey's biweight [18], but all developments are applicable to any ρ -function that satisfies conditions (R1)–(R3) given in Appendix A.

Let us denote by $\widehat{\beta}_p$ and $\widehat{\theta}_p$ the constrained S-estimators of β and θ based on the corresponding ρ -function and assume that model (2) satisfies the following regularity conditions:

- C1** Matrices $\mathbf{A}_i' \mathbf{V}^{-1} \mathbf{A}_i$ are nonsingular for every $i = 1, \dots, M$.
- C2** Matrices $\mathbf{V} = \mathbf{B} \mathbf{D}(\theta) \mathbf{B}'$ and $\frac{\partial(\text{vec}(\mathbf{V}(\theta)))'}{\partial \theta}$ and $\frac{\partial(\text{vec}(\mathbf{V}(\theta)))}{\partial \theta}$ are nonsingular.
- C3** There exists a nonsingular matrix $\Omega = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{i=1}^M \mathbf{A}_i' \mathbf{V}^{-1} \mathbf{A}_i$.

If conditions R1–R3 (Appendix A) and C1–C3 hold, then it is shown (Appendix B) that as $M \rightarrow \infty$,

$$\sqrt{M} \left(\begin{bmatrix} \widehat{\beta}_p' \\ \widehat{\theta}_p' \end{bmatrix} - \begin{bmatrix} \beta' \\ \theta' \end{bmatrix} \right) \xrightarrow{L} N \left(\mathbf{0}, \text{Diag} \left(\frac{\alpha_1}{\alpha_2^2} \Omega^{-1}, \Omega_\theta \right) \right), \quad (7)$$

where $\Omega_\theta = H \frac{\partial(\text{vec}(\mathbf{V}))'}{\partial \theta} \Omega_V \frac{\partial(\text{vec}(\mathbf{V}))}{\partial \theta} H$, $H = \left\{ \frac{\partial(\text{vec}(\mathbf{V}))'}{\partial \theta}, \frac{\partial(\text{vec}(\mathbf{V}))}{\partial \theta} \right\}^{-1}$,

$$\Omega_V = \sigma_1 (\mathbf{I} + \mathbf{K}_{p,p}) (\mathbf{V} \otimes \mathbf{V}) + \sigma_2 \text{vec}(\mathbf{V}) \text{vec}(\mathbf{V})', \quad (8)$$

$\mathbf{K}_{p,p}$ is a $(p^2 \times p^2)$ -block matrix with (i,j) -th block $\mathbf{K}(i,j)$ of size $(p \times p)$, which has 1 at entry (i,j) and 0 otherwise ($\mathbf{K}(i,j)_{ij} = \delta_{ij}$), $\otimes$ is the Kronecker product of two matrices, and

constants α_1 , α_2 , σ_1 , and σ_2 are given in Appendix B. The asymptotic distribution of $\widehat{\beta}_p'$ is the same as in [9–10] for the multivariate regression problem and in [19] for the LME model. Since matrix $\mathbf{V}$ is constrained as $\mathbf{V} = \mathbf{V}(\theta)$, the results for multivariate regression problem are not applicable, and the asymptotic covariance matrix Ω_θ in (7) is derived in Appendix B using the arguments similar to [11].

3. COMPUTING CONSTRAINED S-ESTIMATORS IN LME MODELS

To minimize the determinant of $\mathbf{V}$ subject to constraint (4), one can use the Lagrange multipliers

$$L = \log(\det(\mathbf{V})) - \lambda \left[M^{-1} \sum_{i=1}^M \rho(d_i) - b_0 \right].$$

The S-estimator must satisfy the system, which includes $q + l + 1$ equations, $\frac{\partial L}{\partial \beta} = 0$, $\frac{\partial L}{\partial \theta} = 0$ and (4), where

$$\frac{\partial L}{\partial \beta} = \lambda M^{-1} \sum_{i=1}^M u(d_i) \mathbf{A}_i' \mathbf{V}^{-1} (\mathbf{y}_i - \mathbf{A}_i \beta) = \mathbf{0} \quad (9)$$

$$\frac{\partial L}{\partial \theta} = \text{tr} \left[\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta} \right] + \frac{\lambda}{2M} \sum_{i=1}^M u(d_i) (\mathbf{y}_i - \mathbf{A}_i \beta)^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta} \mathbf{V}^{-1} (\mathbf{y}_i - \mathbf{A}_i \beta) = \mathbf{0} \quad (10)$$

and $u(d) = d^{-1} \rho'(d)$ is the weighting function corresponding to the ρ -function.

For the biweight, the weighting function is

$$u_B(d) = \begin{cases} (1 - (d/c_0)^2)^2, & \text{if } d \leq c_0 \\ 0, & \text{if } d \geq c_0 \end{cases},$$

and for the translated biweight, the weighting function is

$$u_R(d) = \begin{cases} 1, & 0 \leq d \leq S \\ \left(1 - \left(\frac{d-S}{c}\right)^2\right)^2, & S \leq d \leq S+c \\ 0, & d > S+c, \end{cases}.$$

System (4)&(9)&(10), may be re-expressed (see Appendix C) as

$$\begin{aligned} \sum_{i=1}^M u(d_i) \mathbf{A}_i' \mathbf{V}^{-1} (\mathbf{y}_i - \mathbf{A}_i \beta) &= \mathbf{0} \\ \mathbf{Q}(\theta) \theta - \mathbf{U}(\beta, \theta) &= \mathbf{0} \\ M^{-1} \sum_{i=1}^M \rho(d_i) - b_0 &= 0 \end{aligned} \quad (11)$$

where $\mathbf{Q}(\theta)$ the $l \times l$ matrix and $\mathbf{U}(\beta, \theta)$ is the $l \times 1$ vector with entries

$$\begin{aligned} [\mathbf{Q}(\theta)]_{jk} &= \text{tr} [\mathbf{V}^{-1} \mathbf{B} \mathbf{L}_j \mathbf{B}' \mathbf{V}^{-1} \mathbf{B} \mathbf{L}_k \mathbf{B}'] \\ [\mathbf{U}(\beta, \theta)]_j &= \left[\frac{1}{p} \sum_{i=1}^M u(d_i) d_i^2 \right]^{-1} \sum_{i=1}^M u(d_i) (\mathbf{y}_i - \mathbf{A}_i \beta)^T \mathbf{V}^{-1} \mathbf{B} \mathbf{L}_j \mathbf{B}' \mathbf{V}^{-1} (\mathbf{y}_i - \mathbf{A}_i \beta) \end{aligned}$$

Every solution of (11) also satisfies the non-equivalent system of $q + l$ estimating equations:

$$\begin{aligned} \sum_{i=1}^M u(d_i) \mathbf{A}_i' \mathbf{V}^{-1} (\mathbf{y}_i - \mathbf{A}_i \beta) - \mathbf{c}_\beta \left[M^{-1} \sum_{i=1}^M \rho(d_i) - b_0 \right] &= \mathbf{0} \\ \mathbf{Q}(\theta) \theta - \mathbf{U}(\beta, \theta) - \mathbf{c}_\theta \left[M^{-1} \sum_{i=1}^M \rho(d_i) - b_0 \right] &= \mathbf{0} \end{aligned} \quad (12)$$

where $\mathbf{c}_\beta$ is $q \times 1$ vector and $\mathbf{c}_\theta$ is $l \times 1$ vector of arbitrary non-zero constants. If $\mathbf{c}_\beta$ and $\mathbf{c}_\theta$ are zero vectors then system (12) is linearly dependent for any given β and θ . System (12) has the following properties: (i) every solution of the system (11) is also a solution of system (12); (ii) every solution of (12), which satisfies the constraint (4), is also a solution of (11). System (12) may not be reduced to the one derived in [11], which is dependent.

System (12) yields the following iterative equations for β and θ :

$$\begin{aligned}\beta &= \left(\sum_{i=1}^M u(d_i) \mathbf{A}_i' \mathbf{V}^{-1} \mathbf{A}_i \right)^{-1} \left\{ \sum_{i=1}^M u(d_i) \mathbf{A}_i' \mathbf{V}^{-1} \mathbf{y}_i - \mathbf{c}_\beta \left[M^{-1} \sum_{i=1}^M \rho(d_i) - b_0 \right] \right\} \\ \theta &= \mathbf{Q}(\theta)^{-1} \left\{ \mathbf{U}(\beta, \theta) + \mathbf{c}_\theta \left[M^{-1} \sum_{i=1}^M \rho(d_i) - b_0 \right] \right\}\end{aligned}\quad (13)$$

The constrained S-estimators of β and θ are found as the solution of (11) with the smallest $\det(\mathbf{V})$. Our strategy is to find many solutions of (12) using a large number of random starting values for (13). Then there is a higher chance that identified solutions of (12) would include solutions of (11) and the global minimum of $\det(\mathbf{V})$. The random starting values are obtained by subsampling the data and computing the standard restricted maximum likelihood (REML) estimates of β and θ . The non-zero constants $\mathbf{c}_\beta$ and $\mathbf{c}_\theta$ may be chosen arbitrarily, just to make system (12) linearly independent. For the purpose of computing desired S-estimators, it is proposed to use constants $\mathbf{c}_\beta$ and $\mathbf{c}_\theta$ as tuning parameters to facilitate convergence of the iterative algorithm. The actual choice of $\mathbf{c}_\beta$ and $\mathbf{c}_\theta$ would depend on particular LME model and data. The explicit algorithm for computing the S-estimators of β and θ is given in Appendix D.

4. SIMULATION STUDY

Finite sample performance of the modified constrained S-estimators was evaluated in a small simulation study. The simulated data were similar in structure to the cholesterol data example, with covariance components expressed as matrices with possibly correlated dimensions. That is, 3-dimensional vectors of correlated observations $\mathbf{y}_{ikt}$ (k^{th} replicate in subject i on treatment t) were randomly generated from the model

$$\mathbf{y}_{ikt} = \beta_{t(k)} + \gamma_i + \epsilon_{ik}, \quad (14)$$

where $\beta_{t(k)}$ is the 3-dimensional mean vector of correlated responses (HDL, LDL, and TRIG measures in the data example) on treatment t , $t = 1, 2, 3, 4$. As in the data example, the first replicate ($k = 1$) corresponds to $t = 1$ or $t = 2$, and the second replicate ($k = 2$) corresponds to $t = 3$ or $t = 4$. Further, γ_i is a 3×1 vector of random effects, $\gamma_i \sim \mathbf{N}_3(\mathbf{0}, \mathbf{D}_0)$, in subject i , $i = 1, \dots, M$, and ϵ_{ik} is a 3×1 vector of random errors in k^{th} replicate, $\epsilon_{ik} \sim \mathbf{N}_3(\mathbf{0}, \mathbf{D}_1)$. The means were assumed to be the same for all treatments, $\beta_{t(k)} = \beta_0 = [60, 120, 85]'$, $t = 1, 2, 3, 4$. Similar to the covariance parameter estimates in the data example, we assumed $\mathbf{D}_1 = \text{Diag}(49, 441, 900)$ and

$$\mathbf{D}_0 = \begin{bmatrix} 256 & 48 & -269 \\ 48 & 900 & 288 \\ -269 & 288 & 576 \end{bmatrix}, \text{ with correlation matrix } \begin{bmatrix} 1 & 0.1 & -0.7 \\ 0.1 & 1 & 0.4 \\ -0.7 & 0.4 & 1 \end{bmatrix}.$$

Performance of the constrained S-estimators based on Tukey's original biweight (COBS) and translated biweight (CTBS) was compared to the performance of the standard REML estimators for the data generated under normal model (14) with and without contaminations. Detailed description of the contamination scenarios is provided in Table I. In every contaminated scenario, 4 of 40 subjects (10%) had contaminations either in the third dimension only (scenarios 1–6) or in two dimensions (scenarios 10–28). Contaminations were introduced by deterministically replacing the mean(s) of the main model by the contaminated mean(s) such that the difference between the two means was approximately equal to 6 standard deviations in the corresponding dimension(s). Two dimensions were

chosen either with negative correlation coefficient -0.7 (first and third, scenarios 10–13 and 20–23) or with positive correlation coefficient 0.4 (second and third, scenarios 15–18 and 25–28). Scenarios with the same second digit had the same contamination pattern, but different breakdown parameter was used, 0.5 for scenarios 10–18 and 0.2 for scenarios 20–28.

Parameters for Tukey's biweight, $\rho_B(d; c_0)$, were selected to correspond the breakdown parameter $\varepsilon^* = 0.5$ ($c_0 = 5.125$, $b_0 = 2.189$) or $\varepsilon^* = 0.2$ ($c_0 = 9.000$, $b_0 = 2.701$). Similarly, for the translated biweight, $\rho_R(d; c, S)$, parameters corresponded to either the breakdown parameter $\varepsilon^* = 0.5$ and $ARP = 0.01$ ($c = 1.583$, $S = 2.517$, $b_0 = 2.856$) or to the breakdown parameter $\varepsilon^* = 0.2$ and $ARP = 0.00001$ ($c = 0.601$, $S = 5.153$, $b_0 = 2.999$). Figure 2 shows the corresponding weighting functions. It is not possible to find c and S such that $\varepsilon^* = 0.2$ and $ARP = 0.01$, as $ARP = 0.01$ is possible only with the breakdown parameter $\varepsilon^* > 0.35$.

For each scenario listed in Table I, 500 data sets were simulated. Model was fitted using the standard LME model (REML estimators) and the proposed algorithm with 300 random starts (2 subjects were randomly selected from each treatment pair group to compute the starting values). The maximum allowed number of iterations was 100 and convergence criterion was $\delta_C = 10^{-3}$ and $\delta_V = 10^{-4}$. The constant vectors $\mathbf{c}_\beta$ and $\mathbf{c}_\theta$ were chosen as $\mathbf{c}_\beta = 0.01\mathbf{J}_q$ and $\mathbf{c}_\theta = 0.01\mathbf{J}_I$, where $\mathbf{J}_q$ and $\mathbf{J}_I$ are $q \times 1$ and $I \times 1$ vectors of ones. This choice of $\mathbf{c}_\beta$ and $\mathbf{c}_\theta$ was by trial and error to ensure convergence of the iterative algorithm to parameter estimates that would satisfy the constraint.

In the context of the motivating data example, the primary multivariate hypothesis of interest is in comparing the change in cholesterol measures between the replicate 1 ($t=1,2$) and 2 ($t=3,4$), after switching the drugs ($\mathbf{H}_0 : (\beta_1 + \beta_2)/2 = (\beta_3 + \beta_4)/2$). With suitably re-ordered (first by dimension and then by treatment) elements of the fixed effects parameter vector $\beta = [\beta_1, \beta_2, \beta_3, \beta_4]$, the null hypothesis may be written as $\mathbf{H}_0 : \mathbf{G}\beta = \mathbf{0}$, where $\mathbf{G} = \mathbf{U}' \otimes \mathbf{H}$, $\mathbf{U} = \mathbf{I}_3$ and $\mathbf{H} = [-\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}]$. Thus, $\mathbf{G} = \mathbf{U}' \otimes \mathbf{H}$ is the 3×12 contrast matrix with $\text{rank}(\mathbf{G}) = 3$. For the pair-wise treatment comparisons, the null hypotheses have matrix $\mathbf{H}$ with 1 and -1 for the corresponding pair of treatments and zero for others. These hypotheses may be written in the form $\mathbf{H}_0 : (\mathbf{U}' \otimes \mathbf{H})\beta = \mathbf{0}$, where $\mathbf{H}$ is the $c \times q$ contrast matrix of between-cluster comparisons and $\mathbf{U}$ is the $t \times u$ contrast matrix of within-cluster comparisons.

In the absence of contamination, the conservative way to test the multivariate hypothesis of interest in data with a small number of independent vector observations of high dimension is to use the approximate MANOVA-type F-test [20]. Under the model with uncontaminated normal distribution, the test statistics

$$\mathbf{F}_{LH} = \frac{\min\{c, u\} (M - q - u - 1) + 2}{cu \min\{c, u\}} \frac{\mathbf{T}^2}{M - q}, \quad (15)$$

where $\mathbf{T}^2$ is given by

$$\mathbf{T}^2 = \widehat{\beta}' \mathbf{G}' [\mathbf{G} \{\widehat{\text{Var}}(\widehat{\beta})\} \mathbf{G}']^{-1} \mathbf{G} \widehat{\beta}, \quad (16)$$

has an approximate F-distribution with $\min\{c, u\} (M - q - u - 1) + 2$ denominator and $cu = \text{rank}(\mathbf{G})$ numerator degrees of freedom [21]. In the standard LME models, this test is more suitable for data with moderate number of independent clusters than asymptotic tests, such as previously considered robust tests. For example, Copt and Victoria-Feser [11] use the

asymptotic robust score-type test [22], while Copt and Heritier [19] proposed asymptotic robust likelihood ratio (LRT) and robust Wald type tests based on MM estimator. Motivated by relatively small (34) number of patients in our data example, we considered “robustifying” the approximate MANOVA-type F-test by using the S-estimates to compute the test statistic. This approach is admittedly ad hoc, and corresponding F-statistic is not expected to have the postulated F-distribution. Still, in small samples, one may expect to have more conservative test by “robustifying” the MANOVA-type F-test than using an asymptotic robust test, because MANOVA-type F-test penalizes small numbers of independent vector observations of high dimension by limiting the degrees of freedom. In order to investigate whether it is reasonable to use such “robustified” MANOVA-type F-test for our data example, we investigated its performance in our simulation study, which design mimics the data structure and contamination pattern of the real data. With the total of 40 subjects, the MANOVA-type F-test statistics is $\mathbf{F}_{LH} = \frac{\mathbf{T}^2 (M-6)}{3 (M-4)}$, with 3 numerator and 34 denominator degrees of freedom.

Figure 3 shows the distribution of the relative bias in fixed effects parameter estimates. Since contaminations were essentially symmetric for treatments 1 and 3, the relative bias is shown only for treatment 1 group. Notably, for all directly contaminated means, the REML estimates produced large relative bias of about 40–55%. The constrained S-estimator based on the original biweight (COBS) and the one based on the translated biweight (CTBS) yielded the relative bias close to zero, except for the directly contaminated dimensions in scenarios 20–21 and 25–26 with the breakdown parameter 0.2. In these scenarios, the COBS-estimator yielded lower and less variable relative bias than the CTBS-estimator (Figure 3).

As a general rule, the bias of REML and robust estimates was higher for components in θ that correspond to the contaminated dimensions. The relative bias distribution for such covariance parameter estimates is shown in Figures 4 and 5. With the breakdown parameter 0.5, the bias in all covariance parameter estimates is well controlled by either constrained S-estimator. In the absence of contamination (scenario 0), the empirical relative efficiency of covariance parameter estimates (computed as the ratio of the corresponding root mean squared errors) ranged (depending on scenario) from 0.78 to 0.86 for the CTBS-estimators and from 0.84 to 0.91 for the COBS-estimators.

Table II presents empirical rejection rates of the “robustified” MANOVA-type F-tests for the primary multivariate hypothesis of interest and for the pair-wise treatment comparisons. The F-tests based on the REML estimators had unacceptably high type I errors for the primary hypothesis and for all comparisons that involve contaminated dimensions. The overall F-tests based on the COBS-estimators exhibited empirical levels either close to the nominal 5% level or lower when the breakdown parameter was 0.5. The empirical levels were generally close to the nominal levels also for the pair-wise treatment comparisons, except for $\mathbf{H}_0 : \beta_1 = \beta_3$, because all but scenarios 0 (uncontaminated) and 5 had contaminations specifically in the group that was switched from $t=1$ to $t=3$. Inflation of type I error occurred when all 10% of contaminated subjects were in this group. With the breakdown parameter 0.5, the results of the F-tests based on the CTBS-estimators were similar to the ones based on COBS-estimators, although the results based on COBS-estimators were always slightly more conservative (Table II). For the breakdown parameter 0.2, the F-tests were often anti-conservative for scenarios with positively correlated contaminated dimensions (25–28), with more type I error inflation for the tests based on the CTBS- than COBS-estimators. Overall, our simulation results suggest that “robustified” MANOVA-type F-test based on COBS with the breakdown parameter 0.5 may be expected to yield empirical levels close to the nominal alpha level of 0.05 for the data structure and

contamination pattern similar to our data example, which is not the case with the classical MANOVA-type F-test.

5. ANALYSIS OF LIPIDS DATA

Initial analysis of cholesterol levels was performed by fitting model (14) using the standard REML estimation approach as implemented in SAS 9.2 PROC MIXED. Resulting β parameter estimates with the corresponding standard errors are shown in Table II. MANOVA-type F-tests based on these REML estimates indicated significant differences between draw 1 and draw 2 ($p=0.001$) and between each pair of treatments before and after the switch ($p=0.014$ for CBZ vs. LTG; $p=0.014$, for CBZ vs. LEV; $p=0.037$ for PHT vs. LTG; and $p=0.020$ for PHT vs. LEV). However, examination of residuals indicated multiple potential outliers. The line plots in Figure 1 also indicate potential outliers in random subject effects in each cholesterol measure. Furthermore, only TRIG measures have a visible decrease from draw 1 to draw 2, which is consistent across all patients groups defined by the drug 1/drug 2 combination, and that decrease is largest in the 3 patients that are potential outliers. For the LDL measures, the overall decrease from draw 1 to draw 2 is clearly visible only for the group of patients that were switched from CBZ to LTG. Therefore, robust analysis is necessary to provide better understanding of these data.

Thus, the data were analyzed using the COBS- and CTBS-estimators with the breakdown parameters 0.2 and 0.5 (for CTBS, $ARP=0.00001$ if $\epsilon^* = 0.5$ and $ARP=0.01$ if $\epsilon^* = 0.2$), similar to the simulation study. The resulting estimates of β are shown in Table III. As may be expected, the largest discrepancies between robust and REML estimates were observed for the TRIG measures. The mean LDL estimates were generally similar for all treatments except LTG, for which robust estimates were lower by 6–13% than the REML estimates. As compared to the REML estimates based on all data, the robust estimates with breakdown parameter 0.5 yielded the mean TRIG estimates lower by 23%–25% for CBZ and by 37%–41% for PHT treatment (Table III, in bold). The robust estimates of TRIG means based on the breakdown parameter 0.2 were similar to the REML estimates, failing to down weight the outliers sufficiently. The “robustified” MANOVA-type F-tests based on robust COBS estimators with the breakdown parameter 0.5 yielded significant overall difference between draw 1 and draw 2 ($p=0.021$) and significant difference between CBZ and LTG ($p=0.039$). Using these “robustified” F-tests, we do not detect as significant the differences between other individual drugs before and after the switch that were significant using the standard REML approach (CBZ vs. LEV; PHT vs. LTG; and PHT vs. LEV). Line plots in Figure 1 suggest that significance of these comparisons using non-robust REML approach may be driven by high end outliers (red lines for CBZ to LEV switching patients, black lines for PHT to LTG switching patients; blue lines for PHT to LEV switching patients). When outliers are down weighted using robust estimation, the mean difference estimates become smaller and not detectable as significant, possibly due to the small size of the corresponding patient subgroups (5 patients switching from CBZ to LEV; 4 patients switching from PHT to LTG; and 10 patients switching from PHT to LEV), since the study was not powered for the pair-wise treatment comparisons. Note that the largest subgroup of 15 patients was switched from CBZ to LTG. Thus, based on robust analyses results, we may conclude that the average change in all cholesterol measures was significant between draw 1 and draw 2 driven by the difference between CBZ and LTG. The largest decreases were observed in TRIG measures, but moderate decreases in LDL and HDL measures (Table III) also contributed to the significance of the multivariate comparisons.

6. DISCUSSION

The original purpose of this work was to develop a modification of the constrained S-estimator [11] applicable to LME models for vector outcomes with correlated dimensions. Earlier proposed BIEE estimators [5–6] may be applied to the multivariate LME models, but they do not necessarily have good global robust properties. Our modification also extends application of the constrained S-estimator to the LME models for longitudinal data with possible serial correlations. Theoretical developments of the modified constrained S-estimator parallel the work of Copt and Victoria-Feser [11], but our algorithm for computing constrained S-estimators is substantially different. The key differences include (i) random subsampling approach to selecting multiple starting values; (ii) computation of multiple local minima of the $\det(V)$, which allows selecting a global minimum; (iii) explicit incorporation of constraints into estimating equations; (iv) the use of tuning parameters in the iterative formulas to achieve convergence to the points that satisfy the constraint. Random subsampling is also used in most recent algorithms for computing regression S-estimators [23–24], but these algorithms are not applicable here because they do not provide means for estimating parameters of a structured covariance matrix with the covariance components corresponding to error and random effects vectors.

Our simulation study addressed the finite sample performance of the constrained S-estimators based on the original and translated biweight in data sets with relatively high contaminations (deterministically 10% of the data), which involved multiple negatively or positively correlated dimensions. The standard REML estimators performed grossly unsatisfactorily under these contaminations, while the constrained S-estimators with breakdown parameter 0.5 were virtually unbiased. Similar performance of COBS- and CTBS-estimators (with $ARP=0.01$) with the breakdown parameter 0.5 may be expected if one compares the corresponding weighting functions (Figure 2), which have similar magnitude for Mahalanobis distances between 4 and 8. Meanwhile, with the breakdown parameter 0.2, the weighting functions for COBS and CTBS ($ARP=0.00001$) behave quite differently for Mahalanobis distances greater than 4. With the lower breakdown parameter 0.2, we observed an advantage of the COBS estimator in terms of smaller relative bias. The advantage of the COBS estimator was most apparent in scenarios with outliers in positively correlated dimensions.

We also considered a “robustification” of the approximate MANOVA-type F-test [20–21], in which the REML estimates are replaced by the S-estimates, for the purpose of comparison of estimators at the inferential level and for possible analysis of the real data. Our main motivation to use such a “robustified” test was the concern about using the asymptotic robust tests for the data with a relatively small number of independent observation vectors of relatively high dimension (34 subjects with 6-dimensional observation vectors in our data example). For the standard LME models with multivariate responses, the approximate MANOVA-type F-test is at worst conservative for small sample sizes and high dimensions of responses, which are penalized by limiting the degrees of freedom. We wanted to preserve these properties while using the robust S-estimates for computing the test statistic. However, this “robustified” MANOVA-type F-test is ad hoc and the corresponding F-test statistic does not necessarily have the postulated F-distribution (as suggested by the simulation results without contamination). Therefore, such a “robustified” MANOVA-type F-statistic can not be recommended for general practice without additional work addressing its proper distribution. Such investigation is beyond the scope of this paper and is a separate research topic. Meanwhile, our rather thorough investigation using simulations suggested that “robustified” MANOVA-type F-test based on COBS with the breakdown parameter 0.5 may be expected to have empirical levels near the nominal alpha level of 0.05 for the data structure and contamination pattern of our real data. Therefore, this “robustified” F-test was

applied to the data example to yield much more conservative and realistic results than application of the MANOVA-type F-test based on the standard REML estimates.

Acknowledgments

The authors thank Dr. Scott Mintzer of the Jefferson Comprehensive Epilepsy Center, Department of Neurology, Thomas Jefferson University for permission to use cholesterol data as an example. We are also grateful to two anonymous reviewers and Associate Editor for the comments that helped to improve this manuscript. This work was supported by the grant AR054596 from NIH/NIAMSD.

References

1. Mintzer S, Skidmore CT, Abidin CJ, Morales MC, Chervoneva I, Capuzzi DM, Sperling MR. Effects of antiepileptic drugs on lipids, homocysteine, and C-reactive protein. *Annals of Neurology*. 2009; 65:448–456. [PubMed: 19296463]
2. Heritier, S.; Cantoni, E.; Copt, S.; Victoria-Feser, M. *Robust Methods in Biostatistics*. John Wiley & Sons; New York: 2009.
3. Huggins RM. A robust approach to the analysis of repeated measures. *Biometrics*. 1993; 49:715–720.
4. Richardson AM, Welsh AH. Robust restricted maximum likelihood in mixed linear models. *Biometrics*. 1995; 51:1429–1439.
5. Richardson AM. Bounded influence estimation in the mixed linear model. *Journal of the American Statistical Association*. 1997; 92:154–161.
6. Welsh, AH.; Richardson, AM. Approaches to the robust estimation of mixed models. In: Maddal, GS.; Rao, CR., editors. *Handbook of Statistics*. Vol. 15. Elsevier Science; Amsterdam: 1997. p. 343–383.
7. Huber, PJ. *Robust statistics*. John Wiley & Sons; New York: 1982. p. 127–152.
8. Maronna RA. Robust M-estimators of multivariate location and scatter. *The Annals of Statistics*. 1976; 4:51–67.
9. Bilodeau M, Duchesne P. Robust estimation of the SUR model. *The Canadian Journal of Statistics*. 2000; 28:277–288.
10. Van Aelst S, Willems G. Multivariate regression S-estimators for robust estimation and inference. *Statistica Sinica*. 2005; 15:981–1001.
11. Copt S, Victoria-Feser M-P. High-breakdown inference for mixed linear models. *Journal of the American Statistical Association*. 2006; 101:292–300.10.1198/016214505000000772
12. Lopuhaä HP. On the relation between S-estimators and M-estimators of multivariate location and covariance. *The Annals of Statistics*. 1989; 17:1662–1683.
13. Rousseeuw, PJ.; Yohai, VJ. Robust Regression by Means of S-Estimators. In: Franke, J.; Hardle, W.; Martin, RD., editors. *Robust and Nonlinear Time Series Analysis*. Springer-Verlag; New York: 1984. p. 256–272.
14. Ruppert D. Computing S-estimators for regression and multivariate location/dispersion. *Journal of Computational and Graphical Statistics*. 1992; 1:253–270.
15. He X, Fung WK. High breakdown estimation for multiple populations with applications to discriminant analysis. *Journal of Multivariate Analysis*. 2000; 72:151–162.
16. Maronna RA, Yohai VJ. Robust regression with both continuous and categorical predictors. *Journal of Statistical Planning and Inference*. 2000; 89:197–214.
17. Beaton AE, Tukey JW. The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*. 1974; 16:147–185.
18. Rocke DM. Robustness properties of S-estimators of multivariate location and shape in high dimension. *The Annals of Statistics*. 1996; 24:1327–1345.
19. Copt S, Heritier S. Robust alternatives to the F-test in mixed linear models based on MM-estimates. *Biometrics*. 2007; 63:1045–1052.10.1111/j.1541-0420.2007.00804.x [PubMed: 17484776]

20. Vonesh EF, Carter RL. Efficient inference for random-coefficient growth curve models with unbalanced data. *Biometrics*. 1987; 43:617–628. [PubMed: 3663819]
21. Vonesh, EF.; Chinchilli, VM. *Linear and Nonlinear Models for the Analysis of Repeated Measures*. Marcel Dekker; New York: 1997. p. 417-418.
22. Heritier S, Ronchetti E. Robust bounded-influence tests in general parametric models. *Journal of the American Statistical Association*. 1994; 89:897–904.
23. Rousseeuw, PJ.; Van Driessen, K. An algorithm for positive-breakdown regression based on concentration steps. In: Gaul, W.; Opitz, O.; Schader, M., editors. *Data Analysis: Scientific Modeling and Practical Application*. Springer Verlag; New York: 2000. p. 335-346.
24. Salibian-Barrera M, Yohai VJ. A fast algorithm for S-regression estimates. *Journal of Computational and Graphical Statistics*. 2006; 15:414–427.10.1198/106186006X113629

APPENDIX A: ρ -FUNCTIONS

The constrained S-estimator may be based on any redescending ρ -function that satisfies the following regularity conditions from [12]:

- R1** $\rho(d)$ is symmetric with continuous derivative and $\rho(0) = 0$
- R2** $\exists c_0 > 0$ such that $\rho(d)$ is strictly increasing on $[0, c_0]$ and constant on $[c_0, \infty)$
- R3** $d^{-1} \rho'(d)$ and $\rho''(d)$ are bounded and continuous.

The original or Tukey's biweight ρ -function is given by

$$\rho_B(d; c_0) = \begin{cases} d^2/2 - d^4/(2c_0^2) + d^6/(6c_0^4), & \text{if } d \leq c_0 \\ c_0^2/6, & \text{if } d > c_0. \end{cases}$$

Constant c_0 is chosen as a solution of (6) with $\rho(d) = \rho_B(d; c_0)$ and some $\varepsilon^* = 0.5$.

The translated biweight ρ -function [18] is given by

$$\rho_R(d; c, S) = \begin{cases} \frac{1}{2}d^2, & \text{if } 0 \leq d \leq S \\ \frac{1}{2}S^2 - \frac{S^2(S^4 - 5S^2c^2 + 15c^4)}{30c^4} + d^2\left(\frac{1}{2} + \frac{S^4}{2c^4} - \frac{S^2}{c^2}\right) + d^3\left(\frac{4S}{3c^2} - \frac{4S^3}{3c^4}\right) + d^4\left(\frac{3S^2}{2c^4} - \frac{1}{2c^2}\right) - \frac{4S}{5c^4}d^5 + \frac{1}{6c^4}d^6, & \text{if } S < d \leq S+c \\ \frac{1}{2}S^2 + \frac{c(5c+16S)}{30}, & \text{if } d > S+c, \end{cases}$$

where c and S are the constants chosen to control asymptotic rejection probability ζ and breakdown parameter ε^* . Under the p -variate normal model (3), the asymptotic rejection probability (ARP) is defined [18] as $\zeta = 1 - F_{\chi_p^2}(S+c)$, where χ_p^2 is the chi-square distribution with p degrees of freedom and distribution function $F_{\chi_p^2}(x)$. That is, the ARP is the probability that squared Mahalanobis distance exceeds $S+c$ under the assumption of uncontaminated multivariate normal model. To guarantee that the ARP is at least ζ and the

breakdown parameter is ε^* , the constants c and S are chosen so that $S+c = \left(F_{\chi_p^2}^{-1}(1-\zeta)\right)^{1/2}$ and, for a given $\varepsilon^* = 0.5$, $\rho_R(d; c, S)$ satisfies equation (6).

APPENDIX B: ASYMPTOTIC DISTRIBUTION OF S-ESTIMATORS

When $\mathbf{V}$ is an arbitrary positive definite symmetric matrix, the results of Lopuhaa [12], Bilodeau and Duchesne [9], and Van Aelst and Willems [10] imply that $\widehat{\beta}_\rho$ and $\widehat{\mathbf{V}}_\rho$ are asymptotically independent with

$$\sqrt{M} \left(\begin{bmatrix} \text{vec}(\widehat{\mathbf{V}}_\rho) \end{bmatrix}' - [\text{vec}(\mathbf{V})]' \right) \xrightarrow{L} N(\mathbf{0}, \mathbf{\Omega}_\mathbf{V})$$

where operator $\text{vec}(\mathbf{V})$ stacks the columns of $\mathbf{V}$ on the top of each other and $\mathbf{\Omega}_\mathbf{V}$ is given in (8). Since $\mathbf{V} = \mathbf{V}(\boldsymbol{\theta})$ is a linear function of $\boldsymbol{\theta}$, then $\widehat{\beta}_\rho$ and $\widehat{\theta}_\rho$ are also asymptotically independent and

$$\sqrt{M} \left(\begin{bmatrix} \widehat{\beta}_\rho' & \widehat{\theta}_\rho' \end{bmatrix}' - [\beta', \theta']' \right) \xrightarrow{L} N(\mathbf{0}, \text{Diag}(\mathbf{\Omega}_\beta, \mathbf{\Omega}_\theta))$$

for some $\mathbf{\Omega}_\theta$ and $\mathbf{\Omega}_\beta$. Since the derivatives with respect to $\boldsymbol{\beta}$ are the same for the multivariate LME and multivariate regression model, matrix $\mathbf{\Omega}_\beta$ is the same as for the multivariate

regression model as derived in [9], $\mathbf{\Omega}_\beta = \frac{\alpha_1}{\alpha_2} \{E[\mathbf{A}_i' \mathbf{V}^{-1} \mathbf{A}_i]\}^{-1}$, where constants α_1 and α_2 are the same as in [12],

$$\begin{aligned} \alpha_1 &= p^{-1} E_{\chi_p^2} [d^2 \{u(d)\}^2] \\ \alpha_2 &= E_{\chi_p^2} [u(d) + p^{-1} du'(d)]. \end{aligned}$$

Under the assumption **C3**,

$$E[\mathbf{A}_i' \mathbf{V}^{-1} \mathbf{A}_i] = \mathbf{\Omega} = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{i=1}^M \mathbf{A}_i' \mathbf{V}^{-1} \mathbf{A}_i,$$

which implies that $\sqrt{M} \left(\begin{bmatrix} \widehat{\beta}_\rho' \end{bmatrix}' - \beta' \right) \xrightarrow{L} N\left(\mathbf{0}, \frac{\alpha_1}{\alpha_2} \mathbf{\Omega}^{-1}\right)$.

Since $\mathbf{V} = \mathbf{V}(\boldsymbol{\theta})$ is continuously differentiable function of $\boldsymbol{\theta}$,

$$\mathbf{\Omega}_\mathbf{V} = \frac{\partial(\text{vec}(\mathbf{V}(\boldsymbol{\theta})))}{\partial \boldsymbol{\theta}} \mathbf{\Omega}_\theta \frac{\partial(\text{vec}(\mathbf{V}(\boldsymbol{\theta})))'}{\partial \boldsymbol{\theta}}$$

and under assumption **C2**, we obtain

$$\mathbf{\Omega}_\theta = H \frac{\partial(\text{vec}(\mathbf{V}(\theta)))'}{\partial \theta} \mathbf{\Omega}_V \frac{\partial(\text{vec}(\mathbf{V}(\theta)))}{\partial \theta} H$$

$$H = \left\{ \frac{\partial(\text{vec}(\mathbf{V}(\theta)))'}{\partial \theta} \frac{\partial(\text{vec}(\mathbf{V}(\theta)))}{\partial \theta} \right\}^{-1}.$$

The constants σ_1 and σ_2 in (8) are also the same as in [12],

$$\sigma_1 = \frac{p(1+p)E_{\chi_p^2}[d^4\{u(d)\}^2]}{E_{\chi_p^2}[d^2(d\frac{\partial}{\partial d}u(d)+u(d))+(p+1)d^2u(d)]}$$

$$\sigma_2 = -\frac{2}{p}\sigma_1 + \frac{4E_{\chi_p^2}[(\rho(d)-b_0)^2]}{E_{\chi_p^2}[d^2u(d)]}.$$

APPENDIX C: DERIVATION OF SYSTEM (11)

First, we exclude λ from (9)&(10). Since $\mathbf{V} = \mathbf{BDB}'$, where $\mathbf{D} = \sum_{g=1}^l \theta_g \mathbf{L}_g$, we can write $\mathbf{V} = \sum_{g=1}^l \theta_g \mathbf{B} \mathbf{L}_g \mathbf{B}'$. Hence $\frac{\partial \mathbf{V}}{\partial \theta_j} = \mathbf{B} \mathbf{L}_j \mathbf{B}'$. Then multiplying each equation $\frac{\partial \mathbf{L}}{\partial \theta_j}$ in (10) by θ_j and summing $\theta_j \times \frac{\partial \mathbf{L}}{\partial \theta_j}$ over j gives

$$\begin{aligned} & \sum_{j=1}^l \theta_j \text{tr} \left[\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_j} \right] \\ & + \sum_{j=1}^l \theta_j \frac{\lambda}{2M} \sum_{i=1}^M u(d_i) (\mathbf{y}_i - \mathbf{A}_i \beta)' \mathbf{V}^{-1} \mathbf{B} \mathbf{L}_j \mathbf{B}' \mathbf{V}^{-1} (\mathbf{y}_i \\ & - \mathbf{A}_i \beta) \\ & = \text{tr} \left[\mathbf{V}^{-1} \sum_{g=1}^l \theta_g \mathbf{B} \mathbf{L}_g \mathbf{B}' \right] \\ & + \frac{\lambda}{2M} \sum_{i=1}^M u(d_i) (\mathbf{y}_i - \mathbf{A}_i \beta)' \mathbf{V}^{-1} \sum_{j=1}^l \theta_j \mathbf{B} \mathbf{L}_j \mathbf{B}' \mathbf{V}^{-1} (\mathbf{y}_i - \mathbf{A}_i \beta) \\ & = \text{tr}[\mathbf{V}^{-1} \mathbf{V}] \\ & + \frac{\lambda}{2M} \sum_{i=1}^M u(d_i) (\mathbf{y}_i - \mathbf{A}_i \beta)' \mathbf{V}^{-1} (\mathbf{y}_i \\ & - \mathbf{A}_i \beta) \\ & = p + \frac{\lambda}{2M} \sum_{i=1}^M u(d_i) d_i^2 = 0. \end{aligned}$$

Hence $\lambda = -2pM \left(\sum_{i=1}^M u(d_i) d_i^2 \right)^{-1}$ and (10) may be written as

$$tr\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta}\right) - \frac{\sum_{i=1}^M u(d_i) (\mathbf{y}_i - \mathbf{A}_i\boldsymbol{\beta})' \mathbf{V}^{-1} \frac{\partial\mathbf{V}}{\partial\theta} \mathbf{V}^{-1} (\mathbf{y}_i - \mathbf{A}_i\boldsymbol{\beta})}{\sum_{i=1}^M \frac{1}{p} u(d_i) d_i^2} = 0.$$

Further, we can express

$$tr\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_j}\right) = tr\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_j}\mathbf{V}^{-1}\mathbf{V}\right) = \sum_{k=1}^l \theta_k tr[\mathbf{V}^{-1}\mathbf{B}\mathbf{L}_j\mathbf{B}'\mathbf{V}^{-1}\mathbf{B}\mathbf{L}_k\mathbf{B}']$$

Hence, $tr\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta}\right) = \mathbf{Q}(\boldsymbol{\theta})\boldsymbol{\theta}$, where $\mathbf{Q}(\boldsymbol{\theta})$ the $l \times l$ matrix with entries $\mathbf{Q}(\boldsymbol{\theta})_{jk} = tr[\mathbf{V}^{-1}\mathbf{B}\mathbf{L}_j\mathbf{B}'\mathbf{V}^{-1}\mathbf{B}\mathbf{L}_k\mathbf{B}']$ and (10) is written as $\mathbf{Q}(\boldsymbol{\theta})\boldsymbol{\theta} = \mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\theta})$, where $l \times 1$ vector $\mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\theta})$ has entries

$$\mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\theta})_j = \frac{\sum_{i=1}^M u(d_i) (\mathbf{y}_i - \mathbf{A}_i\boldsymbol{\beta})' \mathbf{V}^{-1} \mathbf{B}\mathbf{L}_j\mathbf{B}' \mathbf{V}^{-1} (\mathbf{y}_i - \mathbf{A}_i\boldsymbol{\beta})}{\sum_{i=1}^M \frac{1}{p} u(d_i) d_i^2}.$$

Thus, the proposed S-estimators must satisfy (11).

APPENDIX D: ALGORITHM FOR COMPUTING CONSTRAINED S-ESTIMATORS IN LME MODELS

The following steps are repeated for computing the constrained S-estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$.

- S1** Randomly select a sub-sample $\{\mathbf{y}_{j_s}, s = 1, \dots, S\}$, of independent highest level clusters $\mathbf{y}_j, j = 1, \dots, M$. The sub-samples $\{\mathbf{y}_{j_s}\}$ have to be small enough to include sub-samples completely free from contamination and large just enough to make all parameters in the model estimable.
- S2** From $\{\mathbf{y}_{j_s}, s = 1, \dots, S\}$, compute the standard REML estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ to be used as the starting values for (13). Alternative robust estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ may be used instead.
- S3** Perform pre-determined number of iterations N_I using (13) while evaluating the absolute value of the relative change in $\det(\mathbf{V})$, $|\Delta\det(\mathbf{V})|/\det(\mathbf{V})$, between two consecutive iterations. Record estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ and $\det(\mathbf{V})$ only if $|\Delta\det(\mathbf{V})|/\det(\mathbf{V}) < \delta_V$ and $|M^{-1} \sum_{i=1}^M \rho(d_i) - b_0| < \delta_C$ for some sufficiently small δ_C and δ_V after some number of iterations not exceeding N_I .

The candidate estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, which satisfy (11), are obtained by repeating steps S1–S3 for pre-determined number of random starts. Then the candidate estimator which yields the global minimum of $\det(\mathbf{V})$ is the constrained S-estimator of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$.

Figure 1.

Repeated lipids measures in 34 epilepsy patients. The first row shows scatter plots for all treatments combined (CBZ with black squares; PHT with blue circles; LEV with red diamonds; and LTG with green triangles). The second row shows the individual changes in HDL, LDL, and TRIG after the switch from CBZ to LTG (15 patients, green lines), from CBZ to LEV (5 patients, red lines), from PHT to LTG (4 patients, black lines), and from PHT to LEV (11 patients, blue lines).

Figure 2.

Weighting functions for original Tukey's biweight (OB) and translated biweight (TB) that correspond to breakdown parameter $e=0.5$ and $e=0.2$. Asymptotic rejection probability is 0.01 for TB with $e=0.5$ and 0.00001 with $e=0.2$.

Figure 3. Distribution of relative bias in β parameter estimates for contaminated CBZ treatment group. Scenarios 10–13 and 20–23 have 10% of subjects with negatively correlated contaminations in dimensions HDL and TRIG with correlation coefficient -0.7 . Scenarios 15–18 and 25–28 have 10% of subjects with positively correlated contaminations in dimensions LDL and TRIG with correlation coefficient 0.4 . The breakdown parameter is 0.5 for scenarios 10–13 and 0.2 for scenarios 20–23

Figure 4.

Distribution of relative bias in covariance parameter estimates corresponding to contaminated dimensions 1 (HDL) and 3 (TRIG), which are negatively correlated with correlation coefficient -0.7 . The breakdown parameter is 0.5 for scenarios 10–13 and 0.2 for scenarios 20–23.

Figure 5.

Distribution of relative bias in covariance parameter estimates corresponding to contaminated dimensions 2 (LDL) and 3 (TRIG), which are positively correlated with correlation coefficient 0.4. The breakdown parameter is 0.5 for scenarios 15–18 and 0.2 for scenarios 25–28.

Numbers of subjects with contaminated observations (the difference between the contaminated and “clean” model means was equal to 6 “clean” model standard deviations) in one or two responses. In parenthesis is the treatment (either in the first or the second replicate), which was contaminated.

Table 1

e*	Scenario	Contaminated in TRIG						Contaminated in HDL						Contaminated in LDL					
		CBZ	CBZ	PHT	PHT	LTG	PHT	CBZ	CBZ	LTG	PHT	PHT	LTG	CBZ	CBZ	LTG	PHT	PHT	LTG
0.5	1	4(CBZ)																	
0.5	2	2(CBZ)	2(CBZ)																
0.5	3	4(LEV)																	
0.5	4	2(LEV)		2(LEV)															
0.5	5		2(CBZ)			2(PHT)													
0.5	6	1(CBZ)	1(CBZ)	1(PHT)		1(PHT)													
0.5	10	4(CBZ)						4(CBZ)											
0.5	11	2(CBZ)	2(CBZ)					2(CBZ)2(CBZ)											
0.5	12	4(LEV)						4(LEV)											
0.5	13	2(LEV)		2(LEV)				2(LEV)		2(LEV)									
0.5	15	4(CBZ)												4(CBZ)					
0.5	16	2(CBZ)	2(CBZ)											2(CBZ)	2(CBZ)				
0.5	17	4(LEV)												4(LEV)					
0.5	18	2(LEV)		2(LEV)										2(LEV)					2(LEV)
0.2	20	4(CBZ)						4(CBZ)											
0.2	21	2(CBZ)	2(CBZ)					2(CBZ)2(CBZ)											
0.2	22	4(LEV)						4(LEV)											
0.2	23	2(LEV)		2(LEV)				2(LEV)		2(LEV)									
0.2	25	4(CBZ)												4(CBZ)					
0.2	26	2(CBZ)	2(CBZ)											2(CBZ)	2(CBZ)				
0.2	27	4(LEV)												4(LEV)					
0.2	28	2(LEV)		2(LEV)										2(LEV)					2(LEV)

Table II

Empirical Type I errors of the MANOVA-type F-tests with nominal alpha 0.05. For each scenario 500 data sets were simulated. Bold and bold italic fonts indicate rejection rates significantly lower or higher, respectively, than 5%.

Scen	Σ	draw 1 vs draw 2			CBZ vs LEV			CBZ vs LTG			PHT vs LEV			PHT vs LTG		
		RMLE	COBS	CTBS	RMLE	COBS	CTBS	RMLE	COBS	CTBS	RMLE	COBS	CTBS	RMLE	COBS	CTBS
0	0.5	0.044	0.057	0.092	0.038	0.034	0.083	0.024	0.034	0.066	0.034	0.077	0.092	0.040	0.060	0.085
1	0.5	0.138	0.026	0.041	0.502	0.081	0.094	0.372	0.044	0.053	0.016	0.039	0.051	0.012	0.018	0.024
2	0.5	0.150	0.057	0.059	0.422	0.043	0.067	0.460	0.059	0.077	0.018	0.036	0.050	0.020	0.036	0.052
3	0.5	0.124	0.037	0.046	0.494	0.085	0.091	0.022	0.041	0.044	0.380	0.043	0.055	0.012	0.021	0.040
4	0.5	0.174	0.042	0.059	0.424	0.053	0.065	0.016	0.028	0.040	0.460	0.060	0.069	0.020	0.042	0.065
5	0.5	0.096	0.035	0.047	0.038	0.025	0.030	0.050	0.058	0.057	0.038	0.028	0.051	0.046	0.051	0.070
6	0.5	0.086	0.036	0.038	0.042	0.045	0.049	0.036	0.043	0.057	0.028	0.031	0.034	0.052	0.040	0.047
10	0.5	0.534	0.041	0.062	1.000	0.063	0.104	0.594	0.025	0.036	0.006	0.041	0.054	0.006	0.025	0.044
11	0.5	0.464	0.047	0.047	0.960	0.038	0.058	0.956	0.035	0.058	0.010	0.050	0.061	0.010	0.035	0.044
12	0.5	0.548	0.020	0.047	1.000	0.058	0.075	0.008	0.044	0.052	0.626	0.044	0.057	0.008	0.023	0.047
13	0.5	0.500	0.039	0.053	0.952	0.051	0.061	0.004	0.021	0.045	0.962	0.039	0.067	0.006	0.042	0.045
15	0.5	0.502	0.037	0.059	0.996	0.052	0.085	0.430	0.052	0.057	0.006	0.023	0.047	0.010	0.023	0.043
16	0.5	0.428	0.045	0.049	0.888	0.058	0.068	0.884	0.053	0.072	0.010	0.019	0.035	0.002	0.034	0.042
17	0.5	0.490	0.035	0.030	0.996	0.073	0.058	0.012	0.025	0.039	0.434	0.025	0.042	0.014	0.025	0.028
18	0.5	0.392	0.031	0.046	0.850	0.057	0.072	0.006	0.039	0.058	0.876	0.057	0.055	0.006	0.021	0.036
20	0.2	0.506	0.014	0.044	1.000	0.014	0.064	0.616	0.000	0.024	0.024	0.010	0.016	0.016	0.006	0.010
21	0.2	0.424	0.006	0.026	0.940	0.010	0.052	0.938	0.006	0.042	0.004	0.002	0.008	0.006	0.002	0.004
22	0.2	0.554	0.008	0.048	1.000	0.014	0.086	0.010	0.004	0.006	0.588	0.006	0.050	0.004	0.002	0.008
23	0.2	0.478	0.006	0.040	0.950	0.004	0.048	0.012	0.010	0.012	0.956	0.008	0.064	0.016	0.006	0.006
25	0.2	0.492	0.050	0.262	0.994	0.266	0.508	0.428	0.042	0.232	0.008	0.008	0.010	0.006	0.002	0.000
26	0.2	0.396	0.014	0.170	0.898	0.068	0.340	0.858	0.082	0.342	0.004	0.000	0.004	0.006	0.002	0.004
27	0.2	0.486	0.040	0.242	1.000	0.256	0.508	0.008	0.012	0.012	0.460	0.034	0.278	0.008	0.004	0.010
28	0.2	0.420	0.034	0.180	0.890	0.102	0.381	0.004	0.008	0.008	0.878	0.098	0.353	0.002	0.006	0.008

Table III
Fixed effects parameter estimates (with standard errors, SE) for the cholesterol data example

Drug	Dimension	Breakdown parameter 0.2				Breakdown parameter 0.5			
		REML(SE)	CTBS	COBS	CTBS	COBS	CTBS	COBS	REML(SE) &
CBZ	HDL	65 (3.4)	67 (3.4)	67 (3.5)	67 (3.8)	68 (3.8)			66 (3.4)
LEV	HDL	56 (3.5)	57 (3.6)	57 (3.6)	58 (3.9)	60 (4.0)			59 (3.4)
LTG	HDL	60 (3.4)	59 (3.5)	59 (3.5)	61 (3.8)	61 (3.9)			56 (3.5)
PHT	HDL	58 (3.5)	59 (3.6)	59 (3.7)	61 (4.0)	63 (4.0)			58 (3.6)
CBZ	LDL	124 (8.3)	132 (8.2)	130 (8.3)	129 (9.6)	125 (8.4)			115 (6.9)
LEV	LDL	121 (9.2)	122 (8.7)	121 (8.8)	121 (10.1)	120 (9.0)			121 (7.1)
LTG	LDL	132 (8.4)	124 (8.3)	122 (8.4)	117 (9.7)	115 (8.6)			129 (7.7)
PHT	LDL	129 (9.5)	128 (8.9)	126 (9.0)	129 (10.4)	122 (9.2)			121 (8.3)
CBZ	TRIG	120 (19.7)	119 (19.5)	108 (18.1)	92 (10.0)	90 (9.5)			90 (9.5)
LEV	TRIG	93 (22.3)	92 (21.0)	90 (19.5)	87 (10.9)	84 (10.3)			85 (9.6)
LTG	TRIG	85 (20.1)	88 (20.0)	82 (18.5)	71 (10.3)	71 (9.8)			93 (10.8)
PHT	TRIG	172 (23.0)	171 (21.6)	157 (20.0)	108 (11.2)	101 (10.6)			116 (11.4)

& Data trimmed from potential outliers