

Linear Mixed Effects Models under Inequality Constraints with Applications

Laura Farnan¹, Anastasia Ivanova², Shyamal D. Peddada^{3*}

1 Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina, United States of America, **2** Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, United States of America, **3** Biostatistics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, United States of America

Abstract

Constraints arise naturally in many scientific experiments/studies such as in, epidemiology, biology, toxicology, etc. and often researchers ignore such information when analyzing their data and use standard methods such as the analysis of variance (ANOVA). Such methods may not only result in a loss of power and efficiency in costs of experimentation but also may result poor interpretation of the data. In this paper we discuss constrained statistical inference in the context of linear mixed effects models that arise naturally in many applications, such as in repeated measurements designs, familial studies and others. We introduce a novel methodology that is broadly applicable for a variety of constraints on the parameters. Since in many applications sample sizes are small and/or the data are not necessarily normally distributed and furthermore error variances need not be homoscedastic (i.e. heterogeneity in the data) we use an empirical best linear unbiased predictor (EBLUP) type residual based bootstrap methodology for deriving critical values of the proposed test. Our simulation studies suggest that the proposed procedure maintains the desired nominal Type I error while competing well with other tests in terms of power. We illustrate the proposed methodology by re-analyzing a clinical trial data on blood mercury level. The methodology introduced in this paper can be easily extended to other settings such as nonlinear and generalized regression models.

Citation: Farnan L, Ivanova A, Peddada SD (2014) Linear Mixed Effects Models under Inequality Constraints with Applications. PLoS ONE 9(1): e84778. doi:10.1371/journal.pone.0084778

Editor: Xiaofeng Wang, Cleveland Clinic Lerner Research Institute, United States of America

Received: June 17, 2013; **Accepted:** November 18, 2013; **Published:** January 21, 2014

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: The research of SDP was supported by the Intramural Research Program of the NIEHS, National Institutes of Health (Z01 ES102345-04). LF and AI work was supported in part by National Institutes of Health grant RO1 CA120082-01A1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Dr. Peddada is a PLOS ONE Editorial Board member and also currently serving as an Academic Editor of PLOS ONE. Dr. Peddada confirms that this does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: peddada@niehs.nih.gov

Introduction

In many applications researchers are typically interested in testing for trends or patterns in mean response among two or more experimental or study groups rather than just testing if the groups are different or not. For instance, toxicologists are typically interested in detecting dose-related trends in mean response such as trends in tumor incidence as the dose of a toxin increases [1–2]. In cell and circadian biology researchers are often interested in the phase order of genes participating in cell cycle or the circadian clock [3–10]. In all such situations researchers are interested in testing for an order among statistical parameters. By performing standard statistical procedures, e.g. the F-test in the case of Euclidean space data, all a researcher can conclude is that there is at least one group which is different from the rest of the experimental groups (in the parameter of interest). Such a conclusion is not very useful if the researcher is interested in proving that the mean response is increasing with dose of the toxin. Furthermore, as noted from the simulation study reported in Figure 1, the power of the standard ANOVA based F-test can be substantially smaller than the power of a test that is designed to detect trend in mean response. For example, note that for a total sample size of 48 (i.e. $n = 12$ per group), the standard ANOVA based F-test yields of power of 0.58 whereas the Williams'

constrained test for trend (which will be described in detail later in this article) yields a power 0.81. Conversely, as also noted in Figure 1, the sample size needed by Williams' constrained test for a power of approximately 0.80 is 48 whereas the F-test would require a total sample size of 76. This simple motivating example illustrates our point that, in addition to being scientifically relevant, a greater efficiency can be achieved by using statistical methods that take into consideration the investigator's true hypothesis of interest rather than performing a generic methodology, as often done in scientific literature.

The field of statistics that deals with statistical methods designed to test ordered or constrained hypotheses is commonly called order restricted inference or constrained statistical inference. There exists a very large body of literature on order restricted (constrained inference) spanning nearly sixty years with four books written on the subject, including a recent book by Silvapulle and Sen [11]. Furthermore, for testing for some commonly encountered inequalities in the absence of any covariates in high dimensional data (e.g. gene expression studies), a freely downloadable software called ORIOGEN (Order Restricted Inference for Ordered Gene Expression) (<http://www.niehs.nih.gov/research/resources/software/biostatistics/oriogen/>) was developed in [12,13].

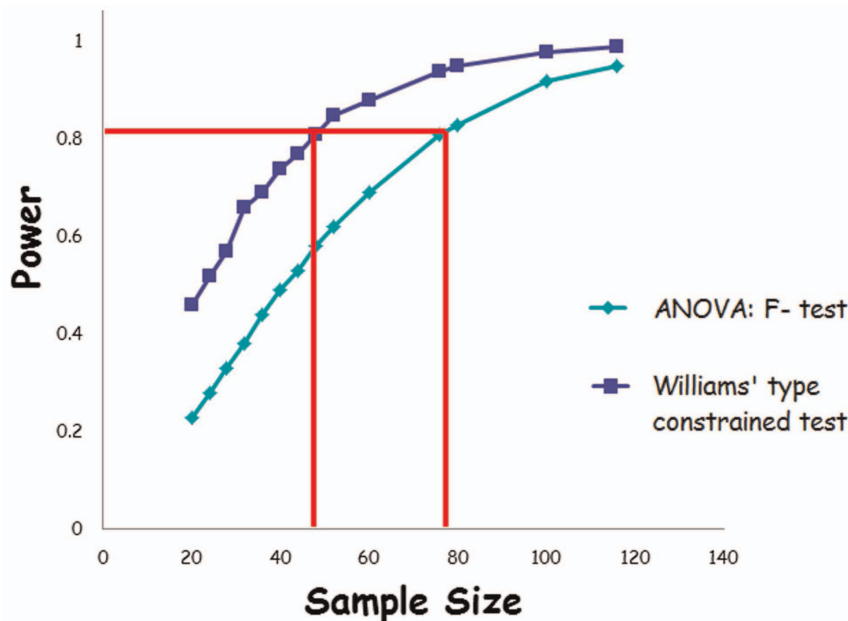


Figure 1. Results of a simulation study to compare power and sample sizes of F-test in One-way ANOVA with the constrained inference Williams’ type test where the critical values are derived using 10,000 bootstrap samples. The power of the Williams test was estimated by averaging 1000 simulated where the critical values are estimated using 10,000 bootstrap samples. The power for F-test was determined using PROC POWER in SAS (9.0). The null hypothesis was that the means of the four dose groups were equal (and zero) and the alternative hypothesis was that the means of the four dose groups have an increasing trend with dose. Data representing the four dose groups were simulated from normal populations with dose means taken to be 0, 0.1, 0.5 and 1, respectively. The actual values of the doses are irrelevant for the two methods described here. The population standard deviation for the four populations was taken to be 1. Corresponding to the 14 different patterns of total sample sizes, namely, 20, 24, 28, 32, 36, 40, 44, 48, 52, 60, 76, 80, 100, 116, the powers of the two methods are plotted. The Type I error was set to 0.05. doi:10.1371/journal.pone.0084778.g001

In many applications it is common for researchers to be interested in comparing the population means of two or more experimental conditions or groups after adjusting for covariates. Depending upon the study design, as in repeated measurement designs, it is common to use linear mixed effects models to account for the underlying dependence structure as well as the covariates. There exists several decades of literature on statistical inference in linear mixed effects models and numerous books have been written on the subject [14]. The standard statistical test for comparing several population means in a linear mixed effects model framework is the classical *F*-test which is widely used and is available in most standard software packages such as SAS, R and others. However, as noted earlier, by rejecting the null hypothesis using the standard *F*-test one can only infer that at least one population mean is different from the others and hence it may not be ideal for ordered alternatives.

In the absence of any covariates, especially continuous covariates, Mukerjee [15] noted that the usual tests for order restrictions on the means of independent normal populations can be extended to the case when normal populations are correlated as in a repeated measurements design. Later Silvapulle [16] generalized the methodology of [15] to some unbalanced designs with incomplete data. He noted that within-subject correlations make it difficult to generalize some tests into repeated measurement models. Earlier, Singh and Wright [17] considered order-restricted inference on fixed effects in a two-factor mixed model. They presented an analogue to the usual *F*-test for homogeneity and obtained several closed-form results.

It was not until [18–20] that statistical inference under inequality constraints in linear mixed effects models was formally addressed. In particular [18] developed an asymptotic likelihood

ratio test (LRT) for linear mixed effects model under homoscedastic errors. Since the asymptotic null distribution of LRT depends upon nuisance parameters, they also provided suitable bounds for the distribution using central chi-square distributions with appropriate degrees of freedom. Although these bounds are convenient, our simulation studies suggest that they could potentially be too conservative, especially when the sample sizes are small. Also, the basic assumption made in [18] is that the random effects as well as random errors of the linear model are homoscedastic and normally distributed. In practice this may not necessarily be true. Motivated by the need for a methodology that is robust to non-normality and heteroscedasticity, in Section 2 we provide a general framework using the MINQUE ([14], [21,22]) and EBLUP type residual based bootstrap [23] to test for any arbitrary linear inequality among means in a linear mixed effects model with possibly heteroscedastic errors.

We present the methodology in Section 2. Results of our simulation study are reported in Section 3. The proposed methodology is illustrated in Section 4 using data from a clinical trial comparing succimer treatment to placebo in children exposed to mercury [24].

Statistical inference under constraints

2.1 The model and notations

Let

$$Y = \mathbf{X}_1\theta_1 + \mathbf{X}_2\theta_2 + \mathbf{U}\xi + \varepsilon \tag{1}$$

denote a linear mixed effects models where θ_1 is the vector of treatment effects of the order $p_1 \times 1$, \mathbf{X}_1 is a design matrix of the order $N \times p_1$ consisting of 0s and 1s, \mathbf{X}_2 is a known matrix of covariates of the order $N \times p_2$ with corresponding (unknown) regression parameter vector θ_2 of the order $p_2 \times 1$, and \mathbf{U} is a $N \times c$ matrix of known design constants. For convenience, we denote $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$ and $\mathbf{U} = (\mathbf{U}_1 : \mathbf{U}_2 : \dots : \mathbf{U}_q)$, where \mathbf{U}_i is of order $N \times c_i$, with $\sum_{i=1}^q c_i = c$, and $\theta = (\theta_1', \theta_2')$ of order $p \times 1$, where $p = p_1 + p_2$. The observation vector Y is of the order $N \times 1$ and the unobservable random vectors $\xi = (\xi_1' : \xi_2' : \dots : \xi_q')$ and ε are independently distributed with mean 0 and covariance matrices \mathbf{T} and $\mathbf{\Sigma}$, respectively, with $\mathbf{T} = Cov(\xi_1' : \xi_2' : \dots : \xi_q') = diag(\tau_1^2 \mathbf{I}_{c_1} : \dots : \tau_q^2 \mathbf{I}_{c_q})$. Each ξ_i , $i = 1, 2, \dots, q$, is a random vector of order $c_i \times 1$. Motivated by applications and for generality we assume a heteroscedastic error structure for $\mathbf{\Sigma}$, where $\mathbf{\Sigma} = diag[\sigma_1^2 \mathbf{I}_{n_1} : \sigma_2^2 \mathbf{I}_{n_2} : \dots : \sigma_k^2 \mathbf{I}_{n_k}]$ and $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$ are unknown variances with $\sum_{i=1}^k n_i = N$.

Let \mathbf{A} denote a $r \times p$ matrix of known constants, such that $\eta = \mathbf{A}\theta$ is an $r \times 1$ estimable linear function (i.e. $C(\mathbf{A}') \subseteq C(\mathbf{X}')$, where C denotes the column space of a matrix). The problem of interest is to test hypotheses of the form:

$$H_0 : \eta = 0 \text{ versus } H_A : \eta \geq 0, \quad (2)$$

where the inequalities are component-wise, with at least one strict inequality. For example, if one is interested in testing a simple order $H_A : \theta_{11} \geq \theta_{12} \geq \dots \geq \theta_{1p_1}$, among the components of θ_1 , then $\mathbf{A} = [\mathbf{A}_1 : \mathbf{0}]$ where $\mathbf{0}$ is the null matrix of suitable order and

$$\mathbf{A}_1 = \begin{pmatrix} 1 & -1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \\ \vdots & & & & & \\ \vdots & & & & & \\ 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix}.$$

The matrix \mathbf{A} can be suitably defined in the case of simple tree order $\theta_{11} \geq \theta_{1i}$, $i \geq 2$, or the umbrella order $\theta_{11} \geq \theta_{12} \geq \theta_{13} \dots \geq \theta_{1r} \leq \theta_{1r+1} \leq \theta_{1r+2} \dots \leq \theta_{1p_1}$, for some r , etc.

2.2 The likelihood ratio test

Suppose ξ and ε are independently and multivariate normally distributed with log-likelihood function denoted by $L(\theta, \mathbf{T}, \mathbf{\Sigma})$. Let $(\hat{\theta}, \hat{\mathbf{T}}, \hat{\mathbf{\Sigma}})$ denote the maximum likelihood estimators (MLE) of $(\theta, \mathbf{T}, \mathbf{\Sigma})$ under no constraints and let $(\tilde{\theta}, \tilde{\mathbf{T}}, \tilde{\mathbf{\Sigma}})$ denote the restricted MLE (RMLE) under the alternative hypothesis (2). Then the likelihood ratio test (LRT) statistic is given by $S_{lrt} = 2[L(\tilde{\theta}, \tilde{\mathbf{T}}, \tilde{\mathbf{\Sigma}}) - L(\hat{\theta}, \hat{\mathbf{T}}, \hat{\mathbf{\Sigma}})]$. Following the arguments in [18] asymptotically, under the null hypothesis, we obtain the following limiting distribution for LRT:

$$\lim_{n \rightarrow \infty} P(S_{lrt} > c) = P\left(\sum_{i=0}^m w_i \chi_i^2 > c\right). \quad (3)$$

The weights w_i in (3) involve unknown variance components $\tau_1^2, \tau_2^2, \dots, \tau_q^2$ and $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$. Since in practice one does not know the variance components, one may plug in the estimated values for the unknown variance components in the above limiting distribution [11]. Such plug-in estimators do not perform well unless the sample size is very large. Furthermore, as the number of treatment groups increases, the computation of the weights w_i in (3) is a challenging problem. Recognizing this challenge, simple central chi-square distribution based bounds for the limiting probability in (3) were derived in [18]. These probabilities can be used for deriving bounds for the asymptotic p-values.

2.3 The residual bootstrap based test

There are three differences between our approach and the above LRT approach. Firstly, rather than using the RMLE which, at each step of the iteration, projects the unconstrained estimator $\hat{\theta}$ onto the set of constraints in the alternative hypothesis, we use the algorithm described in [25]. The estimation algorithm in [25] is identical to the pool adjacent violators algorithm (PAVA) when the constraint is a simple order, otherwise it is a modification of PAVA. Our choice of PAVA is motivated by the fact that RMLE can potential fail under some conditions [25]. Briefly, the PAVA is implemented as follows. Consider a sequence of three numbers X_1, X_2 and X_3 , and suppose in theory these numbers are expected to be non-decreasing. However, suppose the observed values are 3, 1 and 4, respectively. The expected order is violated by the first two numbers and therefore the PAVA averages the two violators, resulting in the ordered numbers $(3+1)/2, (3+1)/2$ and 4. Depending upon the situation, one could use weighted averages rather than simple average. As noted in the flow chart provided in Figure 2, our constrained estimation algorithm is flexible as one can replace PAVA type estimator [25] by the orthogonal projection estimator along the lines of [18]. Details of our algorithm (Algorithm A.1) are provided in the Appendix S1. Secondly, as noted above, because the asymptotic distribution of the LRT is non-trivial to use in practice we use a computationally simple non-parametric bootstrap to derive the p-values. Lastly, as an alternative to LRT statistic we use a Williams' type test statistic [26–28]. Intuitively, the Williams' test for monotonic order is a generalization to the idea of standard t-test with the exception that the numerator is the difference between the estimates of the largest parameter and smallest parameter, where the estimates are obtained by using the PAVA.

Let $\hat{\theta}_{1p_1}^{PAVA} = (\hat{\theta}_{11}^{PAVA}, \hat{\theta}_{12}^{PAVA}, \dots, \hat{\theta}_{1p_1}^{PAVA})'$ denote the PAVA type constrained estimator derived according to the algorithm proposed in this paper (Figure 2). Then as in [29] we may define Williams' type test statistic [26–28] for various order restrictions. In the following we provide the test statistic for three commonly encountered order restrictions. In each case the null hypothesis is $H_0 : \theta_{11} = \theta_{12} = \dots = \theta_{1p_1}$. For the others one may appeal to the general framework provided in [26].

Simple order ($H_a : \theta_{11} \leq \theta_{12} \leq \dots \leq \theta_{1p_1}$)

$$W = \frac{\hat{\theta}_{1p_1}^{PAVA} - \hat{\theta}_{11}^{PAVA}}{\sqrt{Var(\hat{\theta}_{1p_1} - \hat{\theta}_{11})}}, \quad (4)$$

where $Var(\hat{\theta}_{1p_1} - \hat{\theta}_{11})$ is the estimated variance of the contrast $(\hat{\theta}_{1p_1} - \hat{\theta}_{11})$.

A. Flow chart for constructing test statistic

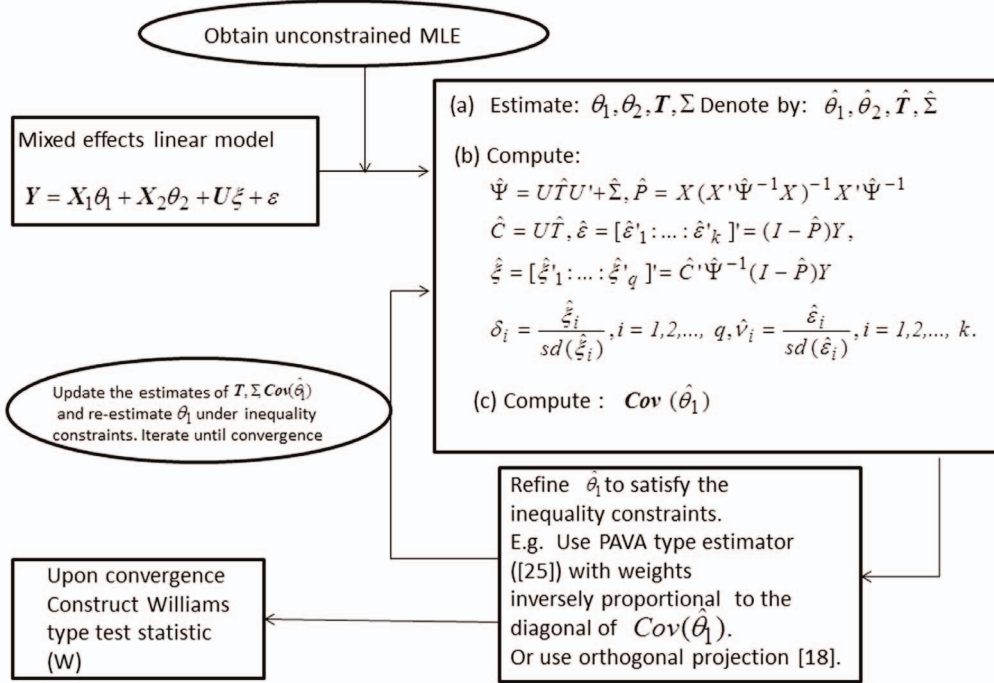


Figure 2. Flow chart for constructing test statistic.
 doi:10.1371/journal.pone.0084778.g002

B. Flow chart for deriving bootstrap data under the null hypothesis

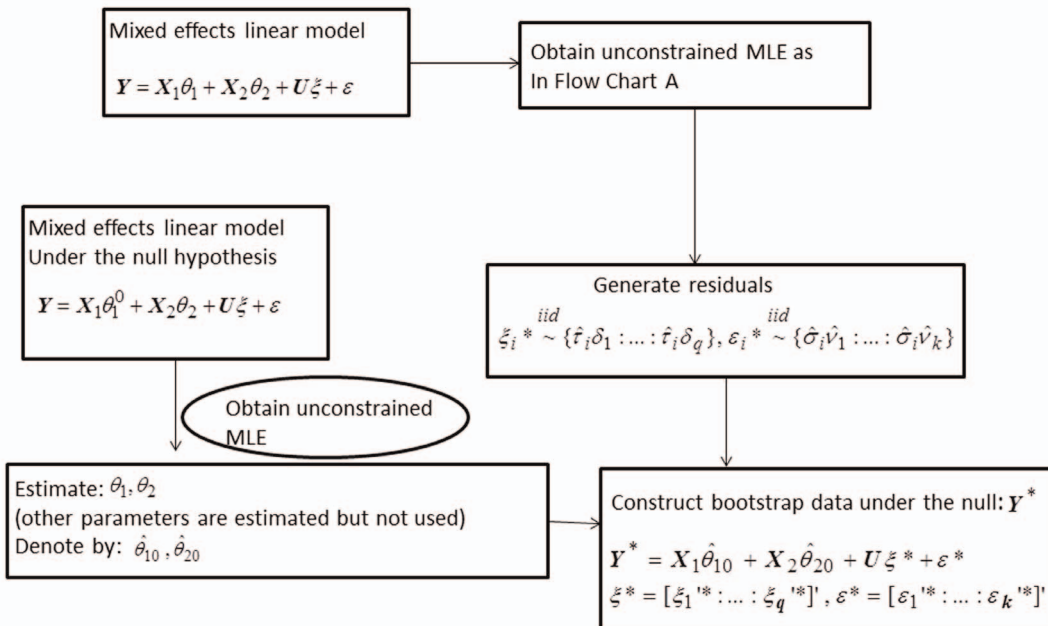


Figure 3. Flow chart for deriving Bootstrap data under the null hypothesis.
 doi:10.1371/journal.pone.0084778.g003

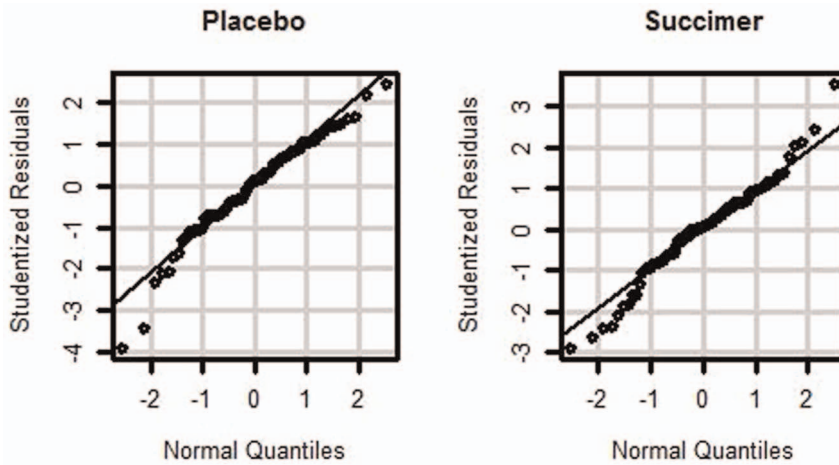


Figure 4. Normal quantile-quantile plots of studentized residuals from regressing log organic mercury in the placebo and succimer groups.

doi:10.1371/journal.pone.0084778.g004

Simple tree order ($H_a : \theta_{11} \leq \theta_{1i}, i \geq 2$).

$$W = \max_{i \geq 2} \frac{\hat{\theta}_{1i}^{PAVA} - \hat{\theta}_{11}^{PAVA}}{\sqrt{Var(\hat{\theta}_{1i} - \hat{\theta}_{11})}} \quad (5)$$

where $Var(\hat{\theta}_{1i} - \hat{\theta}_{11})$ is the estimated variance of the contrast $\hat{\theta}_{1i} - \hat{\theta}_{11}$.

Umbrella order ($H_a : \theta_{11} \leq \theta_{12} \leq \dots \leq \theta_{1r} \geq \theta_{1r+1} \dots \geq \theta_{1p_1}$).

$$W = \max \left\{ \frac{\hat{\theta}_{1r}^{PAVA} - \hat{\theta}_{11}^{PAVA}}{\sqrt{Var(\hat{\theta}_{1r} - \hat{\theta}_{11})}}, \frac{\hat{\theta}_{1r}^{PAVA} - \hat{\theta}_{1p_1}^{PAVA}}{\sqrt{Var(\hat{\theta}_{1r} - \hat{\theta}_{1p_1})}} \right\} \quad (6)$$

where $Var(\hat{\theta}_{1j} - \hat{\theta}_{1k})$ is the estimated variance of the contrast $\hat{\theta}_{1j} - \hat{\theta}_{1k}$.

The bootstrap methodology to obtain the p-values is described in the flow chart in Figure 3 and the details are provided in Algorithm A.2 in the Appendix S1. Although in this paper we are using the Williams' type statistic W , one may use a likelihood ratio

type statistic (or any other constrained test statistic of user's choice) instead of W but use residual bootstrap described in this paper for deriving the p-values. Such a strategy may result in a better power for some patterns of the mean parameters. Thus the framework developed in this paper is fairly flexible.

Simulation study

3.1 Study design

We evaluated the Type I error and power of the proposed EBLUP bootstrap test for the case of simple order using the proposed statistic (4). We compared our method with the asymptotic likelihood ratio test [18] using the upper bound in the equation (4.6) of [18] for deriving the critical values. We considered a variety of patterns of parameters and covariance matrices as follows:

Normally distributed data. The data were simulated using model (1) with the number of subjects per each treatment $n \equiv n_1 = n_2 = \dots = n_{p_1}$, $N = p_1 n$, $\mathbf{X}_1 = \mathbf{I}_{p_1} \otimes \mathbf{J}_n$, \mathbf{a} $N \times p_1$ design matrix consisting of 0's and 1's, $\mathbf{X}_2 = \mathbf{J}_{p_1} \otimes \mathbf{u}$, where u was generated as a $n \times 1$ vector, with its components uniformly distributed in $[0, 2]$, $\theta_2 = 2$, a corresponding regression parameter, $\mathbf{U} = \mathbf{J}_{p_1} \otimes \mathbf{I}_n$, ξ , a $n \times 1$ vector of independent subject random effects, and \mathbf{J}_k , a $k \times 1$ vector of 1s. The random vectors ξ, ε were independently and normally distributed with means 0 and covariance matrices \mathbf{T} and Σ , where $\mathbf{T} = Cov(\xi) = \tau^2 \mathbf{I}_n$. Simulations were performed for $p_1 = 3, 5$ treatment groups, $n = 10, 50$ subjects per each treatment and five different patterns of treatment means θ_1 : (P1) $(0, 0, \dots, 0)$, (P2) $(0, \dots, 0, a)$, (P3) $(0, a, \dots, a)$, (P4) $(a, 2a, \dots, p_1 a)$, (P5) $(0, a, \dots, a, 2a)$.

Here different values of a in the interval $[0, 2]$ were chosen arbitrarily to get a sense of power for a variety of patterns. Note that pattern (P1) corresponds to the null pattern for computing Type I error. Patterns (P2) to (P5) are patterns of θ_1 where the components satisfy a simple order constraint $\theta_1 \leq \theta_2 \leq \dots \leq \theta_p$ for evaluating the power of the two test procedures. Two different structures of Σ were considered, namely, homoscedastic error structure with $\Sigma = \sigma^2 \mathbf{I}_N$ and heteroscedastic error structure with $\Sigma = diag[\sigma_1^2 \mathbf{I}_n : \sigma_2^2 \mathbf{I}_n : \dots : \sigma_{p_1}^2 \mathbf{I}_n]$, where p is the number of treatments, and $\sigma_1^2, \sigma_2^2, \dots, \sigma_{p_1}^2$ are unknown variances with

Table 1. Type I errors for homoscedastic normally distributed data.

p_1	n	τ^2	Asymp-LRT	Proposed method
3	10	1	0.03	0.05
3	50	1	0.01	0.03
3	10	0.2	0.05	0.04
3	10	2	0.04	0.05
3	50	2	0.01	0.03
5	10	1	0.02	0.03
5	10	0.2	0.02	0.04
5	10	2	0.02	0.04

doi:10.1371/journal.pone.0084778.t001

Table 2. Power for homoscedastic normally distributed data.

p_1	n	θ_1					τ^2	Asymp-LRT	Proposed method
3	10	0	0.00	1.25			1	0.82	0.84
3	10	0	1.26	1.26			1	0.82	0.84
3	50	0	0.55	0.55			1	0.85	0.89
3	10	0	0.73	1.45			1	0.86	0.89
5	10	0	0.00	0.00	0.00	1.27	1	0.65	0.86
5	10	0	1.24	1.24	1.24	1.24	1	0.58	0.86
5	10	0	0.37	0.74	1.11	1.48	1	0.80	0.90
5	10	0	0.81	0.81	0.81	1.62	1	0.74	0.93

doi:10.1371/journal.pone.0084778.t002

$N = p_1 n$. The patterns of T and Σ considered in our simulations are as follows:

- *Homoscedastic case:* We fixed $\sigma = 1$ and considered three patterns of τ^2 , namely $\tau^2 = 0.2, 1, 2$.
- *Heteroscedastic case:* We fixed $\tau^2 = 1$ and, as commonly done, we chose the error variance to be a function of the mean, namely $\sigma^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_{p_1}^2)$, where $\sigma_i^2 = \theta_i^2$. In the case of null hypothesis, i.e., $\theta_i = 0 \forall i$, we chose $\sigma_i^2 = 0.1 \forall i$.

Log-normally distributed data. All results are based on 500 simulation runs. We used 500 bootstrap runs to generate the null distribution for the proposed bootstrap test. In all simulations the nominal value for Type I error was taken to be 0.05.

3.2 Results

Complete simulation scenarios and results are presented in Tables 1–6. Both tests operate at the desired nominal Type I error rate, although sometimes the asymptotic likelihood ratio test tends to be slightly conservative. Generally the proposed test is substantially more powerful than the asymptotic likelihood ratio test. In some cases the gains in power are as much as 0.36, e.g. 0.78 vs. 0.42 for heteroscedastic normally distributed data with $p_1 = 5, n = 10, \theta_1 = (0, 0, 0, 0, 0.4)$.

Table 3. Type I errors for heteroscedastic normally distributed data.

p_1	n	σ^2			τ^2	Asymp-LRT	Proposed method		
3	10	0.1	0.10	2.37	1	0.04	0.03		
3	10	0.1	0.20	0.20	1	0.03	0.04		
3	10	0.1	0.09	0.36	1	0.03	0.03		
3	50	0.1	0.10	0.01	1	0.01	0.03		
3	50	0.1	0.02	0.02	1	0.02	0.04		
5	10	0.1	0.10	0.10	0.10	1	0.01	0.04	
5	10	0.1	0.20	0.20	0.20	1	0.01	0.04	
5	10	0.1	0.11	0.44	0.99	1	0.01	0.04	
5	10	0.1	0.11	0.11	0.11	0.45	1	0.02	0.05

doi:10.1371/journal.pone.0084778.t003

Table 4. Power for heteroscedastic normally distributed data.

p_1	n	θ_1					τ^2	Asymp-LRT	Proposed method
3	10	0	0.00	1.54			1	0.82	0.88
3	10	0	0.45	0.45			1	0.81	0.82
3	10	0	0.30	0.60			1	0.82	0.80
3	50	0	0.00	0.10			1	0.70	0.74
3	50	0	0.15	0.15			1	0.86	0.92
3	50	0	0.08	0.16			1	0.95	0.93
5	10	0	0.00	0.00	0.00	0.40	1	0.42	0.78
5	10	0	0.45	0.45	0.45	0.45	1	0.68	0.81
5	10	0	0.33	0.66	1.00	1.33	1	0.96	0.88
5	10	0	0.34	0.34	0.34	0.67	1	0.71	0.82

doi:10.1371/journal.pone.0084778.t004

Illustration

We illustrate the proposed methodology using data from patients in a randomized placebo-controlled, double-blind clinical trial the Treatment of Lead-exposed Children trial, or TLC [30]. In TLC, 384 children aged 12–33 months were assigned to the placebo group and 396 to the succimer group. In the succimer group, up to three 26-day courses were administered. Cao et al.

Table 5. Type I errors for log-normally distributed data.

p_1	n	σ^2			τ^2	Asymp-LRT	Proposed method		
3	10	0.10	0.10	0.04	1	0.02	0.03		
3	10	0.10	0.20	0.20	1	0.03	0.05		
3	10	0.10	0.01	0.04	1	0.03	0.03		
3	50	0.10	0.02	0.02	1	0.01	0.01		
3	50	0.10	0.01	0.03	1	0.01	0.01		
5	10	0.10	0.10	0.10	0.10	1	0.02	0.04	
5	10	0.10	0.04	0.04	0.04	1	0.02	0.03	
5	10	0.10	0.01	0.04	0.09	1	0.02	0.05	
5	10	0.10	0.01	0.01	0.01	0.04	1	0.03	0.03

doi:10.1371/journal.pone.0084778.t005

Table 6. Power for log-normally distributed data.

p_1	n	θ_1			τ^2	Asymp-LRT	Proposed method		
3	10	0	0.00	1.54	1	0.90	0.98		
3	10	0	0.45	0.45	1	0.81	0.85		
3	10	0	0.30	0.60	1	0.85	0.90		
3	50	0	0.00	0.20	1	0.92	0.88		
3	50	0	0.15	0.15	1	0.57	0.68		
3	50	0	0.08	0.16	1	0.89	0.69		
5	10	0	0.00	0.00	0.00	0.40	1	0.44	0.77
5	10	0	0.45	0.45	0.45	0.45	1	0.66	0.83
5	10	0	0.10	0.20	0.30	0.40	1	0.79	0.75
5	10	0	0.34	0.34	0.34	0.67	1	0.72	0.94
5	50	0	0.00	0.00	0.00	0.20	1	0.97	0.96
5	50	0	0.20	0.20	0.20	0.20	1	0.66	0.92
5	50	0	0.08	0.08	0.08	0.16	1	0.87	0.74

doi:10.1371/journal.pone.0084778.t006

[24] measured and analyzed mercury levels in pre-treatment to investigate whether succimer, a mercaptan compound known to reduce blood lead concentration in children, also reduces blood mercury concentration. At the baseline, blood mercury levels were quantified in 657 samples (338 succimer and 319 placebo). At 1-week post treatment, total mercury concentration was quantified in 623 samples (313 succimer and 310 placebo). After 5 months of treatment, blood mercury levels were quantified in 61 samples: 30 succimer treated children and 31 placebo treated children. To investigate the efficacy of succimer treatment relative to placebo, Cao et al. [24] tested for an increasing trend over time in the difference between the mean mercury concentrations in the succimer group and the placebo group, after adjusting for child's age, sex, race and the study center. The authors concluded that there was no significant trend in the mean difference in mercury concentrations in the succimer group and the placebo group. Since their analysis ignored the covariance structure induced by repeated measurements, we reanalyzed the data using the proposed methodology which accounts for the repeated measurements in the data. We fitted a linear mixed effects model with log organic mercury as a dependent variable, treatment, race, gender, center and age as fixed effects, and subject as a random effect. The normal quantile-quantile plots (Figure 4) of resulting studentized residuals in the placebo and succimer groups suggest that the data are potentially non-normally distributed.

Denoting the difference in the mean log mercury levels at the t th time between placebo and succimer groups as by θ_t ($t=1$ for baseline, $t=2$ for 1-week and $t=3$ for 5-months) we tested the following hypothesis:

References

- Dinse GE (1985) Testing for a trend in tumor prevalence rates: I. Nonlethal tumors. *Biometrics*, 41, 751–770.
- Peddada SD, Dinse G, Kissling G (2007) Incorporating Historical Control Data When Comparing Tumor Incidence Rates. *J. Amer. Stat. Assoc.*, 102: 1212–1220.
- Lelandais G, Saint-Georges Y, Geneix C, Al-Shikhley L, Dujardin G, et al. (2009) Spatio-Temporal Dynamics of Yeast Mitochondrial Biogenesis: Transcriptional and Post-Transcriptional mRNA Oscillatory Modules. *PLoS Comput Biol* 5(6): e1000409. doi:10.1371/journal.pcbi.1000409.

$$H_0 : \theta_1 = \theta_2 = \theta_3 \text{ versus}$$

$$H_a : \theta_1 \leq \theta_2 \leq \theta_3,$$

with at least one strict inequality. The three UMLEs were 0.59, 0.92, 0.73 and the MINQUE-based constrained estimates were 0.59, 0.82, 0.82. Using the proposed test we obtained a bootstrap p-value of 0.109. Thus, we are not able to reject the null hypothesis in favor of a trend in the difference in means. These results are consistent with conclusions of [24], that succimer chelation for low level organic mercury exposure in children has limited efficacy.

Summary and Concluding Remarks

Inequality constraints arise naturally in many applications, such as toxicology, where researchers are interested in studying dose-response of a chemical, or gene expression studies in oncology, where a researcher may be interested in understanding the changes in gene expression according to cancer stage. We proposed a new method to test for inequality constraints. Since the method uses Rao's MINQUE theory (cf. [22]) for estimating variance components and PAVA for estimating the means, it does not necessarily require normality. In the simple order restriction, our extensive simulation studies suggest that the proposed methodology provides a better control of type I error than the asymptotic likelihood ratio test of [18] when the data are non-normally distributed. Our proposed methodology seems to control the Type I error at the desired nominal level.

In the manuscript we considered one-dimensional order constraints. In toxicology, researchers are often interested in dose \times time response surfaces, which results in a two-way classification that can be expressed as a constrained inference problem with constraints on rows and columns of a matrix. The proposed testing procedure can be easily extended to deal with such cases.

Supporting Information

Appendix S1 Description of Algorithms A.1 and A.2. (DOCX)

Acknowledgments

Authors thank Grace Kissling and Bahjat Qaqish for their careful reading of the manuscript and for their comments that helped improve the presentation of the paper. Authors also thank Drs. Rogan and Cao for providing us the data that were analyzed in this paper.

Author Contributions

Conceived and designed the experiments: SDP. Performed the experiments: LF. Analyzed the data: LF. Wrote the paper: SDP LF AI.

8. Fernandez M, Rueda C, Peddada SD (2012) Identification of a core set of signature cell-cycle genes whose relative order of time to peak expression is conserved across species. *Nucleic Acids Research*, 40: 2823–32. Epub 2011/12/03. doi: 10.1093/nar/gkr1077.
9. Kondratova AA, Kondratov RV (2012) The circadian clock and pathology of the ageing brain. *Nature Reviews Neuroscience*, doi:10.1038/nrn3208
10. Hastings MH, Reddy AB, Maywood ES (2003) A clockwork web: circadian timing in brain and periphery, in health and disease. *Nature Reviews Neuroscience* 4: 649–661.
11. Silvapulle MJ, Sen PK (2005) *Constrained statistical inference: Inequality, order, and shape restrictions*. Hoboken, N.J.: Wiley-Interscience.
12. Peddada SD, Lobenhofer L, Li L, Afshari C, Weinberg C, et al. (2003) Gene selection and clustering for time-course and dose-response microarray experiments using order-restricted inference. *Bioinformatics*, 19: 834–841.
13. Peddada SD, Harris S, Zajd J, Harvey E (2005) ORIOGEN: Order Restricted Inference for Ordered Gene Expression data. *Bioinformatics*, 21: 3933–3934.
14. Khuri AI, Mathew T, Sinha BK (1998) *Statistical tests for mixed linear models*. New York: Wiley.
15. Mukerjee H (1988) Order restricted inference in a repeated measures model. *Biometrika* 75: 616–617.
16. Silvapulle MJ (1997) On order restricted inference in some mixed linear models. *Statistics & Probability Letters* 36: 23–27.
17. Singh B, Wright FT (1990) Testing for and against an order restriction in mixed-effects models. *Statistics & Probability Letters* 9: 195–200.
18. Davidov O, Rosen S (2011) Constrained inference in mixed-effects models for longitudinal data with application to hearing loss. *Biostatistics* 12: 327–340.
19. Rosen S, Davidov O (2012) Order-restricted inference for multivariate longitudinal data with applications to the natural history of hearing loss. *Statistics in Medicine* 31: 1761–1773.
20. Davis KA, Park CG, Sinha SK (2012) Testing for generalized linear mixed models with cluster correlated data under linear inequality constraints. *Canadian Journal of Statistics* 40: 243–258.
21. Rao CR (1972) Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association* 67: 112–115.
22. Rao CR, Kleffe J (1988) *Estimation of variance components and applications*. Amsterdam; New York; New York, N.Y., U.S.A.: North-Holland.
23. Searle SR, Casella G, McCulloch CE (1992) *Variance components*. New York: Wiley.
24. Cao Y, Chen A, Jones RL, Radcliffe J, Diethrich KN, et al. (2011) Efficacy of succimer chelation of mercury at background exposures in toddlers: A randomized trial. *The Journal of Pediatrics* 158: 480–485.
25. Hwang JTG, Peddada SD (1994) Confidence Interval Estimation Subject to Order Restrictions. *Annals of Statistics*, 22: 67–93.
26. Williams DA (1971) A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics* 27: 103–117.
27. Williams DA (1972) The comparison of several dose levels with a zero dose control. *Biometrics* 28: 519–531.
28. Williams DA (1977) Some inference procedures for monotonically ordered normal means. *Biometrika* 64: 9–14.
29. Peddada SD, Prescott KE, Conaway M (2001) Tests for order restrictions in binary data. *Biometrics* 57: 1219–1227.
30. Rogan WJ (1998) The treatment of lead-exposed children (TLC) trial: Design and recruitment for a study of the effect of oral chelation on growth and development in toddlers. *Paediatric and Perinatal Epidemiology* 12: 313–333.