



Modelling Repeated Measures with Multiple Endpoints: A New Approach

Sandipan Samanta¹, Ranjit Kumar Paul¹ and Arjun Roy²

¹*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*

²*AB Science, Paris-75008, France*

Received 19 November 2015; Revised 24 January 2017; Accepted 03 February 2017

SUMMARY

In this article, the concept of O'Brien's GLS test for repeated measure design has been introduced in heteroscedastic case. The paper is devoted to modify O'Brien's test statistic and propose a methodology to analyze treatment effect using Kronecker product structure where multiple responses are recorded repeatedly over time. Additionally, unbalanced data can also be easily modeled with the proposed methodology. The proposed methodology is illustrated using orthopedic dataset. The size and power of the new test statistic is computed using bootstrap methodology for different parametric configurations. A comparison with doubly repeated measure design is also performed with existing methodology. It is concluded that the proposed methodology is better than Minimum Variance Quadratic Unbiased Estimation (MIVQUE) method and Restricted Maximum Likelihood (REML) procedure.

Keywords: Bootstrap, Doubly repeated measure design, Heteroscedasticity, Kronecker product, O'Brien's GLS test, Multiple responses.

1. INTRODUCTION

Most of the trials in biomedical, pharmaceutical, agricultural and animal sciences are conducted to compare the test group with a control group to evaluate the relative efficacy of the treatment over the control for longer periods of time. Very often, lots of information is to be collected on multiple responses and the treatment is expected to have a positive effect on all responses. In the interest of controlling the overall type I error rate at α level, an adjustment for multiple comparison is necessary. Where a multiple adjustment is required; various methods viz. Bonferroni or Hochberg & Tamhane are available in literature. However, these methods do not consider the correlation among endpoints. Multivariate techniques would be more appropriate to test the treatment effect when it has any effect on multiple responses. Hotelling's T^2 test can be applied to solve this problem but it has less power because it does not consider the direction of a response which is known in advance. O'Brien (1982) proposed more powerful ordinary least squares

(OLS) and generalized least squares (GLS) tests to demonstrate an overall positive treatment effect on all multiple responses. Pocock *et al.* (1987) proposed an ad-hoc extension of O'Brien's GLS test statistics to heteroscedastic case assuming that covariance matrices are known in advance for treatment groups. Logan and Tamhane (2004) derived the extension of O'Brien's OLS and GLS statistics in heteroscedastic condition and investigated a better approximation to the degrees of freedom of the t -distribution. Dallow *et al.* (2008) set up the concept of an operational effect size to evaluate power using different correlation matrices. They also extended O'Brien statistics to handle covariates as well as missing data.

In common practice, multiple measurements of a single response is repeatedly collected over time on each subject or unit. This type of design is commonly known as univariate repeated measure design. In doubly repeated measure data, multiple measurements are collected over time on more than one response variable on each subject or unit (Khattree and Naik

1999). Multivariate longitudinal data provides a unique opportunity in studying the joint evolution of multiple response variables over time. The analysis of multivariate longitudinal data can be challenging as compare to traditional univariate longitudinal data because not only error variances are likely to be different for different response variables but also errors are likely to be correlated for the same response variable measured at different occasions, and errors are also likely to be correlated among response variable measured at the same time. Galecki (1994) introduced the Kronecker product structure for modelling covariance structures for repeated measures specified by more than one repeated factor. Naik and Rao (2001) discussed the method to analyze unbalanced data with Kronecker product structure under double repeated measure design. Using Satterthwaite type approximation they developed the multivariate tests for testing the hypothesis. Rao and Khattree (2005) developed an alternative algorithm for likelihood ratio test statistics to consider the convergence problem of testing the Kronecker product structured. Aloh and Fritsch (2002) proposed an extension of O'Brien method for longitudinal setup by reducing the dimension of data. They reduced the k -dimensional data on each subject to a single value and finally applied MANOVA for repeated measure design.

The objective of this paper is to extend O'Brien methodology in doubly repeated measure design under heteroscedastic set up i.e. treatment and reference has different covariance structure. In Section 2, the O'Brien's OLS and GLS tests are explained. Doubly repeated measure design also describe in this section in brief. The estimation procedure of repeated measures with multiple endpoints using Kronecker product structure is discussed in Section 3. Simulation results are presented in Section 4. In Section 5, application of the methodology is illustrated using an orthopedic data set. Finally, the concluding remarks are represented in Section 6 on the merits of the proposed methodology and delineation of some problems for future research work.

2. O'BRIEN'S TEST STATISTICS

O'Brien (1984) used the sample correlation matrix of responses to compute OLS as well as GLS statistics. Let x_{ijk} denotes the measurement on the k^{th} endpoint for the j^{th} subject in the i^{th} treatment group and

$\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijm})'$ where $k = 1, \dots, m; j = 1, \dots, n_i; i = 1, 2$ are assumed to be independent and identically distributed random vectors from a multivariate normal distribution with mean vector μ_i and covariance matrix Σ_i . In case of homoscedastic covariance structure, let $\Sigma_1 = \Sigma_2 = \Sigma$ (say) and the corresponding correlation matrix be denoted by $R_j, j=1,2$. The OLS test statistic proposed by O'Brien's is

$$t_{OLS} = \sqrt{\frac{n_1 n_2}{(n_1 + n_2)}} \left[\mathbf{j}' (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) / \sqrt{\mathbf{j}' \hat{\mathbf{R}} \mathbf{j}} \right] \quad (2.1)$$

Where \mathbf{j} is a vector of 1's of an appropriate dimension, $\bar{\mathbf{x}}_i = (\bar{x}_{i,1}, \bar{x}_{i,2}, \dots, \bar{x}_{i,m})'$ and $\bar{x}_{i,k} = \sum_{j=1}^{n_i} x_{ijk} / n_i$.

The O'Brien's GLS test statistic is

$$t_{GLS} = \sqrt{\frac{n_1 n_2}{(n_1 + n_2)}} \left[\mathbf{j}' \hat{\mathbf{R}}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) / \sqrt{\mathbf{j}' \hat{\mathbf{R}}^{-1} \mathbf{j}} \right] \quad (2.2)$$

The test statistics (2.1) and (2.2) are standardized weighted sums of individual t -statistics for the m endpoints. The OLS statistic uses equal weights, while GLS statistic uses unequal weights determined by the sample correlation matrix $\hat{\mathbf{R}}$. In both the test statistics, it is necessary to standardize the data in common units via a transformation to relative deviates. The convergence of t_{GLS} to the standard normal distribution is slower than the t_{OLS} because of the use of estimated sample correlation matrix $\hat{\mathbf{R}}$. O'Brien (1984) proposed to approximate the OLS and GLS test statistics by t -distribution with $n_1 + n_2 - 2md.f$. However, for large sample sizes, the standard normal distribution may be used as an approximation.

Let y_{ijkt} denotes the measurement on the t^{th} visit of k^{th} response for the j^{th} subject in the i^{th} treatment group. The data for different subjects are assumed to be independent but different responses for the same subject are assumed to have multivariate normal distribution. Let $\mathbf{y}_{ij} = (y'_{ij1}, y'_{ij2}, \dots, y'_{ijm})'$ and then \mathbf{y}_{ij} is $mt \times 1$ random observational vector corresponding to the j^{th} observation in the i^{th} group for $j = 1, \dots, n_i; i = 1, 2, \dots, g$. Let $cov(\mathbf{y}_{ij}) = \Psi_i \otimes \Sigma_i$ (say Ω_i), where both $\Psi_i = (\psi_{ij}) : t \times t$ and $\Sigma_i = (\sigma_{ij}) : m \times m$ are unknown but it is assumed that Ω_i is a positive definite matrix where \otimes denotes Kronecker product. Using a general linear model of the form $\mathbf{Y} = \mathbf{D}\theta + \mathbf{E}$, any

linear hypothesis about the effect of treatment factor, time factor, or the interaction between them i.e. the elements of \mathbf{D} can be formulated in the form of general linear hypothesis $\mathbf{H}_0: \mathbf{FDG} = \mathbf{0}$, with \mathbf{F} and \mathbf{G} known and full rank matrices. Using Wilks' Λ or any other standard multivariate tests, the null hypothesis \mathbf{H}_0 can be tested. In the above linear model, \mathbf{Y} represents the observation vector, \mathbf{D} is a known design matrix, $\boldsymbol{\theta}$ is collection of unknown parameters and assuming \mathbf{E} independently follow multivariate normal distribution with zero mean and covariance matrix $\bigoplus_{i=1}^2 \ddot{\mathbf{\Omega}}_i$ (say $\ddot{\mathbf{A}}$); where \bigoplus denotes direct sum operator.

3. METHODOLOGY

Without losing the generality, we assume that two treatment groups are to be compared with repeated measurement design for multiple responses. Additionally, we assume n_1 individuals receive one treatment and n_2 individuals receive the other and on each of whom m responses are measured for T visits. As per the common practice, it is presumed that the realization of repeated measurement for both treatment groups is same as well as both groups are homogenous relating to all essential prognostic factors for a response variable of interest.

O'Brien (1984) considered a simplified version of testing the treatment effects using a univariate regression framework after re-parameterizing in the design matrix (\mathbf{D}) to obtain a full-rank model. For our proposed methodology, O'Brien's design matrix structure is followed to estimate the effects; however, it is not full-rank matrix hence, Moore-Penrose inverse is used instead of ordinary inverse. The standardization is performed by subtracting the pooled mean from each observation and dividing by the pooled standard deviation within responses and visits group. We formulate the design matrix for first group as

$$\mathbf{D}_1 = (\mathbf{j}_{mT} \quad \mathbf{I}_m \otimes \mathbf{j}_T \mathbf{j}_m' \otimes \mathbf{I}_T \quad \mathbf{j}_m \otimes \mathbf{I}_T),$$

and, for the second group, it is

$$\mathbf{D}_2 = (-\mathbf{j}_{mT} \quad \mathbf{I}_m \otimes \mathbf{j}_T \mathbf{j}_m' \otimes \mathbf{I}_T \quad -\mathbf{j}_m \otimes \mathbf{I}_T)$$

Finally, the design matrix (\mathbf{D}) is

$$\mathbf{D} = \begin{pmatrix} \mathbf{j}_{n_1} \otimes \mathbf{D}_1 \\ \mathbf{j}_{n_2} \otimes \mathbf{D}_2 \end{pmatrix}$$

Where \mathbf{j} and \mathbf{I} are the vectors of 1's and identity

matrix with appropriate dimension respectively. The GLS estimate of $\boldsymbol{\theta}$ is given by $(\mathbf{D}'\ddot{\mathbf{A}}^{-1}\mathbf{D})^{-1}\mathbf{D}'\ddot{\mathbf{A}}^{-1}\mathbf{Y}$, where $\ddot{\mathbf{A}}$ is the covariance matrix of \mathbf{E} which has a block diagonal structure given by

$$\ddot{\mathbf{A}} = \begin{pmatrix} \ddot{\mathbf{\Omega}}_1 & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \ddot{\mathbf{\Omega}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \ddot{\mathbf{\Omega}}_2 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \ddot{\mathbf{\Omega}}_2 \end{pmatrix}$$

Where $\ddot{\mathbf{\Omega}}_i = \ddot{\mathbf{\Psi}}_i \otimes \ddot{\mathbf{\Sigma}}_i$. Let the correlation structure of $\ddot{\mathbf{A}}$ be denoted as \mathbf{A} .

As we re-parameterize the design matrix, only the first row of $(\mathbf{D}'\mathbf{A}^{-1}\mathbf{D})^{-1}$ and $\mathbf{D}'\mathbf{A}^{-1}\mathbf{Y}$ are required to compute the difference of treatment effect. The last part of $\tilde{\boldsymbol{\theta}}$ is

$$\begin{aligned} \mathbf{D}'\mathbf{A}^{-1}\mathbf{Y} &= [\mathbf{j}'_{n_1} \otimes (\mathbf{D}'_1\mathbf{\Omega}_1^{-1}) \quad \mathbf{j}'_{n_2} \otimes (\mathbf{D}'_2\mathbf{\Omega}_2^{-1})] \mathbf{Y} \\ &= n_1 \mathbf{D}'_1\mathbf{\Omega}_1^{-1}\bar{\mathbf{y}}_1 + n_2 \mathbf{D}'_2\mathbf{\Omega}_2^{-1}\bar{\mathbf{y}}_2 \end{aligned}$$

Finally, the above expression is

$$l = \begin{pmatrix} n_1 \mathbf{j}'_{mT}\mathbf{\Omega}_1^{-1}\bar{\mathbf{y}}_1 - n_2 \mathbf{j}'_{mT}\mathbf{\Omega}_2^{-1}\bar{\mathbf{y}}_2 \\ n_1 (\mathbf{I}_m \otimes \mathbf{j}'_T)\mathbf{\Omega}_1^{-1}\bar{\mathbf{y}}_1 + n_2 (\mathbf{I}_m \otimes \mathbf{j}'_T)\mathbf{\Omega}_2^{-1}\bar{\mathbf{y}}_2 \\ n_1 (\mathbf{j}'_m \otimes \mathbf{I}_T)\mathbf{\Omega}_1^{-1}\bar{\mathbf{y}}_1 + n_2 (\mathbf{j}'_m \otimes \mathbf{I}_T)\mathbf{\Omega}_2^{-1}\bar{\mathbf{y}}_2 \\ n_1 (\mathbf{j}'_m \otimes \mathbf{I}_T)\mathbf{\Omega}_1^{-1}\bar{\mathbf{y}}_1 - n_2 (\mathbf{j}'_m \otimes \mathbf{I}_T)\mathbf{\Omega}_2^{-1}\bar{\mathbf{y}}_2 \end{pmatrix} \tag{3.1}$$

Now,

$$\begin{aligned} \mathbf{D}'\mathbf{A}^{-1}\mathbf{D} &= [(\mathbf{j}_{n_1} \otimes \mathbf{D}_1)' \quad (\mathbf{j}_{n_2} \otimes \mathbf{D}_2)'] \\ &\begin{pmatrix} \mathbf{\Omega}_1^{-1} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{\Omega}_1^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{\Omega}_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Omega}_2^{-1} \end{pmatrix} \begin{bmatrix} (\mathbf{j}_{n_1} \otimes \mathbf{D}_1) \\ (\mathbf{j}_{n_2} \otimes \mathbf{D}_2) \end{bmatrix} \end{aligned}$$

Finally, it reduces to

$$n_1 \mathbf{D}'_1\mathbf{\Omega}_1^{-1}\mathbf{D}_1 + n_2 \mathbf{D}'_2\mathbf{\Omega}_2^{-1}\mathbf{D}_2 \tag{3.2}$$

where $\mathbf{\Omega}_i^{-1} = \mathbf{\Psi}_i^{-1} \otimes \mathbf{\Sigma}_i^{-1}$.

But the expression (3.2) does not have full column rank. Using QR decomposition, it can be shown that the rank of (3.2) will be $m + 2T - 1$. Let us denote that Moore Penrose Inverse of $D'A^{-1}D$ is C^+ and the first row of C^+ is C_1^+ .

Finally, the GLS estimate of the difference of treatment effect (τ) is C_1^+l and its standard deviation (SD (τ)) equals to $[C_1^+ D'A^{-1}D C_1^+]^{1/2}$.

Hence, the GLS test statistic is

$$t_{sr} = \frac{\tau}{SD(\tau)} = \frac{C_1^+l}{[C_1^+ D'A^{-1}D C_1^+]^{1/2}} \tag{3.3}$$

It should be noted that the above statistic is equivalent to (2.2) in homoscedastic setup for a single visit. We have shown these facts with one real data set in Section 5. O'Brien (1984) evaluated the performance of Rank sum, OLS, GLS and modified GLS tests in repeated measure setting and concluded that GLS procedure was outperformed over Rank sum, OLS and modified GLS tests. Readers are requested to go through the Technical Report No. 25; Mayo Clinic for full discussion of the above findings. The above methodology can be used to estimate the treatment effect in repeated measure setting for single endpoints as described in Technical Report 25. However, to test the other effects L can be chosen in such a way that

$$F_{sr} = \frac{(L'\hat{\theta} - m)[L'(D'A^{-1}D)^{-1}L]^{-1}(L'\hat{\theta} - m)}{p(Y-D\hat{\theta})A^{-1}(Y-D\hat{\theta})} \tag{3.4}$$

where p is the ratio between rank of $[L'(D'A^{-1}D)^{-1}L]$ and error $d.f.$

4. SIMULATION RESULTS

In order to compute the size of the test statistic 100,000 experiments has been conducted for different parametric configuration with small to moderate sample size. In order to calculate size of the test, n_i *i.i.d.* data vectors for the treatment group y_{ij} are generated under Null Hypothesis, the active and control treatments follow the same distribution i.e. $H_0: \zeta_1 = \zeta_2 = \zeta$ from an MVN ($\zeta, \hat{\Omega}_i$) distribution where $\hat{\Omega}_{ii} = \hat{\omega}_i, \hat{\Psi}_{ij} = \hat{\psi}_i$ and $\hat{\Sigma}_{ij} = \hat{\sigma}_i$ for $i \neq j$ and t_{sr} has been computed. In each iteration, 100,000 datasets are generated by bootstrap method (Efron and Tibshirani 1995) and corresponding $t_{sr}^{(b)}$, $b = 1, \dots, 100,000$ are calculated. Now from the bootstrap distribution we compute bootstrap p -value as $\hat{p}^*(t_{sr}) = \sum_{b=1}^{100,000} I(t_{sr}^{(b)} > t_{sr})/100,000$ where

$I(\cdot)$ is an indicator variable and rejection of H_0 is performed based on the basis whenever $\hat{p}^*(t_{sr}) < \alpha$, (say, 0.05 or 0.01). The above procedure is repeated 100,000 times and the proportion of rejection of H_0 would give the size of the test (Huque and Sankoh 1997). In Table 1 the simulated type I error of the test statistics given in (3.3) has been represented.

A comparison with doubly repeated measure results using SAS Proc Mixed procedure has also been made. As in most of the simulation data, the likelihood does not converge, therefore, Minimum Variance Quadratic Unbiased Estimation (MIVQUE) method has been used for comparison. The value of $\hat{p}^*(t_{sr})$ has been compared with the average of 100,000 p -values obtained by MIVQUE method for each iteration. Finally, we compute the proportions to see the number of cases in which the average MIVQUE p -values are less than $\hat{p}^*(t_{sr})$. It is observed that the values more than 0.5 in \hat{t}^* in Table 1 show better results of our proposed methodology.

In order to compute the power of the test, data have been simulated under H_0 and using bootstrap simulation $(1-\alpha)^{th}$ quantile value (say $c_{1-\alpha}^*$) have been computed from the bootstrap distribution for different sample sizes. In the next step, 100,000 bootstrap estimates ($t_{sr}^{(b)}$) of t_{sr} under H_1 are generated and power of the test is calculated as $\sum_{b=1}^{100,000} I(t_{sr}^{(b)} > c_{1-\alpha}^*)/100,000$ (Yuan and Hayashi 2003). However, it is observed that only difference in mean vector or adding some constant in one group is not enough to generate data under H_1 . O'Brien's procedure has been followed to compute the power of the test for different sample size and distributions using the correlation structures defined as follows.

$$\Psi_A = \begin{pmatrix} 1 & 0.052 & 0.007 \\ 0.052 & 1 & -0.717 \\ 0.007 & -0.717 & 1 \end{pmatrix},$$

$$\Sigma_A = \begin{pmatrix} 1 & 0.321 \\ 0.321 & 1 \end{pmatrix}$$

and

$$\Psi_B = \begin{pmatrix} 1 & 0.004 & 0.374 \\ 0.004 & 1 & 0.459 \\ 0.374 & 0.459 & 1 \end{pmatrix},$$

$$\Sigma_B = \begin{pmatrix} 1 & 0.413 \\ 0.413 & 1 \end{pmatrix}$$

Table 1. Observed Type I Error of t_{sr} and comparison of p -value with MIVQUE method

Distribution	n_1	n_2	m	T	$\ddot{\omega}_1 = 1; \ddot{\psi}_1 = \ddot{\sigma}_1 = 0$			$\ddot{\omega}_1 = 1; \ddot{\psi}_1 = 0.4; \ddot{\sigma}_1 = 0.4$			$\ddot{\omega}_1 = 1; \ddot{\psi}_1 = -0.2; \ddot{\sigma}_1 = 0.5$		
					$\ddot{\omega}_2 = 1; \ddot{\psi}_2 = \ddot{\sigma}_2 = 0$			$\ddot{\omega}_2 = 1; \ddot{\psi}_2 = 0.1; \ddot{\sigma}_2 = 0.1$			$\ddot{\omega}_2 = 1; \ddot{\psi}_2 = -0.1; \ddot{\sigma}_2 = 0.1$		
					$\alpha = 0.05$	$\alpha = 0.01$	t^*	$\alpha = 0.05$	$\alpha = 0.01$	t^*	$\alpha = 0.05$	$\alpha = 0.01$	t^*
Normal	15	15	3	3	0.053	0.011	0.571	0.055	0.023	0.618	0.055	0.015	0.637
	30	30	3	3	0.056	0.012	0.680	0.052	0.02	0.606	0.057	0.012	0.575
	60	60	3	3	0.048	0.012	0.586	0.048	0.013	0.593	0.047	0.011	0.630
	15	30	3	3	0.058	0.014	0.645	0.051	0.012	0.604	0.057	0.012	0.641
	30	60	3	3	0.057	0.013	0.599	0.051	0.012	0.543	0.051	0.013	0.589
Normal	15	15	3	4	0.059	0.013	0.613	0.067	0.011	0.572	0.041	0.008	0.605
	30	30	3	4	0.057	0.014	0.581	0.076	0.012	0.582	0.051	0.009	0.598
	60	60	3	4	0.055	0.012	0.630	0.071	0.012	0.623	0.058	0.012	0.622
	15	30	3	4	0.071	0.012	0.583	0.059	0.013	0.498	0.046	0.012	0.660
	30	60	3	4	0.057	0.009	0.556	0.079	0.013	0.560	0.049	0.014	0.535
Cauchy	15	15	3	3	0.022	0.001	0.468	0.029	0.001	0.505	0.039	0.001	0.572
	30	30	3	3	0.032	0.001	0.554	0.032	0.001	0.615	0.032	0.001	0.497
	60	60	3	3	0.027	0.001	0.571	0.032	0.001	0.568	0.032	0.001	0.597
	15	30	3	3	0.036	0.001	0.491	0.024	0.003	0.527	0.034	0.001	0.485
	30	60	3	3	0.035	0.001	0.581	0.038	0.004	0.564	0.040	0.001	0.531

Note: t^* is the average ratio between the $\hat{p}^*(t_{sr})$ and average p -value computed by MIVQUE.

Data is generated under homoscedastic set up with the main aim to compare with MIVQUE method.

Table 2. Observed power of t_{sr} test statistics

Distribution	n_1	n_2	m	T	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
Normal	15	15	2	3	0.881	0.743	0.663
	30	30	2	3	0.744	0.576	0.336
	60	60	2	3	0.981	0.870	0.764
	15	30	2	3	0.727	0.572	0.417
	30	60	2	3	0.772	0.522	0.455
Exponential	15	15	2	3	0.881	0.702	0.653
	30	30	2	3	0.737	0.618	0.408
	60	60	2	3	0.877	0.728	0.688
	15	30	2	3	0.742	0.643	0.462
	30	60	2	3	0.757	0.658	0.482
Cauchy	15	15	2	3	0.765	0.455	0.397
	30	30	2	3	0.669	0.481	0.372
	60	60	2	3	0.922	0.875	0.802
	15	30	2	3	0.643	0.551	0.340
	30	60	2	3	0.745	0.668	0.496

Note: Data is generated under heteroscedastic set up and the computed p -value of likelihood ratio test for equality of covariance matrices is 0.016.

To generate the data under H_1 , 0.5 has added to each observation in first group and further, add $0.1k$, $k = 2, 3$ and $0.2t$, $t = 2$ to each off diagonal in Ψ_A and Σ_A respectively. In most of the literature it has been implicitly assumed that for all response variables, the covariance matrices among the visits are same and for all time points, the covariance structures among all response variables are equal (Roy and Khattree 2005). The pooled covariance structure can be used for this purpose. In this paper, pooled covariance structure have formulated as,

$$\Psi_i(\text{pooled}) = \sum_{k=1}^m \Psi_{ik}/m \text{ and}$$

$$\Sigma_i(\text{pooled}) = \sum_{t=1}^T \Sigma_{it}/T$$

5. REAL DATA ANALYSIS

To illustrate our methodology, data has been used from Table 6.6.2 (Timm 2002), which was also analyzed by Naik and Rao (2001). Roy and Khattree (2005) also applied their algorithm on the same dataset. The data was provided by Dr. Thomas Zullo in the School of Dental Medicine, University of Pittsburgh. Total of 18 subjects were selected from the class of dental students and assigned to two orthopedic treatment groups with equal allocation. Each group was evaluated by studying the effectiveness of three measurements ($m = 3$) over three time points ($T = 3$).

Univariate as well as multivariate analysis have been done for the dental study data set using unstructured covariance matrix. It is observed that the mean difference vector between the groups is (4.6, 4.0, 4.4, 1.1, 2.0, 1.6, -0.7, -1.0, -0.8). SAS Proc Mixed procedure is used to analyze the data and the results are provided in Table 3. None of the responses are

Table 3. *F*-value and *P*-value of Treatment Effect for Orthopedic Data

Responses	MIVQUE		REML	
	<i>F</i> -Value	<i>P</i> -Value	<i>F</i> -Value	<i>P</i> -Value
Response 1	1.73	0.098*, 0.195**	1.73	0.098*, 0.195**
Response 2	0.47	0.247*, 0.494**	0.47	0.247*, 0.494**
Response 3	0.09	0.384*, 0.767**	0.09	0.384*, 0.767**
Overall	0.41	0.262*, 0.523**	0.12	0.363*, 0.725**

Note: * represent single-tail and ** represent two-tail *P*-value. Overall represent doubly repeated measure design.

significant at 5% level of significance. For one tail test, only response 1 is significant at 10% level of significance for the tests: MIVQUE and REML.

To apply our methodology, the estimated correlation matrices for both the groups are given by

$$\hat{\Psi}_A = \begin{pmatrix} 1 & 0.907 & 0.937 \\ 0.907 & 1 & 0.979 \\ 0.937 & 0.979 & 1 \end{pmatrix},$$

$$\hat{\Sigma}_A = \begin{pmatrix} 1 & 0.772 & 0.054 \\ 0.772 & 1 & 0.369 \\ 0.054 & 0.369 & 1 \end{pmatrix}$$

and

$$\hat{\Psi}_B = \begin{pmatrix} 1 & 0.989 & 0.987 \\ 0.989 & 1 & 0.997 \\ 0.987 & 0.997 & 1 \end{pmatrix},$$

$$\hat{\Sigma}_B = \begin{pmatrix} 1 & 0.925 & 0.351 \\ 0.925 & 1 & 0.607 \\ 0.351 & 0.607 & 1 \end{pmatrix}$$

Before applying our methodology in dental study dataset, the t_{sr} has been computed for each visit. Likelihood ratio test for homogeneity of covariance matrix is performed and computed *p*-values for each visit are reported in Table 4. The *p*-values of LRT for all the visits are highly non-significant and it is clearly observed from Table 4 that the computed value of t_{sr} is almost same with t_{GLS} given by (2.2) under homoscedastic condition. The calculated O'Brien OLS (t_{OLS}) test statistic for all the visits are also reported in Table 4 and it is seen that it is less than all the other test statistics values.

Table 4. O'Brien t_{OLS} , t_{GLS} and t_{sr} estimates for Dental Study Data

Visit	t_{OLS}	t_{GLS}	t_{sr}
1	0.685	0.979	0.929
2	0.706	0.529	0.513
3	0.661	0.846	0.836

Bootstrap methodology has been used to compute the *p*-value for our proposed test statistics. To compute *p*-value we simulate 100,000 bootstrap samples and calculate the *p*-value as mentioned in section 4. We repeat the above experiment for 100,000 times and it is observed that among all the experiments, maximum and minimum calculated *p*-value is 0.13

and 0.01 respectively whereas 0.060, 0.070 and 0.065 are obtained 850, 780 and 740 times out of 100,000 iterations. The histogram of the bootstrap p -values is depicted in Fig. 1. We recommend the weighted average or mid value of the maximum frequency class in histogram of the bootstrap p -values should be considered to make an inference for hypothesis testing. It is found that the weighted average of the computed p -values is 0.067 which implies there is no evidence to reject H_0 at 5% level of significance. The above methodology has been implemented in SAS ver. 9.1.3 with the main aim to compute the bootstrap p -value for easy application of our methodology by readers, and computer program can be obtained on request from the first author. To this end, user can also modify the SAS code to compute t_{GLS} given in (2.3) with covariates or can be obtained from authors.

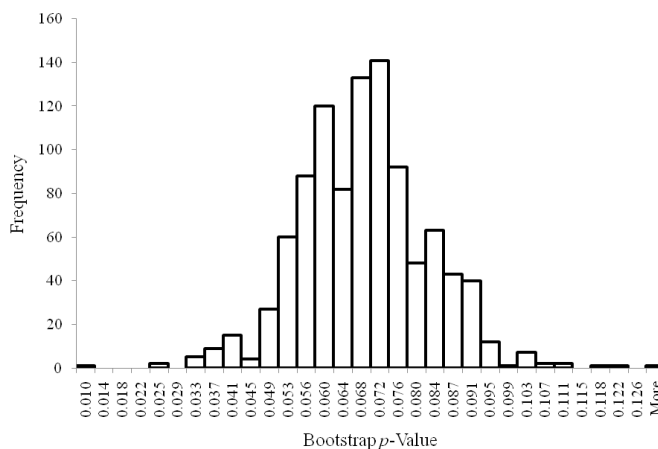


Fig. 1. Empirical distribution of bootstrap p -value for Dental Data

6. CONCLUDING REMARKS

During the last two decades in medical research, it has become very common to compare the effectiveness of two or more drugs in terms of multiple measurements (Dmitrienko *et al.* 2010). Generally multiple responses are required when disease conditions are difficult to epitomize by only one response variable. It is not easy to select only one single endpoint to entirely reveal the unobservable true condition for therapeutic areas like anti-viral, photo damage etc. The problem gets more complicated when multiple responses are collected repeatedly over time. In early phase of clinical trials, data is collected on multiple responses in smaller number of subjects. In this case, multivariate method should be applied to maintain type-I error rate. In this paper, the extension of O'Brien GLS test in repeated

measure design for heteroscedastic covariance structure setup has been presented. The main advantage of this testing procedure is that no distributional assumptions are needed as bootstrap methodology has been used for inference. It has been clearly observed that the proposed methodology is better than MIVQUE method and Restricted Maximum Likelihood (REML) procedure. Additionally, unbalanced data can also be easily modeled with the proposed methodology. Several types of covariance structures are available in literature (Hinkelmann and Kempthorne 2008) to model longitudinal data. The proposed methodology is very much flexible with any kind of covariance structure present in the realization. In future research, it would be useful to find the distribution of the proposed test statistics. We have developed our methodology only for the marginal effect model; however future attempts would be made to extend it for mixed effect model also. Effort can also be directed to apply Newton-Raphson technique or more advanced search algorithm viz., Genetic Algorithm, Swarm Optimization etc. to search the global optima and corresponding estimates.

REFERENCES

- Alosh, M. and Fritsch, K. (2002). Testing for treatment effect in clinical trials with multidimensional categorical data with repeated measurements. *Proceedings of the Joint Statistical Meetings, Biometrics Section*, **913**, 45-50.
- Dallow, N.S., Leonov, S.L. and Roger, J.H. (2008). Practical usage of O'Brien's OLS and GLS statistics in clinical trials. *Pharm. Statist.*, **7**, 53-68.
- Dmitrienko, A., Tamhane, A.C. and Bretz, F. (2010). *Multiple Testing Problems in Pharmaceutical Statistics*. Chapman & Hall, New York.
- Efron, B. and Tibshirani, R.J. (1995). *An Introduction to Bootstrap*. Chapman and Hall, New York.
- Galecki, A.T. (1994). General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Comm. Statist.-Theory Methods*, **23(11)**, 3105-3119.
- Hinkelmann, K. and Kempthorne, O. (2008). *Design and Analysis of Experiments*. Vol-1, 2nd Ed., John Wiley & Sons. Inc., New Jersey.
- Huque, M.F. and Sankoh, A.J. (1997). A reviewer's perspective on multiple endpoint issue in clinical trials. *J. Biopharm. Statist.*, **7(4)**, 545-564.
- O'Brien, P.C. (1984). Procedures for comparing samples with multiple end-points. *Biometrics*, **40**, 1079-1087.
- Khattree, R. and Naik, D.N. (1999). *Applied Multivariate Statistics with SAS Software*. 2nd Ed., John Wiley & Sons. Inc., New York.

- Logan, B.R. and Tamhane, A.C. (2004). On O'Brien's OLS and GLS tests for multiple endpoints. *IMS, Lecture Notes-Monograph Series*, **47**, 76-88.
- Naik, D.N. and Rao, S.S. (2001). Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix. *J. Appl. Statist.*, **28**, 91-105.
- Pocock, S.J., Geller, N.L. and Tsiatis, A.A. (1987). The analysis of multiple endpoints in clinical trials. *Biometrics*, **43**, 487-498.
- Roy, A. and Khattree, R. (2005). Testing the hypothesis of a kronecker product covariance matrix in multivariate repeated measure data. *Processing of the 30th Annual SAS Users Group International Conference (SUGI)*.
- Timm, N.H. (2002). *Applied Multivariate Statistics*. Springer-Verlag, New York.
- Yuan, K.H. and Hayashi, K. (2003). Bootstrap approach to inference and power analysis based on three test statistics for covariance structure models. *British J. Math. Statist. Psycho.*, **56**, 93-110.