JP Journal of Biostatistics
Volume 2, Number 2, 2008, Pages 155-168
Published Online: August 13, 2008
This paper is available online at http://www.pphmj.com

ANALYZING REPEATED MEASUREMENTS DATA: A PRACTICAL COMPARISON OF METHODS IN THE STUDY ON EFFICACY OF LEVOCETRIZINE, DESLORATIDINE AND FEXOFENADINE IN HISTAMINE INDUCED WHEAL SUPPRESSION

ROYES JOSEPH and ANIL C. MATHEW

Department of Statistics
St. Thomas College
Pala, Kerala, India

Department of Community Medicine
Division of Biostatistics
PSG Institute of Medical Sciences and Research
Coimbatore, Tamil Nadu, India
e-mail: dranilmathew@rediffmail.com

Abstract

Variety of methods are available for analyzing repeated measurements data in clinical research. However, there is little information on how established methods, such as summary statistics approach, unstructured multivariate approach, profile analysis, and repeated measures ANOVA compare in practice. In this paper we compared the results by application to a clinical trial data set. The aim of this paper is to exemplify the use of these methods, and directly compare their results by application to a clinical trial data. The focus is on practical aspects rather than technical issues. The data considered were taken from a clinical trial on efficacy of Levocetrizine, Desloratidine and Fexofenadine in histamine induced wheal suppression at 1, 2 and 3 hours. The findings are presented and discussed.

2000 Mathematics Subject Classification: 62P10.

Keywords and phrases: clinical trials, repeated measurements.

Received April 14, 2008

1. Introduction

In clinical trials and other experimental studies, one frequently encounters repeated observations of an outcome, measured on every subject at several time points. The objectives of repeated measures data analysis are to examine and compare response trend over time. This can involve comparison of times within a treatment. The feature of repeated measures experiments that requires special attention in data analysis is the correlation pattern among the responses on the same subjects over time. In this paper, we used some of the suggested methods for the analysis of repeated measurements in comparing the pattern of histamine induced wheal suppression between the three treatments Levocetrizine, Desloratidine and Fexofenadine which are second generation antihistamine commonly used in the treatment of urticaria and perennial allergic rhinitis. The purpose of this paper is to exemplify a range of commonly used methods for the analysis of repeated measurements and to compare their results in the analysis of the above mentioned clinical trial data set. Focus is on aspects relevant to applied medical statisticians rather than technical issues. Comparisons are made in terms of both substantive conclusions and estimated treatment effects.

Section 2 describes the clinical trial data. In Section 3, each method is introduced in turn, with comments on its basis, limitations and strengths and is applied to the analysis in the clinical trial data. The paper concludes with discussion and practical implications of the comparison of methods.

2. Data Collection

The study was done on 30 healthy volunteers (18-50 years) after obtaining an informed consent. The volunteers were not on antihistamines, steroid and immunosuppressant for seven days prior to the study (acetemizole for six weeks). None of them had history of atopy, drug hypersensitivity or use of alcohol. Pregnant and lactating women were excluded.

Volunteers were administrated Levocetrizine 5 mg, Desloratidine 5 mg and Fexofenadine 180 mg at weekly intervals to prevent any carry over

effect of the drugs. A prick test was performed before administration of the drugs by the standard method using histamine 0.1% w/v solution. A drop of 0.1% w/v of histamine solution was placed on the flexor aspect of the forearm. The skin was pricked through the histamine solution with a lancet. The tip of lancet was kept parallel to the skin surface and the skin lifted by tenting the lancet by 45-60°.

After one minute the test site was wiped with filter paper to remove the excess histamine solution. The size of the wheal was calculated by measuring the maximum diameter of the wheal and the orthogonal diameter with a transparent scale. Skin prick test was repeated at 1, 2 and 3 hours after each drug at weekly intervals. The test was performed in a marked square area of 1×1 cm at different site each time, on the flexor aspect of the forearm. The maximum size of the histamine induced wheal (in nearest mm) was recorded at each time. These data were collected by two senior medical doctors, in the Department of Dermatology, who have undergone substantial training in data collection before the study started. The aim of the study was to assess the pattern of wheal suppression in relation to time for each treatment and to compare the efficacy of these three treatments in reducing wheal suppression at various time intervals.

3. Choice of Method

3.1. Univariate summary statistic method

The simplest approach to the analysis of repeated measurements is to reduce the vector of multiple measurements from each experimental unit to a single measurement (Frison and Pocock [5], Crowder and Hand [1], Everitt [4]). This avoids the issue of correlation among the repeated measurements from a subject (Dawson [3]). The principal advantages of this method are its technical simplicity and the ease with which it can be communicated to non-biostatisticians.

In the histamine induced wheal suppression study, the relationship between the size of histamine induced wheal and for each subject can be adequately summarized by the slope of the least square regression line. Assuming that the estimated slopes are normally distributed, the one-

sample t test was used to determine the wheal suppression is affected by time in each treatment. Also one-way ANOVA test was used to determine whether the pattern of change over time is the same across the three treatments.

3.2. Unstructured multivariate approach

3.2.1. One-sample repeated measurements

Let y_{ij} denote the response from subject i at time j, for $i = 1, 2, ..., n$, $j = 1, 2, ..., t$. The vectors, $y_i = (y_{i1}, y_{i2}, ..., y_{it})'$, $i = 1, 2, ..., n$, are a random sample from $N_t(\mu, \Sigma)$, where $\mu = (\mu_1, \mu_2, ..., \mu_t)'$.

To assess whether the size of the histamine induced wheal changes with time, i.e., we have to test $H_0 : \mu_1 = \mu_2 = \cdots = \mu_t$.

Let $y^*_{ij} = y_{ij} - y_{ij-1}$ for $j = 1, 2, 3$. Then $y^*_i = (y^*_{i1}, y^*_{i2}, ..., y^*_{i(t-1)})'$ vectors are a random sample from $N_{t-1}(\mu^*, \Sigma^*)$, where $\mu^* = (\mu_1 - \mu_2, ..., \mu_{t-1} - \mu_t)$.

The hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_t$ is then equivalent to $H_0{}^* : (0, 0, ..., 0)'$.

The test of $H_0{}^*$ can be carried out using the T^2 statistic computed from the sample mean vector and covariance matrix of the y^*_{ij} values. The T^2 statistic is given by

$$T^2 = n\bar{y}^* s^{*-1} \bar{y}^{*\prime} \sim T^2_{(t-1, n-1)}$$

and the F statistic,

$$F = \frac{(n - t + 1)}{(n - 1)(t - 1)} T^2$$

has the $F_{(t-1, n-t+1)}$ distribution if $H_0{}^*$ is true.

To assess whether the relationship between wheal suppression and time is linear and since the four measurements are equally spaced, the

test of nonlinearity can be carried out using orthogonal polynomial coefficients (Pearson and Hartley [9]). The hypothesis that the nonlinear (quadratic and cubic) effects of time on size of wheal suppression are jointly equal to zero is assessed by testing $H_0 : C\mu = 0_2$, where

$$C = \begin{bmatrix} 1 & -1 & -1 & 1 \\ -1 & 3 & -3 & 1 \end{bmatrix}.$$

Thus, $T^2 = n(C\bar{y})'(CSC')^{-1}(C\bar{y})$ and $F = \frac{(n-c)}{(n-1)c}T^2 \sim F_{(c,n-c)}$ distribution when H_0 is true, where c is the rank of C.

3.2.2. Two-sample repeated measurements

The extension of the unstructured multivariate approach to the situation when repeated measurements at t time points are obtained from two independent groups of subjects is straightforward (Davis [2]). Let $y_{hi} = (y_{hi1}, y_{hi2}, ..., y_{hit})$ denote the vector of observations from the ith subject in group h for $i = 1, ..., n_h$, $h = 1, 2$. We assume that the vectors $y_{11}, y_{12}, ..., y_{1n1}$ are an independent random sample from the $N_t(\mu_1, \Sigma)$ distribution, where $\mu_1 = (\mu_{11}, \mu_{11}, ..., \mu_{1t})'$. We similarly assume that the vectors $y_{21}, y_{22}, ..., y_{2n2}$ are an independent random sample for the $N_t(\mu_2, \Sigma)$ distribution, where $\mu_2 = (\mu_{21}, \mu_{21}, ..., \mu_{2t})'$. Note that the covariance matrices of the two distributions are assumed equal. One hypothesis of general interest is $H_0 : \mu_1 = \mu_2$.

Based on the properties of linear combinations of multivariate normal random vectors,

$$\bar{y}_1 - \bar{y}_2 \sim N\left(\mu_1 - \mu_2, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\Sigma\right).$$

The pooled estimator of the covariance matrix Σ is given by

$$S = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n + n_2 - 2},$$

where $S_h = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(y_{hi} - \bar{y}_h)(y_{hi} - \bar{y}_h)'$ is the sample covariance matrix in group h.

Therefore the statistic

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{y}_1 - \bar{y}_2)' S^{-1} (\bar{y}_1 - \bar{y}_2)$$

and the F statistic is

$$F = \frac{n_1 + n_2 - t - 1}{(n_1 + n_2 - 1)t} T^2 \sim F_{t, n_1+n_2-t-1} \text{ distribution when } H_0 \text{ is true.}$$

If $H_0 : \mu_1 = \mu_2$ is rejected, a weaker, and often more realistic, hypothesis is that the mean profiles in the two groups are parallel; that is, that the μ_1 and μ_2 profiles differ only by a constant vertical shift. This hypothesis of parallelism can be expressed in the matrix notation $H_0 : C(\mu_1 - \mu_2) = 0_{t-1}$, where

$$C = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 \\ . & . & . & . & \cdots & . & . \\ . & . & . & . & \cdots & . & . \\ . & . & . & . & \cdots & . & . \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}.$$

3.3. Profile analysis

Suppose that repeated measurements at t time points have been obtained from s groups of subjects. Let n_h denote the number of subjects in group h for $h = 1, 2, ..., s$, and let n denote the total sample size. Let y_{hij} denote the response at time j from the ith subject in group h for $h = 1, 2, ..., s$, $i = 1, 2, ..., n_h$, and $j = 1, 2, ..., t$. We assume that the data vectors $y_{hi} = (y_{hi1}, y_{hi2}, ..., y_{hit})'$ are independent and normally distributed with mean $\mu_h = (\mu_{h1}, ..., \mu_{ht})'$ and common covariance matrix Σ. Thus, $y_{hi} \sim N_t(\mu_h, \Sigma)$.

The profile analysis model is $y_{hij} = \mu_{hj} + e_{hij}$, where e_{hij} is the residual for subject i in group h at time j. The vector $e_{hi} =$

$(e_{hi1}, e_{hi2}, ..., e_{hit})'$ is the vector of residuals for the ith subject in group h. In terms of the multivariate general linear model, $\mathbf{Y} = \mathbf{XB} + \mathbf{E}$, where $\mathbf{Y}$ and $\mathbf{E}$ are $n \times t$ matrices with rows $y'_{11}, y'_{12}, ..., y'_{sns}$ and $e'_{11}, ..., e'_{sns}$, respectively, $\mathbf{X}$ is $n \times s$, and $\mathbf{B}$ is $s \times t$.

Three general hypotheses are of interest in profile analysis:

H_{01}: the profiles for the s groups are parallel (i.e., no group by time interaction);

H_{02}: no differences among groups;

H_{03}: no differences among time points.

Note that H_{01} should be tested first, because acceptance or rejection of this hypothesis affects how the two other hypotheses can be tested.

3.3.1. Test of parallelism

The hypothesis of parallelism is

$$H_{01}: \begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{12} - \mu_{13} \\ \cdot \\ \cdot \\ \cdot \\ \mu_{1t-1} - \mu_{1t} \end{bmatrix} = \begin{bmatrix} \mu_{21} - \mu_{22} \\ \mu_{22} - \mu_{23} \\ \cdot \\ \cdot \\ \cdot \\ \mu_{2t-1} - \mu_{2t} \end{bmatrix} = \cdots = \begin{bmatrix} \mu_{s1} - \mu_{s2} \\ \mu_{s2} - \mu_{s3} \\ \cdot \\ \cdot \\ \cdot \\ \mu_{st-1} - \mu_{st} \end{bmatrix}.$$

Testing this hypothesis is equivalent to carrying out a one-way multivariate analysis of variance (MANOVA) model on the $t-1$ differences between adjacent time points from each sampling unit (O'Brien and Kaiser [8]).

3.3.2. Tests of no differences among groups

Depending on the results of the test of H_{01}, two tests of the hypothesis H_{02} of no differences among groups are possible.

First, if the parallelism hypothesis is reasonable, then the test for differences among groups can be carried out using the sum (or average) of

the repeated observations from each subject. In this case,

$$A_{(s-1)\times s} = (I_{s-1}, -1_{s-1})' \text{ and } C = I_t.$$

Because the s groups are independent, this test of H_{02} is equivalent to that from a one-way ANOVA on the totals (or means) across time from each subject.

A multivariate test for differences among groups can also be carried out without assuming parallelism. In this case, the null hypothesis is

$$H_{02} : \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ . \\ . \\ . \\ \mu_{1t} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ . \\ . \\ . \\ \mu_{2t} \end{bmatrix} = \ldots = \begin{bmatrix} \mu_{s1} \\ \mu_{s2} \\ . \\ . \\ . \\ \mu_{st} \end{bmatrix}.$$

In terms of general hypothesis $H_0 : ABC = D$, where

$$A_{(s-1)\times s} = (I_{s-1}, -1_{s-1})', \; C = I_t \text{ and } D_{(s-1)\times s} = 0.$$

3.3.3. Tests of no differences among time points

Depending on the results of the test of H_{01}, two tests of H_{03} are possible. If the parallelism hypothesis is reasonable, then the test for differences among time points can be carried out using the sum (or average) across groups of the observations at each time point. In this case, the null hypothesis is

$$H_{03} : ABC = D,$$

where $A_{1\times s} = (1, 1, \ldots, 1)$ or $(1/s, \ldots, 1/s)$, $C_{t\times t-1} = \begin{pmatrix} I_{t-1} \\ -1'_{t-1} \end{pmatrix}$ and $D_{1\times t-1} = 0'_{t-1}$.

This is equivalent to a one-sample T^2 test, as described in Subsection 3.2.1.

This procedure weights each of the s groups equally and is usually

appropriate. However, if unequal group sizes result from the nature of the experimental conditions, then it may be desirable to use a weighted average rather than a simple average. In this case, $A = (n_1, ..., n_s)$ or $A = (n_1/n, ..., n_s/n)$ can be used; note that C and D are unchanged.

The hypothesis H_{03} can also be tested without assuming parallelism:

$$H_{03}: \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ . \\ . \\ . \\ \mu_{s1} \end{bmatrix} = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ . \\ . \\ . \\ \mu_{s2} \end{bmatrix} = \cdots = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \\ . \\ . \\ . \\ \mu_{st} \end{bmatrix}.$$

In this case, $A_{s\times s} = I_s$, $C_{t\times(t-1)} = \begin{pmatrix} I_{t-1} \\ -1'_{t-1} \end{pmatrix}$ and $D_{s\times(t-1)} = 0$.

3.4. Repeated measures analysis of variance

Traditional technique like repeated measures ANOVA can model time as either fixed or random, but violation of the sphericity assumption is a central concern. Sphericity dictates that the pair wise difference scores between observations have equal variance. Failure to meet the sphericity assumptions results in a serious liberal bias, or Type 1 error. Technique such as the Greenhouse-Geisser correction can compensate for sphericity violations (Greenhouse and Geisser [7]).

Suppose that repeated measurements at t time points are obtained from s groups of subjects. Let n_h denote the number of subjects in group h. Let y_{hij} denote the response at time j from the ith subject in group h for $h = 1, ..., s$, $i = 1, ..., n_h$, and $j = 1, ..., t$. The model is

$$y_{hij} = \mu + \gamma_h + \tau_j + (\gamma\tau)_{hj} + \pi_{i(h)} + e_{hij}.$$

In this model, μ is the overall mean and γ_h is the fixed effect of group h, with $\sum_{h=1}^{s} \gamma_h = 0$. In addition, τ_j is the fixed effect of time j, with $\sum_{j=1}^{t} \tau_j = 0$, and $(\gamma\tau)_{hj}$ is the fixed effect for the interaction of the hth group with the

jth time. The constraints on the interaction parameters are $\sum_{h=1}^{s}(\gamma\tau)_{hj} = \sum_{j=1}^{t}(\gamma\tau)_{hj} = 0$.

The e_{hij} parameters are independent random error terms with $e_{hij} \sim N(0, \sigma_e^2)$.

Under the assumption repeated measures ANOVA can be used to test for the effects due to treatments, time and the interaction between treatment and time.

4. Results and Discussion

Mean and standard deviation (S.D.) of wheal suppression at Pre, 1hr, 2hr and 3hr for the three treatments (in mm) are given in Table 1.

Table 1

Treatment	Pre	Hour 1	Hour 2	Hour 3
	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)
Levocetrizine	5.033 (0.72)	3.73 (0.79)	2.47 (0.86)	1.47 (1.11)
Desloratidine	5.17 (0.87)	4.05 (1.04)	3.40 (0.82)	2.50 (0.86)
Fexofenadine	5.23 (0.77)	3.60 (0.68)	2.23 (0.90)	1.50 (1.38)

Means plotted in Figure 1 show decrease in wheal size over time for all the three treatments.

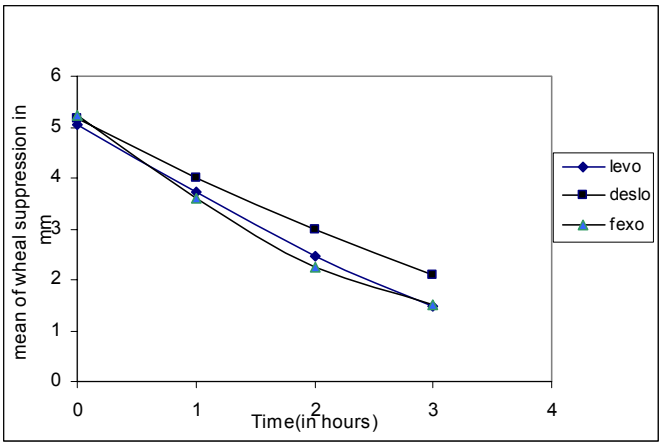

Figure 1. Pattern of wheal suppression between Levocetrizine, Desloratidine and Fexofenadine.

In summary statistic approach, the relationship between time and wheal suppression for each subject can be adequately summarized by the slope of the least square regression line. The sample mean and standard deviation of the slopes for Levocetrizine −0.8267 and 0.2776, for Desloratidine −1.087 and 0.3021, and for Fexofenadine −0.7584 and 0.3238, respectively. The one-sample t test gives p-value of 0.000 for each treatment. This indicating that the size of wheal tends to decrease as time increases. The one-way ANOVA test for compare the distribution of the estimated slopes in the three treatments gives a F statistic of 9.88 with (2, 87) df (p = 0.000). That is, there is sufficient evidence to conclude that the reduction in wheal size in three treatment groups differ significantly.

Unstructured multivariate approach for one-sample repeated measurement analysis also showed that the mean wheal suppression at 4 time points differ significantly for each treatment at 5% level (p-value for each treatment equal to 0.000). In the same model, when the departure from linearity was tested, it is found that the relationship between wheal suppression and time appears to be linear in Levocetrizine (p = 0.2252) and Desloratidine (p = 0.1970). But for Fexofenadine, the departure from linearity was statistically significant (p = 0.0037).

Unstructured multivariate approach for two-sample repeated measurement analysis showed that profiles for Levocetrizine and Desloratidine are not the same ($p = 0.000$). Test for parallelism showed that the profiles of Levocetrizine and Desloratidine are not parallel ($p = 0.0001$).

In profile analysis, the F statistic is 5.92 with 6 and 170 df ($p = 0.000$). This showed that the responses for the three treatments are not parallel. Then the test of no differences among treatments gives F statistic of 5.53 with 8 and 168 df and the test of no differences among time points gives F statistic of 30.27 with 9 and 278. Thus at 5% level of significance, one can conclude that the profiles for the 3 treatments are not the same ($p = 0.000$) and the means at the four time points are highly significantly different ($p = 0.000$).

In repeated measures ANOVA, under the assumption that the within-group covariance matrices are equal and that the sphericity condition is satisfied, there is a highly significant interaction effect between treatment and time ($p = 0.000$). Thus, the shapes of the profiles are not the same across the three treatments. The test of sphericity, however, is rejected (chi-square approximation to Mauchly's criterion is 46.04 with 5 df ($p = 0.000$)). In this situation, we preferred the unstructured multivariate approach.

5. Conclusion

The aim of this paper is to increase the accessibility to applied statisticians of a variety of methods for the analysis of repeated measurements data, based on a comparison of three methods on a real data set. Responses measured on the same subject are correlated because they contain a common contribution from the subject. Moreover, measures on the same subject close in time tend to be more highly correlated than measures far apart in time. Also, variance of repeated measures often changes with time. These potential patterns of correlation and variation may combine to produce a complicated covariance structure of repeated measures. Special methods of statistical analysis are needed for repeated measures data because of the covariance structure.

In this paper we reviewed a number of methods for the analysis of repeated measurement data. These include the summary statistic method, unstructured multivariate approach, profile analysis and repeated measurement analysis of variance. The advantages of using a summary statistic method, principally its simplicity, are discussed in this paper. In unstructured multivariate approach, rather than reducing the vector of repeated measurements from each subject to a single measurement, all of the data are used and there is no covariance structure assumption.

Repeated measurement ANOVA may be a useful alternative approach to multivariate approach. Many studies using repeated measures ANOVA were reported despite being the fact that sphericity assumption is violated (Girden [6]). On the other hand, even when the normality assumption of the multivariate approach is violated, such violations are generally regarded as less serious than violations of the sphericity assumption. Therefore, when the researcher's concern is committing a Type I or a Type II error and several assumptions hold, the multivariate approach is suggested.

In conclusion, we have presented various models related to repeated measures and compared the results in studying the pattern of three treatments on histamine induced wheal suppression. Each model has its own assumptions to meet. Considering the various models, it is concluded that pattern of wheal suppression at various time intervals was statistically significant in all the three treatments. Pattern of wheal suppression between the three treatments was also statistically significant. Pattern in wheal suppression of Levocetrizine and Desloratidine is not parallel. Pattern of Fexofenadine is not linear over time. The study provides a basis for encouraging further efforts in this area.

Acknowledgements

The authors are grateful to Dr. G. K. Shukla, Professor Emeritus Statistics, Indian Institute of Management, Lucknow for his valuable comments which helped in improving the quality of this paper. The authors are thankful to Dr. S. Ramalingam, Principal and Dr. Thomas V.

Chacko, Professor and Head of Department of Community Medicine, PSG Institute of Medical Sciences and Research for their support in doing this study. We are also thankful to Dr. C. R. Srinivas, Professor and Head of the Department of Dermatology and Dr. Reena Rai, Professor of Dermatology, PSG Hospitals for providing the necessary data to do this study. We are also thankful to Rev. Dr. Mathew John Kokkatt, Principal and Dr. K. K. Jose, Professor and Head of Department of Statistics, St. Thomas College, Pala for their encouragement in this study.

References

[1] M. J. Crowder and D. J. Hand, Analysis of Repeated Measures, Chapman and Hall, London, 1990.

[2] C. S. Davis, Statistical Methods for the Analysis of Repeated Measurements, Springer, New York, 2002.

[3] J. D. Dawson, Comparing treatment groups on the basis of slopes, area-under-the curve, and other summary measures, Drug Inform. J. 28 (1994), 723-732.

[4] B. S. Everitt, The analysis of repeated measures: a practical review with examples, Statistician 44 (1995), 113-135.

[5] L. Frison and S. J. Pocock, Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design, Stat. Med. 11 (1992), 1685-1704.

[6] E. R. Girden, ANOVA: Repeated Measures, Sage, Newbury Park, CA, 1992.

[7] S. W. Greenhouse and S. Geisser, On methods in the analysis of profile data, Psychometrika 24 (1959), 95-112.

[8] R. G. O'Brien and M. K. Kaiser, Manova methods for analyzing repeated measures designs: an extensive primer, Psychological Bull. 97 (1985), 316-333.

[9] E. S. Pearson and H. O. Hartley, Biometrika Tables for Statisticians, Vol. I., Cambridge University Press, Cambridge, 1996.

■