Rank-Based Analysis of Unbalanced Repeated Measures Data

M. Mushfiqur Rashid*
Division of Biostatistics IV
Office of Biostatistics, Center for Drug Evaluation and Research
Food and Drug Administration, U.S.A
Mushfiqur.Rashid@fda.hhs.gov

Joseph W. McKean
Department of Statistics
Western Michigan University, U.S.A.

John D. Kloke
Department of Statistics and Medical Informatics
University of Wisconsin, U.S.A.

Abstract

In this article, we have developed a rank (intra-subject) based analysis of clinical trials with unbalanced repeated measures data. We assume that the errors within each patient are exchangeable and continuous random variables. This rank-based inference is valid when the unbalanced data are missing either completely at random or by design. A drop in dispersion test is developed for general linear hypotheses. A numerical example is given to illustrate the procedure.

Keywords: Clinical trials; Dispersion function; JR-Estimators; MR-Estimators; Interactions.

1. Introduction

In a typical clinical trial with repeated measures, a set of patients is chosen according to selection criteria defining a patient population. These patients are randomized into two or more treatment groups and are then observed at successive time points until the end of the trial. Even though trials might be designed with the intent to collect the same set of measurements from each patient, there are many practical situations where the investigators have to deal with observation (per patient) vectors of unequal length. Some patients might not have all measurements collected at all planned visits or time points. Patients could also be discontinued from the trial for one or more reasons. Over the course of the study, patients may miss visits, stop taking their assigned treatment, or be removed from treatment. Whatever the reasons for dropouts, drug related or not,

withdrawals of patients from the trial result in unbalanced data. In this article, we have assumed that these unbalanced data are missing either completely at random or by design. In multivariate settings where missing values occur on more than one variable, the noncompleters are often a substantial portion of the entire data set. If so, ignoring them may be inefficient, causing a large amount of information to be discarded. Moreover, discarding them from the analysis will tend to introduce bias, to the extent that the incompletely observed cases differ systematically from the completely observed cases. The completely observed cases will be unrepresentative of the population for which the inference is usually intended: the population of all cases, rather than the population of cases with no missing data. Therefore, the results of this article have potential application in pharmaceutical, medical, and clinical research.

In this article, using a sum of Jaeckel (1972) type dispersion functions we develop a rank (intra-patient) based analysis of the time effects of the repeated measure design. Our analysis includes estimation of the time effects and their standard error of the estimates. Further, we develop a test based on the drop in dispersion for general linear hypotheses concerning the time effects, including an asymptotically distribution-free test of the interaction between the time effects and treatment groups. As discussed in Sections 4 and 5 our proposed analysis is robust. Furthermore, our analysis is completely invariant to the random effects due to the patients. and it allows each patient to have an observation vector of unequal length.

Kloke, McKean, and Rashid (2009) developed a rank-based analysis of a general mixed model which can be used for this repeated measures design. In contrast to the analysis discussed in this paper, it is not invariant to the random effect due to the patients; although, its inference is adjusted for the random effect. Similar to the theory for the method discussed in this paper, the theory assumes that the number of patients become large. We compare this analysis to our proposed analysis in Section 3.

Also, Rashid, McKean, and Kloke (2012) proposed an analysis for a multicenter clinical design, which is also based on a sum of Jaeckel dispersion functions. Their design was a mixed model with centers treated as random. Hence, for the situation of this paper, centers correspond to a patient. The asymptotic theory for this analysis assumes that the sample size of a center becomes large while the number of centers remains fixed. In contrast, the methodology developed in this paper assumes that the number of patients become large while the number of repeated measures remains finite. Hence, the theories for the two analyses are quite different. The analysis of Rashid et al. (2012) is briefly discussed in Section 3, also.

In Section 2 we present the repeated measures model discussed in this paper. In Section 3, we discuss the details of our proposed method and in Sections 4 and 5 present its asymptotic theory. In Section 6 we propose a consistent estimate of a necessary scale parameter for our procedure. An example which illustrates our methodology is presented in Section 7.

2. Repeated Measures Model

Consider an experiment with q treatment groups, n_j subjects belong to the jth $(j = 1, 2, \ldots, q)$ treatment group and the kth subject in the jth group is supposed to be observed at p time points. A model for this kind of experiment is:

$$Y_{ijk} = \mu + \alpha_i + \delta_j + \gamma_{ij} + \beta_{jk} + e_{ijk}, \tag{2.1}$$

where Y_{ijk} is the observation corresponding to the $ith(i = 1, 2, \ldots, p)$ time treatment, the $jth(j = 1, 2, \ldots, q)$ group treatment and the $kth(k = 1, 2, \ldots, n_j)$ patient, $\mu =$ overall mean (or median), $\alpha_i =$ main effect ($\sum_{i=1}^{p} \alpha_i = 0$) of the ith time treatment, $\delta_j =$ main effect ($\sum_{j=1}^{q} \delta_j = 0$) of the jth group treatment, $\gamma_{ij} =$ ($\sum_{i=1}^{p} \gamma_{ij} = 0, \sum_{j=1}^{q} \gamma_{ij} = 0$ and $\sum_{i=1}^{p} \sum_{j=1}^{q} \gamma_{ij} = 0$) interaction between the ith time treatment and the jth group treatment, e_{ijk} is the error term corresponding to the (ijk)th observation, and β_{jk} is the random subject effect for the kth subject in the jth group. We assume that e_{ijk} are iid random variables with mean (or median) zero and scale parameter σ_e^2, and β_{jk} are iid random variables with mean (or median) 0 and scale parameter σ_β^2. Let $g(t)$ denote the common probability density function (pdf) of e_{ijk}. We discuss tests of general linear hypotheses, but our main hypothesis of interest is that of interaction between the the group and times effects, i.e.,

$$H_0 : \gamma_{ij} \equiv 0 \; versus \; H_A : \gamma_{ij} \neq 0, \; for \; some \, i, j. \tag{2.2}$$

Model (2.1) can be rewritten as a repeated measures model as follows:

$$Y_{ijk} = \mu + \alpha_i + \delta_j + \gamma_{ij} + \varepsilon_{ijk}, \tag{2.3}$$

where $\varepsilon_{ijk} = \beta_{jk} + e_{ijk}$. Let $\underline{Y}_{jk} = (Y_{1jk}, Y_{2jk}, \ldots, Y_{pjk})'$ be the vector of observations for the kth patient who receives the jth group treatment $(j = 1, 2, \ldots, q; k = 1, 2, \ldots, n_j)$ and $\underline{\varepsilon}_{jk}$ be the error vector corresponding to $\underline{Y}_{jk}$. In normal theory inference, one may assume that $\underline{\varepsilon}_{jk}(j = 1, 2, \ldots, q; k = 1, 2, \ldots, n_j)$ are independent random vectors with equicorrelated covariance matrix $\sigma^2[(1-\rho)I_{p \times p} + \rho \underline{1}_p \underline{1}'_p]$, where $var(\varepsilon_{ijk}) = \sigma^2 = \sigma_e^2 + \sigma_\beta^2$ $(0 < \sigma^2 < \infty)$ and $\rho = \sigma_\beta^2 / \sigma^2$ is the correlation coefficient between any two components of $\underline{\varepsilon}_{jk}$, $I_{p \times p}$ is an identity matrix of order p and $\underline{1}_p$ is a $p \times 1$ vector of unity.

If the normality assumption is not reasonable, the experimenter would prefer to use a distribution-free procedure or an asymptotically distribution free procedure. For the theory of such procedures in this paper, we make the following assumptions,

B1: The error vectors $\underline{\varepsilon}_{jk}$ $(j = 1, 2, \ldots, q; k = 1, 2, \ldots, n_j)$ for individual subjects are independent and continuous random vectors with the c.d.f. $F(x_1, x_2, \ldots, x_p)$ and the elements of each $\underline{\varepsilon}_{jk}$ are exchangeable random variables.

B2: The scale parameter $\tau = (1 / [\sqrt{12} \int_{-\infty}^{\infty} f(x, x) dx]) < \infty$ where $f(., .)$ is the bivariate density corresponding to the joint c.d.f. $F(., \ldots, .)$.

Note that under the assumption of exchangeable errors the scale parameter τ can be expressed as

$$\tau = 1 / \left[\sqrt{12} \int_{-\infty}^{\infty} g^2(t) dt \right], \tag{2.4}$$

where $g(t)$ is the pdf of the iid errors e_{ijk}.

Assumption [B.1] is standard for repeated measures data. It reflects the dependency of the data within each subject. Exchangeability of the components in ε_{jk} also implies there is no bias among the components. As discussed below, the scale parameter τ is the constant of proportionality in the standard errors of the rank-based estimates and it is the standardizing factor in the associated test statistics of general linear hypotheses. From the expression (2.4), it follows that τ plays a role similar to $Var(e_{ijk}) = \sigma_e^2 = \sigma^2(1 - \rho)$ in normal theory inference.

Let $I_{ijk} = 1$ if the kth patient in the jth group is measured at the ith time point and $I_{ijk} = 0$ otherwise. Then $\sum_{i=1}^{p} I_{ijk} = m_{jk} (\leq p)$ is the number of levels of the time treatment applied on the kth patient in the jth group treatment. Our description of the model (1.3) is analogous to that of Anderson & Bancroft (1952) (e.g., see page 250, section 19.2). Wei and Lachin (1984) used a similar indicator function (I_{ijk}) for incorporating missing values in the context of incomplete repeated measures data

3. Rank-Based Invariant Estimation

Let $\theta_{ij} = \mu + \alpha_i + \delta_j + \gamma_{ij}$, $\underline{\theta}_j = (\theta_{1j}, \theta_{2j}, \ldots, \theta_{pj})'$, $\Psi = (\underline{\theta}_1, \underline{\theta}_2, \ldots, \underline{\theta}_q)$ and $W_{ijk} = W_{ijk}(\Psi) = Y_{ijk} - \theta_{ij}$. Note that W_{ijk} is the (ijk)th residual corresponding to the (ij)th treatment combination on the kth patient. To incorporate possible missing data, let $W_{ijk}^*(\Psi) = I_{ijk} W_{ijk}(\Psi)$.

Kloke, McKean, and Rashid (2009) extended the usual rank-based estimates of Jaeckel (1972) to general mixed models with covariates, including Model (2.1), for general scores functions. In this paper, we consider Wilcoxon scores only. For the Wilcoxon scores, using the missing data indicator I_{ijk}, Jaeckel's objective function is given by

$$D_{JR}(\Psi) = \sum_{j=1}^{q} \sum_{k=1}^{n_j} \sum_{i=1}^{m_{jk}} \sqrt{12} \left[\frac{W_{ijk}^*(\Psi)}{n+1} - \frac{1}{2} \right] W_{ijk}^*(\Psi), \tag{3.5}$$

where the subscript JR denotes the joint ranking of all of the residuals and n denotes the total sample size; i.e., $n = \sum_{j=1}^{q} \sum_{k=1}^{n_j} m_{jk}$. The dispersion function $D_{JR}(\Psi)$ is a continuous convex function of Ψ and the rank-based estimator of Ψ is the minimizing value which we denote by $\hat{\Psi}_{JR}$. Kloke et al. (2009) developed the asymptotic theory of these estimates and their corresponding tests of general linear hypotheses for the general mixed model. Note that unlike a rank transform, the involved rankings are ranks of residuals not ranks of responses; see McKean and Vidmar (1992) for discussion. The estimates of the scale parameters involved in the associated JR inference are adjusted for the random effects, so the JR inference is asymptotically distribution-free. The JR estimator, though, is not invariant to the random effects.

For multicenter clinical trials, Rashid, McKean, and Kloke (2012) proposed an R estimator which is invariant to the random center effect. This procedure forms a Jaeckel-dispersion function for each center by ranking the residuals within that center. Then the overall dispersion function is the sum of these center-specific dispersion functions. The resulting minimizer is called the MR estimator, where M denotes multiple. The asymptotic theory for the MR estimator is based on a fixed number of centers with the center sample sizes converging to infinity at relatively the same rate. In the situation of this paper, however, the number of repeated measures is fixed at p and for many practical cases p is small. Hence, asymptotic theory should cover the case where the number of subjects within each group gets large. For example, these are the assumptions underlying the theory for the JR estimator. Hence, the MR estimator is not appropriate in this case. As we show next, though, a similar type estimator is invariant to the random subject effects.

3.1 Rank-Based Invariant Estimates

Instead of doing a joint ranking of residuals, as in the JR estimator, we consider ranking the residuals within each subject. For the jk th subject, rank the residuals $W_{1jk}^{*}, \ldots, W_{m_{jk}jk}^{*}$ from 1 to m_{jk} and denote these ranks by $R_{1jk}^{*}, \ldots, R_{m_{jk}jk}^{*}$. Then form the dispersion functions

$$D_{jk}(\underline{\theta}_j) = \sqrt{12} \sum_{i=1}^{p} I_{ijk} \left[\frac{R_{ijk}^{*}}{m_{jk}+1} - \frac{1}{2} \right] W_{ijk}^{*}.$$

Note that the rank of the residual W_{ijk} is invariant to the random effect β_{jk}. Furthermore, because

$$\sum_{i=1}^{p} I_{ijk} \left[\frac{R_{ijk}^{*}}{m_{jk}+1} - \frac{1}{2} \right] = 0,$$

for all combinations j, k, we can write

$$\sqrt{12} \sum_{i=1}^{p} I_{ijk} \left[\frac{R[W^{*}ijk - \beta_{jk}]}{m_{jk}+1} - \frac{1}{2} \right] [W_{ijk}^{*} - \beta_{jk}] = D_{jk}(\underline{\theta}_j);$$

that is, the dispersion functions D_{jk} are invariant to the random effects for all combinations j, k. Further, for each patient this dispersion function is a nonnegative, continuous, and convex function of residuals W_{ijk}^*. An overall dispersion function based on all the patients in the q groups is defined as the sum of all $D_{jk}(\underline{\theta}_j)$ $(j = 1, 2, \ldots, q; k = 1, 2, \ldots, n_j)$:

$$D(\Psi) = \sum_{j=1}^{q} \sum_{k=1}^{m_{jk}} D_{jk}(\underline{\theta}_j). \tag{3.6}$$

Note that $D(\Psi)$ is also a non-negative, location-invariant, continuous, piece-wise linear and convex function of the W_{ijk}, because the sum of several convex functions is again a convex function. Furthermore, $D(\Psi)$ is invariant to all the random effects β_{jk}, $j = 1, \ldots, q$ and $k = 1, \ldots, n_j$.

Although, we have written $D(\Psi)$ as a function of Ψ, note that just as the dispersion function is invariant to the random effect it is also invariant to the group effects δ_j. But it is not invariant to the time effects α_i or the interaction effects γ_{ij}. Let Θ be the subvector of Ψ which includes these time and interaction effects. Then our rank-based-(random effect)-invariant (RBRI) estimator of Θ is

$$\widehat{\Theta} = ArgminD(\Theta). \tag{3.7}$$

Asymptotic theory for $\widehat{\Theta}$ is given in Section 4. This includes standard errors which can be used to determine asymptotically distribution free confidence intervals for the time and interaction effects.

Consider, next, the hypotheses of interaction (2.2). Denote the minimizing value of the full model, (2.1) by

$$D(FULL) = D(\widehat{\Theta}), \tag{3.8}$$

where FULL denotes the full model. Let $D(RED)$ denote the minimizing value for the fit of the reduced model; i.e., Model (2.1) constrained by $\gamma_{ij} \equiv 0$. Large values of the reduction (drop) in dispersion $RD = D(RED) - D(FULL)$ indicate H_A. Consider as a test statistic,

$$H = \frac{RD}{\hat{\tau}/2}, \tag{3.9}$$

where τ is the scale parameter (2.4). As shown in Section 5, this test statistic has an asymptotically χ^2-distribution under H_0 with $(q-1)(p-1)$ degrees of freedom. Hence, an asymptotic level α decision rule is to reject H_0 if

$$H \geq \chi_\alpha^2[(q-1)(p-1)], \tag{3.10}$$

where $\chi_\alpha^2[(q-1)(p-1)]$ is the upper α critical point of a χ^2-distribution with $(q-1)(p-1)$ degrees of freedom. This test is robust; see the discussion of Section 5.

For computation, the R algorithm discussed in Rashid et al. (2012) for the MR estimates can be used to obtain the RBIR estimates and the minimizing values of the dispersion function.

3.2 Special Case

In this subsection, we will examine how the R-estimator is related to the actual observations for a special case when $q = 1$ (only one grouping factor) and $p = 2$ (two levels of the repeated measures factor). Let $\theta_{11} = \mu_1$ and $\theta_{21} = \mu_2$. Further let $\mu_1 = \alpha$ and $\mu_2 = \alpha + \delta.$

Then the dispersion function for the kth subject is given by

$$D_k(\mu_1, \mu_2) = D_k(\alpha, \alpha + \delta) = D_k(0, \delta) = \sqrt{12} \sum_{i=1}^{2} R_{ik}^* W_{ik}$$

where $R_{1k}^* = \{\dfrac{R(W_{1k})}{2+1} - 1/2\}$ is the intra-patient standardized ranks corresponding to $W_{1k} = Y_{1k}$, $R_{2k}^* = \{\dfrac{R(W_{2k})}{2+1} - 1/2\}$ is the intra-patient standardized ranks corresponding to $W_{2k} = Y_{2k} - \delta$, $R(W_{1k})$ is the intra-patient rank of $W_{1k} = Y_{1k}$, and $R(W_{2k})$ is the intra-patient rank of $W_{2k} = Y_{ik} - \delta.$

An overall dispersion function for all the subjects is given by

$$D(\delta) = D_k(0, \delta) = \sqrt{12} \sum_{k=1}^{n} \sum_{i=1}^{2} R_{ik}^* W_{ik}.$$

We obtain the R estimate of δ by minimizing the above dispersion function with respect to δ.

Note the partial derivative of $D(\delta)$ is given by

$$-\sqrt{12} \sum_{k=1}^{n} R_{2k}^*.$$

Thus, the R-estimate of δ can be obtained by solving the following equation

$$-\sqrt{12} \sum_{k=1}^{n} R_{2k}^* = -\sqrt{12} \sum_{k=1}^{n} \left\{ \frac{R(W_{2k})}{2+1} - 1/2 \right\} = 0.$$

The above equation is equivalent to

$$\sum_{k=1}^{n} R(W_{2k}) = 3n/2$$

$$\sum_{k=1}^{n} R(W_{2k}) = \sum_{k=1}^{n} [1 + I_{(y_{2k} - \delta) > y_{1k}}] = 3n/2.$$

where I is an indicator function which takes a value 1 if $(y_{2k} - \delta) > y_{1k}$ and 0 if $(y_{2k} - \delta) < y_{1k}$.

The above equation reduces to

$$\sum_{k=1}^{n} I_{(y_{2k} - y_{1k} > \delta)}] = n/2.$$

Therefore $\hat{\delta}$ is median of n differences $y_{2k} - y_{1k}$, $k = 1, 2, \ldots, n$. In general (when $p > 2$), no closed form R-estimate of the contrasts $\theta_{ij} - \theta_{i'j}$ $(i \neq i' = 1, 2, \ldots, p)$ exists for fixed j.

A competing R estimate, in this simple case, is the Hodges-Lehmann estimator of δ, which is the median of the Walsh averages of the differences $y_{2k} - y_{1k}$. Asymptotic theory for the Hodges-Lehmann estimate requires symmetry of the distribution of these differences, which follows from exchangeability of the errors e_{ijk}. We illustrate this discussion with an example.

Table 3.1: Before and After EKG Readings of Patients

Subject	Predrug	Postdrug	Difference=Predrug - Postdrug
1	6	5	1
2	9	2	7
3	17	0	17
4	22	0	22
5	7	2	5
6	5	1	4
7	5	0	5
8	14	0	14
9	9	0	9
10	7	0	7
11	9	13	-4
12	51	0	51

Example 3.1.1 *Berry (1990) discusses a study involving twelve patients. These patients experienced frequent premature ventricular contractions (PVCs) and were administered a drug with antiarrhythmic properties. One minute EKG readings were taken before and after drug administration. The PVCs were counted on both recordings. The results are shown in Table 3.1.*

Let μ_1 be the median corresponding to the baseline and μ_2 be the median corresponding to the post-drug. Let $\delta = \mu_1 - \mu_2$. Our estimate of δ is the median of the differences which is 7. The LS estimate is the mean of the differences which is 11.5. This disparity in the estimates is due to the large outlier, the postdrug PVC response for the 12 *th* patient. If this patient's values are omitted, the median of the differences remains the same, but

the mean of the differences is 7.91. The LS estimator is highly influenced by outliers, while the rank-based estimate is robust.

4. Asymptotic Distribution Theory of the RBRI Estimators

As we outline in this section, with more details in the appendix, the asymptotic theory for the RBRI estimator follows from the asymptotic theory of the gradient of the dispersion function, much like the theory for rank-based estimators in general which is detailed, for example, in Chapters 3 and 5 of Hettmansperger and McKean (2011). So our discussion is brief.

As discussed at the end of Section 3, for the asymptotic theory, we assume that $n_j \to \infty$, where n_j is the number of patients in the jth group, $j = 1, \ldots, q$. We assume that these n_js grow at similar rates; that is, for some postive constants λ_j

$$\frac{n_j}{N} \to \lambda_j > 0, \quad j = 1, \ldots, q,$$

where $N = \sum_{j=1}^{q} n_j$ and $\sum_{j=1}^{q} \lambda_j = 1$.

Let $\underline{S}(\Theta) = [\underline{S}_1'(\underline{\theta}_1), \underline{S}_2'(\underline{\theta}_2), \ldots, \underline{S}_q'(\underline{\theta}_q)]'$ be the negative of the gradient of $D(\Theta)$. Note that for fixed Y_{ijk}, D_{jk} is piece-wise linear function of W_{ijk}; hence, the partial derivatives of $D_{jk}(\cdot)$ exist almost everywhere and are given by the coefficients of W_{ijk}. Therefore, $\underline{S}_j(\underline{\theta}_j) = [s_{1j}(\underline{\theta}_j), s_{2j}(\underline{\theta}_j), \ldots, s_{pj}(\underline{\theta}_j)]'$ where

$$s_{ij}(\underline{\theta}_j) = \sqrt{12} \sum_{k=1}^{n_j} I_{ijk} R_{ijk}^*, i = 1, 2, \ldots, p; j = 1, 2, \ldots, q.$$

Let Θ^0 be the true value of Θ. Without loss of generality, we assume that $\Theta^0 = 0$.

Theorem 4.1.2 *Under exchangeablity of errors within each patient and assuming* $n_j \to \infty$, $n_j / N \to \pi_j \varepsilon(0,1)$,

$$N^{-1/2} \underline{S}(0) \text{ is } MVN[0, \Sigma] \tag{4.11}$$

as $N \to \infty$, where MVN stands for a multi-variate normal distribution,

$$\Sigma = diag[\pi_1 \Sigma_1^*, \pi_2 \Sigma_2^*, \ldots, \pi_q \Sigma_q^*], \tag{4.12}$$

$$\pi_j = \lim_{n_j \to \infty} (n_j / N) \, (0 < \pi_j < 1, j = 1, 2, \ldots, q),$$

$$\Sigma_j^* = \lim_{n_j \to \infty} (\sigma_{ii}^{(j)}),$$

$$\sigma_{ii'}^{(j)} = -\sum_{k=1}^{n_j} \{I_{ijk} I_{i'jk}\} / \{n_j (m_{jk} + 1)\} \text{ if } i \neq i',$$

$$\sigma_{ii'}^{(j)} = \sum_{k=1}^{n_j} \{(m_{jk} - 1)I_{ijk}\} / \{n_j(m_{jk} + 1)\} \ \ if \ \ i = i'.$$

Theorem 4.2 *Under exchangeablity of errors within each patient and assuming* $n_j \to \infty, \ n_j / N \to \pi_j \varepsilon(0,1),$

$$P_0[Sup \|\sqrt{N}\{\hat{\Theta} - (\tau/N)\Sigma^{*-}\underline{S}(\underline{0})\}\| \ge \varepsilon : \underline{\hat{\Theta}} \in \Gamma] \to 0. \tag{4.13}$$

By expression (4.13) it follows that

$$N^{1/2}(\hat{\Theta}) \ and \ (N^{-1/2}\tau)(\Sigma^*)^{-}\underline{S}(0) \ are \ asymptotically \ equivalent,$$

where $(\Sigma^*)^{-}$ is a generalized inverse of Σ^* satisfying $(\Sigma^*)^{-}\Sigma^*(\Sigma^*)^{-} = (\Sigma^*)^{-}$ and τ is the scale parameter defined by (B.2); see, also, (2.4).

It follows from (4.11) and (4.13) that

$$\hat{\Theta} \ has \ an \ asymptotic \ MVN[\Theta^0, \tau^2 \frac{1}{N}\Sigma^{-}] \ distribution. \tag{4.14}$$

We assume that the rank of $\Sigma_j (j = 1, 2, \ldots, q)$ is $p-1$. In other words, the levels of the timing treatment (repeated measures factor) within the j th grouping treatment are connected. Without this assumption, we are unable to perform the test for either the interaction between the grouping treatment and the timing treatment or the equality of the levels of the timing treatment.

Note that Theorem 4.2 yields the asymptotic representation of the estimator $\hat{\Theta}$; i.e.,

$$\sqrt{N}\hat{\Theta} = (\tau/N)\Sigma^{*-}\underline{S}(\underline{0}) + o_p(1).$$

Based on this, as discussed in Chapter 3 of Hettmansperger and McKean (2011), it follows that the influence function of $\hat{\Theta}$ is bounded in response space. Hence, since the design matrix only contains dummy variables, the estimator $\hat{\Theta}$ is bounded in both response and factor space. Therefore, the estimator is robust.

5. Drop in Dispersion Tests

Drop in dispersion or reduction in dispersion tests are the rank-based analogues of the traditional log likelihood tests; see McKean and Hettmansperger (1976) or Chapter 3 of Hettmansperger and McKean (2011) for a review. In this subsection, we outline these tests based on the RBRI estimators for general linear hypotheses of the form

$$H_0 : C\Theta = 0 \ versus \ H_A : C\Theta \neq 0, \tag{5.15}$$

where C is a $r \times pq$ matrix of rank r such that $C\underline{1}_{pq} = \underline{0}$ where $\underline{1}_{pq}$ is a $pq \times 1$ vector of unity and $\underline{0}$ is an $r \times 1$ vector of zeroes.

Let $\hat{\Theta}_H$ minimize $D(\Theta)$ under the restriction $C\Theta = \underline{0}$. Then $\hat{\Theta}_H$ is called the reduced model RBRI of Θ. By using arguments similar to those given in Rashid (1996), it can be shown that for large N

$$\sqrt{N}\hat{\Theta}_H \approx \sqrt{N}[\hat{\Theta} - \Sigma^- C'\{C\Sigma^- C'\}^-]\hat{\Theta}, \tag{5.16}$$

where $\approx$ stands for asymptotically equivalent. It can be seen that the reduced model RBRI estimate, $\hat{\Theta}_H$, is asymptotically equivalent to a linear function of the full model RBRI estimator $\hat{\Theta}$; which is generally true for rank-based estimators; see Rashid (1996) or, for a review, Chapter 3 of Hettmansperger and McKean (2011).

Suppose $C\Theta = \underline{0}$. Let Θ_H^0 be the true value of Θ under the hypothesis. It follows from (4.14) and (5.16) that for large N, that

$$\sqrt{N}[\hat{\Theta}_H - \Theta^0] \ \textit{follows a MVN}[\underline{0}, \tau^2 B\Sigma^- B'], \tag{5.17}$$

$$B = I_{pq \times pq} - \Sigma^- C'[C\Sigma^- C']^- C$$

and Σ^- is a generalized inverse of Σ satisfying

$$\Sigma^- = \Sigma^- \Sigma \Sigma^-.$$

Theorem 5.1. 3*Under exchangeablity of errors within each patient and assuming* $n_j \to \infty$, $n_j / N \to \pi_j \varepsilon(0,1)$,

$$D^*(\hat{\tau}) = 2[D(\hat{\Theta}_H) - D(\hat{\Theta})] / \hat{\tau} \ \textit{follows} \ \chi_r^2, \tag{5.18}$$

as $N \to \infty$ and where χ_r^2 stands for a chi-square distribution with r degrees of freedom.

The drop-in-dispersion test for general linear hypotheses depends on the R-estimates in the same way that the classical normal theory F test depends on the generalized least squares estimates. It is also analogous to $-2log\Lambda$ in maximum likelihood procedure and has a similar interpretation. Note that the D^* test is a mixed procedure involving both the ranks and the residuals. As discussed on page 197 of Hettmansperger and McKean (2011), however, the influence function of the drop in dispersion test statistic is bounded in response space and for this repeated measures design also in factor space too. Hence, the drop in dispersion test statistic is robust. On the other hand, the maximum likelihood procedure is not robust. It is quite sensitive to outliers as the example in Section 7 illustrates.

6. Estimation of τ

In this section, we briefly sketch another consistent estimator of the scale parameter τ. For our discussion, we can assume without loss of generality that the true regression coefficients θ_{ik} are 0.

Denote the residuals for the k th subjects of the j th group by

$$r_{ijk} = Y_{ijk} - \hat{\theta}_{ij}, \quad i = 1, \ldots, m_{jk}.$$

Note that we can write the differences as these residuals as

$$
\begin{aligned}
r_{i'jk} - r_{ijk} &= Y_{i'jk} - Y_{ijk} - (\hat{\theta}_{i'j} - \hat{\theta}_{ij}) \\
&= [e_{i'jk} - e_{ijk}] - [\hat{\theta}_{i'j} - \hat{\theta}_{ij}].
\end{aligned}
\tag{6.19}
$$

Each bracketed term is invariant to the random effects, so the differences of these residuals are invariant to the random effects. Furthermore, the pdf of difference of the random errors in the first bracket is symmetric about 0.

Form the set of these sets of these residuals; i.e,, the set

$$
r_{ii'jk}^* = r_{i'jk} - r_{ijk}, \quad i \neq i'; i, i' = 1, \ldots, m_{jk}; k = 1, \ldots, n_j : j = 1, \ldots, q.
\tag{6.20}
$$

Let M denote the number of such residuals and, for convenience, renumber them as 1 to M. Next form the Walsh averages of these residuals, i.e., $\{w_{st} = (r_s^* + r_t^*)/2$ for $s \leq t$. Denote the the number of Walsh averages by $M^* = M(M+1)/2$. Let $\{w_{(1)}^* < w_{(2)}^* < \cdots w_{(M^*)}^*\}$ denote the set of ordered Walsh averages. McKean and Hettmansperger (1976) proposed estimating τ by a standardized length of a confidence interval for the intercept given by

$$
\hat{\tau}_{HM} = \frac{\sqrt{M}[r_{(M^*-c)}^* - r_{(c+1)}^*]}{2 z_{\alpha/2}},
\tag{6.21}
$$

where

$$
c = \frac{M(M+1)}{4} - z_{\alpha/2}\sqrt{\frac{M(M+1)(2M+1)}{24}} - \frac{1}{2}
$$

and $0 < \alpha < 1$. The generally recommended value of α is 0.05. For a fixed effects linear model with symmetric errors and residuals based on a Wilcoxon fit, $\hat{\tau}_{MH}$ is a consistent estimator of τ. In our situation, the errors are symmetrically distributed. but due to the random effects there are dependencies of the random errors within a subject.
One way of avoiding this is to first consider the case of no missing data. Then consider the $\binom{p}{2}$ sets of residuals, one for each pairing; that is, for each pair $i, i' = 1, \ldots, p; i < i'$ form the set

$$
r_{i'jk}^* = r_{i'jk} - r_{ijk}, \quad k = 1, \ldots, n_j : j+1, \ldots, q.
\tag{6.22}
$$

Table 7.2:   True Cell Medians for Times and Their Contrasts

	Cell Medians				Contrasts		
Group	1	2	3	4	$\alpha_2 - \alpha_1$	$\alpha_3 - \alpha_1$	$\alpha_4 - \alpha_1$
Placebo	11.2	11.3	10.9	11.4	0.1	−0.3	0.2
Group	10.4	11.3	11.6	11.5	0.9	1.2	1.1

For each of these sets obtain the McKean-Hettmansperger consistent estimator of τ. Then as a final estimator take the average of these estimators. Since each of the $\binom{p}{2}$ estimators is consistent and there are only a finite number of them, this average is

consistent also. This was the estimator used in the example of Section 7. In practice, as long as the missing is within practical bounds, this estimator is consistent.

7. Example

One of us consulted on a data set that would serve as an example but due to confidentiality matters it cannot be used. Instead, we have simulated a data set which is similar to real data set in most aspects. The design for the real data set was a repeated measures design with two groups (Placebo and Treated) and the primary outcome (a lipid) measured at 4 time periods on a total of 81 patients (22 in the Placebo arm and 59 in the Treated arm). There were many outliers in the real data set. The interaction between the groups and times was the effect of interest.

For our simulated data set, we used the same design, including sample sizes and the number of repeated measures. We set the cell medians (means) at the values in Table 7.2, which are close to the estimated cell medians of the real data set. This table also shows the true values of the contrasts $\alpha_i - \alpha_1$, for $i = 2, 3, 4$. Hence, there are time and interaction effects.

We generated data with a high intraclass correlation coefficient. The random errors were simulated from the Laplace distribution with variance set at 1.30, while the random effects are generated normal variates with variance set at 16 also. So the intraclass correlation coefficient is 0.924. In the cell with Group at level 2 and Time at level 1 two points were contaminated.

Using the algorithm discussed in Kloke et al. (2009), we computed the JR-fit for the vector of parameters Ψ. Figure 7.1 show the q–q and Interaction Plots based on the JR Fit. The q–q clearly shows the two contaminated data points. Actually, except for these two points, the rest of the data are good. The profile plots clearly indicate interaction between the group and time. The top profile is that of the treated group, while the bottom profile is based on the fit for placebo group. The interaction plot detected the drop in response between times 2 and 3 for the placebo group.
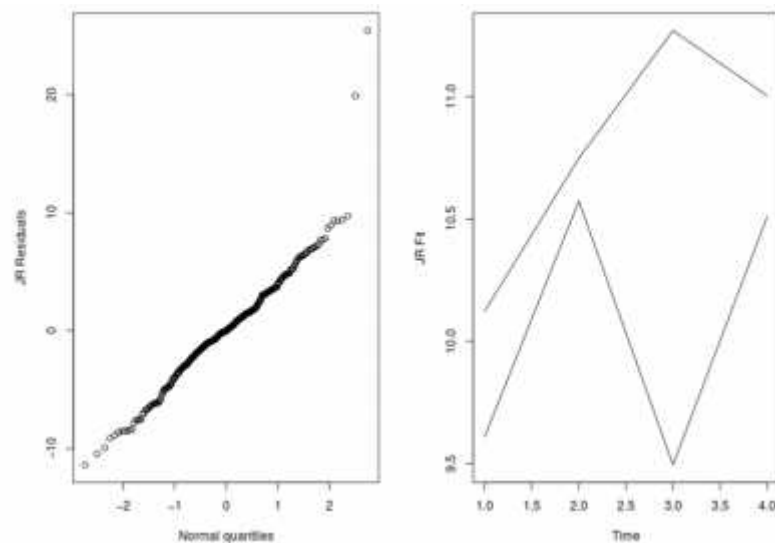
Pak.j.stat.oper.res. Vol.VIII No.3 2012 pp719-735

731

Figure 7.1: *q–q* and Interaction Plots of the JR Fit. The top and bottom profiles are the profiles for the treated and placebo groups, respectively.

Using the MR algorithm, as discussed in the last paragraph of Section 3.1, we obtained the RBRI fit of the vector of parameters Θ. Based on this fit, the McKean-Hettmansperger estimate of τ is 1.479, as discussed in the last paragraph of Section 6. Table 7.3 displays these time effects estimates and their standard errors. For comparison, we also included the JR estimates and their standard errors. The estimates and standard errors for the two methods are similar. They both detect the patterns in the table of the true contrasts. The RBRI seems to be a little more precise.

Table 7.3: **Estimates of Contrasts and Standard Errors for the time Effects for the RBRI and JR Procedures.**

Effect	RBRI Procedure			JR Procedure		
	$\alpha_2 - \alpha_1$	$\alpha_3 - \alpha_1$	$\alpha_4 - \alpha_1$	$\alpha_2 - \alpha_1$	$\alpha_3 - \alpha_1$	$\alpha_4 - \alpha_1$
Placebo	.502 (.355)	−.194 (.355)	.682 (.355)	.965 (.852)	−.114 (.852)	.923 (.852)
Treated	.861 (.253)	1.075 (.253)	.964 (.253)	.626 (.322)	1.15 (.322)	.881 (.322)

Table 7.4: **Estimates of Contrasts and Standard Errors for the time Effects for the RBRI and JR Procedures.**

Results on Tests for Interaction			
Analysis	*df*	Test Statistic	*p*-value
RBRI	3	14.446	0.0024
JR	3	8.325	0.0398
MLE	3	1.682	0.1716

For the test of the hypotheses of interaction (2.2), Table 7.4 displays the test statistics and their p-values in the next table. Results are given for the RBRI and JR-analyses and also the maximum likelihood analysis (assuming normal error distributions). The mle computation was performed by the R function lme.

Both robust tests (RBRI and JR) found interaction to be highly significant. They easily detected the pattern in the table of population medians. On the other hand, the mle analysis was impaired by the two outliers. Practical interpretations would differ between the robust and traditional (mle) analyses. In most applications, with this p-value, the traditional analysis would accept no interaction and concentrate on the main effects. On the other hand, the robust analyses would turn to the estimation of contrasts similar to those in Table 7.3.

8. Conclusion

When the response variable is continuous, the assumption of multivariate normal is not always reasonable. In addition, in other situations involving a continuous variable, the actual distribution may be unknown. Thus, the use of standard parametric procedures is patient to criticism regarding both optimality and validity.

In this article, we have developed rank based inferences for unbalanced repeated measures models with incomplete observations. These inferences are robust alternatives to parametric analyses which are invariant to the random effect due to patients. Note that nonparametric methods may also be of interest as a means of confirming the results of a parametric analysis or in assessing the sensitivity of the results to the assumptions of the selected parametric model.

Appendix

Proof of Theorem 4.1

Let $\underline{\theta}_j^0$ be the true value of $\underline{\theta}_j$. Without loss of generality, we assume that $\underline{\theta}_j^0 = \underline{0}$. Let $R_{ijk} = R(W_{ijk})$. For fixed j and k, under exchangeablity of errors within each patient, R_{ijk} are uniformly distributed over the integers from 1 through m_{jk}.

By using arguments similar to those given in Rashid (2000), it can be shown that

$$n_j^{-1/2} \underline{S}_j(\underline{0}) \text{ is } MVN[\underline{0}, \Sigma_j^*].$$

where

$$\Sigma_j^* = \lim_{n_j \to \infty} (\sigma_{ii'}^{(j)})$$

$$\sigma_{ii'}^{(j)} = -\sum_{j=1}^{n_j} \{I_{ijk} I_{i'jk}\} / \{n_j(m_{jk} + 1)\} \text{ if } i \neq i'$$

and

Pak.j.stat.oper.res. Vol.VIII No.3 2012 pp719-735

733

$$\sigma_{ii}^{(j)} = \sum_{k=1}^{n_j}\{(m_{jk}-1)I_{ijk}\}/\{n_j(m_{jk}+1)\}\ if\ \ i=i^{'}.$$

The covariance matrix of $N^{-1/2}\underline{S}_j(\underline{0})$ is given by

$$Cov_0[N^{-1/2}\underline{S}_j(\underline{0})] = (n_j/N)\Sigma_j$$

where $\Sigma_j = (\sigma_{ii^{'}}^{(j)})$.

Let

$$\lim_{N\to\infty}N^{-1/2}Cov[\underline{S}_1(\underline{0}),\underline{S}_2(\underline{0}),\dots,\underline{S}_q(\underline{0})]^{'} = diag[\pi_1\Sigma_1^*,\dots,\pi_j\Sigma_j^*,\dots,\pi_q\Sigma_q^*] = \Sigma$$

where $\pi_j = \lim_{n_j\to\infty}(n_j/N)$.

Therefore,

$$N^{-1/2}\underline{S}(\underline{0})\ is\ MVN[0,\Sigma], \tag{8.23}$$

as $N\to\infty$.

Proof of Theorem 4.2

Following Rashid (1996), it can be shown that $D(\Theta)$ can be approximated by a quadratic function of $\underline{\theta}$:

$$Q(\Theta) = D(\underline{0})-\Theta^{'}\underline{S}(\underline{0})+\{n/(2\tau)\}\Theta^{'}\Sigma\Theta. \tag{8.24}$$

Note that $\tilde{\Theta} = (\tau/n)\Sigma^{-}\underline{S}(\underline{0})$ minimizes $Q(\Theta)$, where Σ^{-} is a generalized inverse of Σ (defined in (**Error! Reference source not found.**) satisfying $\Sigma^{-}\Sigma\Sigma^{-}=\Sigma^{-}$.

Hence following Jaeckel (1972) we have for given $\varepsilon > 0$

$$P_0[Sup\,\|\sqrt{n}\{\hat{\Theta}-(\tau/n)A^{-}\underline{S}(\underline{0})\}\|\ge\varepsilon:\hat{\Theta}\in\Gamma] = 0, \tag{8.25}$$

where $\Gamma = [\Theta:D(\Theta)=a\ minimum]$. Hence, (4.12) follows from (8.25).

Proof of Theorem 5.1

Suppose $\tilde{\Theta}$ minimizes $Q(\Theta)$ and $\tilde{\Theta}_H$ minimizes $Q(\Theta)$ under the restriction $A\Theta=\underline{0}$. By using arguments as those given in Rashid(1996), it can be shown that the asymptotic behavior of $D(\hat{\Theta})-D(\hat{\Theta}_H)$ can be determined by that of $Q(\tilde{\Theta}_H)-Q(\tilde{\Theta})$. Hence

$$D(\hat{\Theta}_H)-D(\hat{\Theta})\approx Q(\tilde{\Theta}_H)-Q(\tilde{\Theta}).$$

It can be shown that,

$$Q(\tilde{\Theta}_H)-Q(\tilde{\Theta}) = \frac{1}{N}\underline{S}^{'}(\Theta^0)\Sigma^{-}A^{'}[A\Sigma^{-}A^{'}]^{-}A\Sigma^{-}\underline{S}(\Theta^0).$$

Therefore

$$2[D(\hat{\Theta}_H) - D(\hat{\Theta})] / \tau \approx \frac{1}{N} \underline{S}'(\Theta^0) \Sigma^- A' [A\Sigma^- A']^- A\Sigma^- \underline{S}(\Theta^0) + o_p(1).$$

But by (A.1), under the null hypothesis,

$$\frac{1}{N} \underline{S}'(\Theta^0) \Sigma^- A' [A\Sigma^- A']^- A\Sigma^- \underline{S}(\Theta^0)$$

asymptotically follows a chi-square distribution with r degrees of freedom.

Therefore, $2[D(\hat{\Theta}_H) - D(\hat{\Theta})] / \tau$ follows a chi-square distribution with r degrees of freedom under the null hypothesis. Let $\hat{\tau}$ be a consistent estimate of τ. Therefore, by Slutsky's Theorem, $D^*(\hat{\tau})$ follows a chi-square distribution with r degrees of freedom.

References

1. Anderson, R.L. and Bancroft, T.A. (1952). *Statistical Theory in Research*. McGraw-Hill, New York, p 250.

2. Berry, D.A. (1990), Basic principles in designing and analyzing clinical studies, p 35, *Statistical Methodology in the Pharmaceutical Sciences*, Editor D.A. Berry, New York: Marcel Dekker, Inc.

3. Hettmansperger, T. P. and McKean J. W. (2011). *Robust Nonparametric Statistical Methods, 2nd Ed.*, Boca Raton, FL: Chapman-Hall.

4. Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of residuals. *Annals of Mathematical Statistics*, 43, 1449 - 1458.

5. Kloke, J., McKean, J. W. and Rashid, M. M. (2009), Rank-Based Estimation and Associated Inferences for Linear Models with Cluster Correlated Errors, *Journal of the American Statistical Association*, 104, 384-390.

6. McKean, J. W. and Hettmansperger, T. P. (1976), Tests of Hypotheses of the General Linear Model Based on Ranks, *Communications in Statistics A5*, 693-709.

7. McKean, J. W. and Vidmar, T. J. (1994), A Comparison of Two Rank Based Methods for the Analysis of Linear Models, *The American Statistician*, 48, 220-229.

8. Rashid, M. M. (1996). Inference based on ranks for two-way models with a grouping factor and a repeated measures factor. *Journal of Nonparametric Statistics*, 6, 27 - 42.

9. Rashid, M. M. (2000). Comparisons of rank-based methods and normal theory methods for unbalanced one-way repeated measures data. *Journal of the Italian Statistical Society*, 9, 199 - 218.

10. Rashid, M., McKean, J. W. and Kloke, J., D. (2012), R Estimates and associated inferences for mixed models with covariates in a multi-center clinical trial, *Statistics in Biopharmaceutical Research*, In press.

11. Wei, L. J. and Lachin, J. M. (1985). Combining dependent tests with incomplete measurements. *Biometrika*, 72, 359-364.

Pak.j.stat.oper.res. Vol.VIII No.3 2012 pp719-735

735