Rank-based multiple test procedures and simultaneous confidence intervals

Frank Konietschke

Department of Medical Statistics
University of Göttingen
Humboldtallee 32
37073 Göttingen, Germany
e-mail: fkoniet@gwdg.de

Ludwig A. Hothorn

Institute of Biostatistics
Leibniz University Hannover
Herrenhaeuser Strasse 2
30419 Hannover, Germany
e-mail: hothorn@biostat.uni-hannover.de

and

Edgar Brunner

Department of Medical Statistics
University of Göttingen
Humboldtallee 32
37073 Göttingen, Germany
e-mail: brunner@ams.med.uni-goettingen.de

Abstract: We study simultaneous rank procedures for unbalanced designs with independent observations. The hypotheses are formulated in terms of purely nonparametric treatment effects. In this context, we derive rank-based multiple contrast test procedures and simultaneous confidence intervals which take the correlation between the test statistics into account. Hereby, the individual test decisions and the simultaneous confidence intervals are compatible. This means, whenever an individual hypothesis has been rejected by the multiple contrast test, the corresponding simultaneous confidence interval does not include the null, i.e. the hypothetical value of no treatment effect. The procedures allow for testing arbitrary purely nonparametric multiple linear hypotheses (e.g. many-to-one, all-pairs, changepoint, or even average comparisons). We do not assume homogeneous variances of the data; in particular, the distributions can have different shapes even under the null hypothesis. Thus, a solution to the multiple nonparametric Behrens-Fisher problem is presented in this unified framework.

Keywords and phrases: Rank statistics, multiple comparisons, nonparametric Behrens-Fisher problem.

Contents

1. Introduction

In many experiments more than two treatment groups are involved. Hereby the global null hypothesis, i.e. no impact of the factor "treatment" on the response, is often not the main question. In statistical practice, however, the traditional method for making inferences on the effects of interest is achieved in three steps: First the global null hypothesis is tested by an appropriate procedure (e.g. ANOVA). If the global null hypothesis is rejected, multiple comparisons are commonly used to test the different sub-hypotheses. In the last step of the analysis, confidence intervals for the effects of interest are computed.

Although stepwise procedures using different approaches on the same data are pretty common in practice, they may have the undesirable property that the global null hypothesis may be rejected, but none of the individual hypotheses and vice versa. This means, the global test procedure and the multiple testing procedure may be non-consonant to each other (Gabriel 1969 [14]). Further the confidence intervals may include the null, i.e. the value of no treatment effect, even if the corresponding individual null hypotheses have been rejected. This means, the individual test decisions and the corresponding confidence intervals may be *incompatible* (Bretz, Genz, and Hothorn 2001) [3]. In randomized clinical trials, the computation of *compatible* simultaneous confidence intervals (SCI), i.e. confidence intervals which always lead to the same test decisions as the multiple comparisons, is consequently required by regulatory authorities: *"Estimates of treatment effects should be accompanied by confidence intervals, whenever possible. . ."* (ICH E9 Guideline 1998, chap. 5.5, p. 25 [24]). It is well known that the classical Bonferroni adjustment can be used to perform

multiple comparisons as well as for the computation of compatible SCI. This approach, however, has a low power, particulary when the test statistics are not independent. In practice, the computation of correlation accounted compatible SCI is often neglected, because no adequate procedures exist. Thus, there is a desirable demand for statistical procedures, which can be used for (i) testing the global null hypothesis, (ii) performing multiple comparisons, and (iii) for the computation of compatible SCI by taking the correlation between the test statistics into account. Such procedures are particularly of practical importance.

In recent years, parametric *multiple contrast test procedures* (MCTP) with accompanying compatible SCI for linear contrasts in terms of the expectations of homoscedastic normal samples were derived by Mukerjee, Robertson, and Wright (1987) [28] and Bretz, Genz, and Hothorn (2001) [3]. These MCTP take the correlation between the test statistics into account. Hereby the procedures by Bretz et al. (2001) [3] can be used for testing arbitrary contrasts, e.g. many-to-one, all-pairs, average, or even changepoint comparisons. Thus, MCTP provide an extensive tool for the computation of compatible SCI. They are calculated by establishing the exact joint multivariate t - distribution of different test statistics with correlation matrix $\mathbf{R}$ for testing individual hypotheses and control the familywise type I error rate (FWER) in the strong sense (Hochberg and Tamhane 1987 [22]). The results by Bretz et al. (2001) [3] were extended to general linear models by Hothorn, Bretz, and Westfall (2008) [23] and to heteroscedastic models by Hasler and Hothorn (2008) [20] and Herberich, Sikorski, and Hothorn (2010) [21]. For a comprehensive overview of existing parametric methods we refer to Bretz, Hothorn, and Westfall (2010) [4]. We note that the parametric SCI use the critical values from the extreme tail portion of the multivariate t-distribution, which is the portion most sensitive to nonnormality (Gao et al. 2008 [17]). Therefore, when the normality assumption is violated, the problem of robustness will be more serious for SCI compared to individual intervals. In practice, however, skewed data, or even ordered categorical data occur in a natural way. Thus, there is a desirable demand for nonparametric MCTP and compatible SCI, particularly in the case of ordinal data and discontinuous distributions.

Munzel and Hothorn (2001) [30], Wolfsegger and Jaki (2006) [40] and Ryu (2009) [35] derive nonparametric SCI for the treatment effects $p_{ij} = P(X_i < X_j) + 1/2P(X_i = X_j)$, where X_i and X_j are independently distributed. In the literature, p_{ij} is known as relative effect (Brunner and Puri 2001 [10]; Gao et al. 2008 [17]), or, ordinal effect size measure in case of ordered categorical data (Ryu and Agresti 2008 [36]; Ryu 2009 [35]). We note that p_{ij} is an intransitive measure, i.e. it may occur $p_{12} \leq p_{23} \leq p_{31}$ (Brown and Hettmansperger 2002 [5]). Therefore, the use of pairwise defined relative effects in multiple comparisons may result in paradox statements in terms of Efron's dice (see, e.g. Gardner 1974 [18]; Thangavelu and Brunner 2006 [39]) and should be avoided.

The purpose of this article is to propose rank-based MCTP and compatible SCI for transitive relative effects in unbalanced one-way designs with independent observations, a fixed number of levels, and arbitrary contrasts. Their derivation requires the asymptotic joint distribution of rank statistics under ar-

bitrary alternatives. The covariance structure of the rank statistics turns out to result in a quite involved representation (Puri 1964 [32]; Rust and Fligner 1984 [33]), who consider only continuous distributions. Here, we will represent the structure of the covariance matrix in a simple and unified way allowing for discontinuous distributions and moreover we will provide procedures for multiple comparisons and SCI. For an user friendly application of the proposed methods, the freely available R-software package *nparcomp* was developed.

We further note that the MCTP proposed here do not assume homogeneous variances. The distributions can have different shapes even under the null hypothesis. Thus a solution to the *multiple nonparametric Behrens-Fisher problem* will be presented in a closed form. The new procedures are generalizations of the well-known multiple rank sum tests by Dunn (1964) [11], Steel (1959 [37], 1960 [38]), and Gao et al. (2008) [17], to heteroscedastic designs.

The paper is organized as follows. The nonparametric model, treatment effects, and hypotheses are presented in Section 2. A unified estimation approach for relative effects and the asymptotic multivariate normality of linear rank statistics in this general setup are derived in Section 3, where also consistent estimators of the parameters of this distribution are given. In Section 4, MCTP and compatible SCI are derived. A modification of the test statistics and approximations to their finite-sample distributions are presented in Section 5. As a practical example, a real data set is analyzed in Section 6. All procedures follow from a general asymptotic theory, which is presented in the Appendix.

Throughout the article we use the following notation. By $\mathbf{1}_a$ we denote the $a \times 1$ column vector of 1's and by $\mathbf{I}_a = diag\{1, \ldots, 1\}$ the a-dimensional unit matrix. Further $\otimes$ denotes the Kronecker product and $\oplus$ the Kronecker sum of two matrices, respectively. Finally $vec(\cdot)$ denotes the vector operator of a matrix, which stacks the columns of a matrix on top of each other.

2. Nonparametric model and hypotheses

Let X_{ik} be the kth (independent) replicate in the ith group among the total a groups. Let n_i be the sample size within the ith group, and $N = \sum_{i=1}^{a} n_i$ is the total sample size. Let $F_i(x) = P(X_{ik} < x) + 1/2 P(X_{ik} = x)$ denote the normalized version of the distribution function, which is the average of the left and right continuous version of the distribution function. In the context of nonparametric models, the normalized version of the distribution function $F_i(x)$ was first mentioned by Lévy (1925) [27]. Later on, it was used by Ruymgaart (1980) [34], Akritas, Arnold, and Brunner (1997) [1], Munzel (1999) [29], Gao et al. (2008) [17], among others, to derive asymptotic results for rank statistics including the case of ties in a unified way. We note that the F_i may be arbitrary distributions, with the exception of the trivial case of one-point distributions. The general model specifies only that

$$X_{ik} \sim F_i, \ k = 1, \ldots, n_i, \tag{2.1}$$

and does not require that the distributions are related in any parametric way; in particular it does not require homoscedasticity (Akritas et al. 1997 [1]). Factorial

designs can be described in this setup by putting a "factor pattern" on the index i in the same way as in the theory of linear models. The vector of the distributions is denoted by $\mathbf{F} = (F_1, \ldots, F_a)'$.

The general model (2.1) does not entail any parameters by which a difference between the distributions could be described. Therefore, the distribution functions $F_i(x)$ are used to define treatment effects by

$$p_j = \int G dF_j, \; j = 1, \ldots, a, \tag{2.2}$$

where $G = \sum_{i=1}^{a} w_i F_i$ denotes a mean distribution in its weighted $(w_i = n_i/N$; Akritas et al. 1997 [1]) or unweighted $(w_i = 1/a$; Brunner and Puri 2001 [10]; Gao et al. 2008 [17]) form. If $p_i < p_j$, then the values from F_i tend to be smaller than those from F_j. In case of $p_i = p_j$, none of the observations from F_i and F_j tend to be smaller or larger. Gao et al. (2008, p. 2576) [17] state that the unweighted relative effect *"has the advantage of not being influenced by the allocation of sample sizes in the data"*. Therefore, we will mainly concentrate on this effect throughout the paper. Paradox statements in terms of Efron's dice cannot occur, because each comparison in (2.2) refers to a fixed reference distribution G (Thangavelu and Brunner 2006 [39]). We will rewrite p_j as a linear combination

$$p_j = \sum_{i=1}^{a} w_i p_{ij}, \; j = 1, \ldots, a, \tag{2.3}$$

of $p_{ij} = \int F_i dF_j$. Let $\mathbf{p} = (p_1, \ldots, p_a)' = \int G d\mathbf{F}$ denote the vector of the relative treatment effects. The representation of p_j in (2.3) enables a simple representation of the covariance matrix of linear rank statistics under arbitrary alternatives. We further note that the weights w_i may be arbitrarily chosen under the constraint $\sum_{i=1}^{a} w_i = 1$. To have a reasonable interpretation of hypotheses in terms of these generalized relative effects, the weights should not depend on the sample sizes.

In the nonparametric setup discussed above, Akritas et al. (1997) [1] propose to formulate hypotheses by the distribution functions as $H_0^F : \mathbf{CF} = \mathbf{0}$, where $\mathbf{C}$ denotes an arbitrary contrast matrix, i.e. $\mathbf{C1}_a = \mathbf{0}$, and derive global test procedures for H_0^F. Gao et al. (2008) [17] consider the family of hypotheses

$$\mathbf{\Omega}^F = \{H_0^F : \mathbf{c}_\ell' \mathbf{F} = 0, \; \ell = 1, \ldots, q\}, \tag{2.4}$$

where $\mathbf{c}_\ell' = (c_{\ell 1}, \ldots, c_{\ell a})$ denotes an arbitrary contrast. They derive multiple test procedures for many-to-one and all-pairs comparisons. All test procedures for H_0^F, however, are limited to testing problems and cannot be used to construct confidence intervals for the underlying treatment effects $\delta_\ell = \mathbf{c}_\ell' \mathbf{p}$. In this paper, we consider the family of hypotheses

$$\mathbf{\Omega}^p = \{H_0^p : \mathbf{c}_\ell' \mathbf{p} = 0, \; \ell = 1, \ldots, q\}, \tag{2.5}$$

and we derive MCTP for $\mathbf{\Omega}^p$ and compatible SCI for the effects $\delta_\ell = \mathbf{c}_\ell' \mathbf{p}$.

We note that the hypothesis in the classical Behrens-Fisher model is contained in this general setup as a special case. This is easily seen from the fact that $p_{ij} = 1/2$ if F_i and F_j are both symmetric distributions with the same center of symmetry. The nonparametric hypothesis $H_0^F : \mathbf{CF} = \mathbf{0}$ is very general and implies $H_0^p : \mathbf{Cp} = \mathbf{0}$: $H_0^F : \mathbf{CF} = \mathbf{0} \Rightarrow H_0^p : \mathbf{Cp} = \mathbf{C} \int Gd\mathbf{F} = \int Gd(\mathbf{CF}) = \mathbf{0}$. In the special case of quite restrictive location models $F_i(x) = F(x - \mu_i), i = 1, \ldots, a$, the nonparametric and parametric hypotheses in terms of the location parameters μ_i are equivalent. For a detailed discussion of the hypotheses formulated above we refer to Akritas et al. (1997) [1] and Brunner and Munzel (2000) [8].

3. Asymptotic normality of linear rank statistics

Rank estimators of the quantities p_j defined in (2.3) are derived by replacing the unknown distribution functions $F_i(x)$ by their empirical counterparts $\widehat{F}_i(x) = n_i^{-1} \sum_{k=1}^{n_i} c(x - X_{ik})$, $i = 1, \ldots, a$, where $c(x) = 0, 1/2, 1$ according as $x < 0, x = 0, x > 0$, respectively (Ruymgaart 1980 [34]). Note that

$$\widehat{F}_i(X_{jk}) = \frac{1}{n_i} \left(R_{jk}^{(ij)} - R_{jk}^{(j)} \right), \ i \neq j, \tag{3.1}$$

where $R_{jk}^{(ij)}$ is the rank of observation X_{jk} among all $n_i + n_j$ observations in the combined sample (i, j), and $R_{jk}^{(j)}$ is the internal rank of X_{jk} among all n_j observations in sample j. If there are no ties, then $R_{jk}^{(ij)}$ is the usual rank of X_{jk}. In the presence of ties, $R_{jk}^{(ij)}$ is the midrank of X_{jk}. The quantities $n_i \widehat{F}_i(X_{jk})$ are also called placements (Orban and Wolfe 1982 [31]).

To estimate the relative effect p_{ij} used in (2.3), we use the normed placements given in (3.1) by

$$\widehat{p}_{ij} = \int \widehat{F}_i d\widehat{F}_j = \frac{1}{n_j} \sum_{k=1}^{n_j} \widehat{F}_i(X_{jk}) = \frac{1}{n_i} \left(\overline{R}_{j.}^{(ij)} - \frac{n_j + 1}{2} \right). \tag{3.2}$$

Here, $\overline{R}_{j.}^{(ij)} = n_j^{-1} \sum_{k=1}^{n_j} R_{jk}^{(ij)}$ is the mean of the ranks in sample j. Thus, one obtains an estimator of p_j in (2.3) as a linear combination of $\widehat{p}_{ij}$ in (3.2) by

$$\widehat{p}_j = \sum_{i=1}^{a} w_i \widehat{p}_{ij}. \tag{3.3}$$

This representation of the estimator provides an unified approach for both the usual ranks as well as the so-called *pseudo ranks* (Gao and Alvo 2005 [15], 2008 [16]; Gao et al. 2008 [17]), since $R_{jk} = \frac{1}{2} + N \sum_{i=1}^{a} w_i \widehat{F}_i(X_{jk})$. The usual ranks are obtained by letting $w_i = n_i/N$ and the pseudo ranks by letting $w_i = 1/a$. Let $\widehat{\mathbf{p}} = (\widehat{p}_1, \ldots, \widehat{p}_a)'$ denote the vector of the estimators $\widehat{p}_j$ in (3.3). We note that $\widehat{p}_j$ is an unbiased and consistent estimator of p_j, which follows from the unbiasedness and consistency of $\widehat{p}_{ij}$ (Brunner et al. 2002 [9]).

The asymptotic equivalence stated in the next theorem will facilitate the representation of the asymptotic covariance matrix in a simple and unified form.

Theorem 1. *Let $\widehat{\mathbf{F}} = (\widehat{F}_1, \ldots, \widehat{F}_a)'$ denote the vector of the empirical distributions, and let $\widehat{G} = \sum_{i=1}^{a} w_i \widehat{F}_i$ denote the empirical weighted mean distribution. Under the assumption that $N \to \infty$ such that $N/n_i \leq N_0 < \infty$, for $i = 1, \ldots, a$,*

$$\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}) \doteq \sqrt{N}\left(\int G d\widehat{\mathbf{F}} - \int \mathbf{F} d\widehat{G} + \mathbf{1}_a - 2\mathbf{p}\right), \tag{3.4}$$

where $\doteq$ denotes the asymptotic equivalence of two sequences of random variables.

For a proof of Theorem 1 we refer to Gao et al. (2008) [17], Lemma 2.1. Akritas et al. (1997) [1] and Gao et al. (2008) [17] state that the asymptotic covariance matrix of $\sqrt{N}\mathbf{C}(\widehat{\mathbf{p}} - \mathbf{p})$ leads to a simple representation under the hypothesis $H_0^F : \mathbf{C}\mathbf{F} = \mathbf{0}$. From (3.4) it follows immediately that

$$\sqrt{N}\mathbf{C}(\widehat{\mathbf{p}} - \mathbf{p}) \overset{H_0^F}{\doteq} \sqrt{N}\mathbf{C} \int G d\widehat{\mathbf{F}} = \sqrt{N}\mathbf{C}\overline{\mathbf{Y}}_{.},$$

where $\overline{\mathbf{Y}}_{.} = (\overline{Y}_{1.}, \ldots, \overline{Y}_{a.})'$ is a vector of independent random variables $\overline{Y}_{i.} = n_i^{-1} \sum_{k=1}^{n_i} G(X_{ik})$. Thus, under the hypothesis H_0^F, the asymptotic covariance matrix is given by $\mathbf{V}_N^F = \mathbf{C}\mathbf{\Sigma}^F\mathbf{C}'$, where $\mathbf{\Sigma}^F = \bigoplus_{i=1}^{a} N/n_i \sigma_i^2$ and $\sigma_i^2 = Var(G(X_{i1}))$. For the computation of consistent estimates for the unknown quantities σ_i^2, and a comprehensive overview of different global test procedures for H_0^F, we refer to Akritas et al. (1997) [1] and Brunner and Puri (2001) [10]. Multiple test procedures for $\mathbf{\Omega}^F$ defined in (2.4) are presented by Gao et al. (2008) [17].

The representation of the covariance matrix of (3.4) by means of $\widehat{G}$ leads to the same obstinate structure as the representation by Puri (1964) [32]. Therefore, we first rewrite the right-hand side of (3.4) and by some simple algebraic arguments we obtain for the jth component

$$\begin{aligned}
\sqrt{N}(\widehat{p}_j - p_j) &\doteq \sqrt{N}\left(\int G d\widehat{F}_j - \int F_j d\widehat{G} + 1 - 2p_j\right) \\
&= \sqrt{N}\left(\int \left(\sum_{i=1}^{a} w_i F_i\right) d\widehat{F}_j - \int F_j d\left(\sum_{i=1}^{a} w_i \widehat{F}_i\right) + 1 - 2p_j\right) \\
&= \sqrt{N}\sum_{i=1}^{a} w_i Z_{ij},
\end{aligned}$$

where $Z_{ij} = n_j^{-1} \sum_{k=1}^{n_j} Y_{ijk} - n_i^{-1} \sum_{k=1}^{n_i} Y_{jik} + 1 - 2p_{ij}$ is a sum of the independent random variables $Y_{ijk} = F_i(X_{jk})$ and $Y_{jik} = F_j(X_{ik})$, respectively. For a convenient vectorial representation, we define the random vector $\mathbf{Z} = vec[(Z_{ij})_{i,j=1,\ldots,a}]$. Let $\mathbf{W} = (w_1, \ldots, w_a) \otimes \mathbf{I}_a$ denote the known matrix of weights $w_1, \ldots, w_a$. Thus, an equivalent representation of the asymptotic equivalent sums defined in (3.4) is given by $\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p}) \doteq \sqrt{N}\mathbf{W}\mathbf{Z}$. The asymptotic

covariance matrix of $\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p})$ is given by

$$\mathbf{V}_N = Cov(\sqrt{N}\mathbf{WZ}) = \mathbf{W\Sigma W}', \tag{3.5}$$

where $\mathbf{\Sigma} = (\sigma_{ij,rs})_{i,j,r,s=1,\ldots,a} = Cov(\sqrt{N}\mathbf{Z})$. Here, $\sigma_{ij,ij} = Var(\sqrt{N}Z_{ij})$ denotes the variance of $\sqrt{N}Z_{ij}$, and $\sigma_{ij,rs} = Cov(\sqrt{N}Z_{ij}, \sqrt{N}Z_{rs})$ denotes the covariance of $\sqrt{N}Z_{ij}$ and $\sqrt{N}Z_{rs}$, respectively. For convenience define $\theta_{ij,rs} = Cov(Y_{ij1}, Y_{rs1})$. By independence of Y_{ijk} and $Y_{ijk'}, k \neq k'$,

$$\sigma_{ij,rs} = \begin{cases} 0, & i = j \vee r = s, & 0, & i \neq r, s \wedge j \neq r, s, \\ \frac{N}{n_i}\theta_{ji,si}, & i = r \wedge j \neq s, & \frac{N}{n_j}\theta_{ij,rj}, & i \neq r \wedge j = s, \\ -\frac{N}{n_i}\theta_{ji,ri}, & i = s \wedge j \neq r, & -\frac{N}{n_j}\theta_{ij,sj}, & i \neq s \wedge j = r, \\ \sigma_{ij,ij}, & i = r \wedge j = s, & \sigma_{ij,ji}, & i = s \wedge j = r, \end{cases} \tag{3.6}$$

where $\sigma_{ij,ij} = N/n_j\theta_{ij,ij} + N/n_i\theta_{ji,ji}$ and $\sigma_{ij,ji} = -\sigma_{ij,ij}$. Note that the representation of p_j as a linear combination of p_{ij} leads to the simple representation of the structure of $\mathbf{V}_N$ as given in (3.5). The asymptotic normality of the linear rank statistic $\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p})$ will be given in the next theorem.

Theorem 2. *Let $\mathbf{V}_N$ as given in (3.5). If $N \to \infty$ such that $\frac{N}{n_i} \leq N_0 < \infty$ and if $\mathbf{V}_N \to \mathbf{V}$ such that $rank(\mathbf{V}_N) = rank(\mathbf{V}) \; \forall N \geq M_0 < \infty$, then, the linear rank statistic $\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p})$ has, asymptotically, as $N \to \infty$, a multivariate normal distribution with expectation $\mathbf{0}$ and covariance matrix $\mathbf{V}_N$.*

A consistent estimator of $\mathbf{V}_N$ will be provided in the next section.

3.1. Estimation of the covariance matrix

Note that it is sufficient to derive consistent estimators of the variances $\theta_{ij,ij} = Var(Y_{ij1})$ and the covariances $\theta_{ij,rs} = Cov(Y_{ij1}, Y_{rs1})$ as given in (3.6). If the random variables Y_{ijk} were observable, then a natural estimator of the covariance $\theta_{ji,si}$, for example case 3 in (3.6), would be given by the empirical covariance

$$\widetilde{\theta}_{ji,si} = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (Y_{jik} - \overline{Y}_{ji\cdot})(Y_{sik} - \overline{Y}_{si\cdot}). \tag{3.7}$$

The random variables Y_{ijk}, however, are not observable, and, for the computation of an estimator, they must be replaced by observable random variables, which are "close enough" to the originals in an appropriate norm. To this end, let $\widehat{Y}_{ijk} = \widehat{F}_i(X_{jk})$ denote the normed placements as given in (3.1) and define the centered placements

$$D_{ijk} = \widehat{Y}_{ijk} - \frac{1}{n_j}\sum_{k=1}^{n_j}\widehat{Y}_{ijk} = \frac{1}{n_i}\left(R_{jk}^{(ij)} - R_{jk}^{(j)} - \overline{R}_{j\cdot}^{(ij)} + \frac{n_j + 1}{2} \right). \tag{3.8}$$

Then, an estimator of $\mathbf{V}_N = \mathbf{W}\boldsymbol{\Sigma}\mathbf{W}'$ is given by $\widehat{\mathbf{V}}_N = \mathbf{W}\widehat{\boldsymbol{\Sigma}}_N\mathbf{W}'$, where $\widehat{\boldsymbol{\Sigma}}_N$ denotes the matrix $\boldsymbol{\Sigma}$ with $\sigma_{ij,rs}$ being replaced by the estimators

$$
\widehat{\sigma}_{ij,rs} = \begin{cases}
0 & i = j \vee r = s, & 0 & i \neq r, s \wedge j \neq r, s, \\
\frac{N}{n_i}\widehat{\theta}_{ji,si}, & i = r \wedge j \neq s, & \frac{N}{n_j}\widehat{\theta}_{ij,rj}, & i \neq r \wedge j = s, \\
-\frac{N}{n_i}\widehat{\theta}_{ji,ri}, & i = s \wedge j \neq r, & -\frac{N}{n_j}\widehat{\theta}_{ij,sj}, & i \neq s \wedge j = r, \\
\widehat{\sigma}_{ij,ij}, & i = r \wedge j = s, & \widehat{\sigma}_{ij,ji}, & i = s \wedge j = r,
\end{cases} \tag{3.9}
$$

where $\widehat{\sigma}_{ij,ij} = N/n_j\widehat{\theta}_{ij,ij} + N/n_i\widehat{\theta}_{ji,ji}$ and $\widehat{\sigma}_{ij,ji} = -\widehat{\sigma}_{ij,ij}$. Here

$$
\widehat{\theta}_{ji,si} = \frac{1}{n_i - 1}\sum_{k=1}^{n_i} D_{jik}D_{sik}, j \neq s, \quad \widehat{\theta}_{ij,rj} = \frac{1}{n_j - 1}\sum_{k=1}^{n_j} D_{ijk}D_{rjk}, i \neq r
$$

$$
\widehat{\theta}_{ji,ji} = \frac{1}{n_i - 1}\sum_{k=1}^{n_i} D_{jik}^2, i \neq j, \qquad \widehat{\theta}_{ij,ij} = \frac{1}{n_j - 1}\sum_{k=1}^{n_j} D_{ijk}^2, i \neq j,
$$

denote the empirical variances of $\widehat{Y}_{ijk}$ and the empirical covariances of $\widehat{Y}_{ijk}$ and $\widehat{Y}_{rsk}$, respectively. Theorem A.4 shows that $\widehat{\mathbf{V}}_N = \mathbf{W}\widehat{\boldsymbol{\Sigma}}_N\mathbf{W}'$ is a consistent estimate for $\mathbf{V}_N$.

The asymptotic distribution of $\sqrt{N}(\widehat{\mathbf{p}} - \mathbf{p})$ and the estimator $\widehat{\mathbf{V}}_N$ can now be used for the derivation of MCTP for $\boldsymbol{\Omega}^p$ and compatible SCI for $\delta_\ell = \mathbf{c}_\ell'\mathbf{p}$.

4. Multiple contrast test procedures

In order to develop MCTP for the family of hypotheses $\boldsymbol{\Omega}^p$ defined in (2.5), we first need to derive the test statistics for each individual hypothesis $H_0^p : \mathbf{c}_\ell'\mathbf{p} = 0$. Define $\widehat{v}_{\ell\ell} = \mathbf{c}_\ell'\widehat{\mathbf{V}}_N\mathbf{c}_\ell$ and let $T_\ell^p = \sqrt{N}\mathbf{c}_\ell'(\widehat{\mathbf{p}} - \mathbf{p})/\sqrt{\widehat{v}_{\ell\ell}}$, $\ell = 1, \ldots, q$. By the asymptotic normality of $\sqrt{N}\mathbf{c}_\ell'(\widehat{\mathbf{p}} - \mathbf{p})$ and Slutsky's theorem, it follows that $T_\ell^p \xrightarrow{d} N(0, 1)$. The test statistics T_ℓ^p are collected in the vector

$$
\mathbf{T} = (T_1^p, \ldots, T_q^p)'. \tag{4.1}
$$

Note that the MCTP and compatible SCI are derived by establishing the asymptotic joint distribution of the test statistics $\mathbf{T}$ defined in (4.1). Although the exact marginal distributions of the test statistics T_ℓ^p depend on the sample sizes and are not identical for unbalanced designs with finite replications, the marginal distribution of T_ℓ^p is asymptotically standard normal, as $N \to \infty$. Further the asymptotic distribution of $\mathbf{T}$ is derived under arbitrary alternatives. Thus, it is completely specified under any configuration of the null hypotheses.

Corollary 1. *Define $v_{\ell\ell} = \mathbf{c}_\ell'\mathbf{V}_N\mathbf{c}_\ell$ and $v_{\ell m} = \mathbf{c}_\ell'\mathbf{V}_N\mathbf{c}_m$. Under the assumptions of Theorem 2, the statistic $\mathbf{T}$ follows a multivariate normal distribution with expectation $\mathbf{0}$ and correlation matrix $\mathbf{R} = (r_{\ell m})_{\ell,m=1,\ldots,q}$, where $r_{\ell m} = v_{\ell m}/\sqrt{v_{\ell\ell}v_{mm}}$, asymptotically.*

Proof. Let $\mathbf{C} = (\mathbf{c}'_\ell)_{\ell=1,\ldots,q}$ denote the contrast matrix obtained by the q single contrasts $\mathbf{c}'_\ell$. The proof follows by the asymptotic normality of $\sqrt{N}\mathbf{C}(\widehat{\mathbf{p}} - \mathbf{p})$ and Slutsky's theorem. $\qquad\square$

Corollary 1 particularly states that the type of contrast is incorporated in the correlation matrix of the vector of test statistics $\mathbf{T}$. For a strong control of the FWER, however, the limiting joint distribution of the test statistics under alternatives is not sufficient. In addition, it must be shown that the limiting distribution of the statistics is completely specified under arbitrary intersections of the null hypotheses, i.e. that the family of hypotheses $\mathbf{\Omega}^p$ and $\mathbf{T}$ constitutes a joint testing family.

Lemma 1. *The family of hypotheses $\mathbf{\Omega}^p$ and the corresponding test statistics $\mathbf{T}$ constitute a joint testing family asymptotically.*

So far, multiple comparison rank procedures compute the variances and covariances of the statistics under the global null hypothesis (Munzel and Hothorn 2001 [30]). Thus, the resulting test statistics do not constitute a joint testing family asymptotically, and do not provide a strong control of the FWER (Hochberg and Tamhane 1987, p. 249 [22]). Next we will derive a simultaneous test procedure (STP) from the joint testing family $\{\mathbf{\Omega}^p, \mathbf{T}\}$.

Let $z_{1-\alpha,2,\mathbf{R}}$ denote the two-sided $(1-\alpha)$-equicoordinate quantile of $N(\mathbf{0}, \mathbf{R})$, i.e.

$$P\left(\bigcap_{\ell=1}^{q} \left\{-z_{1-\alpha,2,\mathbf{R}} \leq X_\ell \leq z_{1-\alpha,2,\mathbf{R}}\right\}\right) = 1 - \alpha \qquad (4.2)$$

for $\mathbf{X} = (X_1, \ldots, X_q) \sim N(\mathbf{0}, \mathbf{R})$ (Bretz et al. 2001 [3]). We write $z_{1-\alpha,2,\mathbf{R}}$ to emphasize that it is the two-sided equicoordinate quantile; one-sided quantiles are written as $z_{1-\alpha,1,\mathbf{R}}$. For bivariate distributions, $z_{1-\alpha,2,\mathbf{R}}$ geometrically forms a cuboid having a square base. The quantiles become smaller with an increasing correlation (see Figure 1). For the numerical computation of $z_{1-\alpha,2,\mathbf{R}}$ we refer to Bretz et al. (2001) [3] and Genz and Bretz (2009) [19].
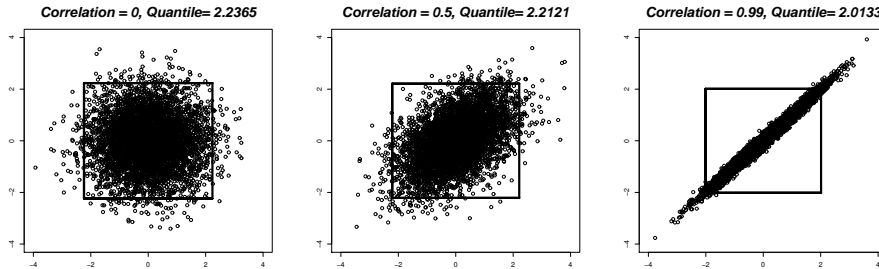

FIG 1. *Two-sided equicoordinate quantiles of different bivariate $N(\mathbf{0}, \mathbf{I}_2 + \rho(\mathbf{1}_2\mathbf{1}'_2 - \mathbf{I}_2))$ distributions with different correlations ρ: $\rho = 0$ (left), $\rho = 0.5$ (middle) and $\rho = 0.99$ (right).*

The asymptotic correlation matrix $\mathbf{R}$, however, is unknown and must be estimated. Let $\widehat{v}_{\ell\ell}$ and $\widehat{v}_{\ell m}$ denote the consistent estimators of $v_{\ell\ell}$ and $v_{\ell m}$ in Corollary 1 replacing $\mathbf{V}_N$ with $\widehat{\mathbf{V}}_N$ as given in Theorem A.4. Then, a consistent estimator of the correlation matrix $\mathbf{R}$ is given by $\widehat{\mathbf{R}} = (\widehat{r}_{\ell m})_{\ell,m=1,\ldots,q}$, where $\widehat{r}_{\ell m} = \widehat{v}_{\ell m}/\sqrt{\widehat{v}_{\ell\ell}\widehat{v}_{mm}}$. Thus, the set $\{\mathbf{\Omega}^p, \mathbf{T}, z_{1-\alpha,2,\widehat{\mathbf{R}}}\}$ of hypotheses, corresponding test statistics and one critical value for all individual hypotheses constitutes an asymptotic STP (Gabriel, 1969 [14]). The strong error control of the proposed method is shown in the next theorem.

Theorem 3. *As $N \to \infty$, the STP $\{\mathbf{\Omega}^p, \mathbf{T}, z_{1-\alpha,2,\widehat{\mathbf{R}}}\}$ controls the FWER in the strong sense.*

For large sample sizes, the individual hypothesis $H_0^p : \mathbf{c}_\ell'\mathbf{p} = 0$ will be rejected at a two-sided multiple level α, if $|T_\ell^{0.5}| \geq z_{1-\alpha,2,\widehat{\mathbf{R}}}$. Asymptotic $(1 - \alpha)$-simultaneous confidence intervals for the treatment effects $\delta_\ell = \mathbf{c}_\ell'\mathbf{p}$ are obtained from

$$\left[\mathbf{c}_\ell'\widehat{\mathbf{p}} - z_{1-\alpha,2,\widehat{\mathbf{R}}}\sqrt{\widehat{v}_{\ell\ell}/N} \; ; \; \mathbf{c}_\ell'\widehat{\mathbf{p}} + z_{1-\alpha,2,\widehat{\mathbf{R}}}\sqrt{\widehat{v}_{\ell\ell}/N}\right]. \tag{4.3}$$

Note that the test decision for $H_0^p : \mathbf{c}_\ell'\mathbf{p} = 0$ and the SCI defined in (4.3) are compatible by construction. This means, whenever an individual hypothesis is rejected, the corresponding confidence interval does not include the null. Further, for large sample sizes, the global null hypothesis $H_0^p : \mathbf{Cp} = \mathbf{0}$ will be rejected at a two-sided multiple level α, if $\max\{|T_1^{0.5}|,\ldots,|T_q^{0.5}|\} \geq z_{1-\alpha,2,\widehat{\mathbf{R}}}$. In practical applications it can be reasonable to consider the one-sided confidence regions

$$\left[-1 \; ; \; \mathbf{c}_\ell'\widehat{\mathbf{p}} + z_{1-\alpha,1,\widehat{\mathbf{R}}}\sqrt{\widehat{v}_{\ell\ell}/N}\right] \text{ or } \left[\mathbf{c}_\ell'\widehat{\mathbf{p}} - z_{1-\alpha,1,\widehat{\mathbf{R}}}\sqrt{\widehat{v}_{\ell\ell}/N} \; ; \; 1\right]. \tag{4.4}$$

For the special cases of trend alternatives and genetic models, compatible SCI based on pairwise rankings are provided by Konietschke and Hothorn (2012) [25] and Konietschke, Libiger, and Hothorn (2012) [26].

4.1. Range preserving confidence intervals

For the following considerations, we assume that the linear contrasts $\mathbf{c}_\ell' = (c_{\ell 1},\ldots,c_{\ell a})$ are normed, i.e. $|c_{\ell i}| \leq 1$. Note that the SCI defined in (4.3) may not be range preserving, i.e. the lower bounds can be smaller than -1 and the upper bounds can be larger than 1. For a detailed discussion about range preserving confidence intervals we refer the reader to Efron and Tibshirani (1993, Section 13.6) [12] and Brunner, Domhof and Langer (2002, Section 4.5, p. 60) [7]. Range preserving SCI for the treatment effects $\delta_\ell = \mathbf{c}_\ell'\mathbf{p}$ can be obtained by applying Cramér's multivariate Delta-Theorem. Let

$$g(x_\ell) = \frac{1}{2}\log\left(\frac{1+x_\ell}{1-x_\ell}\right) \text{ and } g^{-1}(y_\ell) = \frac{\exp(2y_\ell) - 1}{\exp(2y_\ell) + 1}, \; \ell = 1,\ldots,q, \tag{4.5}$$

denote the Fisher-transformation of x_ℓ and its inverse function, and let $\mathbf{g}(\mathbf{x}) = (g(x_1), \ldots, g(x_q))'$. Further let $\mathbf{d} = (\delta_1, \ldots, \delta_q)'$ denote the vector of the treatment effects $\delta_\ell = \mathbf{c}'_\ell \mathbf{p}$ and let $\widehat{\mathbf{d}} = (\widehat{\delta}_1, \ldots, \widehat{\delta}_q)'$ denote the vector of the consistent estimators $\widehat{\delta}_\ell = \mathbf{c}'_\ell \widehat{\mathbf{p}}$. Finally let $\boldsymbol{\Psi} = diag(1 - \delta_1^2, \ldots, 1 - \delta_q^2)^{-1}$ denote the Jacobian matrix of $\mathbf{g}(\mathbf{d})$.

Corollary 2. *Under the assumptions of Theorem 2, the vector $\sqrt{N}(\mathbf{g}(\widehat{\mathbf{d}}) - \mathbf{g}(\mathbf{d}))$ follows a multivariate $N(\mathbf{0}, \boldsymbol{\Gamma})$ distribution, where $\boldsymbol{\Gamma} = \boldsymbol{\Psi}\mathbf{V}_N\boldsymbol{\Psi}'$.*

Proof. The proof follows by Theorem 2 and by Cramér's multivariate Delta-Theorem. □

Define $\widehat{\boldsymbol{\Psi}} = diag(1 - \widehat{\delta}_1^2, \ldots, 1 - \widehat{\delta}_q^2)^{-1}$ and note that the estimator $\widehat{\boldsymbol{\Gamma}} = \widehat{\boldsymbol{\Psi}}\widehat{\mathbf{V}}_N\widehat{\boldsymbol{\Psi}}'$ is a consistent estimator of $\boldsymbol{\Gamma}$ defined in Corollary 2. To test the individual hypothesis $H_0^p : \mathbf{c}'_\ell \mathbf{p} = 0$, define the test statistic $\widetilde{T}_\ell^p = \sqrt{N}(g(\widehat{\delta}_\ell) - g(\delta_\ell))/\sqrt{\widehat{\gamma}_{\ell\ell}}$, where $\widehat{\gamma}_{\ell\ell}$ denotes the ℓth diagonal element of $\widehat{\boldsymbol{\Gamma}}$. The test statistics $\widetilde{T}_\ell^p$ are collected in the vector

$$\widetilde{\mathbf{T}} = (\widetilde{T}_1^p, \ldots, \widetilde{T}_q^p)'. \tag{4.6}$$

Under the assumptions of Corollary 1, the vector $\widetilde{\mathbf{T}}$ follows, asymptotically, as $N \to \infty$, a multivariate normal distribution with expectation $\mathbf{0}$ and correlation matrix $\mathbf{R}$. Since the Fisher-transformation is monotone, $\{\boldsymbol{\Omega}^p, \widetilde{\mathbf{T}}\}$ consitutes a joint testing family and $\{\boldsymbol{\Omega}^p, \widetilde{\mathbf{T}}, z_{1-\alpha,1,\widehat{\mathbf{R}}}\}$ controls the FWER in the strong sense (see Theorem 3). Thus, range preserving $(1 - \alpha)$-simultaneous confidence intervals for the treatment effects $\delta_\ell = \mathbf{c}'_\ell \mathbf{p}$ are obtained from $p^*_{\ell,L} = g^{-1}(p^g_{\ell,L})$ and $p^*_{\ell,U} = g^{-1}(p^g_{\ell,U})$, where

$$p^g_{\ell,L}, \; p^g_{\ell,U} = g(\widehat{\delta}_\ell) \mp z_{1-\alpha,2,\widehat{\mathbf{R}}}\sqrt{\widehat{\gamma}_{\ell\ell}/N}.$$

Since $g(x_\ell)$ and $g^{-1}(y_\ell)$ are both strictly monotone transformations, the range preserving SCI are compatible to the individual test decisions, by construction.

5. Small sample approximations and simulation results

The procedures considered in the previous section are valid for large sample sizes. The quality of the approximations by multivariate normal distributions of the proposed methods were investigated by simulation studies for different numbers of factor levels, sample sizes, and different kinds of contrasts. The simulations indicate that the convergence of $\mathbf{T}$ defined in (4.1) to its asymptotic multivariate normal distribution is rather slow. In general, the approximation is worse for a large number of factor levels and smaller sample sizes. Thus we also consider a small sample modification of this statistic. We adopt the Box-type approximation (Box 1954 [2]) proposed by Brunner, Dette and Munk (1997) [6] and Gao et al. (2008) [17] to approximate the distribution of $\mathbf{T}$ by a multivariate

$T(\nu, \mathbf{0}, \widehat{\mathbf{R}})$ distribution with ν degrees of freedom, expectation $\mathbf{0}$ and correlation matrix $\widehat{\mathbf{R}}$.

For each linear contrast $\mathbf{c}'_\ell = (c_{\ell 1}, \ldots, c_{\ell a}), \ell = 1, \ldots, q$, define the random variables $A_{\ell i k} = c_{\ell i}(G(X_{ik}) - w_i F_i(X_{ik})) - \sum_{s \neq i} c_{\ell s} w_s F_s(X_{ik})$. By reorganizing the asymptotic equivalent sums of random variables in (3.4), it is easily seen that $\sqrt{N} \mathbf{c}'_\ell(\widehat{\mathbf{p}} - \mathbf{p}) \doteq \sqrt{N} \sum_{i=1}^a n_i^{-1} \sum_{k=1}^{n_i} A_{\ell i k}$. By independence of $A_{\ell i k}$ and $A_{\ell i k'}$, $k \neq k'$, we obtain for the variance $Var(\sqrt{N} \sum_{i=1}^a n_i^{-1} \sum_{k=1}^{n_i} A_{\ell i k}) = N \sum_{i=1}^a \omega_{i\ell}^2/n_i$, where $\omega_{\ell i}^2 = Var(A_{\ell i 1})$. With the same arguments as in the proof of Theorem A.4, the unknown variances $\omega_{\ell i}^2$ can be consistently estimated by the empirical variances $\widehat{\omega}_{\ell i}^2 = (n_i - 1)^{-1} \sum_{k=1}^{n_i} (B_{\ell i k} - \overline{B}_{\ell i \cdot})^2$, where $B_{\ell i k} = c_{\ell i}(\widehat{G}(X_{ik}) - w_i \widehat{F}_i(X_{ik})) - \sum_{s \neq i} c_{\ell s} w_s \widehat{F}_s(X_{ik})$. Following Gao et al. (2008) [17], the distribution of $\mathbf{T}$ can be approximated by a multivariate $T(\nu, \mathbf{0}, \widehat{\mathbf{R}})$ distribution with $\nu = \max\{1, \min_{\ell=1,\ldots,q}\{\nu_1, \ldots, \nu_q\}\}$ degrees of freedom, where

$$\nu_\ell = \left(\sum_{i=1}^a \widehat{\omega}_{\ell i}^2/n_i\right)^2 \Big/ \sum_{i=1}^a \widehat{\omega}_{\ell i}^4/\left(n_i^2(n_i-1)\right). \tag{5.1}$$

The quality of the modifications of the MCTP $\mathbf{T}$ in (4.1) and $\widetilde{\mathbf{T}}$ in (4.6) to their finite-sample distributions by multivariate $T(\nu, \mathbf{0}, \widehat{\mathbf{R}})$ distributions were investigated for Dunnett-type (D), Tukey-type (T), Average-type (A), and change-point (C) comparisons in different one-way layouts with sample sizes:

- Design 1: $n_1 = 7$, $n_2 = 7$, $n_3 = 7$
- Design 2: $n_1 = 20$, $n_2 = 15$, $n_3 = 25$, $n_4 = 25$
- Design 3: $n_1 = 7$, $n_2 = 7$, $n_3 = 7$, $n_4 = 7$, $n_5 = 7$
- Design 4: $n_1 = 25$, $n_2 = 25$, $n_3 = 15$, $n_4 = 20$, $n_5 = 30$.

The corresponding contrast matrices for these four kinds of different contrasts are given in Section A.5.

The results reported here constitute a representative set from a much larger simulation study using R (www.r-project.org). All simulation results were obtained from 10,000 simulation runs. The equicoordinate quantiles were computed with the R-package *mvtnorm* (Genz and Bretz 2009 [19]). We also include the parametric counterpart of $\mathbf{T}$ proposed by Hasler and Hothorn (2008) [20] ($\mathbf{H}$) in the simulation study. This MCTP denotes $\mathbf{T}$ without ranking the data. In case of skewed distributions, $\mathbf{H}$ tends to be quite conservative in case of positive dependent test statistics (e.g. many-to-one comparisons), but very liberal when the test statistics are negatively dependent (average comparisons). The simulation results for the different designs are displayed in Table 1.

The simulation results indicate that the rank-based MCTP $\widetilde{\mathbf{T}}$ controls the FWER quite accurately, even in case of small sample sizes, all considered numbers of factor levels, and arbitrary contrasts for both normal and lognormal distributions (see Table 1). For other distributions, e.g. exponential or even ordered categorical data, the simulation results were quite similar and are not shown here. The MCTP $\mathbf{T}$ tends to be slightly liberal in case of extremely

Simulated FWER ($\alpha = 5\%$) of the MCTP **T**, $\widetilde{\textbf{T}}$, and **H** for different kinds of contrasts (**C**): Dunnett (D) (see A.1), Tukey (T) (see A.2), Average (A) (see A.3), and Changepoint (C) (see A.4).

	Design											
	1			2			3			4		
C	**T**	$\widetilde{\textbf{T}}$	**H**	**T**	$\widetilde{\textbf{T}}$	**H**	**T**	$\widetilde{\textbf{T}}$	**H**	**T**	$\widetilde{\textbf{T}}$	**H**
	Normal Distribution $X_{ik} \sim N(0,1)$											
D	6.3	4.9	4.8	5.2	4.3	4.9	6.6	4.4	4.8	5.2	4.3	4.8
T	7.3	5.6	4.8	5.5	5.1	4.5	8.2	5.3	5.3	5.9	4.8	5.2
A	6.1	4.8	4.1	6.2	4.5	4.6	6.5	5.4	6.7	4.9	4.3	5.4
C	6.0	4.9	5.1	5.3	4.9	5.4	5.0	4.1	6.1	5.3	4.8	5.2
	Lognormal Distribution $X_{ik} \sim LN(\mu, \sigma^2)$											
D	6.3	4.9	1.7	5.1	4.2	2.7	6.4	4.2	1.3	5.4	4.5	3.7
T	7.0	5.2	1.6	5.7	4.5	2.6	8.3	5.2	1.4	6.1	4.7	2.1
A	5.8	4.8	3.5	5.3	4.5	9.0	6.8	5.3	15.2	4.8	4.3	13.4
C	5.7	4.8	3.3	5.2	4.7	6.5	5.6	4.5	6.9	5.3	4.8	6.8

small sample sizes and large numbers of factor levels. The parametric MCTP **H** shows a quite liberal or even quite conservative behavior in case of skewed distributions, depending on the chosen contrast.

The powers of the rank based MCTP **T** and $\widetilde{\textbf{T}}$ were compared with the power of the parametric MCTP **H** for all-pairs comparisons. Hereby both the all-pairs power P("reject all false null hypotheses") and the any-pairs power P("reject any true or false null hypothesis") were investigated for a one-point shift alternative $\boldsymbol{\delta} = (0,0,0,\delta)'$. The simulation results for $a = 4$ levels and equal sample sizes ($n_i = 25$) are shown in Figure 2.

The powers of the tests were investigated for both normal and lognormal distributions. For normal distributions, the powers of the nonparametric rank tests are nearly as powerful (even for relatively small sample sizes) as the parametric MCTP **H**. For lognormal distributions, the powers of the rank tests are considerably higher than the power of the parametric version. Simulation results for other distributions, unbalanced designs, and different kinds of contrasts were quite similar and are not shown here.

6. Example

In this section we apply the MCTP and compatible SCI proposed in the previous sections to a dataset with ordinal data analyzed by Akritas et al. (1997) [1]. Originally, two inhalable test substances (drug 1 an drug 2), each in a concentration of 2 ppm, 5 ppm, and 10 ppm, were compared with regard to their irritative activity in the respiratory tract of the rat after subchronic inhalation. In each level, 20 rats were graded on an ordinal scale: 0=no irritation, 1 = slightly irritation, 2=distinct irritation, and 3=severe irritation. Here, we only analyze the data for drug 1 with the R-software package *nparcomp*. The results and point estimators $\widehat{p}_2, \widehat{p}_5$, and $\widehat{p}_{10}$ of the relative treatment effects p_2, p_5, and p_{10} are displayed in Table 2.
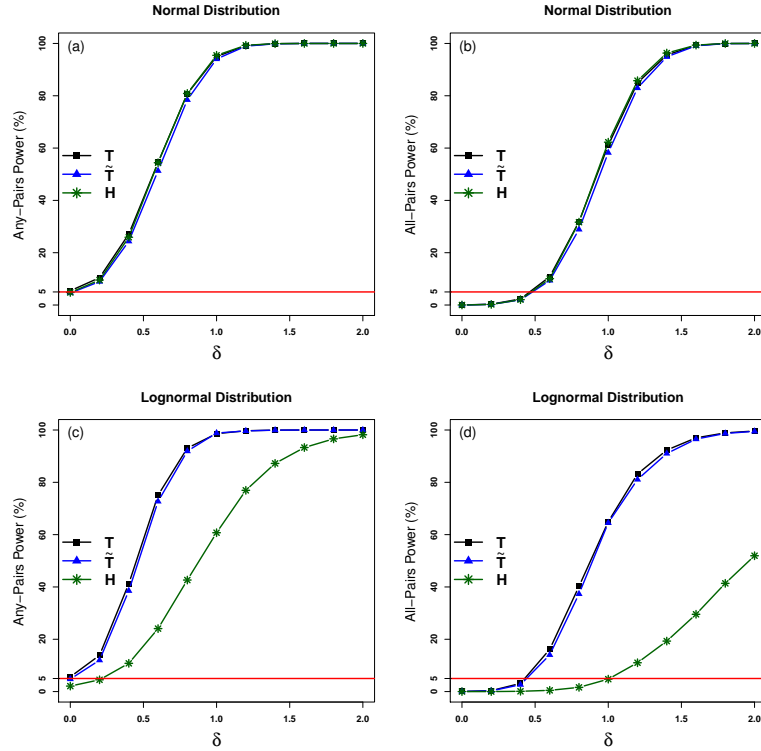
FIG 2. *Empirical any-pairs (left) and all-pairs (right) power curves ($\alpha = 5\%$) of the MCTP* $\mathbf{T}$, $\widetilde{\mathbf{T}}$, *and* $\mathbf{H}$ *for normal distributions ((a) and (b)) and lognormal distributions ((c) and (d)) for all-pairs comparisons and shift alternative* $\boldsymbol{\delta} = (0, 0, 0, \delta)'$.

TABLE 2
Data and estimated effects $\widehat{p}_j$ for the Irritation Trial

Concentration	Rats with scale					Results
	0	1	2	3		$\widehat{p}_j$
2	18	2	0	0		.31
5	12	6	2	0		.45
10	3	7	6	4		.73

From Table 2, it follows that the scores obtained with concentration 2 ppm tend to be smaller ($\widehat{p}_2 = .31$) than the scores in group 5 ppm ($\widehat{p}_5 = .45$), and in group 10 ppm ($\widehat{p}_{10} = .73$), respectively. The MCTP $\widetilde{\mathbf{T}}$ in (4.6) can be used to test the multiple hypotheses $H_0^p : p_2 = p_5$, $H_0^p : p_2 = p_{10}$, and $H_0^p : p_5 = p_{10}$ at multiple level $\alpha = 5\%$ by taking the correlation between the test statistics into account as well as for the computation of compatible SCI. The estimated degree of freedom of the corresponding multivariate t-distribution is given by $\nu = 28.72$. The adjusted p-values for the individual hypotheses $H_0^p : \mathbf{c}_\ell' \mathbf{p} = 0$ are calculated by $1 - \Phi(-|\widetilde{T}_\ell^{0.5}|\mathbf{1}_3, |\widetilde{T}_\ell^{0.5}|\mathbf{1}_3, 28.72, \mathbf{0}, \widehat{\mathbf{R}})$, where $\Phi(\cdot, \cdot, \nu, \mathbf{0}, \widehat{\mathbf{R}})$ denotes

TABLE 3
All-pairs comparisons and 95%-compatible SCI for $\delta_\ell = \mathbf{c}'_\ell \mathbf{p}$ in the Irritation Trial

| Comparison | Estimator | 95%-SCI | | Adjusted p-Value |
		Lower	Upper	
$p_5 - p_2$	.14	$-4 \cdot 10^{-3}$	.28	$6.31 \cdot 10^{-2}$
$p_{10} - p_2$	.42	.28	.50	$< 10^{-6}$
$p_{10} - p_5$	.28	.11	.42	$1.67 \cdot 10^{-3}$

the cumulative distribution function of the multivariate $T(\nu, \mathbf{0}, \widehat{\mathbf{R}})$ distribution. The results are displayed in Table 3. It follows immediately that the irritation of the respiratory tract of the rats in group 2 ppm is milder than the damage effected by 5 ppm, and 10 ppm, respectively ($p < .0001$ and $p = 1.67 \cdot 10^{-3}$). The lower bounds of the 95%-SCI are larger than zero. The data do not provide any evidence to reject the null hypothesis $H_0^p : p_2 = p_5$ at multiple level $\alpha = 5\%$ ($p = 6.31 \cdot 10^{-2}$; 95%-SCI: $[-4 \cdot 10^{-3}; .28]$). We can conclude that the damage gets worse with an increasing concentration of the test substance.

7. Discussion

Recently, Elliot and Hynan (2011) [13] propose a SAS macro implementation of a multiple comparison post hoc test for a Kruskal-Wallis analysis. The procedure is an omnibus test based on two steps: (1) testing the global null hypothesis, and (2) performing multiple comparisons. This nonparametric procedure cannot be used for the computation of confidence intervals for the effects of interest. In this manuscript, rank-based MCTP and compatible SCI for transitive relative effects in unbalanced designs have been introduced. The procedures are based on the asymptotic multivariate normality of linear rank statistics under arbitrary alternatives. Explicit expression for the covariance matrix of the rank statistics, as well as their multivariate normality, are obtained in a technically simple and general framework. Subsequent covariance estimation is achieved in terms of the empirical distribution functions. Under this unified framework, the procedures can be used for testing arbitrary multiple linear hypotheses in terms of relative effects, with an accompanying computation of compatible SCI for the treatment effects. Some simulation results demonstrate the practical benefit of the proposed methods. For a convenient application of the proposed methods, the R-software package *nparcomp* was developed and is available on CRAN.

Acknowledgement

Appendix A: Asymptotic results

A.1. Proof of Theorem 2

Let $\lambda_{i,N}, i = 1, \ldots, a$, denote the eigenvalues of $\mathbf{V}_N$ and let $\lambda_N^* = \min_i\{\lambda_{N,i}|\lambda_{N,i} > 0\}$ denote the smallest eigenvalue larger than zero. Since $\mathbf{V}_N \to \mathbf{V}$ for $N \to \infty$ such that $rank(\mathbf{V}_N) = rank(\mathbf{V})$, there exists a constant $c_0 > 0$ and $M_0 < \infty$ such that $\lambda_N^* \geq c_0 \ \forall N \geq M_0$. Without loss of generality let $\lambda_{1,N}, \ldots, \lambda_{j,N} \to 0$ and let $\lambda_{j+1,N}, \ldots, \lambda_{a,N} \geq c_0 \ \forall N \geq M_0$. From the spectral decomposition theorem, there further exists an invertible matrix $\mathbf{S}$ such that $\mathbf{SV}_N\mathbf{S}' = \mathbf{D} = \bigoplus_{i=1}^2 \mathbf{D}_i$, where $\mathbf{D}_1 = diag\{\lambda_{1,N}, \ldots, \lambda_{j,N}\}$ and $\mathbf{D}_2 = diag\{\lambda_{j+1,N}, \ldots, \lambda_{a,N}\}$. Thus, $\mathbf{D}_1 \to \mathbf{0}$, by assumption. If $j = a$, then $\mathbf{V}_N = \mathbf{0}$, which can be considered as a multivariate normal one-point distribution. The asymptotic multivariate normality of the sums of independent random variables $\sqrt{N}\mathbf{WZ}$ is now established by the Cramer-Wold device. Let $\mathbf{k} = (k_1, \ldots, k_a)'$ denote an arbitrary vector of constants. Since $\mathbf{S}$ is invertable and thus describes a bijektive map, there exists for each $\mathbf{k}$ a vector $\widetilde{\mathbf{k}}$ with $\mathbf{k}' = \widetilde{\mathbf{k}}'\mathbf{S}$. From the Lindeberg-Feller limit theorem it follows that

$$L_N = \frac{\sqrt{N}\mathbf{k}'\mathbf{WZ}}{\sqrt{Var(\sqrt{N}\mathbf{k}'\mathbf{WZ})}} = \frac{N\mathbf{k}'\mathbf{WZ}}{\sqrt{Var(N\mathbf{k}'\mathbf{WZ})}} \xrightarrow{d} N(0,1),$$

because

$$
\begin{aligned}
Var(N\mathbf{k}'\mathbf{WZ}) &= Var(\sqrt{N}\mathbf{k}'\mathbf{W}(\sqrt{N}\mathbf{Z})) = N\mathbf{k}'\mathbf{V}_N\mathbf{k} \\
&= N\widetilde{\mathbf{k}}'\mathbf{SV}_N\mathbf{S}'\widetilde{\mathbf{k}} \geq N\widetilde{\mathbf{k}}'\mathbf{D}_2\widetilde{\mathbf{k}} \geq \min_{s=j+1,\ldots,a}\{\widetilde{k}_s^2\}Nc_0 \to \infty.
\end{aligned}
$$

This means that the sums of the variances of $N\mathbf{k}'\mathbf{WZ}$ diverge for $N \to \infty$ and Linderberg's condition is fulfilled, because the random variables $N/n_i Y_{ijk}$ are uniformly bounded by the assumption that $N/n_i \leq N_0 < \infty$.

A.2. Proof of Lemma 1

Under the assumptions of Theorem 2, $\mathbf{T}$ follows, asymptotically, as $N \to \infty$, a multivariate normal distribution with expectation $\mathbf{0}$ and correlation matrix $\mathbf{R}$. Thus, the asymptotic joint distribution of $\mathbf{T}$ is completely specified under $H_0^p : \bigcap_{\ell=1}^q \{H_0^p : \mathbf{c}_\ell'\mathbf{p} = 0\}$. Each test statistic T_ℓ^p converges, as $N \to \infty$, to the standard normal distribution. In particular, the asymptotic distribution of T_ℓ^p is independent from the distribution of T_m^p ($\ell \neq m$). This means, that under arbitrary intersection hypotheses $H_0^p : \bigcap_{j \in J}\{H_0^p : \mathbf{c}_j'\mathbf{p} = 0\}$, the asymptotic joint distribution of $\mathbf{T}^J = \{T_j^p, j \in J\}$ is completely specified. Here, $J \subseteq \{1, \ldots, q\}$ denotes an arbitrary set of indexes. This completes the proof.

A.3. Proof of Theorem 3

The testing family $\{\mathbf{\Omega}^p, \mathbf{T}\}$ constitutes an asymptotic joint testing family. Further, the STP $\{\mathbf{\Omega}, \mathbf{T}, z_{1-\alpha,2,\mathbf{R}}\}$ is coherent, by construction. Therefore, the assumptions of Theorem 2 in Gabriel (1969) [14] are fulfilled. Next the unknown correlation matrix is replaced by the consistent estimator $\widehat{\mathbf{R}}$. Consider the map $\mathbf{f}(\mathbf{R}) = z_{1-\alpha,2,\mathbf{R}}$. Since $\mathbf{f}$ is continuous, it follows that $z_{1-\alpha,2,\widehat{\mathbf{R}}} - z_{1-\alpha,2,\mathbf{R}} \xrightarrow{p} 0$, as $N \to \infty$. This means, the quantity $z_{1-\alpha,2,\widehat{\mathbf{R}}}$ is a consistent estimator of $z_{1-\alpha,2,\mathbf{R}}$. Therefore, if $N \to \infty$, the STP $\{\mathbf{\Omega}^p, \mathbf{T}, z_{1-\alpha,2,\widehat{\mathbf{R}}}\}$ controls the FWER in the strong sense.

A.4. Estimation of the covariance matrix $\mathbf{V}_N$

Lemma A.2. *Let* $|| \cdot ||_\infty$ *denote the sup norm and let* D_{ijk} *as given in (3.8). Further define* $\widetilde{D}_{ijk} = Y_{ijk} - n_j^{-1} \sum_{k=1}^{n_j} Y_{ijk}$. *If* $n_i \to \infty$, $D_{ijk} - \widetilde{D}_{ijk} \xrightarrow{a.s.} 0$.

Proof.

$$
\begin{aligned}
|D_{ijk} - \widetilde{D}_{ijk}| &\leq \left| \widehat{F}_i(X_{jk}) - F_i(X_{jk}) \right| + \frac{1}{n_j} \sum_{k=1}^{n_j} \left| \widehat{F}_i(X_{jk}) - F_i(X_{jk}) \right| \\
&\leq \left\| \widehat{F}_i - F_i \right\|_\infty + \max_{k=1,\ldots,n_j} \left| \widehat{F}_i(X_{jk}) - F_i(X_{jk}) \right| \\
&\leq 2 \left\| \widehat{F}_i - F_i \right\|_\infty \xrightarrow{a.s} 0, \ n_i \to \infty,
\end{aligned}
$$

where the last step follows from the Glivenko-Cantelli theorem. $\qquad\square$

Theorem A.4. *Let* $\widehat{\mathbf{\Sigma}}_N = (\widehat{\sigma}_{ij,rs})_{i,j,r,s=1,\ldots,a}$ *and let* $\widehat{\mathbf{V}}_N = \mathbf{W}\widehat{\mathbf{\Sigma}}_N\mathbf{W}'$. *If* $\min\{n_1, \ldots, n_a\} \to \infty$ *such that* $\frac{N}{n_i} \leq N_0 < \infty$, *then* $\widehat{\mathbf{V}}_N - \mathbf{V}_N \xrightarrow{a.s.} \mathbf{0}$.

Proof. Since a is bounded, it is sufficient to show consistency elementwise. Consider the covariance $\theta_{ji,si} = Cov(Y_{ji1}, Y_{si1})$ and let $\widetilde{\theta}_{ji,si}$ as given in (3.7). By the strong law of large numbers, $\widetilde{\theta}_{ji,si} - \theta_{ji,si} \xrightarrow{a.s.} 0$, if $n_j, n_s \to \infty$. Therefore it is sufficient to show $|\widehat{\theta}_{ji,si} - \widetilde{\theta}_{ji,si}| \xrightarrow{a.s.} 0$, if $n_j, n_s \to \infty$. By the triangular inequality it follows that

$$
\begin{aligned}
|\widehat{\theta}_{ji,si} - \widetilde{\theta}_{ji,si}| &= \left| \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (D_{jik}D_{sik} - \widetilde{D}_{jik}\widetilde{D}_{sik}) \right| \\
&\leq \frac{1}{n_i - 1} \sum_{k=1}^{n_i} \left| D_{jik}(D_{sik} - \widetilde{D}_{sik}) - \widetilde{D}_{sik}(\widetilde{D}_{jik} - D_{jik}) \right| \\
&\leq \frac{1}{n_i - 1} \sum_{k=1}^{n_i} \left| D_{sik} - \widetilde{D}_{sik} \right| + \frac{1}{n_i - 1} \sum_{k=1}^{n_i} \left| \widetilde{D}_{jik} - D_{jik} \right| \\
&\leq \max_{k=1,\ldots,n_i} |D_{sik} - \widetilde{D}_{sik}| + \max_{k=1,\ldots,n_i} |D_{jik} - \widetilde{D}_{jik}|
\end{aligned}
$$

$$\leq \quad 2 \max_{k=1,\ldots,n_i} |Y_{sik} - \widehat{Y}_{sik}| + 2 \max_{k=1,\ldots,n_i} |Y_{jik} - \widehat{Y}_{jik}|$$

$$\leq \quad 2 \left\|\widehat{F}_s - F_s\right\|_\infty + 2 \left\|\widehat{F}_j - F_j\right\|_\infty \overset{a.s.}{\to} 0, n_j, n_s \to \infty.$$

For the other elements the proof is basically the same and is therefore omitted. The rest of the proof follows by considering linear combinations of the estimators. □

A.5. Simulation settings

The contrast matrices used in the simulation studies (see Section 5) are given by the Dunnett-type (D) contrast matrix

$$\mathbf{C} = \begin{pmatrix} -1 & 1 & 0 & \ldots & 0 & 0 \\ -1 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 0 & 0 & \ldots & \ldots & 1 \end{pmatrix}, \tag{A.1}$$

the Tukey-type (T) contrasts

$$\mathbf{C} = \begin{pmatrix} -1 & 1 & 0 & \ldots & \ldots & 0 & 0 \\ -1 & 0 & 1 & 0 & \ldots & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 0 & 0 & 0 & \ldots & \ldots & 1 \\ 0 & -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & \ldots & \ldots & \ldots & -1 & 1 \end{pmatrix}, \tag{A.2}$$

Average-type (A) contrasts

$$\mathbf{C} = \begin{pmatrix} 1 & -\frac{1}{a-1} & -\frac{1}{a-1} & \cdots & -\frac{1}{a-1} \\ -\frac{1}{a-1} & 1 & -\frac{1}{a-1} & \cdots & -\frac{1}{a-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{1}{a-1} & -\frac{1}{a-1} & -\frac{1}{a-1} & \cdots & 1 \end{pmatrix}, \tag{A.3}$$

and Changepoint (C) comparisons

$$\mathbf{C} = \begin{pmatrix} -1 & +\frac{n_2}{N_{2a}} & \cdots & \cdots & +\frac{n_{a-1}}{N_{2a}} & +\frac{n_a}{N_{2a}} \\ -\frac{n_1}{N_{12}} & -\frac{n_2}{N_{12}} & +\frac{n_3}{N_{3a}} & \cdots & +\frac{n_{a-1}}{N_{3a}} & +\frac{n_a}{N_{3a}} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ -\frac{n_1}{N_{1(a-1)}} & -\frac{n_2}{N_{1(a-1)}} & -\frac{n_3}{N_{1(a-1)}} & \cdots & -\frac{n_{a-1}}{N_{1(a-1)}} & +1 \end{pmatrix}, \tag{A.4}$$

where $N_{ij} = n_i + \ldots + n_j; i < j$. For a comprehensive overview of different contrasts we refer the reader to Bretz et al. (2001) [3].

References

[1] Akritas, M. G., Arnold, S. F. and Brunner, E. (1997). Nonparametric hypotheses and rank statistics for unbalanced factorial designs. *Journal of the American Statistical Association* **92** 258–265. MR1436114

[2] Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics* **25** 290–302. MR0061787

[3] Bretz, F., Genz, A. and Hothorn, L. A. (2001). On the numerical availability of multiple comparison procedures. *Biometrical Journal* **43** 645–656. MR1863491

[4] Bretz, F., Hothorn T. and Westfall, P. (2010). *Multiple Comparisons Using R*, Chapman and Hall, London.

[5] Brown, B. M. and Hettmansperger, T. P. (2002). Kruskal-Wallis, multiple comparisons and Efron dice. *Australian and New Zealand Journal of Statistics* **44** 427–438. MR1934732

[6] Brunner, E., Dette, H. and Munk, A. (1997). Box-Type approximations in nonparametric factorial designs. *Journal of the American Statistical Association* **92** 1494–1502. MR1615259

[7] Brunner, E., Domhof, S. and Langer, F. (2002). *Nonparametric Analysis of Longitudinal Data in Factorial Experiments.* New York, Wiley. MR1865401

[8] Brunner, E. and Munzel, U. (2000). The nonparametric Behrens-Fisher problem - Asymptotic theory and small sample approximations. *Biometrical Journal* **42** 17–25. MR1744561

[9] Brunner E., Munzel U. and Puri M. L. (2002). The multivariate nonparametric Behrens-Fisher problem. *Journal of Statistical Planning and Inference* **108** 37–53. MR1947390

[10] Brunner, E. and Puri, M. L. (2001). Nonparametric methods in factorial designs. *Statistical Papers* **42** 1–52. MR1821004

[11] Dunn, O. J. (1960). Multiple comparisons using rank sums. *Technometrics* **2** 241–252.

[12] Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* New York, Chapman and Hall. MR1270903

[13] Elliott, A. C. and Hynan, L. S. (2011). A SAS macro implementation of a multiple comparison post hoc test for a Kruskal-Wallis analysis. *Computer Methods and Programs in Biomedicine* **102** 75 –80.

[14] Gabriel, K. R. (1969). Simultaneous test procedures - some theory of multiple comparisons. *Annals of Mathematical Statistics* **40** 224–250. MR0240931

[15] Gao, X. and Alvo, M. (2005). A unified nonparametric approach for unbalanced factorial designs. *Journal of the American Statistical Association* **100** 926–941. MR2201020

[16] Gao, X. and Alvo, M. (2008). Nonparametric multiple comparison procedures for unbalanced two-way factorial designs. *Journal of Statistical Planning and Inference* **138** 3674–3686. MR2455956

[17] Gao, X., Alvo, M., Chen, J. and Li, G. (2008). Nonparametric multiple comparison procedures for unbalanced one-way factorial designs. *Journal of Statistical Planning and Inference* **138** 2574–2591. MR2432382

[18] Gardner, M. (1974). On the paradoxical situation that arise from nontransitive relations. *Scientific American* **231** 120–125.

[19] Genz, A. and Bretz, F. (2009). *Computation of multivariate normal and t probabilities. Lecture Notes in Statistics*, New York, Springer. MR2840595

[20] Hasler, M. and Hothorn, L. A. (2008). Multiple contrast tests in the presence of heteroscedasticity. *Biometrical Journal* **50** 793–800. MR2542344

[21] Herberich, E., Sikorski, J. and Hothorn, T. (2010). A robust procedure for comparing multiple means under heteroscedasticity in unbalanced designs. *PLoS ONE*, DOI:10.1371/journal.pone.0009788.

[22] Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures*. Wiley, New York. MR0914493

[23] Hothorn, T., Bretz, F. and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal* **50**, 346–363. MR2521547

[24] ICH (1998). *Statistical Principles for Clinical Trials*. Guideline, Available at http://private.ich.org.

[25] Konietschke, F. and Hothorn, L. A. (2012). Evaluation of toxicological studies using a nonparametric Shirley-Type trend test for comparing several dose levels with a control group. *Statistics in Biopharmaceutical Research*, in press (DOI: 10.1080/19466315.2011.633861).

[26] Konietschke, F., Libiger, O. and Hothorn, L. A. (2012).Nonparametric Evaluation of Quantitative Traits in Population-Based Association Studies when the Genetic Model is Unknown. *PLoS ONE* 7(2): e31242. doi:10.1371/journal.pone.0031242.

[27] Lévy, P. (1925). *Calcul des probabilitées*, Gauthiers-Villars, Paris.

[28] Mukerjee, H., Robertson, T. and Wright, F. T. (1987). Comparison of several treatments with a control using multiple contrasts. *Journal of the American Statistical Association* **82** 902–910. MR0909999

[29] Munzel, U. (1999). Linear rank score statistics when ties are present. *Statistics and Probability Letters* **41** 389–395. MR1666092

[30] Munzel, U. and Hothorn, L. A. (2001). A unified approach to simultaneous rank test procedures in the unbalanced one-way layout. *Biometrical Journal* **43** 553–569. MR1859376

[31] Orban, J. and Wolfe, D. A. (1982). A class of distribution-free two-sample tests based on placements. *Journal of the American Statistical Association* **77** 666–672. MR0675896

[32] Puri, M. L. (1964). Asymptotic efficiency of a class of c-sample tests. *Annals of Mathematical Statistics* **35** 102–121. MR0158482

[33] RUST, S. W. and FLIGNER, M. A. (1984). A modification of the Kruskal-Wallis statistic for the generalized Behrens-Fisher problem. *Communications in Statistics - Theory and Methods* **13** 2013–2027. MR0759955

[34] RUYMGAART, F. H. (1980). A unified approach to the asymptotic distribution theory of certain midrank statistics. *Statistique non Parametrique Asymptotique.* J.P. Raoult (Ed.), Lecture Notes on Mathematics, No. 821, Springer, Berlin. MR0604018

[35] RYU, E. (2009). Simultaneous confidence intervals using ordinal effect measures for ordered categorical outcomes. *Statistics in Medicine* **28** 3179–3188. MR2750413

[36] RYU, E. and AGRESTI, A. (2008). Modeling and inference for an ordinal effect size measure. *Statistics in Medicine* **27** 1703–1717. MR2420336

[37] STEEL, R. D. G. (1959). A multiple comparison rank test: treatments versus control. *Biometrics* **15** 560–572. MR0108869

[38] STEEL, R. D. G. (1960). A rank sum test for comparing all pairs of treatments. *Technometrics* **2** 197–207. MR0113260

[39] THANGAVELU, K. and BRUNNER, E. (2006). Wilcoxon Mann-Whitney test and Efrons paradox dice. *Journal of Statistical Planning and Inference* **137** 720–737. MR2301712

[40] WOLFSEGGER, M. and JAKI, T. (2006). Simultaneous confidence intervals by iteratively adjusted alpha for relative effects in the one-way layout. *Statistics and Computing* **16** 15–23. MR2224186