DE GRUYTER doi 10.1515/ijb-2012-0020 — The International Journal of Biostatistics 2013; 9(1): 63–73

Research Article

Frank Konietschke*, Sandra Bösiger, Edgar Brunner, and Ludwig A. Hothorn

Are Multiple Contrast Tests Superior to the ANOVA?

Abstract: Multiple contrast tests can be used to test arbitrary linear hypotheses by providing local and global test decisions as well as simultaneous confidence intervals. The ANOVA-*F*-test on the contrary can be used to test the global null hypothesis of no treatment effect. Thus, multiple contrast tests provide more information than the analysis of variance (ANOVA) by offering which levels cause the significance. We compare the exact powers of the ANOVA-*F*-test and multiple contrast tests to reject the global null hypothesis. Hereby, we compute their *least favorable configurations* (LFCs). It turns out that both procedures have the same LFCs under certain conditions. Exact power investigations show that their powers are equal to detect their LFCs.

Keywords: analysis of variance, multiple contrast tests, multivariate *t*-distribution, one-way layout, least favorable configuration, sample size computations

***Corresponding author: Frank Konietschke,** Department of Medical Statistics, University Medical Center Göttingen, Humboldtallee 32, Göttingen, Lower Saxony 37073, Germany, E-mail: Frank.Konietschke@medizin.uni-goettingen.de
Sandra Bösiger, Siemens – Siemens Healthcare Diagnostics Products GmbH, Marburg, Germany, E-mail: sandra.boesiger@stud.uni-goettingen.de
Edgar Brunner, Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Lower Saxony, Germany, E-mail: brunner@ams.med.uni-goettingen.de
Ludwig A. Hothorn, Institute of Biostatistics, Leibniz University Hannover, Hannover, Lower Saxony, Germany, E-mail: hothorn@biostat.uni-hannover.de

1 Introduction

In many psychological, biological, and medical trials, more than two treatment groups are involved. In these situations, one is interested in detecting any significant difference among the treatment means $\mu_1, \ldots, \mu_a$, i.e. to test the global null hypothesis $H_0 : \mu_1 = \cdots = \mu_a$, and, particularly, in the detection of specific significant differences, i.e. in performing multiple comparisons according to the computation of simultaneous confidence intervals (SCI). In randomized clinical trials, the computation of SCI 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 [23]). Hereby, the family-wise error rate α should be strongly controlled.

In statistical practice, however, the usual way to detect specific significant differences among the effects of interest, and to compute SCI, consists of three steps: (1) the global null hypothesis H_0 is tested by an appropriate procedure, e.g. analysis of variance (ANOVA), (2) if the global null hypothesis is rejected, multiple comparisons are usually carried out to test individual hypotheses, e.g. the lth partial null hypothesis $H_0^{(\ell)} : \mu_i = \mu_j$, and (3) in the final step, SCI for the treatment 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 [26] and Hsu [21]. 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 [1]. 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, particularly when the test statistics are not independent.

In recent years, multiple contrast test procedures (MCTPs) with accompanying compatible SCI for linear contrasts were derived by Mukerjee et al. [2] and Bretz et al. [1]. The procedures are based on the exact multivariate distribution of a vector of *t*-test statistics, where each test statistic corresponds to an individual null hypothesis, e.g. $H_0^{(\ell)} : \mu_i = \mu_j$. It will be rejected, if the corresponding test statistic exceeds a critical value being obtained from the distribution of the vector of *t*-test statistics. The global null hypothesis will be rejected, if any individual hypothesis is rejected. Therefore, the individual and global test decisions are consonant and coherent. These MCTPs take the correlation between the test statistics into account and can be used for testing arbitrary contrasts, e.g. many-to-one, all-pairs, or even average comparisons [1]. Thus, MCTPs provide an extensive tool for powerful multiple comparisons, for the computation of compatible SCI, and for testing the global null hypothesis. The results by Bretz et al. [1] were extended to general linear models by Hothorn et al. [3], to heteroscedastic models by Hasler and Hothorn [4] and Herberich et al. [5], and for ranking procedures by Konietschke and Hothorn [6], Konietschke et al. [7], and Konietschke et al. [8]. For a comprehensive overview of existing methods, we refer to Bretz et al. [27].

Comparing MCTP and the global testing procedure ANOVA, one notices that both procedures can be used to test the global null hypothesis H_0. From a practical point of view, MCTPs demonstrate their superiority to the ANOVA in terms of providing the information which levels cause the statistical overall significance as well as by offering SCI. In quite restricted homoscedastic normal models, both procedures are exact level α tests. Arias-Castro et al. [9] studied global and multiple testing procedures under sparse alternatives and emphasize *"Because ANOVA is such a well established method, it might surprise the reader – but not the specialist – to learn that there are situations where the Max test, though apparently naive, outperforms ANOVA by a wide margin"* [9, p. 2534]. The evidence of a loss in power of the MCTP to detect global alternatives, if so, has not been investigated yet [25]. Thus, exact power comparisons remain.

It is the aim of this article to investigate the exact power of MCTP and of the ANOVA to detect global alternatives. To give a fair comparison, we restrict our analysis to those linear contrasts which are embedded in the ANOVA, i.e. contrasts which compare each mean μ_i to the overall mean $\bar{\mu}_\cdot$. In particular, we compute the *least favorable configuration* (LFC) of the alternative, i.e. the alternative which is detected with a minimal power of both the ANOVA and the MCTP. The results indicate that the LFCs of both procedures are identical. Exact power calculations show that their powers to detect the LFCs are equal.

2 Statistical model and test statistics

We consider a completely randomized one-way layout

$$Y_{ij} \sim N(\mu_i, \sigma^2),\ i = 1, \ldots, a, \text{ and } j = 1, \ldots, n_i, \qquad [1]$$

where the index i denotes the level of the treatment group, and j denotes the jth unit within the ith group. Throughout this article, let $N = \sum_{i=1}^{a} n_i$ denote the total sample size, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_a)'$ the vector of expectations, $\boldsymbol{\theta} = \boldsymbol{\mu}/\sigma$ its scaled version, and let $\boldsymbol{\Lambda} = \text{diag}(n_1, \ldots, n_a)$ denote the diagonal matrix of the sample sizes. Furthermore, let $\overline{\mathbf{Y}}_\cdot = (\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{a\cdot})'$ denote the vector of means, let $\overline{Y}_{\cdot\cdot} = a^{-1} \sum_{i=1}^{a} \overline{Y}_{i\cdot}$ denote the overall mean, and let $s^2 = (N-a)^{-1} \sum_{i=1}^{a} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i\cdot})^2$ denote the pooled sample variance.

Our aim is to test the null hypothesis $H_0 : \mu_1 = \ldots = \mu_a$ versus the alternative $H_1 : \mu_i \neq \bar{\mu}_\cdot$ for at least one μ_i, where $\bar{\mu}_\cdot = a^{-1} \sum_{i=1}^{a} \mu_i$ is the mean of expectations. The global null hypothesis H_0 can be equivalently written as

$$H_0 : \begin{cases} \mu_1 = \bar{\mu}_. \\ \mu_2 = \bar{\mu}_. \\ \vdots \\ \mu_a = \bar{\mu}_. \end{cases} \Leftrightarrow H_0 : \mathbf{C}\boldsymbol{\mu} = \begin{pmatrix} 1-1/a & -1/a & \dots & -1/a \\ -1/a & 1-1/a & \dots & -1/a \\ \vdots & \vdots & \ddots & \vdots \\ -1/a & -1/a & \dots & 1-1/a \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_a \end{pmatrix} = \mathbf{0}. \quad [2]$$

The contrast matrix $\mathbf{C}$ is also known as the $a \times a$ centering matrix $\mathbf{P}_a = \mathbf{I}_a - \frac{1}{a}\mathbf{J}_a$, where $\mathbf{I}_a$ denotes the $a \times a$ unit matrix, and $\mathbf{J}_a = \mathbf{1}_a\mathbf{1}_{a'}$ denotes the $a \times a$-matrix of 1's. Throughout this article, $\mathbf{C}$ will be called *Grand-mean-type* contrast matrix [10]. Each row vector $\mathbf{c}_i'$ of $\mathbf{C}$ is one contrast and will be used later for testing individual hypotheses $H_0^{(i)} : \mathbf{c}_i'\boldsymbol{\mu} = 0$, i.e. $H_0^{(i)} : \mu_i = \overline{\mu}_.$ for $i = 1, \dots, a$. The ANOVA-F-test

$$F_C = \left\{ \sum_{i=1}^{a} n_i (\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2 / (a-1) \right\} / s^2 \quad [3]$$

is the commonly used statistic for testing H_0. As usually known, F_C is exactly $F(a-1, N-a|\lambda)$-distributed, where $\lambda = \theta'[\mathbf{\Lambda} - N^{-1}\mathbf{\Lambda}\mathbf{J}_a\mathbf{\Lambda}]\theta$ denotes the non-centrality parameter. Clearly, under H_0, λ is equal to zero. It follows from the definition of F_C in eq. [3] that this global testing procedure is the scaled sum of the squared contrasts $\hat{\delta}_i = \mathbf{c}_i'\overline{\mathbf{Y}}_\cdot = \overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}$ in means. Therefore, it cannot provide any information about the means which differ significantly from the overall mean $\overline{Y}_{\cdot\cdot}$. The MCTP by using the contrasts $\mathbf{c}_i'$ on the contrary consists of the vector of t-test type statistics

$$\mathbf{T} = (T_1, \dots, T_a)', \text{ where } T_i = \mathbf{c}_i'\overline{\mathbf{Y}}_\cdot / \left(s\sqrt{\mathbf{c}_i'\mathbf{\Lambda}^{-1}\mathbf{c}_i} \right) = (\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}) / \left(s\sqrt{\mathbf{c}_i'\mathbf{\Lambda}^{-1}\mathbf{c}_i} \right) \quad [4]$$

is the modified t-test statistic for testing $H_0^{(i)} : \mu_i = \bar{\mu}_.$. Thus, $\mathbf{T}$ consists of the scaled single contrasts $\hat{\delta}_i$. We note that the MCTP is not restricted to comparisons to the overall mean. For example, Dunnett-type many-to-one [11] comparisons can be performed by using the contrast matrix in

$$H_0 : \begin{cases} \mu_1 = \mu_2 \\ \mu_1 = \mu_3 \\ \vdots \\ \mu_1 = \mu_a \end{cases} \Leftrightarrow H_0 : \mathbf{C}\boldsymbol{\mu} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ -1 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 0 & 0 & \dots & \dots & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_a \end{pmatrix} = \mathbf{0}. \quad [5]$$

Tukey-type [12] all-pairs comparisons can be conducted using

$$H_0 : \begin{cases} \mu_1 = \mu_2 \\ \mu_1 = \mu_3 \\ \vdots \\ \mu_1 = \mu_d \\ \mu_2 = \mu_3 \\ \vdots \\ \mu_{a-1} = \mu_a \end{cases} \Leftrightarrow H_0 : \mathbf{C}\boldsymbol{\mu} = \begin{pmatrix} -1 & 1 & 0 & \dots & \dots & 0 & 0 \\ -1 & 0 & 1 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 0 & 0 & 0 & \dots & \dots & 1 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & \dots & -1 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \vdots \\ \mu_a \end{pmatrix} = \mathbf{0}. \quad [6]$$

and by replacing the contrasts $\mathbf{c}_i'$ in eq. [4] by the row vectors of the chosen contrast matrix. For a detailed overview of different kinds of contrasts, we refer to Bretz et al. [1]. The ANOVA-F-test, however, is restricted to the comparisons to the overall mean as described in eq. [2]. Therefore, we will only compare the ANOVA with the MCTP $\mathbf{T}$ as given in eq. [4]. As further results, we will also investigate the powers of the MCTP by using the Dunnett-type or Tukey-type contrast matrix $\mathbf{C}$ as given in eq. [5] or [6], respectively. For convenience, we will write the different contrasts in a unified way by a non-specified contrast matrix $\mathbf{C} = (\mathbf{c}_1', \dots, \mathbf{c}_q')'$ throughout this article.

Bretz et al. [1] have shown that $\mathbf{T}$ follows a multivariate $T(\nu, \mathbf{R}, \delta(\boldsymbol{\theta}))$ distribution with $\nu = N - a$ degrees of freedom, correlation matrix $\mathbf{R}$ and non-centrality parameter vector

$$\boldsymbol{\delta}(\boldsymbol{\theta}) = (\delta(\theta_1), \ldots, \delta(\theta_a))', \quad [7]$$

where $\delta(\theta_i) = \mathbf{c}_i'\boldsymbol{\mu}/(\sigma\sqrt{\mathbf{c}_i'\boldsymbol{\Lambda}^{-1}\mathbf{c}_i})$. Under the global null hypothesis $H_0 : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$, the non-centrality parameter vector $\boldsymbol{\delta}(\boldsymbol{\theta})$ is equal to $\mathbf{0} = (0, \ldots, 0)'$. The correlation matrix $\mathbf{R}$ is known and only depends on the sample sizes n_i in model [1]. It can be easily computed by standardizing the covariance matrix $\mathbf{V} = \sigma^2\mathbf{C}\Lambda^{-1}\mathbf{C}'$ of $\mathbf{C}\overline{\mathbf{Y}}_{\cdot}$. For a detailed explanation, we refer to Bretz et al. [1]. The individual null hypothesis $H_0^{(i)} : \mu_i = \bar{\mu}_{\cdot}$ will be rejected at multiple α level of significance, if $|T_i| \geq t_{1-\alpha}(\nu, \mathbf{R})$, where $t_{1-\alpha}(\nu, \mathbf{R})$ denotes the $(1-\alpha)$-equicoordinate quantile from the multivariate $T(\nu, \mathbf{R}, \mathbf{0})$-distribution, that is

$$P\left(\bigcap_{i=1}^{a}\{-t_{1-\alpha}(\nu, \mathbf{R}) \leq T_i \leq t_{1-\alpha}(\nu, \mathbf{R})\}\right) = 1 - \alpha.$$

In particular, compatible $(1-\alpha)$-SCI for the treatment effects $\delta_i = \mu_i - \bar{\mu}_{\cdot}$ are given by

$$CI_i = \left[\mathbf{c}_i'\overline{\mathbf{Y}}_{\cdot} \pm t_{1-\alpha}(\nu, \mathbf{R}) \cdot s\sqrt{\mathbf{c}_i'\Lambda^{-1}\mathbf{c}_i}\right]. \quad [8]$$

The global null hypothesis $H_0 : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ will be rejected, if

$$T_0 = \max\{|T_1|, \ldots, |T_a|\} \geq t_{1-\alpha}(\nu, \mathbf{R}). \quad [9]$$

Apparently, both test statistics F_C in eq. [3] and T_0 in eq. [9] consist of the same contrasts $\hat{\delta}_i$ and the same error estimate s^2. The difference between the procedures is that the ANOVA-F-test uses the scaled sum of the squares of the contrasts and the MCTP uses the maximum of the scaled single contrasts. The impact of these two different principles on the powers of the tests will be investigated in the next section.

3 Power comparisons of the ANOVA and MCTP

It is obvious that the ANOVA-F-test F_C is a squared test statistic, while T_0, or better the single contrasts T_i embedded in T_0, are linear statistics. Roughly speaking, both methods are not comparable analytically. We, therefore, consider the power of the MCTP T_0 to detect the global alternative $H_1 : \mu_i \neq \bar{\mu}_{\cdot}$ for at least one μ_i, $i = 1, \ldots, a$. Due to the abundance of possible alternatives, we will compute the LFC of both ANOVA and the MCTP, i.e. the alternatives which are detected with a minimal power. Next, the powers to detect their LFC can be fairly compared. As pointed out in Section 2, the vector of t-test statistics $\mathbf{T}$ as defined in eq. [4] follows a multivariate $T(\nu, \mathbf{R}, \boldsymbol{\delta}(\boldsymbol{\theta}))$ distribution with $\nu = N - a$ degrees of freedom, correlation matrix $\mathbf{R}$, and non-centrality parameter vector $\boldsymbol{\delta}(\theta) = (\boldsymbol{\delta}(\theta_1), \ldots, \delta(\theta_a))'$. Thus, the power of T_0 to detect H_1 at significance level α can be defined by

$$\begin{aligned}\beta(\theta) &= P_{H_1}(T_0 \geq t_{1-\alpha}(\nu, \mathbf{R})) \\ &= 1 - P_{H_1}(\max_{i=1,\ldots,a}|T_i| \leq t_{1-\alpha}(\nu, \mathbf{R})) \\ &= 1 - P_{H_1}(-t_{1-\alpha}(\nu, \mathbf{R}) \leq T_1 \leq t_{1-\alpha}(\nu, \mathbf{R}), \ldots, -t_{1-\alpha}(\nu, \mathbf{R}) \leq T_a \leq t_{1-\alpha}(\nu, \mathbf{R})).\end{aligned} \quad [10]$$

Note that $\text{rank}(\mathbf{C}) = a - 1$, hence, the correlation matrix $\mathbf{R}$ is singular and the distribution of $\mathbf{T}$ cannot have a density with respect to Lebesgue measure. The exact power of the MCTP as defined in eq. [10], however, can be computed by using the $(a-1)$-variate regular multivariate t-distribution function of the $(a-1)$-statistics $\widetilde{\mathbf{T}} = (T_1, \ldots, T_{a-1})'$ being computed with the $(a-1)$ linear independent contrasts $\mathbf{c}_1', \ldots, \mathbf{c}_{a-1}'$,

respectively, and an appropriate transformation of the integration region, i.e. the probability in eq. [10], can be computed by

$$\begin{aligned}\beta(\boldsymbol{\theta}) &= 1 - P_{H_1}(-t_{1-\alpha}(\nu, \mathbf{R}) \leq T_1 \leq t_{1-\alpha}(\nu, \mathbf{R}), \ldots, -t_{1-\alpha}(\nu, \mathbf{R}) \leq T_a \leq t_{1-\alpha}(\nu, \mathbf{R})) \\ &= 1 - P_{H_1}(u_1 \leq T_1 \leq v_1, \ldots, u_{a-1} \leq T_{a-1} \leq v_{a-1}),\end{aligned}$$

where $\mathbf{u} = (u_1, \ldots, u_{a-1})'$ and $\mathbf{v} = (v_1, \ldots, v_{a-1})'$ denote the new integration bounds. For the computation of $\mathbf{u}$ and $\mathbf{v}$, we refer to Bretz [13], Bretz et al. [1], Bretz and Genz [24] and Genz and Kwong [14].

Now, it is our purpose to consider the two conditions

$$b_1(\boldsymbol{\theta}) = \max_{1 \leq i \leq a} \left|\theta_i - \overline{\theta}_\cdot\right| \geq b \quad \text{or} \quad b_2(\boldsymbol{\theta}) = \max_{1 \leq i, j \leq a} |\theta_i - \theta_j| \geq b \qquad [11]$$

and to establish the configuration of the θ_i for which the power function $\beta(\boldsymbol{\theta})$ is minimized, i.e. we compute the LFC $\boldsymbol{\theta}^*$ of $\boldsymbol{\theta}$ such that

$$\beta(\boldsymbol{\theta}^*) = \min_{\boldsymbol{\theta} \in \mathbb{R}^a : b_i(\boldsymbol{\theta}) \geq b > 0} \beta(\boldsymbol{\theta}), \quad i = 1, 2. \qquad [12]$$

Note that in unbalanced designs, the power of the LFC $\beta(\boldsymbol{\theta}^*)$ cannot be invariant under any permutation of the coordinates of $\boldsymbol{\theta}^*$, which follows from the definition of the multivariate t-distribution. To get a useful result, we, therefore, restrict the computation to balanced designs. The LFCs $\boldsymbol{\theta}^*$ of $\mathbf{T}$ for Grand-mean and Tukey-type MCTPs are given in Theorem 1.

Theorem 1. *Suppose that $n_1 = \cdots = n_a$, let $b \geq 0$ and let $\mathbf{C}$ denote the Grand-mean-type or Tukey-type contrast matrix $\mathbf{C}$ as given in eqs. [2] or [6], respectively. Further let*

1. $\boldsymbol{\theta}^* = (0, \ldots, 0, ba/(a-1))'$, so that $b_1(\boldsymbol{\theta}^*) = b$. Then, if

$$b_1(\boldsymbol{\theta}) \geq b \Rightarrow \beta(\boldsymbol{\theta}) \geq \beta(\boldsymbol{\theta}^*).$$

2. Let $\boldsymbol{\theta}^* = (-b/2, 0 \ldots, 0, b/2)'$, so that $b_2(\boldsymbol{\theta}^*) = b$. Then, if

$$b_2(\boldsymbol{\theta}) \geq b \Rightarrow \beta(\boldsymbol{\theta}) \geq \beta(\boldsymbol{\theta}^*).$$

It follows from Theorem 1 that, under the restrictions $b_1(\boldsymbol{\theta}) \geq b$ or $b_2(\boldsymbol{\theta}) \geq b$, the LFCs $\boldsymbol{\theta}^* = (0, \ldots, 0, ba/(a-1))'$ or $\boldsymbol{\theta}^* = (-b/2, \ldots, 0, b/2)'$, respectively, will be detected with minimal power. In particular, Hayter and Liu [15, 16] compute the LFCs of the ANOVA-F-test under both restrictions $b_1(\boldsymbol{\theta})$ and $b_2(\boldsymbol{\theta})$, respectively. It turns out that both the ANOVA and the MCTP have the same LFCs. The comparisons of the powers to detect their LFCs will be investigated in Section 3.1.

3.1 Numerical comparisons

The computations of the exact powers of both procedures to detect their LFCs under the restrictions $b_1(\boldsymbol{\theta})$ and $b_2(\boldsymbol{\theta})$, respectively, are of particular interest. In Figure 1, the exact power curves (type-I error level $\alpha = 5\%$) of both procedures for $a = 3, 4, 5$ levels with sample sizes $n_i \equiv 5, 10, 15, 20$ are displayed (restriction $b_1(\boldsymbol{\theta})$ upper row; restriction $b_2(\boldsymbol{\theta})$ lower row).

It can be readily seen from Figure 1 that the powers of the ANOVA and the MCTP to detect their LFCs appear to be equal. Under the restriction $b_1(\boldsymbol{\theta})$, the MCTP has a slightly higher power than the ANOVA. Hence, by offering more informations in terms of local test decisions and SCI, MCTPs are preferably applied for statistical inferences.

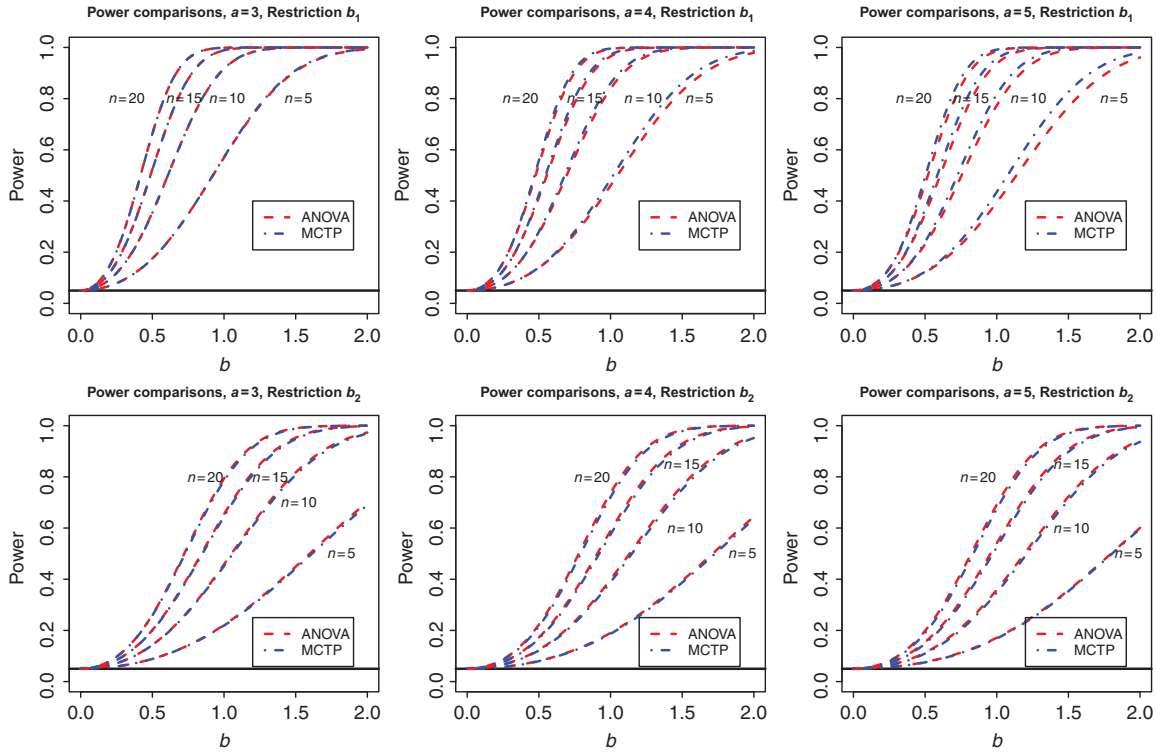

Figure 1 Power comparisons (type-I error level $\alpha = 5\%$) of the ANOVA and MCTP using the Grand-mean-type contrasts in eq. [2]: Restriction b_1 upper row; Restriction b_2 lower row.

Next, we compute the minimal required sample size to detect the LFCs for a given difference $b = 0.9$, different power levels $1 - \beta(\boldsymbol{\theta}^*)$, and different type-I error levels $\alpha = 0.01, 0.05, 0.1$ [20] and [22]. The results under the restriction $b_1(\boldsymbol{\theta})$ for the ANOVA, Grand-mean-type, and Tukey-type MCTP, respectively, are given in Table 1.

Table 1 shows that slightly smaller sample sizes are required to detect the LFC using the Grand-mean-type MCTP than with the ANOVA, particularly for increasing numbers of factor levels and decreasing α under the restriction $b_1(\boldsymbol{\theta})$. For the Tukey-type MCTP, no homogeneous behavior can be detected. In Table 2, the minimal required sample sizes for the LFC detection under the restriction $b_2(\boldsymbol{\theta})$ are displayed. The minimal sample size to detect the LFC using the ANOVA is slightly smaller than using the Grand-mean-type MCTP. The smallest sample size is revealed with the Tukey-type MCTP.

3.2 Power investigations for selected alternatives

The LFCs provide only two possible candidates among an infinite number of alternatives. In this section, we investigate the powers of the two procedures to detect different kinds of alternatives, namely

- alternative 1: $\boldsymbol{\theta} = (b, 0, \ldots, 0, b)'$
- alternative 2: $\boldsymbol{\theta} = (b, 0, \ldots, 0, 2 \cdot b)'$
- alternative 3: $\boldsymbol{\theta} = (-b, 0, \ldots, 0, 2 \cdot b)'$

with varying sample sizes $n \in \{5, 10, 15, 20\}$, numbers of factor levels $a \in \{3, 4, 5\}$, and varying values of $b, 0 \leq b \leq 2$. The results are displayed in Figure 2. It can be readily seen from Figure 2 that the powers of both procedure particularly depends on the chosen kind of alternative. The ANOVA seems to be more powerful in terms of trend patterns (alternative 1 and alternative 2), while being slightly less powerful for

Table 1 Minimal sample sizes of the ANOVA F_C in eq. [3], Grand-mean-type MCTP **T** with **C** in eq. [2], and Tukey-type MCTP in eq. [6] for given $b = 0.9$, and restriction $b_1(\boldsymbol{\theta}) = \max_{1 \le i \le a} |\theta_i - \bar{\theta}_\cdot|$

$a = 3$			$a = 4$				$a = 5$			
α	$1-\beta$	T(2)	F_C	T(6)	T(2)	F_C	T(6)	T(2)	F_C	T(6)
0.01	0.60	9.59	10.00	10.00	10.97	12.00	11.94	11.88	14.00	13.29
	0.70	11.02	11.86	11.54	12.68	14.00	13.81	13.75	16.00	15.40
	0.80	12.92	13.17	13.54	14.80	16.00	16.24	16.11	18.00	18.07
	0.90	15.78	16.00	16.60	18.15	20.00	19.89	19.71	22.00	22.17
	0.95	18.39	19.00	19.35	21.16	23.00	23.24	22.96	26.00	25.83
0.05	0.60	6.25	7.00	6.33	7.35	8.00	7.75	8.15	9.00	8.72
	0.70	7.42	8.00	7.62	8.77	10.00	9.25	9.73	11.00	10.45
	0.80	9.00	9.06	9.23	10.64	12.00	11.25	11.77	13.00	12.68
	0.90	11.45	12.00	11.76	13.51	15.00	14.35	14.92	17.00	16.14
	0.95	13.71	14.00	14.13	16.16	17.00	17.21	17.80	20.00	19.31
0.10	0.60	4.74	5.00	4.80	5.73	6.00	5.90	6.41	7.00	6.71
	0.70	5.81	6.00	5.91	7.00	8.00	7.25	7.86	9.00	8.25
	0.80	7.23	8.00	7.33	8.69	9.00	9.03	9.74	11.00	10.27
	0.90	9.45	10.00	9.64	11.35	12.00	11.84	12.66	14.00	13.41
	0.95	11.55	12.00	11.76	13.80	15.00	14.46	15.37	17.00	16.33

Table 2 Minimal sample sizes of the ANOVA F_C in eq. [3], MCTP **T** with Grand-mean contrasts **C** in eq. [2], and Tukey-type MCTP in eq. [6] for given $b = 0.9$, and restriction $b_2(\mathbf{q}) = \max_{1 \le i, j \le a} |\theta_i - \theta_j|$.

$a = 3$			$a = 4$				$a = 5$			
α	$1-\beta$	T(2)	F_C	T(6)	T(2)	F_C	T(6)	T(2)	F_C	T(6)
0.01	0.60	26.52	26.00	25.51	30.04	29.00	28.13	32.62	32.00	30.08
	0.70	31.21	31.00	29.94	35.20	34.00	32.88	38.00	37.00	35.01
	0.80	37.20	36.00	35.59	41.77	40.00	38.78	44.89	43.00	41.2
	0.90	46.35	45.00	44.25	51.60	49.00	47.88	55.34	53.00	50.65
	0.95	54.70	53.00	52.11	60.60	58.00	56.08	64.64	62.00	59.13
0.05	0.60	16.67	17.00	16.41	19.29	19.00	18.60	21.26	21.00	20.29
	0.70	20.44	20.94	20.07	23.49	23.00	22.59	25.78	25.00	24.51
	0.80	25.34	25.00	24.81	28.91	28.00	27.72	31.54	31.00	29.89
	0.90	33.01	33.00	32.23	37.28	36.00	35.65	40.42	39.00	38.19
	0.95	40.10	40.00	39.06	44.97	44.00	42.89	48.51	47.00	45.75
0.10	0.60	12.45	13.00	12.34	14.59	15.00	14.23	16.26	16.00	15.67
	0.70	15.75	16.00	15.62	18.33	18.00	17.81	20.29	20.00	19.52
	0.80	20.12	20.00	19.89	23.22	23.00	22.49	25.53	25.00	24.49
	0.90	27.03	27.00	26.65	30.84	30.00	29.80	33.63	33.00	32.2
	0.95	33.50	33.00	32.95	37.88	37.00	36.57	41.11	40.00	39.28

umbrella alternatives (alternative 3). Finally, we investigate the powers of the procedures to reject a point alternative of the form

- alternative 4: $\boldsymbol{\theta} = (0, 0, \ldots, 0, 1.35)'$
- alternative 5: $\boldsymbol{\theta} = (1.15, 0, \ldots, 0, 1.15)'$

with varying numbers of groups $a \in \{3, \ldots, 50\}$ and sample size $n = 10$. The results are displayed in Figure 3. It follows from Figure 3 that the powers of the ANOVA to reject the two chosen alternatives are monotonically decreasing in a, while the powers of the MCTP are nearly constant in a.

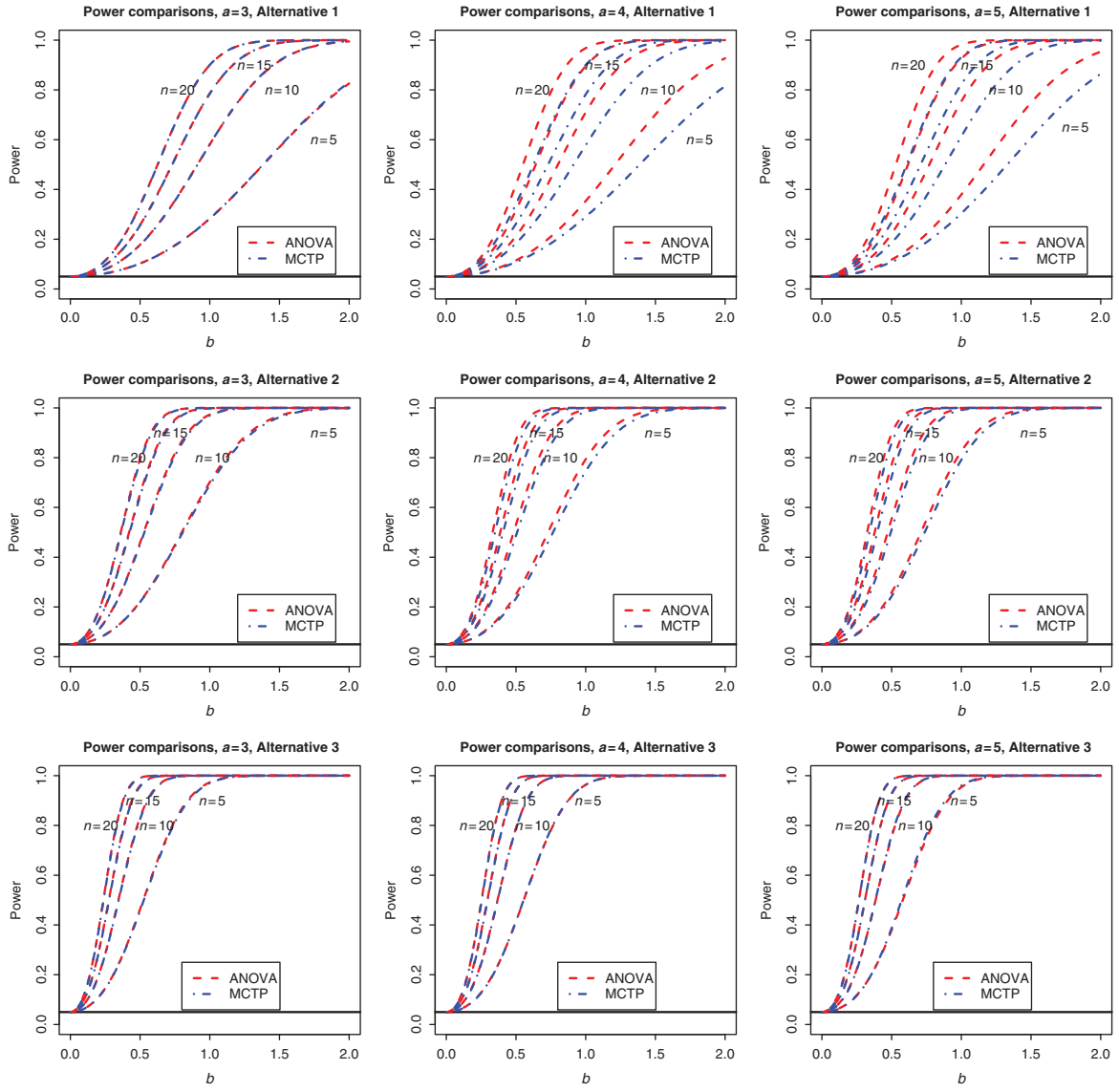

Figure 2 Power comparisons (type-I error level $\alpha = 5\%$) of the ANOVA and MCTP using the Grand-mean-type contrasts in eq. [2] to detect alternative 1 (upper row), alternative 2 (middle row), and alternative 3 (lower row).

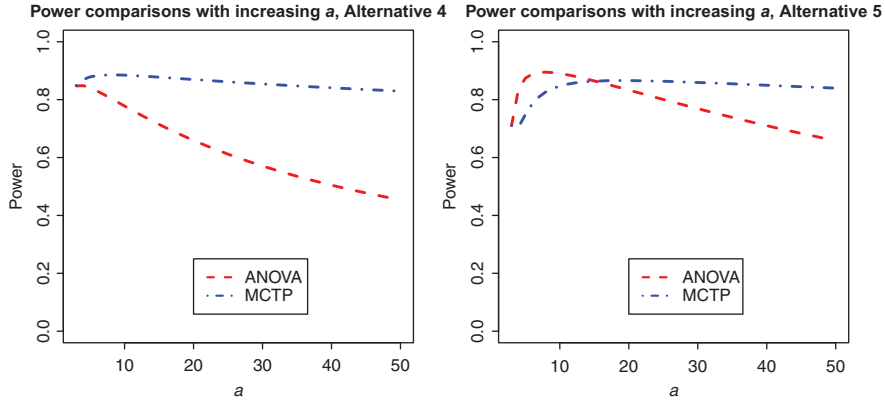

Figure 3 Power comparisons (type-I error level $\alpha = 5\%$) of the ANOVA and MCTP using the Grand-mean-type contrasts in eq. [2] to detect alternative 4: $\theta = (0, 0, \ldots, 0, 1.35)'$ and alternative 5: $\theta = (1.15, 0, \ldots, 0, 1.15)'$, each with $n = 10$, respectively.

4 Discussion

ANOVA procedures are commonly applied in statistical practice, when more than two samples are compared. They can only be used, however, to test the global null hypothesis, which is not often the main question of the practitioners. Specific informations for the local group levels in terms of multiple contrasts, adjusted p-values, and SCI are of particular practical importance. Bretz et al. [1] proposed exact MCTP and SCI which allow for arbitrary user-defined contrasts, e.g. Tukey-type [12], Dunnett-type [11], or even changepoint comparisons. Adjusted p-values and SCI for pre-defined or user-defined contrasts can be easily estimated using the R package *multcomp* [17] and *mvtnorm* [18]. These procedures provide local informations as well as SCI as required by international regulatory authorities. Thus, from a practical point of view, they are preferably applied for making statistical inferences. Since also both the MCTPs and the ANOVA-type procedures can be used to test the same overall null hypothesis, the remaining question is "How much is the price in terms of a loss in power" which needs to be paid for the additional informations offered by the MCTP. For the set of all possible kinds of alternatives, the ANOVA is a uniformly most powerful unbiased and invariant test procedure. In this article, we compared the exact power of both the MCTP and the ANOVA and we computed their LFCs to reject the global null hypothesis under two different restrictions. It turned out that both kinds of procedures have the same LFCs under the restrictions $b_1(\theta)$ and $b_2(\theta)$, respectively. Exact power calculations additionally showed that the power curves of both tests are equal. This gives a reason to claim that MCTPs are not inferior to the ANOVA. Obviously, as the LFCs are a small subset of two alternative configurations among an infinite number of possible candidates, the question "Are MCTPs superior to the ANOVA?" cannot be answered. The ANOVA is sensitive to many shapes – even for convex and concave mean profiles – whereas the MCTPs are mostly sensitive to the pre-specified kind of alternative. The ANOVA, however, cannot provide the information which factor levels cause the statistical difference. Moreover, MCTPs also provide directional decisions, whereas the quadratic form of the F-test provides only two-sided decisions.

We restricted our analysis to one-way normal designs with homoscedastic variances. The investigation of higher-way layouts, e.g. two-way ANOVA models, analysis of covariance models, etc., will be part of future research.

Appendix

Proof of Theorem 1

The proof follows the same ideas as the proof of Theorems 1 and 2 in Hayter and Liu [15]. By conditioning on the value of the random variable s^2, it is apparent that for any $\boldsymbol{\theta}, \boldsymbol{\theta}^* \in \mathbb{R}^a$,

$$W_c(\boldsymbol{\theta}) \leq W_c(\boldsymbol{\theta}^*)\ \forall c \in \mathbb{R} \in \Rightarrow \beta(\boldsymbol{\theta}^*) \leq \beta(\boldsymbol{\theta}),$$

where the function $W_c(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \mathbb{R}^a$ and $c \in \mathbb{R}$ is defined as

$$W_c(\boldsymbol{\theta}) = P(|X_i - \overline{X}_\cdot| \leq c, i = 1, \ldots, a).$$

Here, $X_i, i = 1, \ldots, a$ denote independent normal random variables with variances $1/n$ and means θ_i, respectively. Note that $W_c(\boldsymbol{\theta})$ is the multivariate $N(\boldsymbol{\theta}, R)$ distribution function, which can be computed by using the corresponding $(a-1)$-variate regular multivariate normal distribution. Now, for any $c \in \mathbb{R}$, we have the following four properties for the function $W_c(\boldsymbol{\theta})$.

1. $W_c(\boldsymbol{\theta}) = W_c(-\boldsymbol{\theta})$.
2. $W_c(\boldsymbol{\theta} + \lambda 1) = W_c(\boldsymbol{\theta})$, $\lambda \in \mathbb{R}$.
3. $W_c(\pi(\boldsymbol{\theta})) = W_c(\boldsymbol{\theta})$, where the operator π permutes coordinates.

4. $W_c(\pi(\boldsymbol{\theta}))$ is log-concave [19], i.e. for $0 \leq \gamma \leq 1$, and for all $\boldsymbol{\theta}, \boldsymbol{\theta}^* \in \mathbb{R}^a$,

$$W_c(\gamma\boldsymbol{\theta} + (1-\gamma)\boldsymbol{\theta}^*) \geq W_c^{\gamma}(\boldsymbol{\theta}) W_c^{1-\gamma}(\boldsymbol{\theta}^*).$$

The log-concavity of $W_c(\boldsymbol{\theta})$ implies by induction that for any $m \in \mathbb{N}$

5. $W_c(\sum_{i=1}^m \gamma_i \boldsymbol{\theta}(i)) \geq W_c(\boldsymbol{\theta}^{(1)})$, where $\gamma_i \geq 0$, $\sum_{i=1}^m \gamma_i = 1$ and $W_c(\boldsymbol{\theta}^{(1)}) = \cdots = W_c(\boldsymbol{\theta}^{(m)})$.
6. Properties 1 and 5 imply that $W_c(\rho\boldsymbol{\theta}) \geq W_c(\boldsymbol{\theta})$ for all $|\rho| \leq 1$.

Proof of Theorem 1.1

Suppose that $b_1(\boldsymbol{\theta}) = \theta_i - \overline{\theta}_{\cdot} = \widetilde{b} \geq b$. Let $\boldsymbol{\theta}^{(i)}, i = 1, \ldots, (a-1)!$ denote the vectors obtained by permuting $\theta_1, \ldots, \theta_{a-1}$ and leaving θ_a in place. Let $\overline{\theta}_a = \frac{1}{a-1}\sum_{i=1}^{a-1} \theta_i$ and note that $\theta_a - \overline{\theta}_a = \frac{a}{a-1}(\theta_a - \overline{\theta}) = \frac{a}{a-1}\widetilde{b}$. Now, by properties 1–6, it follows that for any $c \in \mathbb{R}$,

$$\begin{aligned} W_c(\boldsymbol{\theta}) &\overset{3.,5.}{\leq} W_c\left(\tfrac{1}{(a-1)!}\sum_{i=1}^{(a-1)!} \boldsymbol{\theta}^{(i)}\right) = W_c(\overline{\theta}_a, \ldots, \overline{\theta}_a, \theta_a) \overset{2.}{=} W_c(0, \ldots, 0, \theta_a - \overline{\theta}_a) \\ &= W_c\left(0, \ldots, 0, \tfrac{a}{a-1}\widetilde{b}\right) \overset{6.}{\leq} W_c(\boldsymbol{\theta}^*). \ \square \end{aligned}$$

Proof of Theorem 1.2

Suppose that $b_2(\boldsymbol{\theta}) = \theta_a - \theta_1 = \widetilde{b} \geq b$. Let $\boldsymbol{\theta}^{(i)}, i = 1, \ldots, (a-2)!$ denote the vectors obtained by permuting $\theta_2, \ldots, \theta_{a-1}$ and leaving θ_1 and θ_a in place. Let $\overline{\theta}_{1a} = \frac{1}{(a-2)}\sum_{i=2}^{a-1} \theta_i$. Then, by properties 1–6, it follows that for any $c \in \mathbb{R}$,

$$\begin{aligned} W_c(\boldsymbol{\theta}) &\overset{3,5}{\leq} W_c\left(\tfrac{1}{(a-2)!}\sum_{i=1}^{(a-2)!} \boldsymbol{\theta}^{(i)}\right) = W_c(\theta_1, \overline{\theta}_{1a}, \ldots, \overline{\theta}_{1a}, \theta_a) \\ &\overset{1,3}{=} W_c^{1/2}(\theta_1, \overline{\theta}_{1a}, \ldots, \overline{\theta}_{1a}, \theta_a) \cdot W_c^{1/2}(-\theta_a, -\overline{\theta}_{1a}, \ldots, -\overline{\theta}_{1a}, -\theta_1) \\ &\overset{4}{\leq} W_c\left(\tfrac{1}{2}(\theta_1, \overline{\theta}_{1a}, \ldots, \overline{\theta}_{1a}, \theta_a) + \tfrac{1}{2}(-\theta_a, -\overline{\theta}_{1a}, \ldots, -\overline{\theta}_{1a}, -\theta_1)\right) \\ &= W_c\left(-\tfrac{1}{2}\widetilde{b}, 0, \ldots, 0, \tfrac{1}{2}\widetilde{b}\right) \overset{6}{\leq} W_c(\boldsymbol{\theta}^*). \ \square \end{aligned}$$

The proof for Tukey-type comparisons is very similar and is, therefore, omitted, see Hayter and Liu [15].

Acknowledgments: The authors are grateful to an Associate Editor and two anonymous referees for helpful comments which considerably improved the article. This work was supported by the German Research Foundation projects DFG-Br 655/16–1 and Ho 1687/9–1.

References

1. Bretz F, Genz A, Hothorn LA. On the numerical availability of multiple comparison procedures. Biom J 2001;43:645–56.
2. Mukerjee H, Robertson T, Wright FT. Comparison of several treatments with a control using multiple contrasts. J Am Stat Association 1987;82:902–10.
3. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. Biom J 2008;50:346–63.
4. Hasler M, Hothorn LA. Multiple contrast tests in the presence of heteroscedasticity. Biom J 2008;50:793–800.
5. Herberich E, Sikorski J, Hothorn T. A robust procedure for comparing multiple means under heteroscedasticity in unbalanced designs. PLoS One 2010. DOI:10.1371/journal.pone.0009788.
6. Konietschke F, Hothorn LA. Evaluation of toxicological studies using a nonparametric Shirley-type trend test for comparing several dose levels with a control group. Stat Biopharm Res 2012;4:14–27.

7. Konietschke F, Hothorn LA, Brunner E. Rank-based multiple test procedures and simultaneous confidence intervals. Electron J Stat 2012;6:738–59.
8. Konietschke F, Libiger O, Hothorn LA. Nonparametric evaluation of quantitative traits in population-based association studies when the genetic model is unknown. PLoS One 2012;7:e31242. DOI:10.1371/journal.pone.0031242.
9. Arias-Castro E, Candes EJ, Plan Y.. Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. Ann Stat 2011;39:2533–56.
10. Djira GD, Hothorn LA. Detecting relative changes in multiple comparisons with an overall mean. J Qual Control 2009;41:60–5.
11. Dunnett CW. A multiple comparison procedure for comparing several treatments with a control. J Am Stat Association 1955;50:1096–121.
12. Tukey JW. The problem of multiple comparisons. Dittoed manuscript, Department of Statistics, Princeton University, Princeton, NJ, 1953.
13. Bretz F. Powerful modifications of Williams' test on trend. Ph.D. thesis, University of Hannover, 1999.
14. Genz A, Kwong KS. Numerical evaluation of singular multivariate normal distributions. J Stat Comput Simulation 2000;68:1–21.
15. Hayter AJ, Liu W. The power function of the studentised range test. Ann Stat 1990;18:465–8.
16. Hayter AJ, Liu W. A method of power assessment for tests comparing several treatments with a control. Commun Stat-Theory Meth 1992;21:1871–89.
17. Hothorn T, Bretz F, Westfall P. multcomp: simultaneous inference in general parametric models. R package version 0.8–15, 2012. Available at: http://CRAN.R-project.org/
18. Genz A, Bretz F, Tetsuhisa M, Mi X, Leisch F, Scheipl F, et al. mvtnorm: multivariate normal and t distributions. R package version 0.9–9994, 2012. Available at: http://CRAN.R-project.org/
19. Prekopa A. On logarithmic concave measures and functions. Acta Sci Mathematicarum 1973;34:335–43.
20. Horn M, Vollandt R. Sample sizes for comparisons of *k* treatments with a control based on different definitions of power. Biom J 1998;40:589–612.
21. Hsu JC. Multiple comparisons – theory and methods. London: Chapman and Hall, 1996.
22. Liu W. On sample size determination of Dunnett's procedure for comparing several treatments with a control. J Stat Plann Inference 1997;62:255–61.
23. ICH. Statistical principles for clinical trials. Guideline, international conference on harmonization, 1998. Available at: http://private.ich.org
24. Bretz F, Genz A. Numerical computation of multivariate t–probabilities with application to power calculation of multiple contrasts. J Stat Comput Simulation 1999; 63:361–78.
25. Hayter AJ, Hurn M. Power comparisons between the F-test, the studentised range test, and an optimal test of the equality of several normal means. J Stat Comput Simulation 1992;42:173–85.
26. Gabriel, KR. (1969). Simultaneous test procedures – some theory of multiple comparisons. Annals of Mathematical Statistics 40:224–250.
27. Bretz, F., Hothorn, T., Westfall, P. (2010). Multiple Comparisons Using R, CRC Press, Chapman & Hall/CRC Press, Boca Raton, Florida, USA,