Directional multivariate tests rejecting null and negative effects in all variables

Ekkehard Glimm[*,1] and **Jürgen Läuter**[2]

[1] Novartis Pharma AG, Novartis Campus, 4056 Basel, Switzerland
[2] Otto-von-Guericke-Universität Magdeburg, Mittelstr. 2/151, 39114 Magdeburg, Germany

Received 10 October 2009, revised 19 April 2010, accepted 23 May 2010

This paper suggests two directional multivariate tests that aim at establishing superiority of a treatment over a control in at least one of several endpoints that are assumed to have a multivariate normal distribution. One of these tests is a one-sided, scale-invariant version of the classical Hotelling T^2-test. The other is based on a summary score with weights derived from the data. Both tests overcome an important shortcoming of previous "one-sided" multivariate suggestions, namely that the null hypothesis was restricted to a single point in the multidimensional parameter space. The derivation of the tests is supplemented by simulations investigating their performance and by the application in an osteoporosis trial.

Key words: Directional alternatives; Multivariate tests.

Supporting Information for this article is available from the author or on the WWW under http://dx.doi.org/10.1002/bimj.200900254

1 Background and motivation

This paper deals with the problem of testing whether at least one of many variables in a multivariate distribution has a mean which is larger than zero. This problem arises, for example, in clinical trials where treatment success is assessed by many endpoints and it is desired to establish that at least one of them shows a positive response to the treatment. The problem has received considerable attention in recent years (Dunnett and Tamhane, 1992; Tamhane *et al*., 1996; Cai and Sarkar, 2006; Röhmel *et al*., 2006; Chuang-Stein *et al*., 2007). Some of the papers consider situations in which the correlation structure of the endpoints is either completely known, or has a sparse structure described by very few parameters, or is otherwise known to be restricted (e.g. assuming that all pairwise correlations are positive). We do not discuss such situations in this paper. Excluding these methods, the topic can also be approached from a multiple testing perspective by considering univariate tests for the individual endpoints and combining them on the basis of the closed test procedure (Marcus *et al*. 1976) using Bonferroni's inequality. Undoubtedly, this approach has many advantages. Namely, it is simple to implement and if there are few variables, the power loss relative to methods that assume the knowledge of the correlation between endpoints (such as, e.g. Dunnett's test, Dunnett 1955) is small. If the primary interest is in the investigation and interpretation of single variables in isolation, these methods are most appropriate. They are less appropriate and—due to the inherent conservatism of the Bonferroni adjustment—less powerful, if a deviation from the null hypothesis of no treatment effect, say, manifests itself in moderate elevations of the values of several

*Corresponding author: e-mail: ekkehard.Glimm@Novartis.com, phone: +41-61-3240173, Fax: +41-61-3243039

variables without a single dominant one. In such cases, multivariate statistical methods are more appropriate.

Traditionally, multivariate statistical inference has focused on invariant methods. For example, Hotelling's famous T^2-test is the uniformly most powerful invariant test for the mean of a multivariate normal distribution (cf. Anderson 2003). Invariant tests are characterized by the assumption that only the distance of the true parameter value from the null hypothesis, not its direction, are relevant. The advances of bioinformatics with its huge data sets of correlated variables, e.g. in gene expression data, have recently brought about a revival of research into multivariate statistics, in particular in the case of sample sizes n that are smaller than the number of variables p. Older research on this topic (Box 1954, Dempster 1958) has recently been expanded upon by several authors (Srivastava and Fujikoshi 2006, Srivastava and Du 2008) and generalized, for example, to non-parametric statistics (Munzel and Brunner 2000, Oja and Randles 2004, Bathke and Harrar 2008). These suggestions are no longer invariant. However, they are so due to technical limitations, not by purpose. In contrast, our focus in this paper is on purposefully directional multivariate tests. Furthermore, while some of the methods we explore in the sequel can be applied in the case of $n < p$, this is only a secondary aspect of the investigations. Rather, potential applications for the suggested methods will usually have $n > p$ and will also have few variables, maybe up to 10 at most.

The investigations of this paper were in part inspired by a phase II clinical trial in osteoporosis. In this trial, several endpoints measured treatment effects on systemic aspects (such as pain relief and flexibility of joints) and others physiologic aspects (such as joint space narrowing and cartilage volume). It was of course hoped that a benefit could be established regarding both of these aspects, but it was unclear if this would be the case and how strong a benefit (if existent) would manifest itself in the various endpoints. In Section 6, we will discuss the analysis of data from such a trial. For confidentiality reasons, the numbers in that section are not from the real trial.

For simplicity, we introduce our suggestions for the one-sample case. The two-sample case is a straightforward extension (see Section 4). In terms of an application, we can think of the response being a difference between a post-treatment and a baseline measurement in multiple endpoints. Assume that we observe p response variables on each of n individuals. The data are arranged in an $n \times p$ matrix $\mathbf{X}$. A row of $\mathbf{X}$ represents the p responses of an individual; hence, rows are assumed to be stochastically independent and follow a p-dimensional normal distribution with unknown mean vector $\boldsymbol{\mu} = (\mu_i)_{i=1,\ldots,p}$ and unknown covariance matrix $\boldsymbol{\Sigma}$. If there is no treatment effect on any of the endpoints, we have $\boldsymbol{\mu} = \mathbf{0}$.

Hotelling's T^2-test tests the hypothesis $H_0 : \boldsymbol{\mu} = \mathbf{0}$ against the general, "non-directional" alternative $A : \boldsymbol{\mu} \neq \mathbf{0}$. Its power only depends on the Mahalanobis distance $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ of $\boldsymbol{\mu}$ from zero. For practical applications with "directional" questions, this is often an unsuitable property, as positive and negative deviations from the null hypothesis are treated equally. Clinical trial applications of multivariate statistics are often faced with the problem of appropriately generalizing the concept of the univariate *one-sided* test. This has led to the derivation of numerous "directional" multivariate tests. All of these suggestions aim at a restricted alternative hypothesis, some of them explicitly stating the alternative, others doing so only implicitly (e.g. O'Brien 1984, Läuter 1996). Most notably, Kudo (1963), Nüesch (1966) and Perlman (1969) have derived the likelihood-ratio (LR) test of H_0 against the "one-sided" alternative $A: \boldsymbol{\mu} > \mathbf{0}$. Other restricted alternatives have also been considered. Silvapulle and Sen (2004) provide an overview.

Unfortunately, as has been pointed out by Silvapulle (1997), Perlman and Wu (2004) and Röhmel *et al.* (2006), the LR test has some serious drawbacks regarding its practical application. First, it is computationally very demanding. More seriously, the test has an intuitively unappealing property: It can lead to rejection of the null hypothesis in favor of the alternative, if all observed estimates $\bar{x}_i$ of μ_i are negative. This might happen because H_0 is restricted to a single point in space, such that the situation $\mu_i < 0$ is excluded *a priori* from consideration. Silvapulle (1997) gives a nice illustration of this shortcoming which also affects many suggestions that have been made to overcome the com-

putational complications of the LR test (e.g. Schaafsma and Smid 1966, Tang *et al.* 1989, Tang *et al.* 1993, Glimm *et al.* 2002).

In this paper, we suggest two multivariate tests that strictly keep α for the entire negative orthant. Thus, they allow to claim a statistically significant positive effect in at least one of the p response variables. In Sections 2 and 3, respectively, these two tests are introduced. After briefly discussing the two-sample case in Section 4, the power of the suggestions is compared via simulation in Section 5. In Section 6, the application to data from the osteoporosis trial is presented and discussed.

2 Directional Hotelling test

Follmann (Follmann 1995, Follmann 1996) suggested an alternative to the LR test which is particularly easy to implement. The method converts Hotelling's T^2-test into a directional test by requiring the extra condition $\bar{\mathbf{x}}' \cdot \mathbf{1}_p \geq 0$ for rejection, where $\mathbf{1}_p$ is a vector of p ones and $\bar{\mathbf{x}}$ is the usual least-squares estimate of $\boldsymbol{\mu}$. Hotelling's T^2-test accepts $H_0 : \boldsymbol{\mu} = \mathbf{0}$ if $\mathbf{0}$ lies in an ellipsoid with center $\bar{\mathbf{x}}$. By introducing the extra condition, this acceptance region is modified to a half-space and a half-ellipsoid. The T^2-test is performed at level 2α, i.e. with a contracted ellipsoid, to maintain the preassigned test level α.

Obviously, this approach avoids the case where H_0 is rejected with $\bar{x}_i < 0$ for all $i = 1, \ldots, p$. However, the issue is not entirely resolved: One cannot conclude from the rejection of $H_0 : \boldsymbol{\mu} = \mathbf{0}$ that every "shifted" hypothesis $H_0^{\boldsymbol{\mu}} : \mu_i \leq 0$, for all $i = 1, \ldots, p$ where μ_i are the elements of $\boldsymbol{\mu}$, can also be rejected. In addition, Follmann's test is not scale-invariant.

We now give a modification of Follmann's test which (i) renders it scale-invariant and (ii) uses the extended null hypothesis $H_0^{\text{orth}} : \mu_i \leq 0$ for $i = 1, \ldots, p$. Condition (ii) demands that the test level α is kept for each fixed vector $\boldsymbol{\mu}$ of the negative orthant. A weaker version of this test considers the null hypothesis $H_0^{\text{corn}} : u_i < \mu_i \leq 0$ for $i = 1, \ldots, p$ with given fixed values of u_i.

If H_0^{orth} is rejected, we can conclude that not all variables have zero or negative mean values, i.e. in at least one of the p variables there is a positive response. However, the unique identification of such a "positive variable" would demand further testing steps, for example, the application of the closure principle by Marcus *et al.* (1976).

Let

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}'_{(1)} \\ \vdots \\ \mathbf{x}'_{(n)} \end{pmatrix} \sim N_{n \times p}(\mathbf{1}_n \boldsymbol{\mu}', \mathbf{I}_n \otimes \Sigma) \tag{1}$$

be the $n \times p$ matrix of n individuals, each having observations from p endpoints with means $\boldsymbol{\mu} = (\mu_i)_{i=1,\ldots,p}$ and common positive-definite covariance matrix $\boldsymbol{\Sigma}$, where $n > p$. For the sake of convenience, notation does not distinguish between random variables and their realizations.

The usual least-squares estimates are $\bar{\mathbf{x}} = (\bar{x}_i)_{i=1,\ldots,p} = (1/n) \cdot \sum_{j=1}^{n} \mathbf{x}_{(j)}$ for $\boldsymbol{\mu}$ and $\mathbf{S} = (1/(n-1))\mathbf{G} = (1/(n-1)) \sum_{j=1}^{n} (\mathbf{x}_{(j)} - \overline{\mathbf{x}}) (\mathbf{x}_{(j)} - \overline{\mathbf{x}})'$ for $\boldsymbol{\Sigma}$. A minimum-volume $(1-\alpha)$ confidence region for $\boldsymbol{\mu}$ is given by the ellipsoid around $\bar{\mathbf{x}}$

$$C_{1-\alpha}(\bar{\mathbf{x}}, \mathbf{G}) = \{\boldsymbol{\mu}_0 \text{ with } \frac{(n-p)n}{p} (\boldsymbol{\mu}_0 - \bar{\mathbf{x}})' \mathbf{G}^{-1} (\boldsymbol{\mu}_0 - \bar{\mathbf{x}}) < F_{1-\alpha}(p, n-p)\} \tag{2}$$

where $F_{1-\alpha}(p, n-p)$ is the $(1-\alpha)$ quantile of the F-distribution with p and $n-p$ degrees of freedom. Hotelling's T^2-test rejects $H_0^{\boldsymbol{\mu}}$ if and only if $\boldsymbol{\mu} \notin C_{1-\alpha}(\bar{\mathbf{x}}, \mathbf{G})$.

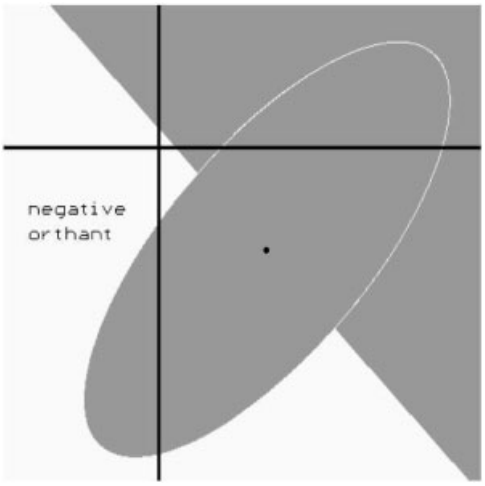

Figure 1 Confidence region corresponding to the directional Hotelling test for H_0^{orth}(Procedure I). The test fails significance, because the ellipse intersects the negative orthant ($p = 2$).

To test the extended null hypothesis H_0^{orth}, we suggest the modified directed confidence region

$$C_{1-2\alpha}(\bar{\mathbf{x}}, \mathbf{G}) \cup \left\{ \boldsymbol{\mu}_0 \text{ with } \sum_{i=1}^{p} \frac{\mu_{0i} - \bar{x}_i}{\sqrt{g_{ii}}} > 0 \right\}, \tag{3}$$

where g_{ij}, $i, j = 1,\ldots,p$, are the elements of the sums-of-products matrix $\mathbf{G}$. This way, a scale-invariant test arises. Each of the p endpoints is standardized with its corresponding standard deviation. Region (3) consists of a half-space and a half-ellipsoid (Fig. 1). The corresponding multivariate test is given by

Procedure I: H_0^{orth} is rejected if no vector $\boldsymbol{\mu}$ of the negative orthant lies in confidence region (3).

The probability that $\boldsymbol{\mu}$ is not in the directed confidence region (3) is exactly α because

(i) the probability of $\boldsymbol{\mu} \notin C_{1-2\alpha}(\bar{\mathbf{x}}, \mathbf{G})$ is 2α (for $\alpha \leq 0.5$),
(ii) the probability of $\boldsymbol{\mu} \notin \{\boldsymbol{\mu}_0 \text{ with } \sum_{i=1}^{p} (\mu_{0i} - \bar{x}_i)/\sqrt{g_{ii}} > 0\}$ is 0.5,
(iii) the boundary line of the half-space $\{\boldsymbol{\mu}_0 \text{ with } \sum_{i=1}^{p} (\mu_{0i} - \bar{x}_i)/\sqrt{g_{ii}} > 0\}$ goes through the center $\bar{\mathbf{x}}$ of the ellipsoid such that $C_{1-2\alpha}(\bar{\mathbf{x}}, \mathbf{G})$ is cut into two halves with probability mass $(1-2\alpha)/2$. Consequently, the region excluded by (3) is $0.5-(1-2\alpha)/2 = \alpha$.

This reasoning is valid for any dimension p. (ii) holds because the matrix $\mathbf{G}$ is stochastically independent of the mean vector $\bar{\mathbf{x}}$. The "multiple" rejection condition for H_0^{orth}(i.e. requiring that (3) excludes *all* $\boldsymbol{\mu}$ with $\mu_i \leq 0$) results in a further reduced significance level for each fixed $\boldsymbol{\mu}$ in the negative orthant. Hence, the test of H_0^{orth} always keeps the significance level α.

In order to reject H_0^{orth}, procedure **I** requires that $\bar{\mathbf{x}} \neq 0$ and $\sum_{i=1}^{p} \bar{x}_i/\sqrt{g_{ii}} \geq 0$ hold, such that $\bar{\mathbf{x}}$ must not be in the negative orthant. In the following, it is assumed that these conditions are met. In addition, we need to check that

$$\min_{\boldsymbol{\mu} \in \text{neg. orthant}} \frac{(n-p)n}{p} (\boldsymbol{\mu} - \bar{\mathbf{x}})' \mathbf{G}^{-1} (\boldsymbol{\mu} - \bar{\mathbf{x}}) \geq F_{1-2\alpha}(p, n-p). \tag{4}$$

Since the quadratic form in this expression is a convex function of $\boldsymbol{\mu}$, its unrestricted unique minimum is at $\boldsymbol{\mu} = \bar{\mathbf{x}}$ and its values are monotonously increasing in all directions if one moves away

from this minimum. Furthermore, convexity implies that the restricted minimum value in (4) must be on the boundary of the negative orthant, i.e. at least one element μ_i must be 0. In addition, this restricted minimum is unique. Consequently, we can find the minimum of the quadratic form within the negative orthant by repeating the following steps:

(i) Fix some μ_i to be 0.
(ii) Obtain the minimum of $(n-p)n/p(\boldsymbol{\mu}-\bar{\mathbf{x}})'\mathbf{G}^{-1}(\boldsymbol{\mu}-\overline{\mathbf{x}})$ with this restriction. This may result in some or all of the non-null μ_i's being positive.

We have to do this for all 2^p-1 possible combinations of zeros in places of μ_i's. The one ($\boldsymbol{\mu}^*$, say) that provides the minimum value among the solutions with all components $\mu_i \leq 0$ is the minimum sought in (4). In special cases, this may be $\boldsymbol{\mu}^* = \mathbf{0}$. The p-value of the test is the solution α^* of

$$\frac{(n-p)n}{p}(\boldsymbol{\mu}^*-\bar{\mathbf{x}})'\mathbf{G}^{-1}(\boldsymbol{\mu}^*-\bar{\mathbf{x}}) = F_{1-2\alpha^*}(p, n-p).$$

Obviously, we reject H_0^{orth} if $\alpha^* \leq \alpha$.

Each of the single minimization problems is easy to solve. If the p variables are partitioned into two subsets corresponding to

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \bar{\mathbf{x}} = \begin{pmatrix} \bar{\mathbf{x}}_1 \\ \bar{\mathbf{x}}_2 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{pmatrix},$$

then the minimum of the quadratic form subject to $\boldsymbol{\mu}_2 = \mathbf{0}$ is given by $\boldsymbol{\mu}_1^* = \bar{\mathbf{x}}_1 - \mathbf{G}_{12}\mathbf{G}_{22}^{-1}\bar{\mathbf{x}}_2$. This result is obtained from well-known properties of the conditional normal distribution (e.g. Mardia *et al.* (1979), Chapter 3).

It follows that in the special case of $p = 2$, H_0^{orth} can be rejected if $\bar{\mathbf{x}} \neq \mathbf{0}$, $\bar{x}_1/\sqrt{g_{11}} + \bar{x}_2/\sqrt{g_{22}} \geq 0$ and none of the three points

$$\boldsymbol{\mu} = \mathbf{0}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mathbf{0} \\ \bar{\mathbf{x}}_2 - \mathbf{g_{21}}\mathbf{g_{11}^{-1}}\,\bar{\mathbf{x}}_1 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \bar{\mathbf{x}}_1 - \mathbf{g_{12}}\mathbf{g_{22}^{-1}}\,\bar{\mathbf{x}}_2 \\ \mathbf{0} \end{pmatrix}$$

is in the negative orthant as well as in the $(1-2\alpha)$ ellipse.

In comparison with Follmann's test, our proposal extends the test's acceptance region by certain points $\boldsymbol{\mu}$ in the intersection of ellipsoid and negative orthant (see Fig. 1). Naturally, this is only possible at the expense of some power. In some cases, the weaker null hypothesis H_0^{corn} mentioned above can be applied. Namely, if we know in advance that there are lower limits u_i for the parameters μ_i, we only have to check whether a vector $\boldsymbol{\mu}$ of the corresponding negative corner lies in the directed confidence region (3). Thus the power of the test can be increased.

3 Standardized sum test for H_0^{orth}

Läuter and co-workers (Läuter 1996, Läuter *et al.* 1996, Läuter *et al.* 1998) introduced the concept of spherical multivariate tests. These are exact multivariate tests of $H_0 : \boldsymbol{\mu} = \mathbf{0}$ based on low-dimensional scores calculated from the observed data. Their main advantage is that they can be used with arbitrarily high dimension p. In particular, p may be larger than the sample size n. The tests are not affine-invariant. Just like all the tests discussed in the introduction, however, these tests cannot be considered as tests of H_0^{orth}. With them it is also possible that for given data $\mathbf{X}$, H_0 is rejected, but $H_0^{\boldsymbol{\mu}_0} : \boldsymbol{\mu} = \boldsymbol{\mu}_0$is not for some $\boldsymbol{\mu}_0$ of the negative orthant.

Here, we will consider only one class of spherical tests, the so-called standardized sum (SS) tests. We assume that the covariance matrix $\mathbf{\Sigma}$ has positive diagonal elements $\sigma_{11},\ldots,\sigma_{pp}$. Positive definiteness of $\mathbf{\Sigma}$ is not required. The usual SS test in the one-sided version rejects H_0 if

$$t^0 = \sqrt{n-1}\frac{\sqrt{n}\bar{\mathbf{x}}'\mathbf{d}^0}{\sqrt{\mathbf{d}^{0\prime}\mathbf{G}\mathbf{d}^0}} \geq t_{1-\alpha}(n-1) \tag{5}$$

where $t_{1-\alpha}\ (n-1)$ is the $(1-\alpha)$ quantile of the t-distribution with $n-1$ degrees of freedom and $\mathbf{d}^0 = \left(1/(\sqrt{g_{ii}+n\,\bar{x}_i^2})\right)_{i=1,\ldots,p}$. The test is scale-invariant and has high power if all variables have roughly the same positive deviation from the null hypothesis in their respective scale and roughly equal correlations with each other (Läuter *et al.* 1996).

The corresponding test for $H_0^{\boldsymbol{\mu}_0} : \boldsymbol{\mu} = \boldsymbol{\mu}_0$is

$$t = \sqrt{n-1}\frac{\sqrt{n}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{d}}{\sqrt{\mathbf{d}'\mathbf{G}\mathbf{d}}} \geq t_{1-\alpha}(n-1) \tag{6}$$

with $\mathbf{d} = \left(1/\sqrt{g_{ii}+n(\bar{x}_i-\mu_{0i})^2}\right)_{i=1,\ldots,p}$. Unfortunately, t does not generally increase if the parameters μ_{0i} decrease. Hence, rejection of H_0 does not translate into rejection of $H_0^{\boldsymbol{\mu}_0}$ for all $\boldsymbol{\mu}_0$ of the negative orthant.

Additional conditions are necessary to establish rejection of H_0^{orth}. In the following, a corresponding modification of test (5) will be derived. Let $\xi_1,\ldots,\xi_p$ be angles defined for $g_{ii}>0$ by

$$0<\xi_i<\pi, \quad \sin\xi_i = \sqrt{g_{ii}}d_i, \quad \cos\xi_i = \sqrt{n}(\bar{x}_i - \mu_{0i})d_i \tag{7}$$

for $i=1,\ldots,p$. Then the test (6) can be written as:

$$t = \frac{\sqrt{n-1}\sum_{i=1}^p \sqrt{n}(x_i-\mu_{0i})d_i}{\sqrt{\sum_{i=1}^p\sum_{h=1}^p d_h\sqrt{g_{hh}}r_{hi}\sqrt{g_{ii}}d_i}} = \frac{\sqrt{n-1}\sum_{i=1}^p cos\ \xi_i}{\sqrt{\sum_{i=1}^p\sum_{h=1}^p sin\ \xi_h r_{hi} sin\ \xi_i}} \geq t_{1-\alpha}(n-1). \tag{8}$$

Here, $r_{hi} = g_{hi}/\sqrt{g_{hh}g_{ii}}$ denotes the correlation coefficients from the residual matrix $\mathbf{G}$. Writing $\mathbf{R} = (r_{hi})_{h,i=1,\ldots,p}$, $\cos\xi = (\cos\xi_i)_{i=1,\ldots,p}$ and $\sin\xi = (\sin\xi_i)_{i=1,\ldots,p}$, (8) becomes

$$t = \sqrt{n-1}\frac{(\cos\xi)'1_p}{\sqrt{(\sin\xi)'\,\mathbf{R}\sin\xi}} \geq t_{1-\alpha}(n-1). \tag{9}$$

It is important to note that the parameters μ_{0i} are contained in the angles ξ_i, but do not appear otherwise. The test of H_0^{orth} is significant if inequality (9) holds for all $\boldsymbol{\mu}_0$ of the negative orthant.

As can be seen from (7), $\cot\xi_i = cos\ \xi_i/sin\ \xi_i$ increases and ξ_i decreases for fixed $\mathbf{G}$ if $\bar{x}_i - \mu_{0i}$ increases. Hence, the numerator $(\cos\xi)'\mathbf{1}_p$ of the t-ratio increases if $\boldsymbol{\mu}_0$ is moved from $\mathbf{0}$ into the negative orthant.

In contrast, the denominator can decrease or increase with such a move of $\boldsymbol{\mu}_0$, depending on the specific values of $\mathbf{R}$ and $\bar{\mathbf{x}}$. If all r_{hi} and all $\bar{x}_i$ are non-negative in an application, then $\xi_i \leq \pi/2$ holds such that $(\sin\xi)'\mathbf{R}\sin\xi$ decreases as μ_i decreases. Therefore, the rejection of $\boldsymbol{\mu}_0 = \mathbf{0}$ also implies that the rejection of all $\boldsymbol{\mu}_0$ in the negative orthant and the test of H_0^{orth} is finished. If, however, negative values r_{hi} or negative $\bar{x}_i$ arise, additional checks are necessary.

Let ξ^0 denote the vector of angles ξ_i for $\boldsymbol{\mu}_0 = \mathbf{0}$ and let ξ^{0+} be the modified vector obtained by replacing all components of ξ^0 which are larger than $\pi/2$ by $\pi/2$. Likewise, let $\mathbf{R}^+$ be the matrix obtained from $\mathbf{R}$ by replacing all negative elements with 0. Then we have the inequality

$$(\sin\xi)'\mathbf{R}\sin\xi \leq (\sin\xi^{0^+})'\mathbf{R}^+\sin\xi^{0^+}$$

in the negative orthant. Consequently,

$$t^{\text{orth}} = \sqrt{n-1}\frac{(\cos \xi^0)'1_p}{\sqrt{(\sin \xi^{0^+})'\mathbf{R}^+ \sin \xi^{0^+}}} \geq t_{1-\alpha}(n-1). \tag{10}$$

is a sufficient condition for the rejection of $H_0^{\text{orth}}(\alpha < 0.5)$.
To state the test procedure, we rewrite (10) without the angles ξ_i:
Procedure IIa: Reject H_0^{orth} if

$$t^{\text{orth}} = \sqrt{n-1}\frac{\sqrt{n}\bar{\mathbf{x}}'\mathbf{d}^0}{\sqrt{\mathbf{d}^{0^+\prime}\mathbf{G}^+\mathbf{d}^{0^+}}} \geq t_{1-\alpha}(n-1). \tag{11}$$

This provides a conservative test ($\alpha < 0.5$) of H_0^{orth}. Here, $\mathbf{d}^{0^+}$ is based on $\mathbf{d}^0$ with all $d_i^0 = 1/\sqrt{g_{ii}+n\bar{x}_i^2}$ replaced by $1/\sqrt{g_{ii}}$ if $\bar{x}_i < 0$ and $\mathbf{G}^+$ is $\mathbf{G}$ with all negative elements replaced by 0. This is the desired sharpening modification of the usual SS test (5).

A more conservative simplification of this test is given by
Procedure IIb: Reject H_0^{orth} if

$$\tilde{t}^{\text{orth}} = \sqrt{n-1}\frac{(\cos \xi^0)'\mathbf{1}_p}{\sqrt{\mathbf{1}_p'\,\mathbf{R}^+\mathbf{1}_p}} = \sqrt{n-1}\frac{\sqrt{n}\bar{\mathbf{x}}'\mathbf{d}^0}{\sqrt{\mathbf{1}_p'\,\mathbf{R}^+\mathbf{1}_p}} \geq t_{1-\alpha}(n-1). \tag{12}$$

Röhmel *et al.* (2006) have done a more detailed investigation of the case $p = 2$. They show that in order to establish significance with level α controlled in the whole negative orthant, it is sufficient to check the validity of the inequality (6) only for $\boldsymbol{\mu}_0 = \mathbf{0}$ and additionally, if $g_{12} < 0$, for the "vertices"

$$\boldsymbol{\mu}_0 = \begin{pmatrix} -\infty \\ 0 \end{pmatrix}, \quad \boldsymbol{\mu}_0 = \begin{pmatrix} 0 \\ -\infty \end{pmatrix}. \tag{13}$$

This results in the following modification:

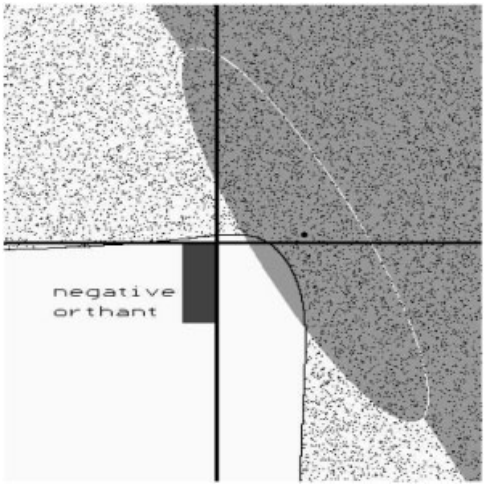

Figure 2 One-sided confidence region of the SS test (dotted area) and of the directional Hotelling test (light-grey half-plane and half-ellipse). Rectangle (dark-grey) corresponding to a restricted null hypothesis H_0^{corn}.

Procedure IIc: ($p = 2$ only) Reject H_0^{orth} if (5) holds and either $g_{12} \geq 0$, or (in case of $g_{12} < 0$)

$$\begin{aligned} t &= \sqrt{n-1}\,\frac{1+\sqrt{n}\,\bar{x}_2\,d_2^0}{\sqrt{g_{22}}d_2^0} \geq t_{1-\alpha}(n-1) \quad \text{and} \\ t &= \sqrt{n-1}\,\frac{\sqrt{n}\,\bar{x}_1\,d_1^0+1}{\sqrt{g_{11}}d_1^0} \geq t_{1-\alpha}(n-1). \end{aligned} \tag{14}$$

Procedure IIc thus provides an "optimal testing rule" in case of $p = 2$. In case of a restricted negative orthant, Procedure IIc can be modified to testing $H_0 : \boldsymbol{\mu}_0 = \mathbf{0}$ with (5) and additionally, if $g_{12} < 0$, the vertices

$$\boldsymbol{\mu}_0 = \begin{pmatrix} u_1 \\ 0 \end{pmatrix}, \quad \boldsymbol{\mu}_0 = \begin{pmatrix} 0 \\ u_2 \end{pmatrix} \tag{15}$$

instead of (13).

Figure 2 shows the one-sided confidence regions and, correspondingly, the directional rejection regions of the SS test and the directional Hotelling test from Section 2 for an example with $p = 2$, $n = 5$. In this example, the covariance g_{12} is negative, and the SS test has a non-monotone behaviour: The small rectangle in the corner of the negative orthant belongs to the rejection region of the SS test so that the corresponding restricted null hypothesis H_0^{corn} is rejected. In contrast, H_0^{orth} cannot be rejected because parts of the whole negative orthant are intersecting with the SS test confidence region. The one-sided confidence region of the Hotelling test, a half-plane and a half-ellipse, does not intersect with the negative orthant. Therefore, the one-sided Hotelling test rejects H_0.

The monotonicity investigations performed here are "pointwise" for fixed values of $\bar{\mathbf{x}}$ and $\mathbf{G}$, not taking into account the multivariate distribution at hand. It is conceivable that the one-sided SS test (5) for $H_0 : \boldsymbol{\mu} = \mathbf{0}$ keeps the α-level when applied to normally distributed data without any modification for any hypothesis $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ with $\boldsymbol{\mu}_0$ in the negative orthant. This question has not yet been settled.

It is also worth noting that in practical applications with few variables and large sample sizes, the additional checks discussed in this section will very rarely play a role. If all correlations between variables are positive, no additional check is necessary. In the simulations done by Röhmel *et al.* (2006) for $p = 2$, there was no case where the additional vertex check (13) was necessary due to a non-monotonicity of the test statistic (5). As Röhmel *et al.* (2006) also discuss, the reason for this is that such monotonicities can only arise in cases with strong negative correlations and one variable having a very large effect.

4 The two-sample case

The previous sections have implicitly covered the two-sample case as well. Suppose there are n_k observations $\mathbf{x}_{(jk)} \sim N_p(\boldsymbol{\mu}_k,\ \boldsymbol{\Sigma})$ in groups $k = 1,\ 2$ and inference is concerned with tests of $H_0^{orth} : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\mu}_0$, where $\boldsymbol{\mu}_0$ is in the negative orthant, and with corresponding confidence regions for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. All methods presented in Sections 2 and 3 are based on the complete, sufficient statistics $\bar{\mathbf{x}}$ and $\mathbf{G}$. In the two-sample case, these are essentially the same, if we re-define $\bar{\mathbf{x}}$ as $\bar{\mathbf{x}} := \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2$ and $\mathbf{G}$ as $\mathbf{G} : = \mathbf{G}_1 + \mathbf{G}_2$, where $\bar{\mathbf{x}}_k = \frac{1}{n_k}\Sigma_{j=1}^{n_k}\mathbf{x}_{(jk)}$ and $\mathbf{G}_k = \Sigma_{j=1}^{n_k}(\mathbf{x}_{(jk)} - \bar{\mathbf{x}}_k)\,(\mathbf{x}_{(jk)} - \bar{\mathbf{x}}_k)'$. The two-sample case is thus handled by using these re-definitions in Sections 2 and 3. The only other modifications necessary are a change of a constant in the total sums-of-products matrix and a change in the denominator degrees of freedom of F- and t-statistics, respectively. Regarding the former, $\mathbf{G} + n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'$ has to be replaced by $\mathbf{G}_1 + \mathbf{G}_2 + (n_1 n_2/(n_1 + n2))(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - \boldsymbol{\mu}_0)(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - \boldsymbol{\mu}_0)'$ in other words, n is replaced by $n_1 n_2/(n_1 + n_2)$. Regarding degrees of freedom, $n-1$ needs to be changed to

n_1+n_2-2. The minor modifications this requires in the previous sections can be summarized as follows:

(i) In Section 2, formula (2), $(n-p)n/p$ has to be changed to $((n_1+n_2-1-p)/p)n_1n_2/(n_1+n_2)$ and the numerator degrees of freedom of the F-quantile from $n-p$ to n_1+n_2-1-p in addition to re-definition of $\bar{\mathbf{x}}$ and $\mathbf{G}$.
(ii) In Section 3, "n" has to be replaced with $n_1n_2/(n_1+n_2)$ and "$n-1$" with n_1+n_2-2 in addition to re-definition of $\bar{\mathbf{x}}$ and $\mathbf{G}$.

5 Simulation results

This section presents results obtained from comparing procedures I and II from Sections 2 and 3 by simulation. As mentioned previously, the procedures are not primarily intended for "huge" dimensions as they occur, for example, in microarray analyses with thousands of variables. First of all, the directional Hotelling test from Section 2 requires $n > p$. The standardized sum test from Section 3 does not require this, but directional hypotheses like H_0^{orth} are rarely relevant in applications with very many variables. The spherical multivariate tests introduced by Läuter and co-workers (Läuter 1996, Läuter et al. 1996, Läuter et al. 1998) are appropriate for large p.

We have tried to set up a few simulation scenarios as a compromise between situations that are likely in practical applications and cases that highlight specific properties of the suggested methods. Thus, we are restricting our attention to cases where all correlations between variables are positive and where the direction of deviations from the null hypothesis is the same in all variables. The results are summarized in the following Tables 1–5. All values are the results of 100 000 simulation runs.

Tables 1 and 2 investigate the power of the suggested tests in the two-sample case with $p=4$ variables. In these two tables, the Mahalanobis distance $\Delta^2=(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$ is set to a fixed value. Table 1 has the results for equally correlated, equally informative variables. Here and subsequently, we use the term "informative" to indicate "distance from the null hypothesis", e.g. in this case, all variables contribute equally to the Mahalanobis-distance from $\boldsymbol{\mu}_1-\boldsymbol{\mu}_2=\mathbf{0}$ (they are "equally far apart" from the null hypothesis). As expected, the SS test performs very well in this situation. For the low sample size of $n_1=n_2=6$, the power of the directional Hotelling test suffers from instability problems that typically occur with the ordinary Hotelling test as well if the dimension is large in comparison with the sample size.

Table 2 has two equally informative variables and two that simply represent additional "noise" (no group differences and no correlation with each other and the two informative variables). In this

Table 1 Power of two-sample tests, $p=4$, $\alpha=0.05$, $\boldsymbol{\mu}_1-\boldsymbol{\mu}_2 \propto \mathbf{1}_p$, $\Sigma=(1-\rho)\cdot\mathbf{I}_p+\rho\cdot\mathbf{1}_p\mathbf{1}_p$

ρ	$n_1=n_2=6$, $\Delta^2=4$			$n_1=n_2=20$, $\Delta^2=1$		
	Direct. T^2	SS proc. IIa	SS proc. IIb	Direct. T^2	SS proc. IIa	SS proc. IIb
0	0.645	0.892	0.847	0.762	0.903	0.896
0.1	0.628	0.919	0.883	0.755	0.919	0.913
0.2	0.616	0.931	0.897	0.749	0.926	0.920
0.4	0.589	0.941	0.904	0.731	0.925	0.919
0.6	0.554	0.942	0.897	0.716	0.928	0.920
0.9	0.466	0.941	0.883	0.657	0.927	0.918

Table 2 Power of two-sample tests, $p = 4$, $\alpha = 0.05$, $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \propto (1\ 1\ 0\ 0)'$, correlation ρ between $\mathbf{x}_1$ and $\mathbf{x}_2$, 0 otherwise.

ρ	$n_1 = n_2 = 8$, $\Delta^2 = 4$			$n_1 = n_2 = 20$, $\Delta^2 = 1$		
	Direct. T^2	SS proc. IIa	SS proc. IIb	Direct. T^2	SS proc. IIa	SS proc. IIb
0	0.797	0.684	0.596	0.701	0.626	0.602
0.1	0.795	0.709	0.615	0.698	0.649	0.625
0.2	0.797	0.733	0.636	0.702	0.673	0.647
0.4	0.794	0.763	0.655	0.703	0.703	0.675
0.6	0.792	0.786	0.666	0.696	0.731	0.700
0.9	0.773	0.813	0.675	0.678	0.754	0.720

Table 3 Rejection probabilities of two-sample tests under H_0: $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, $p = 4$, $= 0.05$, $n_1 = n_2 = 20$.

ρ	balanced covariances (as in Table 1)			Unbalanced covariances (as in Table 2)		
	Direct. T^2	SS proc. IIa	SS proc. IIb	Direct. T^2	SS proc. IIa	SS proc. IIb
0	0.025	0.035	0.033	0.026	0.035	0.033
0.1	0.022	0.044	0.041	0.025	0.036	0.033
0.2	0.019	0.048	0.044	0.024	0.037	0.034
0.4	0.014	0.050	0.046	0.023	0.038	0.035
0.6	0.011	0.049	0.044	0.021	0.039	0.035
0.9	0.005	0.049	0.044	0.018	0.040	0.036

Table 4 Power of two-sample tests, $p = 2$, $\alpha = 0.05$, $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = (2\ 0)'$, $\boldsymbol{\Sigma} = \mathbf{I}_2$.

$n_1 = n_2$	Direct. T^2	SS (5)	SS proc. IIc	SS proc. IIa	SS proc. IIb
2	0.097	0.148	0.101	0.077	0
3	0.333	0.260	0.244	0.190	0.036
4	0.555	0.374	0.370	0.305	0.146
5	0.716	0.470	0.469	0.414	0.255
6	0.827	0.561	0.561	0.507	0.358
7	0.896	0.638	0.638	0.592	0.451
8	0.939	0.704	0.704	0.665	0.535
9	0.966	0.762	0.762	0.726	0.610
10	0.980	0.809	0.809	0.777	0.676

case, the SS test and the directional Hotelling test are similar in their performance with the SS test having slight advantages with highly correlated variables and the directional Hotelling performing a little better when correlations are low.

Table 3 shows the probability of rejection for the covariances from Tables 1 and 2 if $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ is true. The nominal level α is 5%, but all tests investigated here have to keep α for the composite hypothesis H_0^{orth}, not just H_0, so it is no surprise that α is not exhausted at $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. The tests are conservative in the sense that they do not exhaust the α-level anywhere. However, when correlations are high, the SS test procedure IIa comes very close to doing so. In contrast, the rejection prob-

Table 5 Rejection probability of two-sample tests under H_0: $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, $\boldsymbol{\Sigma} = \mathbf{I}_2$, $p = 2$, $\alpha = 0.05$.

$n_1 = n_2$	Direct. T^2	SS (5)	SS proc. IIc	SS proc. IIa	SS proc. IIb
2	0.027	0.050	0.020	0.012	0
3	0.034	0.050	0.046	0.024	0.003
4	0.037	0.050	0.050	0.030	0.011
5	0.038	0.051	0.051	0.034	0.018
6	0.038	0.050	0.050	0.035	0.022
7	0.039	0.050	0.050	0.037	0.026
8	0.039	0.050	0.050	0.040	0.029
9	0.039	0.049	0.049	0.040	0.031
10	0.040	0.049	0.049	0.042	0.034

ability of the directional Hotelling test decreases with increasing positive correlations between the variables.

Table 4 shows how much power is lost by the orthant-related modifications of the directional SS test and the corresponding simplifications discussed in Section 3. The special case of $p = 2$ with one informative variable and one uncorrelated uninformative variable is considered. For $p = 2$, the "optimal" procedure IIc is available. This rule has less power than the original SS test (5) alone for extremely small sample sizes $n_i \leq 5$, but test (5) might not keep the level α for the entire negative orthant. For still very moderate sample sizes of $n_i > 5$, our simulations did not find any power loss due to the additional vertex checks. This is in line with Röhmel et al. (2006). In addition, the table gives the power of the more conservative directional test procedures IIa and IIb. Note that in any practical application, one would be allowed to do the simplest test procedure IIb first, if it is not significant, try procedure IIa and if this does not yield significance either, in case of $p = 2$, finally try procedure IIc. The purpose of Table 4 is not a power comparison between SS- and directional T^2 test. It is clear that in the situation of one informative and one uninformative variable, the SS test is inferior. The simulation results of the directed T^2 test reflect this.

In analogy to Table 3, Table 5 shows the rejection probabilities under H_0 in case of two independent variables with equal variance.

6 Application in a clinical trial

In a phase II clinical trial on an osteoporosis drug with two treatment groups (treatment and control), it was initially unclear whether the benefit of a new treatment over standard treatment would primarily be

(i) physiologic improvement of the knee, measured by joint space width (JSW) in mm,
(ii) better pain relief, measured by a pain score, or
(iii) better functional ability, measured by a function score.

Thus, the focus of this phase II trial was to establish a benefit in at least one of these indicators. A future phase III trial would then focus on the most promising variables.

For the corresponding tests, a level of $\alpha = 0.05$ was selected. The trial was performed with 32 patients per group. The three endpoints were investigated as change from baseline after 3 months of treatment. In all three variables, positive values indicate an improvement. It was expected that the treatment would yield better results in all three endpoints and that all three endpoints would be positively

Table 6 Observed means and covariances in the osteoporosis trial.

	Means			Covariance
Treatment	JSW	Pain score	Function score	
New	0.43	12.1	63.6	$\begin{pmatrix} 0.38 & 17.0 & 43 \\ 17.0 & 2763 & 3257 \\ 43 & 3257 & 12042 \end{pmatrix}$
Control	0.08	14.4	83.0	$\begin{pmatrix} 0.17 & 8.4 & 20 \\ 8.4 & 2752 & 2043 \\ 20 & 2043 & 7572 \end{pmatrix}$
Means : difference Covariance : pooled estimate	0.35	−2.4	−19.4	$\begin{pmatrix} 0.27 & 12.7 & 32 \\ 12.7 & 2758 & 2650 \\ 32 & 2758 & 9807 \end{pmatrix}$

correlated. However, it was suspected that trial duration might be too short for the pain and the function scores, resulting in large variability of them as well as in a lack of positive treatment effect.

Table 6 shows the results of the trial. It is clear that the three endpoints are not on the same scale. Only JSW produced a result in accord with expectations. Contrary to expectations, the results of the two scores turned out to be worse on average in the treatment group than in the control group. The pooled estimate of the correlation between the three endpoints was

$$\begin{pmatrix} 1 & 0.46 & 0.61 \\ 0.46 & 1 & 0.51 \\ 0.61 & 0.51 & 1 \end{pmatrix}.$$

Let $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ denote the treatment and the control mean, respectively, $\mathbf{S}$ the pooled covariance estimate, $\mathbf{G} = (n_1+n_2-2)\mathbf{S}$, and $\bar{\mathbf{x}} = \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2$the mean difference between treatment and control. The T^2-statistic for $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ yields a value of

$$\frac{(n_1 + n_2 - p - 1)}{p(n_1 + n_2 - 1)} \frac{n_1 n_2}{n_1 + n_2} \bar{\mathbf{x}}' \mathbf{S}^{-1} \bar{\mathbf{x}} = 5.37$$

which is larger than the critical value $F_{1-2\alpha}\ (p,\ n_1+n_2-p-1) = 2.18$. The corresponding p-value is 0.0024. Regarding the required additional checks, we have $\bar{\mathbf{x}} \neq 0$ and $\sum_{i=1}^{p} \bar{x}_i / \sqrt{g_{ii}} = 0.422 \geq 0$. Follmann's criterion $\bar{\mathbf{x}}' \cdot \mathbf{1}_p \geq 0$ fails here due to the different scales of the endpoints.

Finally, investigation of $C_{1-2\alpha}(\bar{\mathbf{x}}, \mathbf{G})$ reveals that it does not intersect with the negative orthant. Consequently, we can conclude that the new treatment is superior to the standard treatment in at least one of the three endpoints. The minimal value of the quadratic form in the negative orthant is 2.27. It is attained at $\boldsymbol{\mu}_0 = (0,\ -18.4,\ -59.5)'$. The corresponding p-value of the directed test is 0.0447.

As an additional aspect of the application of the methodology presented here, we must of course verify that the treatment does not cause harm in one of the endpoints. Here, this was covered by separate non-inferiority tests on all three endpoints. These are not discussed in this paper. We note, however, that as a consequence, the directional multivariate tests could have been applied with the "weaker" null hypothesis H_0^{corn} using the non-inferiority margins as u_i's.

The SS test is primarily designed to have high power against alternatives where all variables have approximately the same deviation from the null hypothesis in their respective scales. Thus, it is no surprise that it does not work well in this application. The usual SS test (5) yields a t^0 value of 0.639 here. This corresponds to a p-value of 0.2625. The modified SS test procedure IIa is almost identical with a t^{orth} value of 0.638 and a p-value of 0.2629.

7 Discussion

This paper suggests two new multivariate tests for establishing that at least one of several endpoints in a clinical trial shows a beneficial treatment effect. It therefore fills a gap in existing multivariate test approaches, since these only consider a "single point" null hypothesis (like $H_0 : \boldsymbol{\mu} = 0$) which allows no claim about other undesired parameter constellations (like all $\mu_i < 0$).

In comparison with multiple testing approaches (like the Bonferroni method), the new tests have most power if the treatment effect is roughly equally strong in all variables, for example, if all variables are subject to an underlying common treatment effect. If the treatment effect is not "evenly spread" across all variables in this way, but rather there is a single variable with a strong treatment effect, then multiple testing procedures are superior. The SS test in particular has good power if all variables are equally far away from the null hypothesis and have equal pairwise correlations. This is well known and investigated for multivariate methods in general (Srivastava 2005). Since in this respect, the methods suggested here are no different from other multivariate methods, we did not do extensive simulations of these aspects, but rather concentrated on investigating the price to be paid for extending the multivariate test decision to the negative orthant. Especially for the SS test, this price is very small and in the vast majority of concrete examples, there will be no difference between the unmodified and the modified versions of this test.

In clinical trial applications, a significant result of the new tests allows to conclude that in at least one endpoint there is a *beneficial* treatment effect, and not just an effect. The limits of this interpretational extension should be acknowledged. Since a significant result does not rule out the possibility of a harmful treatment effect in some endpoints, the new suggestions are not appropriate for confirmatory clinical trials which require a positive effect on all co-primary endpoints. Nevertheless, we believe that the extended conclusions facilitated by the new suggestions are of real, practically relevant value in earlier phases of clinical development when there still is a number of candidate endpoints.

Acknowledgements The authors thank two anonymous referees for their insightful comments that helped in improving the paper.

Conflict of Interest
The authors have declared no conflict of interest.

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis* (3rd edn). Wiley, New York.

Bathke, A. C. and Harrar, S. W. (2008). Nonparametric methods in multivariate factorial designs for large number of factor levels. *Journal of Statistical Planning and Inference* **138**, 588–610.

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.

Cai, G. and Sarkar, S. K. (2006). Modified Simes' critical values under positive dependence. *Journal of Statistical Planning and Inference* **136**, 4129–4146.

Chuang-Stein, C., Stryszak, P., Dmitrienko, A. and Offen, W. (2007). Challenge of multiple co-primary endpoints: A new approach. *Statistics in Medicine* **26**, 1181–1192.
Dempster, A. P. (1958). A high dimensional two sample significance test. *Annals of Mathematical Statistics* **29**, 995–1010.
Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* **50**, 1096–1121.
Dunnett, C. W. and Tamhane, A. C. (1992). A step-up multiple test procedure. *Journal of the American Statistical Association* **87**, 162–170.
Follmann, D. (1995). Multivariate Tests for multiple endpoints in clinical trials. *Statistics in Medicine* **14**, 1163–1175.
Follmann, D. (1996). A simple multivariate test for one-sided alternatives. *Journal of the American Statistical Association* **91**, 854–861.
Glimm, E., Srivastava, M. S. and Läuter, J. (2002). Multivariate tests of normal mean vectors with restricted alternatives. *Communications in Statistics—Simulation and Computation* **31**, 589–604.
Kudo, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* **50**, 403–418.
Läuter, J. (1996). Exact t and F tests for analyzing studies with multiple endpoints. *Biometrics* **52**, 964–970.
Läuter, J., Glimm, E. and Kropf, S. (1996). New multivariate tests for data with an inherent structure. *Biometrical Journal* **38**, 5–23.
Läuter, J., Glimm, E. and Kropf, S. (1998). Multivariate tests based on left-spherically distributed linear scores. *Annals of Statistics* **26**, 1972–1988.
Marcus, R., Peritz, E., Gabriel and K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63**, 655–660.
Mardia, K. V., Kent, J. T., Bibby and M. J. (1979). *Multivariate Analysis*. Academic Press, London.
Munzel, U. and Brunner, E. (2000). Nonparametric methods in multivariate factorial designs. *Journal of Statistical Planning and Inference* **88**, 117–132.
Nüesch, P. (1966). On the problem of testing location in multivariate problems for restricted alternatives. *Annals of Mathematical Statistics* **37**, 113–119.
O'Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics* **40**, 1079–1087.
Oja, H. and Randles, R. H. (2004). Multivariate nonparametric tests. *Statistical Science* **19**, 598–605.
Perlman, M. D. (1969). One-sided testing problems in multivariate analysis. *Annals of Mathematical Statistics* **40**, 549–567.
Perlman, M. D. and Wu, L. (2004). A note on one-sided tests with multiple endpoints. *Biometrics* **60**, 276–280.
Röhmel, J., Benda, N., Gerlinger, C. and Läuter, J. (2006). On testing simultaneously non-inferiority in two multiple primary endpoints and superiority in at least one of them. *Biometrical Journal* **39**, 1–18.
Schaafsma, W. and Smid, L. J. (1966). Most stringent somewhere most powerful tests against alternatives restricted by a number of inequalities. *Annals of Mathematical Statistics* **37**, 1161–1172.
Silvapulle, M. J. (1997). A curious example involving the likelihood ratio test against one-sided alternatives. *The American Statistician* **51**, 178–181.
Silvapulle, M. J. and Sen, P. K. (2004). *Constrained Statistical Inference*. Wiley, New York.
Srivastava, M. S. (2005). *Methods of Multivariate Statistics*. Wiley, New York.
Srivastava, M. S. and Fujikoshi, Y. (2006). Multivariate analysis of variance with fewer observations than the dimension. *Journal of Multivariate Analysis* **97**, 1927–1940.
Srivastava, M. S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis* **99**, 386–402.
Tamhane, A. C., Hochberg, Y. and Dunnett, C. W. (1996). Multiple test procedures for dose finding. *Biometrics* **52**, 21–37.
Tang, D. -I., Geller, N. and Pocock, S. J. (1993). On the design and analysis of randomized clinical trials with multiple endpoints. *Biometrics* **49**, 23–30.
Tang, D. -I., Gnecco, C., Geller, N. (1989). An approximate likelihood ratio test for a normal mean vector with nonnegative components with application to clinical trials. *Biometrika* **76**, 577–583.

