Leibniz
Universität Hannover

Reports of the Institute of Biostatistics

No 01 / 2010

Leibniz University of Hannover

Natural Sciences Faculty

Titel: *Multiple contrast tests for multiple endpoints*

Author: *Mario Hasler*[1]

[1] Lehrfach Variationsstatistik, Christian-Albrechts-University of Kiel, Germany
Hermann-Rodewald-Str. 9, D–24118 Kiel, Germany
Tel.: +49 431 8802991
Email: hasler@email.uni-kiel.de

1 Introduction

Experimental trials often do not cover only one single endpoint but many correlated endpoints (see the data of Schulte et al. (2002)). A measurement object may be related to different variables or be observed in the course of time. Multiplicity adjustment must then take the number of endpoints into account, too. Thus, the first strategy is to reduce the number of endpoints to the smallest possible number that is necessary and that still provides the main information about the data. Second, it is useful to divide the endpoints into primary and secondary ones, where the primary endpoints are most important. The guideline on biostatistics according to the ICH E9 Expert Working Group (1999) recommends the selection of one primary endpoint. However, this is often not sufficient from an investigator's point of view. The secondary endpoints are considered only after the primary objective of the trial has been achieved. A possible objection is that such a classification of endpoints according to their importance can be somewhat arbitrary. Like the first, this strategy also reduces the dimension of the problem, but the question, how to handle multiple primary endpoints, remains. The statistical analysis for these endpoints must control the FWE over all of them. On the other hand, their correlations are important. For example, highly correlated endpoints do not give the same amount of information about the data as uncorrelated ones. Effects may be erroneously ignored when analyzing the endpoints separately.

Multiple contrast tests (MCTs) and related simultaneous confidence intervals (SCIs) provide test decisions and parameter estimation, respectively, for each comparison. They control the FWE at level α, and take correlations into account. However, they had been limited to comparisons of treatments on a single endpoint so far. This report presents some aspects concerning the power of MCTs and SCIs for multiple endpoints according to Hasler (2009). Ratios of means are focused here, because SCIs are then comparable also for the different endpoints, which can be assumed to have different scales.

2 Test Procedure

For $h = 1, \ldots, p$, $i = 1, \ldots, k$ and $j = 1, \ldots, n_h$, let X_{hij} denote the jth observation on the ith endpoint under the hth treatment in a one-way layout, and $\sum_{h=1}^{p}(n_h - 1) \geq k$. Each endpoint is hence measured for all $N = \sum_{h=1}^{p} n_h$ objects. Suppose the vectors $(X_{h1j}, \ldots, X_{hkj})'$ to be mutually independent and to follow k-variate normal distributions with mean vectors $\boldsymbol{\mu}_h = (\mu_{h1}, \ldots \mu_{hk})'$ and unknown covariance matrices $\boldsymbol{\Sigma}_h = (\sigma_{h,ii'})_{i,i'}$. Let the means per endpoint, $\mu_{1i}, \ldots, \mu_{pi}$, have the same algebraic sign, i.e., $\text{sign}(\mu_{1i}) = \ldots = \text{sign}(\mu_{pi})$ $(i = 1, \ldots, k)$. Presume possibly different variances and covariances for the endpoints but the same covariance matrices for all treatments, i.e., $\boldsymbol{\Sigma}_1 = \ldots = \boldsymbol{\Sigma}_p = \boldsymbol{\Sigma} = (\sigma_{ii'})_{i,i'}$. That means

$$\{X_{hij} : i = 1, \ldots, k\} \sim \perp \text{N}_k(\boldsymbol{\mu}_h, \boldsymbol{\Sigma}) \quad (h = 1, \ldots, p,\, j = 1, \ldots, n_h).$$

Let $\bar{\boldsymbol{X}}_h = (\bar{X}_{h1}, \ldots, \bar{X}_{hk})'$ and $\hat{\boldsymbol{\Sigma}}_h$ be the sample mean vectors and the sample covariance matrices for the treatments, respectively. The pooled sample covariance matrix $\hat{\boldsymbol{\Sigma}} = (\hat{\sigma}_{ii'})_{i,i'}$ is given by

$$\hat{\boldsymbol{\Sigma}} = \frac{\sum_{h=1}^{p}(n_h - 1)\hat{\boldsymbol{\Sigma}}_h}{\sum_{h=1}^{p}(n_h - 1)}$$

with the estimates $\hat{\sigma}_{ii'}$ $(1 \leq i, i' \leq k)$ for the covariances of the different endpoints. The diagonal elements, which are required for the following test procedure, are hence

$$\hat{\sigma}_{ii} = S_i^2 = \frac{(n_1 - 1)S_{1i}^2 + \cdots + (n_p - 1)S_{pi}^2}{n_1 + \cdots + n_p - p} \quad (i = 1, \ldots, k)$$

with

$$S_{hi}^2 = \frac{1}{n_h - 1} \sum_{j=1}^{n_h} (X_{hij} - \bar{X}_{hi})^2 \quad (h = 1, \ldots, p).$$

From the pooled sample covariance matrix $\hat{\boldsymbol{\Sigma}}$, we then derive the estimation $\hat{\boldsymbol{R}} = (\hat{\rho}_{ii'})_{i,i'}$ of the common correlation matrix of the data $\boldsymbol{R} = (\rho_{ii'})_{i,i'}$. We are interested in the matrix of ratios of contrasts, $\boldsymbol{G} = (\gamma_{li})_{l,i}$, where

$$\begin{aligned} \gamma_{li} &= \frac{\sum_{h=1}^{p} c_{lh}\mu_{hi}}{\sum_{h=1}^{p} d_{lh}\mu_{hi}} \\ &= \frac{\boldsymbol{c}_l'\boldsymbol{\mu}_{,i}}{\boldsymbol{d}_l'\boldsymbol{\mu}_{,i}} \quad (l = 1, \ldots, q,\ i = 1, \ldots, k) \end{aligned}$$

with $\boldsymbol{\mu}_{,i} = (\mu_{1i}, \ldots, \mu_{pi})'$. The vectors $\boldsymbol{c}_l = (c_{l1}, \ldots, c_{lp})'$ and $\boldsymbol{d}_l = (d_{l1}, \ldots, d_{lp})'$ consist of real constants and are the same for all endpoints; they do not depend on the particular value of the index i. Endpoint-specific contrasts are also possible in principle, but we disregard this fact for simplicity. Without loss of generality, the objective is to test the hypotheses

$$H_{0,li} : \gamma_{li} \leq \theta_{li} \quad (l = 1, \ldots, q,\ i = 1, \ldots, k) \tag{1}$$

with contrast- and endpoint-specific relative thresholds $\theta_{li} \in (0, \infty)$. Usually, $\theta_{li} = 1$ for all $l = 1, \ldots, q$ and for all $i = 1, \ldots, k$. If the test direction is reversed for some endpoints, the corresponding test statistics have to be multiplied with minus one. We focus here on ratios of means to enable comparison of the results for the different endpoints, which can be assumed to have different scales. Related SCIs for ratios are on the same relative (e.g., per cent) scale for all contrasts and endpoints, while SCIs for differences are not. On the other hand, for the case of $\theta_{li} = 1$ for all $l = 1, \ldots, q$ and for all $i = 1, \ldots, k$, this test coincides with the difference-based one. Testing problem (1) is a UIT because the overall null hypothesis of interest can be expressed as an intersection of the local null hypotheses, i.e.,

$$H_0 = \bigcap_{l=1}^{q} H_{0l} \quad \text{and} \quad H_{0l} = \bigcap_{i=1}^{k} H_{0,li}.$$

Thus, Theorem the procedure is coherent and consonant according to Gabriel (1969).

The test statistics are given by

$$\begin{aligned} T_{li} &= \frac{\sum_{h=1}^{p} (c_{lh} - \theta_{li}d_{lh})\, \bar{X}_{hi}}{S_i \sqrt{\sum_{h=1}^{p} (c_{lh} - \theta_{li}d_{lh})^2 / n_h}} \\ &= \frac{(\boldsymbol{c}_l - \theta_{li}\boldsymbol{d}_l)'\, \bar{\boldsymbol{X}}_{,i}}{S_i \sqrt{(\boldsymbol{c}_l - \theta_{li}\boldsymbol{d}_l)'\, \boldsymbol{M}\, (\boldsymbol{c}_l - \theta_{li}\boldsymbol{d}_l)}} \quad (l = 1, \ldots, q,\ i = 1, \ldots, k) \end{aligned}$$

where

$$\boldsymbol{M} = \begin{pmatrix} 1/n_1 & & 0 \\ & \ddots & \\ 0 & & 1/n_p \end{pmatrix}.$$

The vectors $\boldsymbol{T}_l = (T_{l1}, \ldots, T_{lk})'$, containing the test statistics for the lth comparison on all endpoints, can be reshaped to

$$\boldsymbol{T}_l = \left(\frac{Y_{l1}}{\sqrt{U_1/\nu}}, \ldots, \frac{Y_{lk}}{\sqrt{U_k/\nu}} \right)' \quad (l = 1, \ldots, q),$$

where under H_{0l}, the vector $(Y_{l1}, \ldots, Y_{lk})'$ follows a k-variate normal distribution with a correlation matrix denoted by $\boldsymbol{R}_{ll}$. The $U_1, \ldots, U_k$ are dependent χ^2 variables with

$$\nu = \sum_{h=1}^{p} (n_h - 1)$$

degrees of freedom. Note that $U_1, \ldots, U_k$ are different random variables but they follow the same distribution. Under H_{0l}, $\boldsymbol{T}_l$ is approximately k-variate t-distributed with ν degrees of freedom and correlation matrix $\boldsymbol{R}_{ll}$, i.e.,

$$\boldsymbol{T}_l \overset{appr.}{\sim} t_k(\nu, \boldsymbol{R}_{ll}).$$

The exact distribution is – strictly speaking – unknown. Consequently, under H_0, the vector of all test statistics,

$$\boldsymbol{T} = (\boldsymbol{T}_1', \ldots, \boldsymbol{T}_q')' = (T_{11}, \ldots, T_{li}, \ldots, T_{qk})',$$

follows (approximately) a qk-variate t-distribution with ν degrees of freedom and a correlation matrix, denoted by $\tilde{\boldsymbol{R}}$, i.e.,

$$\boldsymbol{T} \overset{appr.}{\sim} t_{qk}(\nu, \tilde{\boldsymbol{R}}).$$

The correlation matrix $\tilde{\boldsymbol{R}}$ is given by

$$\tilde{\boldsymbol{R}} = (\boldsymbol{R}_{ll'})_{l,l'} = \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} & \ldots & \boldsymbol{R}_{1q} \\ \boldsymbol{R}_{12} & \boldsymbol{R}_{22} & \ldots & \boldsymbol{R}_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{R}_{1q} & \boldsymbol{R}_{2q} & \ldots & \boldsymbol{R}_{qq} \end{pmatrix}.$$

The submatrices $\boldsymbol{R}_{ll'} = (\rho_{ll',ii'})_{i,i'}$ describe the correlations between the contrasts l and l' for all endpoints. Their elements are

$$\begin{aligned} \rho_{ll',ii'} &= \rho_{ii'} \frac{\sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh}) (c_{l'h} - \theta_{l'i'} d_{l'h}) \frac{1}{n_h}}{\sqrt{\sum_{h=1}^{p} (c_{lh} - \theta_{li} d_{lh})^2 \frac{1}{n_h}} \sqrt{\sum_{h=1}^{p} (c_{l'h} - \theta_{l'i'} d_{l'h})^2 \frac{1}{n_h}}} \\ &= \rho_{ii'} \frac{(\boldsymbol{c}_l - \theta_{li} \boldsymbol{d}_l)' \boldsymbol{M} (\boldsymbol{c}_{l'} - \theta_{l'i'} \boldsymbol{d}_{l'})}{\sqrt{(\boldsymbol{c}_l - \theta_{li} \boldsymbol{d}_l)' \boldsymbol{M} (\boldsymbol{c}_l - \theta_{li} \boldsymbol{d}_l)} \sqrt{(\boldsymbol{c}_{l'} - \theta_{l'i'} \boldsymbol{d}_{l'})' \boldsymbol{M} (\boldsymbol{c}_{l'} - \theta_{l'i'} \boldsymbol{d}_{l'})}} \\ &\quad (1 \leq l, l' \leq q,\ 1 \leq i, i' \leq k), \end{aligned} \tag{2}$$

where the $\rho_{ii'}$ are the elements of the correlation matrix $\boldsymbol{R} = (\rho_{ii'})_{i,i'}$ of the data. It is obvious that for $i = i'$, we recover the correlations of an MCT for ratios of means, see, e.g., Dilba et al. (2006). Hence, the case of only one endpoint ($k = 1$) and several treatments may be incorporated into the present theory rather easily. Furthermore, focusing on one fixed contrast ($l = l'$) and equal thresholds for all endpoints ($\theta_{li} = \theta_l \; \forall \, i = 1, \ldots, k$), the structure of the correlation matrix simplifies according to $\rho_{ll',ii'} = \rho_{ii'}$ and $\boldsymbol{R}_{ll} = \boldsymbol{R}$. Note that neither the matrix $\tilde{\boldsymbol{R}}$ nor the matrix $\boldsymbol{R}_{ll'}$ has a product correlation structure, i.e., the elements do not factorize. Because the common correlation matrix of the data $\boldsymbol{R}$ is not known and must be estimated, we conclude that, under H_0,

$$\boldsymbol{T} \overset{appr.}{\sim} t_{qk}(\nu, \hat{\tilde{\boldsymbol{R}}}),$$

where $\hat{\tilde{\boldsymbol{R}}}$ is the estimation of $\tilde{\boldsymbol{R}}$.

The decision rule for testing problem (1) is to reject $H_{0,li}$ for each ratio of contrasts γ_{li} with

$$T_{li} > t^{l}_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}}),$$

where $t^{l}_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})$ is a lower $(1-\alpha)$-quantile of a related qk-variate t-distribution. If two-sided testing is of interest, the absolute values for T_{li} and quantiles $t^{ts}_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})$ have to be taken. For the computation of these quantiles, one may resort to the numerical integration routines of Genz and Bretz (1999, 2002) (see also Bretz et al. (2001)) mentioned earlier, which are not restricted to special correlation structures. The related adjusted p-values per comparison and endpoint can also be obtained, of course.

3 Simultaneous Confidence Intervals

Let $\boldsymbol{\xi} = (\xi_{11}, \ldots, \xi_{qk})'$ be a point in the parameter space of $\boldsymbol{\gamma} = (\gamma_{11}, \ldots, \gamma_{qk})'$. Assuming that increasing values of the data, X_{hij}, represent a better effect of the treatments, the $(1-\alpha)100\%$ confidence set for the statistical problem (1) is given by

$$\begin{aligned} C\left((x,y)\right) &= \left\{\boldsymbol{\xi} : T_{li}(\xi_{li}) \leq t^{l}_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}}), \quad l = 1, \ldots, q, \, i = 1, \ldots, k\right\} \\ &= \left\{\boldsymbol{\xi} : A_{li}\xi_{li}^2 + B_{li}\xi_{li} + C_{li} \leq 0, \quad l = 1, \ldots, q, \, i = 1, \ldots, k\right\} \end{aligned}$$

where

$$\begin{aligned}
A_{li} &= \left(\sum_{h=1}^{p} d_{lh}\bar{X}_{hi}\right)^2 - \left(t^l_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})\right)^2 S_i^2 \sum_{h=1}^{p} d_{lh}^2/n_h \\
&= \left(\boldsymbol{d}_l'\bar{\boldsymbol{X}}_{,i}\right)^2 - \left(t^l_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})\right)^2 S_i^2 \boldsymbol{d}_l'\boldsymbol{M}\boldsymbol{d}_l, \\
B_{li} &= -2\left(\left(\sum_{h=1}^{p} c_{lh}\bar{X}_{hi}\right)\left(\sum_{h=1}^{p} d_{lh}\bar{X}_{hi}\right) - \left(t^l_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})\right)^2 S_i^2 \sum_{h=1}^{p} c_{lh}d_{lh}/n_h\right) \\
&= -2\left(\left(\boldsymbol{c}_l'\bar{\boldsymbol{X}}_{,i}\right)\left(\boldsymbol{d}_l'\bar{\boldsymbol{X}}_{,i}\right) - \left(t^l_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})\right)^2 S_i^2 \boldsymbol{c}_l'\boldsymbol{M}\boldsymbol{d}_l\right), \\
C_{li} &= \left(\sum_{h=1}^{p} c_{lh}\bar{X}_{hi}\right)^2 - \left(t^l_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})\right)^2 S_i^2 \sum_{h=1}^{p} c_{lh}^2/n_h \\
&= \left(\boldsymbol{c}_l'\bar{\boldsymbol{X}}_{,i}\right)^2 - \left(t^l_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})\right)^2 S_i^2 \boldsymbol{c}_l'\boldsymbol{M}\boldsymbol{c}_l. \qquad (3)
\end{aligned}$$

This approach is based on Fieller's Theorem (Fieller, 1954). As is known, the correlation matrix $\tilde{\boldsymbol{R}}$ depends here on the unknown ratios γ_{li}, $\hat{\rho}_{ll',ii'} = \hat{\rho}_{ll',ii'}(\gamma_{li}, \gamma_{l'i'})$. Application of the plug-in approach of Dilba et al. (2006) corresponds to the use of

$$\begin{aligned}
\hat{\gamma}_{li} &= \frac{\sum_{h=1}^{p} c_{lh}\bar{X}_{hi}}{\sum_{h=1}^{p} d_{lh}\bar{X}_{hi}} \\
&= \frac{\boldsymbol{c}_l'\bar{\boldsymbol{X}}_{,i}}{\boldsymbol{d}_l'\bar{\boldsymbol{X}}_{,i}} \quad (l = 1, \ldots, q,\ i = 1, \ldots, k)
\end{aligned}$$

in Equation (2) instead of θ_{li} (similarly for index $l'i'$). For simplicity, we do not introduce a new symbol for the resulting estimated correlation matrix. The lower limits of the approximate $(1-\alpha)100\%$ SCIs for $(\gamma_{11}, \ldots, \gamma_{qk})'$ are hence given by

$$\hat{\gamma}_{li}^{lower} = \frac{-B_{li} - \sqrt{B_{li}^2 - 4A_{li}C_{li}}}{2A_{li}} \quad (l = 1, \ldots, q,\ i = 1, \ldots, k).$$

If $A_{li} > 0$, then the solution is finite (see, e.g., Buonaccorsi and Iyer (1984) for the case of only one endpoint). The statistical problem (1) can be decided as follows: For a specified level α, we reject $H_{0,li}$ for each contrast γ_{li} with

$$\hat{\gamma}_{li}^{lower} > \theta_{li}.$$

For the two-sided case, we obtain

$$\begin{aligned}
C\left((x,y)\right) &= \left\{\boldsymbol{\xi} : |T_{li}(\xi_{li})| \leq t^{ts}_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}}), \quad l = 1, \ldots, q,\ i = 1, \ldots, k\right\} \\
&= \left\{\boldsymbol{\xi} : A_{li}\xi_{li}^2 + B_{li}\xi_{li} + C_{li} \leq 0, \quad l = 1, \ldots, q,\ i = 1, \ldots, k\right\},
\end{aligned}$$

where the A_{li}, B_{li} and C_{li} are defined as in (3) but with quantiles $t^{ts}_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})$ instead of $t^{l}_{qk,1-\alpha}(\nu, \hat{\tilde{\boldsymbol{R}}})$. The confidence limits are given by

$$\hat{\gamma}_{li}^{lower} = \frac{-B_{li} - \sqrt{B_{li}^2 - 4A_{li}C_{li}}}{2A_{li}} \quad (l = 1, \ldots, q,\ i = 1, \ldots, k),$$

$$\hat{\gamma}_{li}^{upper} = \frac{-B_{li} + \sqrt{B_{li}^2 - 4A_{li}C_{li}}}{2A_{li}} \quad (l = 1, \ldots, q,\ i = 1, \ldots, k).$$

For a specified level α, we reject $H_{0,li}$ for each contrast γ_{li} with

$$\hat{\gamma}_{li}^{lower} > \theta_{li} \quad \text{or} \quad \hat{\gamma}_{\text{li}}^{\text{upper}} < \theta_{\text{li}}.$$

4 Power Considerations

The testing problem (1) is simplified here to the case of equal thresholds, $\theta_{li} = \theta$ for all $l = 1, \ldots, q$ and $i = 1, \ldots, k$. Let higher response values indicate better treatment effects and τ^* denote the greatest irrelevant ratio to the control mean which is to be detected. Define the set of indices $I(\tau^*) = \{(l, i) : \tau_{li} > \tau^*\}$. All ratios of contrasts with τ_{li} values greater than τ^* are relevant. An (approximate) expression for the complete (or all-pairs) power of the statistical problem (1) is given by

$$P\left\{T_{li} > t^{l}_{qk,1-\alpha}(\nu, \tilde{\boldsymbol{R}}) \middle| \psi_{li}, \boldsymbol{\Sigma} \quad \forall (l, i) \in I(\theta^*)\right\}. \tag{4}$$

An (approximate) expression for the minimal (or any-pair) power of the statistical problem (1) is given by

$$P\left\{T_{li} > t^{l}_{qk,1-\alpha}(\nu, \tilde{\boldsymbol{R}}) \middle| \psi_{li}, \boldsymbol{\Sigma} \quad \text{for at least one } (l, i) \in I(\theta^*)\right\}. \tag{5}$$

The probability to reject for any contrast is defined as the *global power*. If one is interested only in the global test decision for statistical problem (1), then this definition is appropriate. An (approximate) expression for the global power of the statistical problem (1) is given by

$$P\left\{T_{li} > t^{l}_{qk,1-\alpha}(\nu, \tilde{\boldsymbol{R}}) \middle| \psi_{li}, \boldsymbol{\Sigma} \quad \text{for at least one } l = 1, \ldots, q \text{ and } i = 1, \ldots, k\right\}. \tag{6}$$

Because the data's correlations are estimated, the quantiles $t^{l}_{qk,1-\alpha}(\nu, \tilde{\boldsymbol{R}})$ in fact are random variables because they depend on the sample values. Therefore, the probabilities (4), (5) and (6) are only approximate ones. The power function (6) can be calculated from a non-central qk-variate t-distribution with ν degrees of freedom and non-centrality parameter $\boldsymbol{\kappa} = (\kappa_{11}, \ldots, \kappa_{li}, \ldots, \kappa_{qk})'$, where

$$\begin{aligned} \kappa_{li} &= \frac{\sum_{h=1}^{p} (c_{lh} - \theta d_{lh})\, \mu_{hi}}{\sqrt{\sigma_{ii} \sum_{h=1}^{p} (c_{lh} - \theta d_{lh})^2 / n_h}} \\ &= \frac{(\boldsymbol{c_l} - \theta \boldsymbol{d_l})' \, \boldsymbol{\mu}_{,i}}{\sqrt{\sigma_{ii} \, (\boldsymbol{c_l} - \theta \boldsymbol{d_l})' \, \boldsymbol{M} \, (\boldsymbol{c_l} - \theta \boldsymbol{d_l})}}. \end{aligned}$$

Figure 1 (2) illustrates Equation (6) for three (five) treatments and the Dunnett contrast depending on the ratio $\gamma_{q1} = \mu_{p1}/\mu_{11}$ (where $q = p-1$). The remaining ratios γ_{li} are fixed and equal. The relative thresholds against which the test is performed are $\theta_{li} = 1$ for all $l = 1, \ldots, q$ and all $i = 1, \ldots, k$. Several equicorrelations (rows) for two, four and eight endpoints (columns) are considered. The total sample size is 60 (100). Three allocations are shown each. The solid line represents the well-known optimal allocation for the Dunnett contrast, i.e., $n_1 = \sqrt{p-1}\, n_h$ $(h = 2, \ldots, p)$. Hence, the sample size for the control group is $n_1 = 24, 12, 6$ $(32, 16, 8)$, and the sample sizes for the non-control groups are balanced. Although the correlations of the endpoints are taken into account, their exact influence it is not clear from Figures 1 and 2. Therefore, this problem is presented by Figure 3 (4). Again, Equation (6) is illustrates for three (five) treatments with a similar background, but now depending on the correlations of the endpoints. The ratio γ_{q1} is set here to 1.25. One-sided and two-sided tests (rows) for two, four and eight endpoints (columns) are considered. The power indeed depends on the correlations. The minimum is achieved for vanishing correlation and increases for increasing absolute correlation values.

The package `multtest` (Pollard et al., 2007) of the statistical software R [2008] provides resampling-based multiple hypothesis testing. Non-parametric bootstrap and permutation tests are implemented. Tests based on t- and F-statistics are included. The main application of this package is gene selection in microarray experiments. The function `MTP` performs test procedures for multiple endpoints by single-step and step-down minP and maxT methods to control the FWE (or other error rates). Tests based on t-tests are restricted to comparisons of two groups, e.g. a treatment and a control. A simulation study has been performed to compare this t-test-based bootstrap approach (Boot.) with the new MCT method (Multiv.). The single-step option `method=''ss.maxT''` has been used for comparability. Hence, only $p = 2$ treatment groups are simulated (and hence $q = 1$), and $\theta_{1i} = 1$ for all $i = 1, \ldots, k$ (for more details, see Hasler (2009)). Figure 5 shows the results of the mentioned power comparison. The rows are related to the different equicorrelations, the columns to the number of endpoints. Minimal and complete power coincide in this case, because the treatment group differs only for the first endpoint. A higher power of the new multivariate method is visible only for high correlations and high numbers of endpoints. Figure 6 shows the minimal power for the case that the mean of the treatment group was changed simultaneously for all endpoints and by the same relative amount. Except for the minimal equicorrelation ρ^{min}, the bootstrap method is better with respect to power than the new method. This difference becomes more pronounced with increasing correlation and with increasing number of endpoints. Figure 7 shows the complete power for the same background (simultaneously changing the mean of the treatment group for all endpoints). The bootstrap method has slightly less power than the multivariate method, but this difference becomes negligible for high correlations. In summary: The power behavior of the competitors is almost equal. The gain in the minimal power for the bootstrap approach is insignificant in view of the properties and flexibility of the new MCT method.

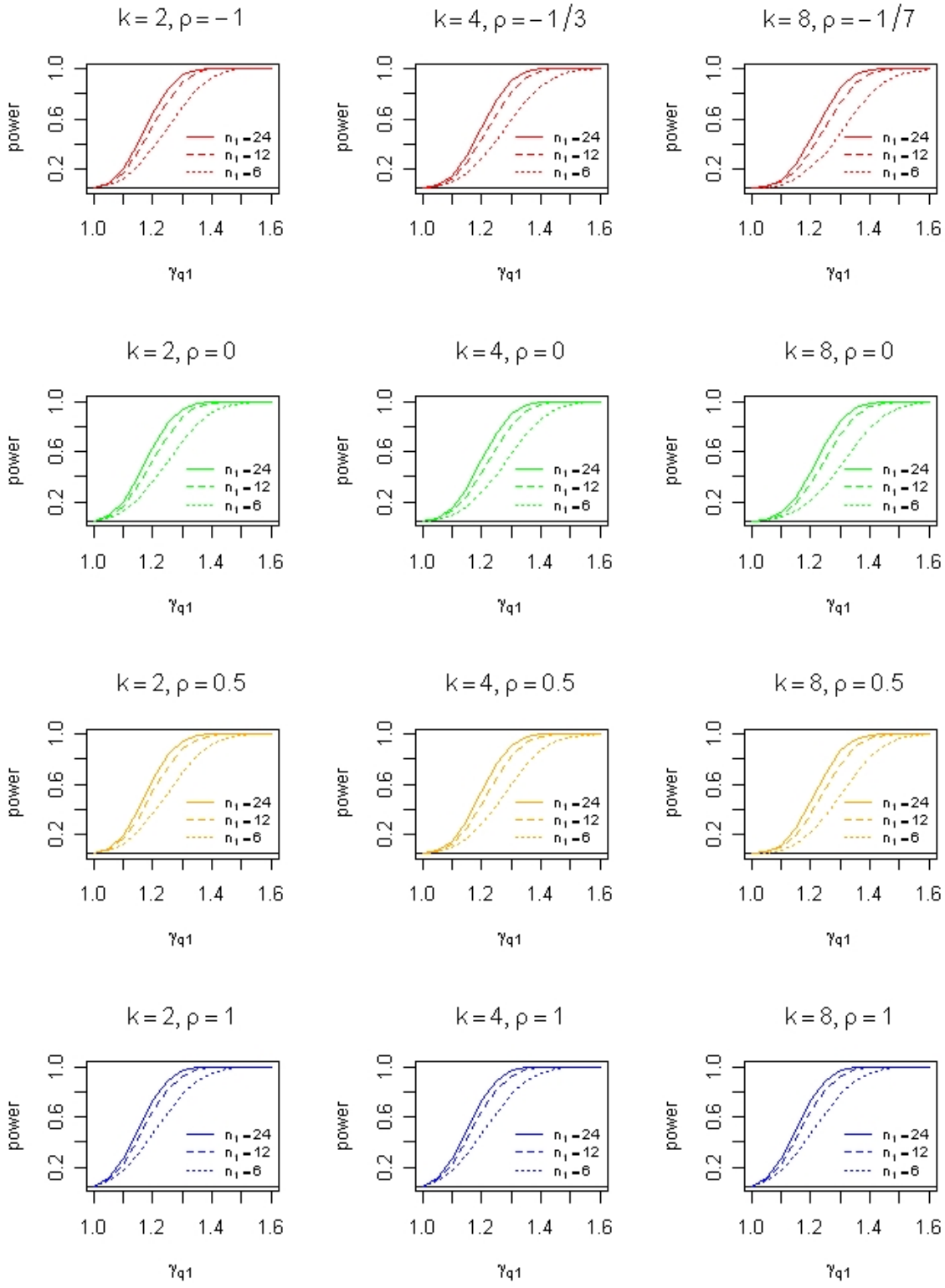

Figure 1: *Global power function of one-sided MCTs for $p = 3$ treatments, the Dunnett contrast, several numbers of endpoints, ratios γ_{q1}, and equicorrelations; $\alpha = 0.05$.*

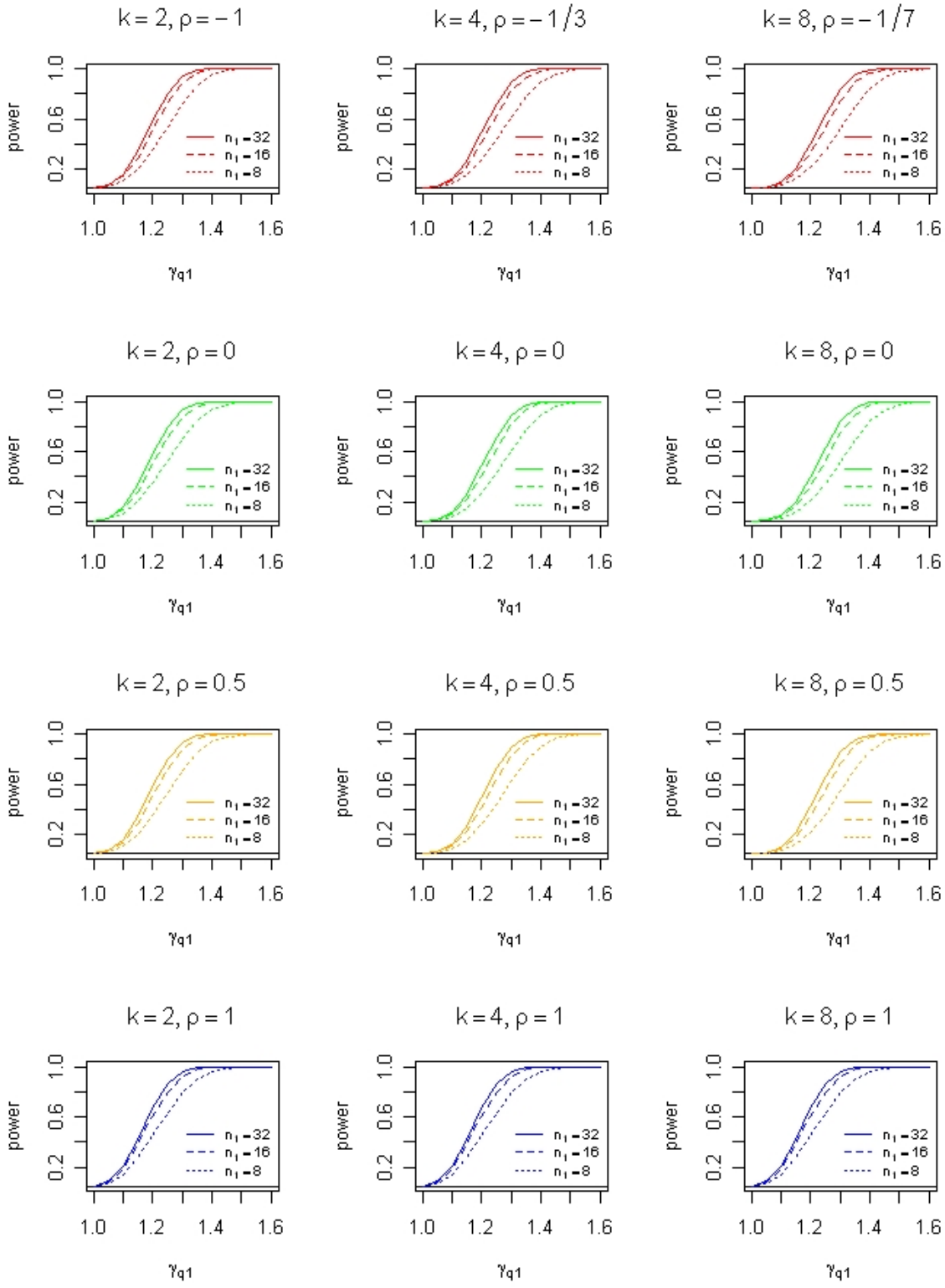

Figure 2: *Global power function of one-sided MCTs for $p = 5$ treatments, the Dunnett contrast, several numbers of endpoints, ratios γ_{q1}, and equicorrelations; $\alpha = 0.05$.*

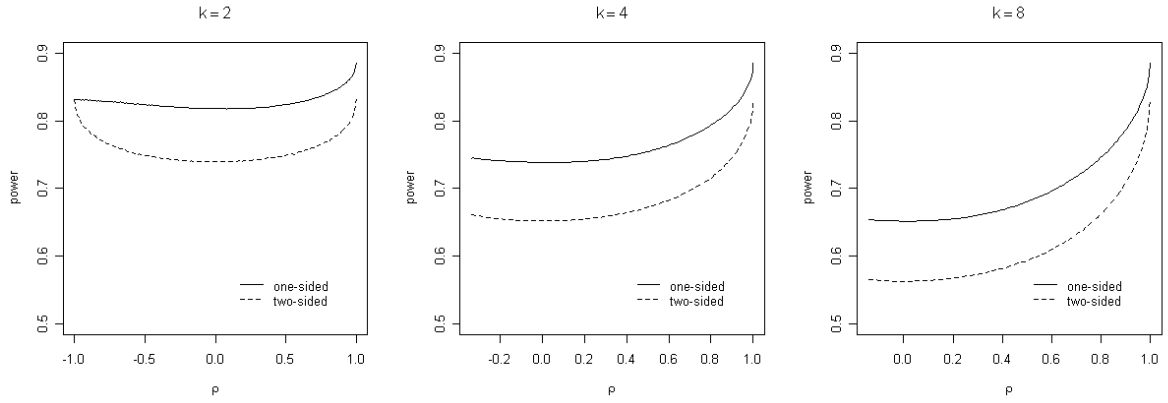

Figure 3: *Global power function of MCTs for $p = 3$ treatments, the Dunnett contrast, several numbers of endpoints, and equicorrelations; $\gamma_{q1} = 1.25$, $\alpha = 0.05$.*

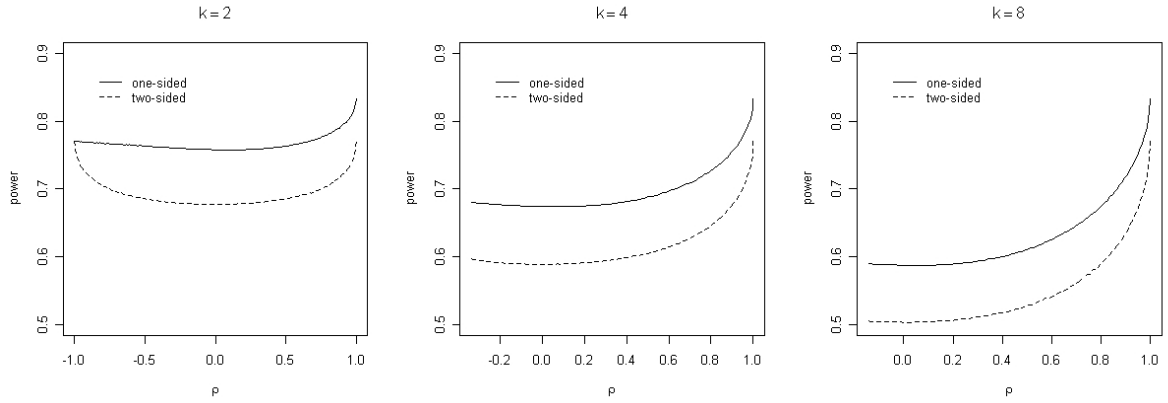

Figure 4: *Global power function of MCTs for $p = 5$ treatments, the Dunnett contrast, several numbers of endpoints, and equicorrelations; $\gamma_{q1} = 1.25$, $\alpha = 0.05$.*

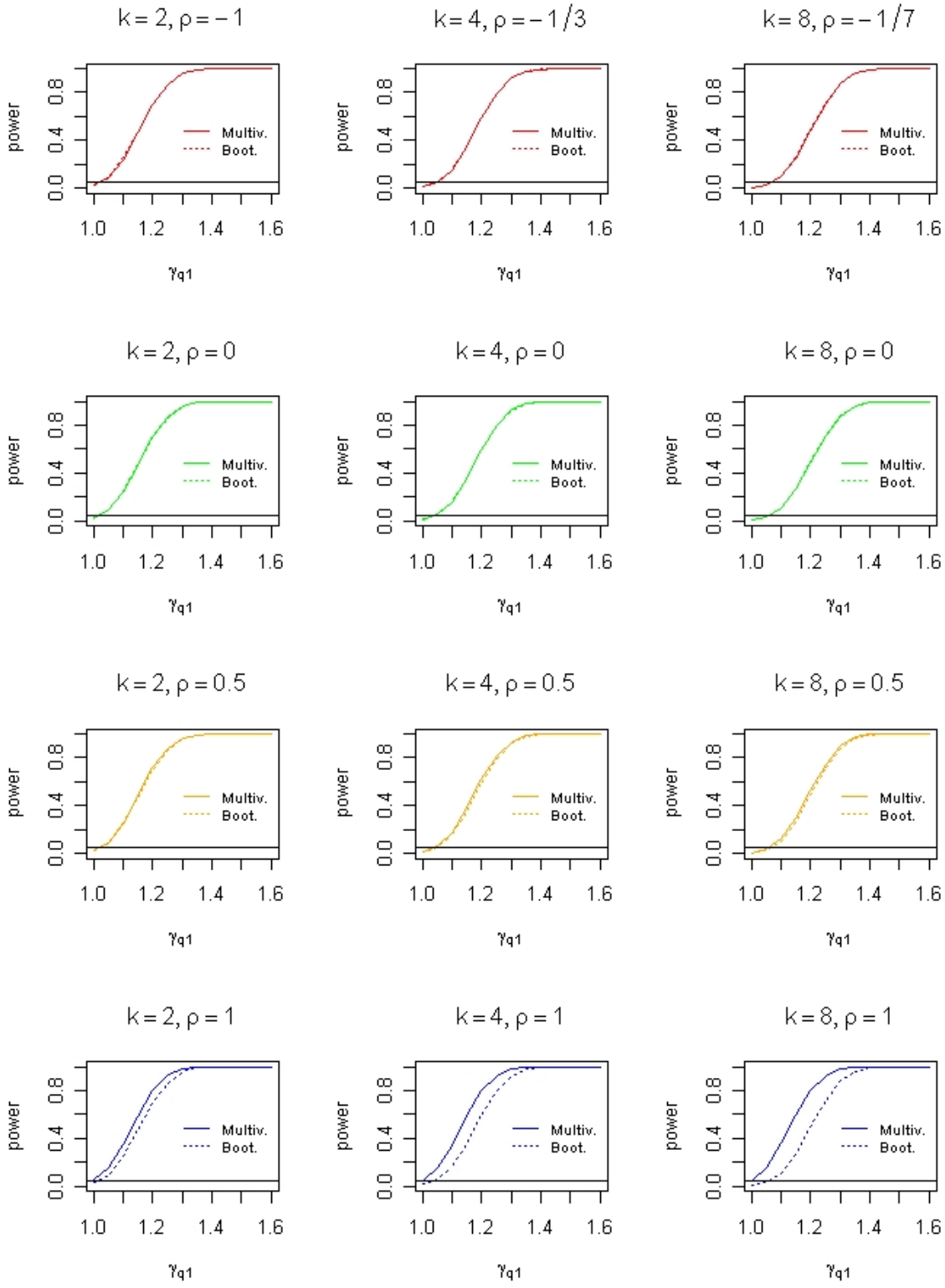

Figure 5: *Minimal and complete power function of one-sided MCTs for $p = 2$ treatments, the Dunnett contrast, several numbers of endpoints, and equicorrelations; $\gamma_{q1} = 1$, $\alpha = 0.05$.*

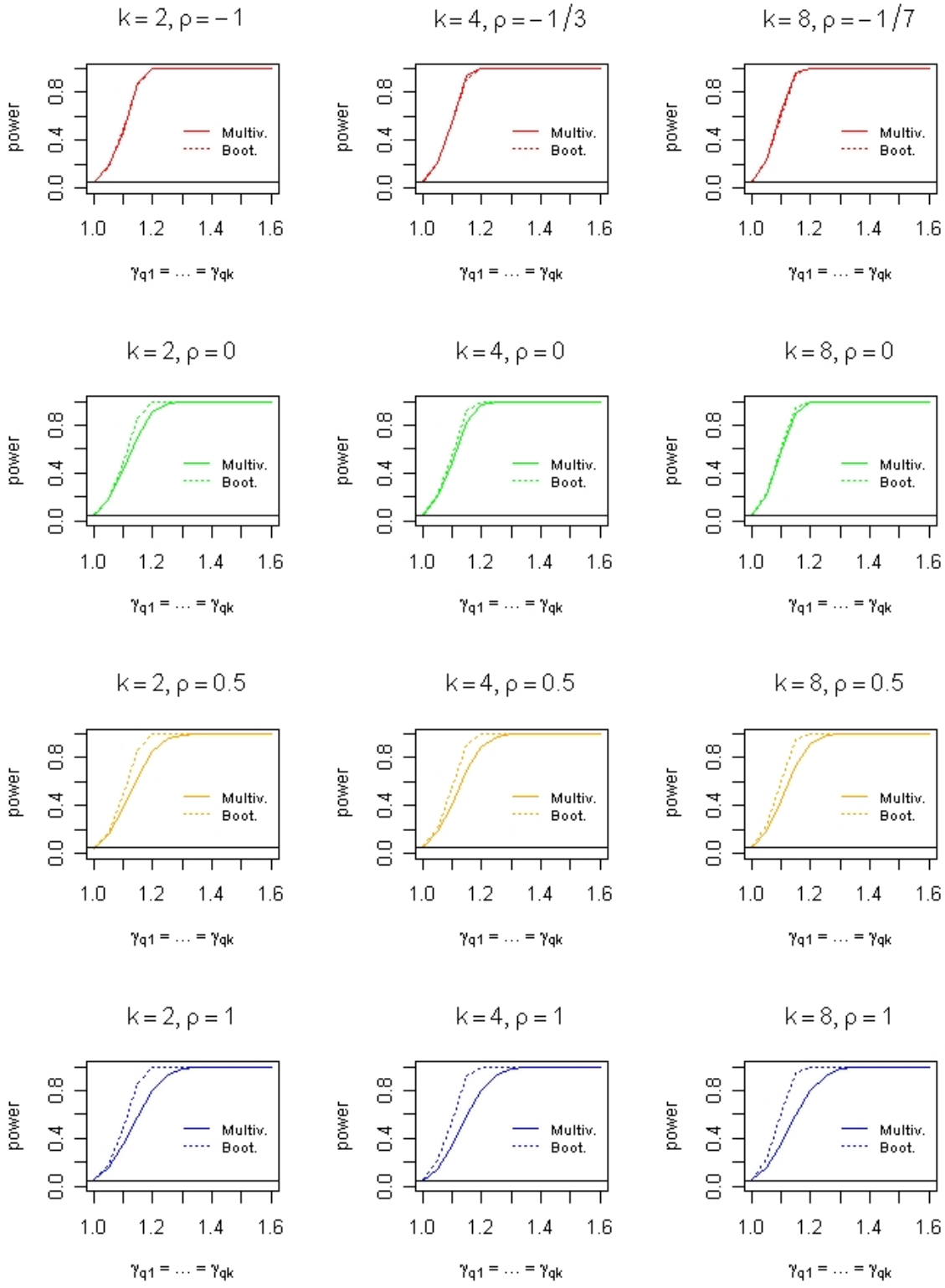

Figure 6: *Minimal power function of one-sided MCTs for* $p = 2$ *treatments, the Dunnett contrast, several numbers of endpoints, and equicorrelations;* $\gamma_{q1} = 1$, $\alpha = 0.05$.

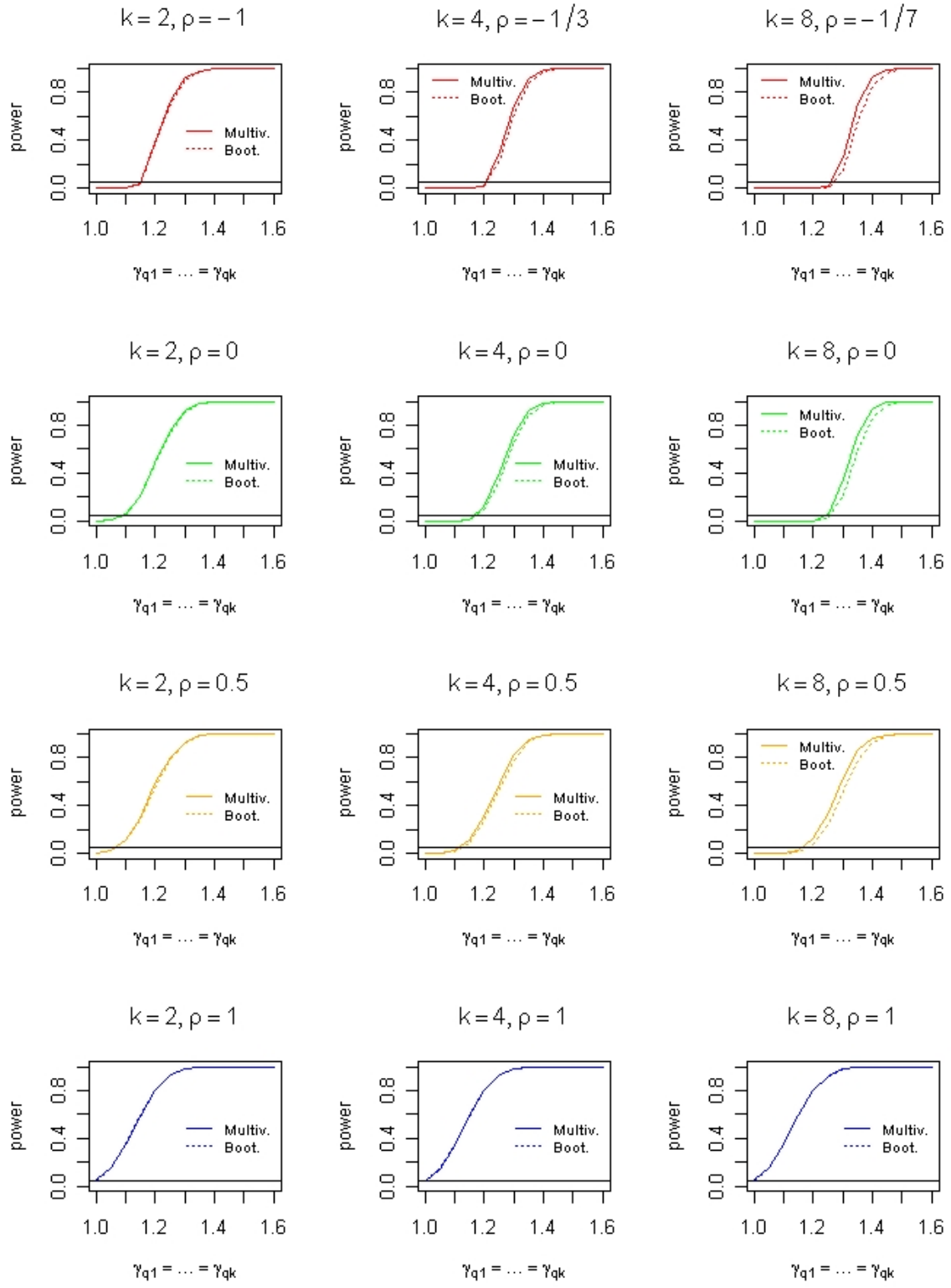

Figure 7: *Complete power function of one-sided MCTs for $p = 2$ treatments, the Dunnett contrast, several numbers of endpoints, and equicorrelations;* $\gamma_{q1} = 1$, $\alpha = 0.05$.

5 Conclusions and recommendations

The problem of many, possibly correlated, endpoints has been investigated. MCTs and related SCIs had been restricted to comparisons on a single endpoint so far. This methodology was extended to the case of an arbitrary number of endpoints by deriving an approximate multivariate t-distribution. Ratios of means have been considered for comparability of the different endpoints which may have different scales. An approach for differences of means has not been focused explicitly, but it can easily be obtained based on this work. If variances or correlations are assumed to differ for the different groups, the PI procedure for heterogeneous variances of Hasler and Hothorn (2008) can be applied. The procedures presented can be shown to maintain the FWE. The version for heterogeneous covariances shows a slight liberalism, but it is in acceptable ranges. Test decisions (e.g., p-values) for all contrasts and all endpoints are available as well as SCIs. For this reason, a fair power comparison with existing methods is not feasible. A resampling-based competitor with the same features exists only for the case of comparisons of only two groups (package `multtest` (Pollard et al., 2007) in R [2008]). Depending on which power is considered, the new method has about the same power properties or it is slightly worse. This is compensated by a gain in flexibility.

References

F. Bretz, A. Genz, and L. A. Hothorn. On the numerical availability of multiple comparison procedures. *Biometrical Journal*, 43:645–656, 2001.

J. P. Buonaccorsi and H. K. Iyer. A comparison of confidence regions and designs in estimation of a ratio. *Communications in Statistics-Theory and Methods*, 13:723–741, 1984.

G. Dilba, F. Bretz, and V. Guiard. Simultaneous confidence sets and confidence intervals for multiple ratios. *Journal Of Statistical Planning And Inference*, 136(8):2640–2658, August 2006.

E. C. Fieller. Some problems in interval estimation. *Journal of the Royal Statistical Society, Series B*, 16:175–185, 1954.

K. R. Gabriel. Simultaneous test procedures - Some theory of multiple comparisons. *Annals Of Mathematical Statistics*, 40(1):224–250, 1969.

A. Genz and F. Bretz. Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation*, 63:361–378, 1999.

A. Genz and F. Bretz. Methods for the computation of multivariate t-probabilities. *Journal of Computational and Graphical Statistics*, 11:950–971, 2002.

M. Hasler. *Extensions of Multiple Contrast Tests*. PhD thesis, Gottfried Wilhelm Leibniz Universität Hannover, 2009.

M. Hasler and L. A. Hothorn. Multiple contrast tests in the presence of heteroscedasticity. *Biometrical Journal*, 50(5):793–800, 2008.

ICH E9 Expert Working Group. Ich harmonised tripartite guideline: Statistical principles for clinical trials. *Statistics In Medicine*, 18(15):1903–1904, August 1999.

K. S. Pollard, Y. Ge, S. Taylor, and S. Dudoit. *multtest: Resampling-based multiple hypothesis testing*, 2007. R package version 1.16.1.

A. Schulte, J. Althoff, S. Ewe, and H. B. Richter-Reichhelm. Two immunotoxicity ring studies according to OECD TG 407 - Comparison of data on cyclosporin A and hexachlorobenzene. *Regulatory Toxicology And Pharmacology*, 36(1):12–21, 2002.