Simultaneous Confidence Regions Corresponding to Holm's Step-Down Procedure and Other Closed-Testing Procedures

Olivier Guilbaud*

AstraZeneca R & D, S-15185 Södertälje, Sweden

Received 2 July 2007, revised 30 April 2008, accepted 18 May 2008

Summary

Holm's (1979) step-down multiple-testing procedure (MTP) is appealing for its flexibility, transparency, and general validity, but the derivation of corresponding simultaneous confidence regions has remained an unsolved problem. This article provides such confidence regions. In fact, simultanenous confidence regions are provided for any MTP in the class of short-cut consonant closed-testing procedures based on marginal p-values and weighted Bonferroni tests for intersection hypotheses considered by Hommel, Bretz and Maurer (2007). In addition to Holm's MTP, this class includes the fixed-sequence MTP, recently proposed gatekeeping MTPs, and the fallback MTP. The simultaneous confidence regions are generally valid if underlying marginal p-values and corresponding marginal confidence regions (assumed to be available) are valid. The marginal confidence regions and estimated quantities are not assumed to be of any particular kinds/dimensions. Compared to the rejections made by the MTP for the family of null hypotheses H under consideration, the proposed confidence regions provide extra free information. In particular, with Holm's MTP, such extra information is provided: for all nonrejected Hs, in case not all Hs are rejected; or for certain (possibly all) Hs, in case all Hs are rejected. In case not all Hs are rejected, no extra information is provided for rejected Hs. This drawback seems however difficult to overcome. Illustrations concerning clinical studies are given.

Key words: Alpha-exhaustive test; Clinical study; Closed-testing procedure; Consonance; Equivalence; Gatekeeping; Holm procedure; Non-inferiority; Superiority.

1 Introduction

Simultaneous confidence regions have been available that correspond to certain stepwise multiple-testing procedures (MTPs), including: (a) the stepwise confidence bounds of Bofinger (1987) and Stefansson, Kim and Hsu (1988, Section 2) that correspond to the step-down version of one-sided Dunnett-type comparisons with a control under normality assumptions; and (b) the stepwise confidence intervals without multiplicity adjustment of Berger and Hsu (1999) that correspond to the fixed-sequence MTP. However, no simultaneous confidence regions have been available that correspond to Holm's (1979) step-down MTP, although this MTP is quite appealing and widely used due to its flexibility, transparency, and general validity. This article provides such simultaneous confidence regions. In fact, simultaneous confidence regions are provided for any MTP in the class of short-cut consonant closed-testing procedures based on marginal p-values and weighted Bonferroni tests for intersection hypotheses considered by Hommel, Bretz and Maurer (2007), which includes Holm's (1979) MTP. The proposed simultaneous confidence regions are quite flexible and generally valid – if underlying marginal p-values and corresponding confidence regions (assumed to be available) are valid. The proposed inferences are of interest when one aims at assertions that will "show" that estimated quantities belong to specific target regions. For example, in confirmatory clinical studies, target regions may corre-

* e-mail: olivier.guilbaud@astrazeneca.com, Phone: +46 855 329 146, Fax: +46 855 328 947

spond to intended claims about "superiority", "non-inferiority" and/or "equivalence"; cf. (6), (7), (9), and Section 4. For convenience, no notational distinction is made between random quantities and the corresponding realizations in this article.

1.1 The general situation considered

Consider the following general situation. Data $\boldsymbol{X}$ are generated from an unknown probability distribution P belonging to some class $\boldsymbol{P}$. This class is not restricted to be of any particular kind, so the model $\{P \in \boldsymbol{P}\}$ can be parametric, non-parametric, or semi-parametric. We are interested in making simultaneous inferences about $m \geq 2$ specified quantities, $\theta_1 = \theta_1(P), \ldots, \theta_m = \theta_m(P)$, based on marginal inferences about these quantities. In particular we are interested in simultaneous assertions of the form "$\theta_i \in R_i$" where R_i is a specified target region for θ_i, and if possible, additional simultaneous assertions about the θ_i's.

It is assumed that for each $i = 1, \ldots, m$: (a) the set $\boldsymbol{\Theta}_i = \{\theta_i(P);\ P \in \boldsymbol{P}\}$ of potential θ_i-values is known, and the specified target region R_i for θ_i satisfies

$$R_i \subset \boldsymbol{\Theta}_i, \qquad R_i \neq \varnothing, \quad \text{and} \quad R_i \neq \boldsymbol{\Theta}_i\,; \tag{1}$$

(b) $p_i = p_i(\boldsymbol{X})$ is an available p-value through which the null hypothesis $H_i:\ \theta_i \notin R_i$ can be tested versus $H_i^c:\ \theta_i \in R_i$ at any given level $0 < u < 1$ by rejecting H_i if and only if $p_i \leq u$, with the property that (Lehmann and Romano, 2005, p. 1140) for any given $0 < u < 1$,

$$\Pr[p_i \leq u] \leq u \quad \text{if} \quad H_i \text{ is true}; \tag{2}$$

(c) a specified family, indexed by $0 < \gamma < 1$, of level-γ confidence regions $C_{i;\gamma} = C_{i;\gamma}(\boldsymbol{X})$ for $\theta_i = \theta_i(P)$ such that

$$\Pr[\theta_i \in C_{i;\gamma}] \geq \gamma \quad \text{for all} \quad 0 < \gamma < 1 \quad \text{and} \quad P \in \boldsymbol{P} \tag{3}$$

is available that corresponds to the p_i-based tests of H_i in that for each $\boldsymbol{X}$ and each $0 < u < 1$,

$$C_{i,1-u}(\boldsymbol{X}) \subset R_i \quad \text{if and only if} \quad p_i(\boldsymbol{X}) \leq u\,, \tag{4}$$

and is such that for each $\boldsymbol{X}$,

$$\begin{aligned} &C_{i;\gamma'}(\boldsymbol{X}) \subset C_{i;\gamma''}(\boldsymbol{X}) \subset \boldsymbol{\Theta}_i \quad \text{for all} \quad 0 < \gamma' < \gamma'' < 1\,, \\ &C_{i;\gamma}(\boldsymbol{X}) = \boldsymbol{\Theta}_i \quad \text{if} \quad \gamma = 1 \quad (\text{by convention})\,. \end{aligned} \tag{5}$$

Illustrations of situations where (1)–(5) hold are given in Sections 2.1 and 4. Neither the θ_i's, R_i's, nor the $C_{i;\gamma}$'s, are restricted to be of any particular kinds/dimensions; and θ_i's, R_i's, and $C_{i;\gamma}$'s with distinct indexes i may be of different kinds/dimensions. This is of interest for clinical studies, where one often has a mixture of different kinds of intended claims and supporting data that concern e.g. efficacy, safety, and/or quality of life.

To fix ideas, one may think of a parallel-group clinical study where two treatments, say A and B, are compared in terms of real-valued θ_i's corresponding to certain response characteristics, so that each $\boldsymbol{\Theta}_i \subset \Re \equiv (-\infty\,,\,\infty)$. Under relevant additional distributional assumptions for the i-th response: (a) θ_i may e.g. be a difference θ_i' of means of two normal distributions, a shift parameter θ_i'' between two continuous distribution functions of unknown common shape, or a difference or ratio θ_i''' of the probability parameters of two binomial distributions, with $\boldsymbol{\Theta}_i = \Re$ for θ_i' and θ_i'', and $\boldsymbol{\Theta}_i$ equal to $(-1\,,\,1)$ or $(0\,,\,\infty)$ for θ_i'''; (b) p_i and $C_{i;\gamma}$ may be based on t-related methods for θ_i', based on a nonparametric methods (Lehmann, 1975) for θ_i'', or based on methods like those in Agresti and Min (2001, Sections 5–7) for θ_i'''; and (c) the target assertion "$\theta_i \in R_i$" aimed at may be

$$\text{"}\theta_i > \theta_{i,0}\text{"}, \quad \text{i.e.} \quad R_i = (\theta_{i,0}\,,\,\infty) \cap \boldsymbol{\Theta}_i, \quad \text{or} \tag{6}$$

$$\text{"}\theta_i < \theta_{i,0}\text{"}, \quad \text{i.e.} \quad R_i = (-\infty\,,\,\theta_{i,0}) \cap \boldsymbol{\Theta}_i\,, \quad \text{or} \tag{7}$$

$$\text{"}\theta_i \neq \theta_{i,0}\text{"}, \quad \text{i.e.} \quad R_i = ((-\infty\,,\,\theta_{i,0}) \cup (\theta_{i,0}\,,\,\infty)) \cap \boldsymbol{\Theta}_i\,, \tag{8}$$

with $\theta_{i,0} \in \boldsymbol{\Theta}_i$ suitably chosen. For example, assuming that positive θ_i-values favor A whereas negative values favor B: (a) "superiority" of A relative to B with respect to θ_i may correspond to a $\theta_{i,0}$-value in (6) equal to 0, or equal to a certain positive value of clinical relevance; whereas (b) "non-inferiority" of A relative to B with respect to θ_i may correspond to a certain negative $\theta_{i,0}$-value in (6) that is so close to 0 that the difference is clinically negligible. The target assertion (8) of non-equality is typically of little practical interest, and should rather be viewed and handled as being composed of (6) and (7) to make directional inferences possible; see Hsu (1996, p. 38) for a relevant discussion. The target assertion "$\theta_i \in R_i$" may also be an equivalence-to-$\theta_{i,0}$ assertion, for example

$$\text{“}|\theta_i - \theta_{i,0}| < \delta_i\text{”}, \quad \text{i.e.} \quad R_i = (\theta_{i,0} - \delta_i\,,\, \theta_{i,0} + \delta_i) \cap \boldsymbol{\Theta}_i\,, \tag{9}$$

with $\theta_{i,0} \in \boldsymbol{\Theta}_i$ and $\delta_i > 0$ suitably chosen; in which case p_i and $C_{i;\gamma}$ may be based on the "two one-sided tests" approach and corresponding (Bofinger, 1992) expanded confidence intervals; see Berger and Hsu (1996, Section 5) for a discussion about such equivalence-related inferences. More generally, θ_i, R_i, and $C_{i;\gamma}$ may be multi-dimensional.

It is important to note that each p_i and $C_{i;\gamma}$ in (2)–(5) is marginal. In particular, (3) concerns only the marginal coverage probability of a single $C_{i;\gamma}$ – no assumption is made about simultaneous coverage probabilities.

1.2 The idea

How can we proceed in such a situation? To illustrate the idea in terms of Holm's (1979) MTP based on p_i's and other quantities in (1)–(5), let $0 < \alpha < 1$ be given, and suppose temporarily that the target assertions "$\theta_i \in R_i$" are given equal weights. It is shown in Section 3 how simultaneous confidence assertions can then be made about $\theta_1, \ldots, \theta_m$ that: (a) have simultaneous confidence level $\geq 1 - \alpha$; (b) consist of target assertions "$\theta_i \in R_i$" for all i's in a certain subset I of $\{1, \ldots, m\}$; and (c) consist of assertions of the type "$\theta_i \in R_i \cup C_{i;\gamma}$" for all i's in the complementary subset $I^c = \{1, \ldots, m\} - I$, with common $\gamma = 1 - \alpha/|I^c|$. Moreover, an additional $1 - \alpha$ confidence assertion that is arbitrary but pre-specified can be made in case the event $I = \{1, \ldots, m\}$ occurs. A key point here is that the number $|I^c|$ of elements in I^c used in the Bonferroni-type adjustment of γ is smaller than or equal to m, possibly much smaller. Each assertion "$\theta_i \in R_i \cup C_{i;\gamma}$" with $i \in I^c$ indicates by how much one missed the target assertion "$\theta_i \in R_i$" aimed at. The probability of making any erroneous assertion with this procedure is $\leq \alpha$.

The set I (set I^c) mentioned in the preceding paragraph consists of the indexes i of H_i's that are rejected (are accepted) by Holm's MTP. The resulting simultaneous confidence assertions thus "correspond" (Hayter and Hsu, 1994, p. 129) to Holm's MTP based on the p_i's in that they imply the same rejections of H_i's, and in addition, provide extra "free" information about θ_i's. Such extra information is provided for all the θ_i's with $i \in I^c$ if $I^c \neq \varnothing$, or for a subset (possibly all) of the θ_i's if $I^c = \varnothing$, depending on how the additional $1 - \alpha$ confidence assertion is pre-specified, cf. Section 3.2.

More generally, it is shown in Section 3 that similar simultaneous confidence assertions can be made for any MTP in the class of short-cut consonant closed-testing procedures based on marginal p-values and weighted Bonferroni tests for intersection hypotheses considered by Hommel et al. (2007, Section 2.2).

Some preliminaries are given in Section 2, including a description of the Hommel et al. (2007, Section 2.2) class of MTPs. Section 3 contains the main results about simultaneous confidence regions in this article. Illustrations are given in Section 4 that concern clinical studies. Some concluding comments are made in Section 5 – in particular about the drawback that no extra "free" information is provided about θ_i's for which H_i is rejected, unless all H_i's are rejected; and about certain closely related confidence regions by Klaus Strassburger and Frank Bretz presented at the 5th International Conference on Multiple Comparison Procedures (MCP 2007) in Vienna, 9–11 July 2007.

2 Preliminaries

The definitions and assumptions made in Section 1.1 are understood in the sequel. Moreover, it is assumed that $0 < \alpha < 1$ is given.

2.1 The p_i-based test of H_i, and the corresponding $C_{i;\gamma}$-based confidence-region test

The relation (4) means that for any given level $0 < u < 1$, the p_i-based test of $H_i: \theta_i \notin R_i$ versus $H_i^c: \theta_i \in R_i$ that rejects H_i if and only if $p_i \le u$ is equivalent to the confidence-region test (Aitchinson, 1964; Hochberg and Tamhane, 1987, p. 23) that rejects H_i if and only if $C_{i;1-u} \subset R_i$. Thus the p-value p_i is related to the $C_{i;\gamma}$'s through

$$p_i(\boldsymbol{X}) = \begin{cases} \inf A_i, & \text{if} \quad A_i \neq \varnothing, \\ 1, & \text{if} \quad A_i = \varnothing, \end{cases} \tag{10}$$

with $A_i = \{u \in (0\,,\,1);\ C_{i;1-u}(\boldsymbol{X}) \subset R_i\}$, where $\inf A_i$, the greatest lower bound for all values $u \in A_i$, is equal to $\min A_i$ if this minimum exists; cf. Lehmann and Romano (2005, p. 1140). Here are two examples illustrating situations where (1)–(5) hold.

Example 1 – one-sided inferences based on *t* statistics Suppose that $\hat{\theta}_i$ and $s_{\hat{\theta}_i}$ are independent with $\hat{\theta}_i \sim N(\theta_i, \sigma^2_{\hat{\theta}_i})$ and $s^2_{\hat{\theta}_i}/\sigma^2_{\hat{\theta}_i} \sim \chi^2_{(f_i)}/f_i$, so that $(\hat{\theta}_i - \theta_i)/s_{\hat{\theta}_i}$ is distributed according to the central t-distribution with $f_i \ge 1$ degrees of freedom. Suppose that $\boldsymbol{\Theta}_i = \Re$, and that as in (6), the specified target region for θ_i equals $R_i = (\theta_{i,0}\,,\,\infty)$, so one is interested in showing that $\theta_i > \theta_{i,0}$. Let p_i be the p-value of an ordinary t-test of $H_i: \theta_i \le \theta_{i,0}$ versus $H_i^c: \theta_i > \theta_{i,0}$ based on the test statistic $(\hat{\theta}_i - \theta_{i,0})/s_{\hat{\theta}_i}$, so that p_i satisfies $t_{f_i,1-p_i} = (\hat{\theta}_i - \theta_{i,0})/s_{\hat{\theta}_i}$, with $t_{f,q}$ equal to the q-quantile of the central t-distribution with f degrees of freedom. Moreover, let $C_{i;\gamma}$ be the ordinary t-based level-γ one-sided confidence region for θ_i given by

$$C_{i;\gamma} = (\hat{\theta}_i - t_{f_i,\gamma} s_{\hat{\theta}_i}\,,\,\infty)\,. \tag{11}$$

It can then be verified that (1)–(5) hold. For example, $C_{i;1-u}$ given by (11) is a subset of $R_i = (\,\theta_{i,0}\,,\,\infty\,)$ if and only if $t_{f_i,1-u} \le (\hat{\theta}_i - \theta_{i,0})/s_{\hat{\theta}_i}$, i.e. if and only if $p_i \le u$; so (4) holds.

Example 2 – one-sided inferences based on Wilcoxon–Mann–Whitney (WMW) statistics. Suppose that $Y_1, \ldots, Y_{n_A}$ and $Z_1, \ldots, Z_{n_B}$ are two independent random samples of sizes $n_A \ge 1$ and $n_B \ge 1$ from a Y-distribution and a Z-distribution that differ only by a shift θ_i in that Y and $Z - \theta_i$ have a common distribution. The only assumption about this distribution is that its distribution function is continuous, so that the probability of getting any tie among the $n_A + n_B$ observations is zero. Suppose that $\boldsymbol{\Theta}_i = \Re$, and that the specified target region for θ_i equals $R_i = (\theta_{i,0}\,,\,\infty)$, as in Example 1. Let $D_{(1)} < \cdots < D_{(n_A n_B)}$ be the ordered $n_A n_B$ differences $Z_{j''} - Y_{j'}$; and for any $a \in (-\infty\,,\,\infty)$, let $W_{Y,Z-a}$ be the number of pairs $(Y_{j'}, Z_{j''})$ with $Y_{j'} < Z_{j''} - a$ $(1 \le j' \le n_A,\ 1 \le j'' \le n_B)$. Let p_i be the p-value of an ordinary WMW-test of $H_i: \theta_i \le \theta_{i,0}$ versus $H_i^c: \theta_i > \theta_{i,0}$ based on the test statistic $W_{Y,Z-\theta_{i,0}}$, so that p_i equals the upper tail probability $c_h = \Pr\,[W_0 \ge h]$ with $h = W_{Y,Z-\theta_{i,0}}$. Here W_0 denotes the random variable with the known null distribution of $W_{Y,Z}$ when $\theta_i = 0$. Moreover, let $C_{i;\gamma}$ be the WMW-based level-γ one-sided confidence region for θ_i given by (Lehmann, 1975, Section 2.6)

$$C_{i;\gamma} = \begin{cases} (D_{(k)},\,\infty), & \text{if} \quad k \ge 1\,, \\ (-\infty\,,\,\infty), & \text{if} \quad k = 0\,, \end{cases} \tag{12}$$

with $k = k(\gamma)$ equal to the largest integer $h \ge 0$ such that $c_h = \Pr\,[W_0 \ge h]$ is $\ge \gamma$, so $k(\gamma)$ can be determined from the known c_h's. It can then be verified that (1)–(5) hold. For example, $C_{i;1-u}$ given by (12) is a subset of $R_i = (\,\theta_{i,0}\,,\,\infty\,)$ if and only if at least $n_A n_B - k(1-u) + 1$ differences $Z_{j''} - Y_{j'}$

are $\geq \theta_{i,0}$, i.e. if and only if $c_h \leq u$ with $h = W_{Y,Z-\theta_{i,0}}$ (because $k(\gamma) \geq h$ if and only if $\gamma \leq c_h$, and by symmetry, $c_h = 1 - c_{h'}$ with $h' = n_A n_B - h + 1$), i.e. if and only if $p_i \leq u$; so (4) holds.

2.2 The Hommel et al. (2007) class of MTPs, and the index-sets I_{Reject} and I_{Accept}

Hommel et al. (2007, Section 2.2) considered an interesting class of consonant closed-testing procedures based on marginal p-values and weighted Bonferroni-type tests for intersection hypotheses. This class is fairly large, and its members admit a short-cut (similar to Holm's (1979) procedure) that simplifies their implementation and interpretation. It includes Holm's (1979) step-down MTP (with and without weights), the fixed-sequence MTP, various recently proposed Bonferroni-based gatekeeping MTPs, and the fallback MTP. An MTP in this class is defined as follows in terms quantities introduced in Section 1.1.

Suppose that for each non-empty index set $I \subset \{1, \ldots, m\}$, weights $w_1(I), \ldots, w_m(I)$ are given that satisfy

$$\begin{aligned} &0 \leq w_i(I) \leq 1\,, \qquad \textstyle\sum_{i=1}^{m} w_i(I) \leq 1\,, \quad \text{and} \\ &w_i(I) \leq w_i(J) \quad \text{for all} \quad i, I, J \quad \text{with} \quad i \in J \quad \text{and} \quad J \subset I\,. \end{aligned} \tag{13}$$

It is important to note that $w_i(I)$'s may equal 0. Null hypotheses H_i are then rejected through the following algorithm with steps 0, 1, 2, ...

Algorithm 1 Step 0: Set $I_1 = \{1, \ldots, m\}$.

Step $r \geq 1$: Set $S_r = \{i \in I_r\,;\ w_i(I_r) > 0 \text{ and } p_i \leq \alpha w_i(I_r)\}$. If $S_r = \varnothing$ then stop; else reject all H_i with $i \in S_r$, set $I_{r+1} = I_r - S_r$, and go to Step $r + 1$.

The stop occurs in the first Step $r \geq 1$ where either: (a) $I_r \neq \varnothing$ but there is no index $i \in I_r$ for which the two inequalities in the definition of S_r are both satisfied, in which case some H_i's are not rejected; or (b) I_r is set equal to $\varnothing$ in the previous step, in which case all $H_1, \ldots, H_m$ are rejected. It is understood that H_i's that are not rejected through Algorithm 1 are accepted. Thus, the two complementary index-sets

$$\begin{aligned} I_{\text{Reject}} &= (\text{set of indexes } i \in \{1, \ldots, m\} \text{ of } H_i\text{s rejected through Algorithm 1})\,, \\ I_{\text{Accept}} &= \{1, \ldots, m\} - I_{\text{Reject}}\,, \end{aligned} \tag{14}$$

are well-defined functions of $p_1, \ldots, p_m$, and $|I_{\text{Reject}}| + |I_{\text{Accept}}| = m$.

The MTP defined by Algorithm 1 has multiple-level α, i.e. controls in the strong sense its type-I family-wise error rate to be $\leq \alpha$ (Hochberg and Tamhane, 1987, p. 3; Westfall and Young, 1993, Section 1.2). This is because this MTP is equivalent to a closed-testing procedure with the level-α Bonferroni-type test for each intersection hypothesis $H_I = \cap_{i \in I} H_i$, $\varnothing \neq I \subset \{1, \ldots, m\}$, that rejects H_I if and only if for some $i \in I$, $w_i(I) > 0$ and $p_i \leq \alpha w_i(I)$.

Remark 1 If the condition $w_i(I_r) > 0$ is removed in the definition of S_r in Algorithm 1, more H_i's may be rejected, namely also H_i's with $p_i = 0$ and $w_i(I_1) = 0$. Also such a modified algorithm has multiple level α, because (2) implies that $\Pr[p_i = 0] = 0$ if H_i is true. However, the present formulation of Algorithm 1: (a) matches the developments in Hommel et al. (2007, Section 2.2), where $p_i/w_i(I)$ is defined on page 4066 to be 1 if $w_i(I) = 0$; and (b) avoids non-standard variants of common MTPs, e.g. of the fixed-sequence MTP with $w_i(I)$'s given by (16) where an H_i with $p_i = 0$ would be rejected even if there are preceding hypotheses that are not rejected.

Remark 2 The MTP given by Algorithm 1 does not depend on the values of $w_i(I)$'s with $i \notin I$, so provided they satisfy (13), such values may be arbitrarily specified. Of course, it follows from (13) that in case $\sum_{i \in I} w_i(I) = 1$, then necessarily, $w_i(I) = 0$ if $i \notin I$. The MTP given by Algorithm 1 will

be called α-exhaustive if $\sum_{i\in I} w_i(I) = 1$ for each non-empty $I \subset \{1, \ldots, m\}$, as in Hommel et al. (2007).

Hommel et al. (2007, Section 2.2) described $w_i(I)$'s satisfying (13) that correspond to various common MTPs. Here are two important special cases.

Weights $w_i(I)$ for Holm's (1979) MTP Suppose that $v_1, \ldots, v_m$ are given positive weights summing up to 1 that are associated with $H_1, \ldots, H_m$, or rather with the target rejection assertions "$\theta_1 \in R_1$", ..., "$\theta_m \in R_m$". These weights may e.g. reflect importance in that weights corresponding to more important target assertions are greater than those corresponding to less important target assertions. Holm's (1979, pp. 69–70) MTP for $H_1, \ldots, H_m$ based on these weights $v_1, \ldots, v_m$ corresponds to Algorithm 1 with $w_1(I), \ldots, w_m(I)$ defined for each non-empty $I \subset \{1, \ldots, m\}$ by

$$w_i(I) = \begin{cases} v_i / \sum_{j\in I} v_j \,, & \text{if} \quad i \in I \,, \\ 0 \,, & \text{otherwise} \,. \end{cases} \tag{15}$$

Clearly, this MTP is α-exhaustive (in the sense described in Remark 2).

Weights $w_i(I)$ for the fixed-sequence MTP Suppose that the sequence in which $H_1, \ldots, H_m$ are stated is relevant and fixed, e.g. with the assertion "$\theta_1 \in R_1$" being of most importance, and the assertion "$\theta_m \in R_m$" being of least importance. The rejection rule of the fixed-sequence MTP is then to reject H_i if and only if $p_j \leq \alpha$ for all $j \leq i$. The rejections made with this rule are the same as those made through Algorithm 1 with $w_1(I), \ldots, w_m(I)$ defined for each non-empty $I \subset \{1, \ldots, m\}$ by

$$w_i(I) = \begin{cases} 1 \,, & \text{if} \quad i \in I \quad \text{and} \quad i \quad \text{is the smallest index in } I \,, \\ 0 \,, & \text{otherwise} \,. \end{cases} \tag{16}$$

Clearly, also this MTP is α-exhaustive.

Remark 3 An MTP defined by Algorithm 1 with $w_i(I)$'s satisfying (13) that is not α-exhaustive can be improved to reject more by replacing its $w_i(I)$'s by $w_i'(I)$'s such that $w_i'(I) \geq w_i(I)$ for $i \in I$ that satisfy (13) and are such that the resulting MTP becomes α-exhaustive (so $w_i'(I) = 0$ if $i \notin I$). Perhaps the simplest example of this is the Bonferroni MTP that in terms of the positive $v_1, \ldots, v_m$ in (15) has weights $w_i(I) = v_i$, $1 \leq i \leq m$. This MTP can be improved to reject more by replacing its $w_i(I)$'s by the weights (15), which leads to Holm's MTP based on the same underlying $v_1, \ldots, v_m$ for $H_1, \ldots, H_m$.

3 Main Results About Confidence Regions

In the sequel it is assumed that a set of weights $w_i(I)$ satisfying (13) is specified, so that the MTP defined by Algorithm 1 with these $w_i(I)$'s is a member of the class of consonant Bonferroni-based closed-testing procedures considered by Hommel et al. (2007, Section 2.2).

3.1 Simultaneous confidence regions for $\theta_1, \ldots, \theta_m$ and θ_{m+1}

In addition to the definitions and assumptions in Section 1.1, let $\theta_{m+1} = \theta_{m+1}(P)$ be any specified quantity of interest, and let $C_{m+1;1-\alpha} = C_{m+1;1-\alpha}(\boldsymbol{X})$ be any specified confidence region for θ_{m+1} that has marginal-level $1 - \alpha$, so that

$$\Pr\left[\theta_{m+1} \in C_{m+1;1-\alpha}\right] \geq 1 - \alpha \quad \text{for all} \quad P \in \boldsymbol{P} \,. \tag{17}$$

Neither θ_{m+1} nor $C_{m+1;1-\alpha}$ are restricted to be of any particular kind/dimension. It is assumed that the set $\boldsymbol{\Theta}_{m+1} = \{\theta_{m+1}(P);\ P \in \boldsymbol{P}\}$ of potential θ_{m+1}-values is known. As will be shown, θ_{m+1} and $C_{m+1;1-\alpha}$ may be chosen to sharpen inferences about $\theta_1, \ldots, \theta_m$.

The main result about confidence regions, Theorem 1, is stated in terms of random regions $C_1^* = C_1^*(\boldsymbol{X}), \ldots, C_{m+1}^* = C_{m+1}^*(\boldsymbol{X})$ for $\theta_1, \ldots, \theta_{m+1}$ defined as follows in terms of: quantities in (1)–(5), the index-sets I_{Reject} and I_{Accept} in (14), and the weights $w_i(I_{\text{Accept}})$ given by the specified set of $w_i(I)$'s satisfying (13) used in Algorithm 1. Recall from (13) that $w_i(I)$'s may equal 0, and from (5) that by convention, $C_{i;\gamma} = \boldsymbol{\Theta}_i$ if $\gamma = 1$. Now, for each $i = 1, \ldots, m+1$, let

$$C_i^* = \begin{cases} R_i\,, & \text{if} \quad i \in I_{\text{Reject}}\,, \\ R_i \cup C_{i;1-\alpha\, w_i(I_{\text{Accept}})}\,, & \text{if} \quad i \in I_{\text{Accept}}\,, \\ C_{m+1;1-\alpha}\,, & \text{if} \quad i = m+1 \quad \text{and} \quad |I_{\text{Reject}}| = m\,, \\ \boldsymbol{\Theta}_{m+1}\,, & \text{if} \quad i = m+1 \quad \text{and} \quad |I_{\text{Reject}}| < m\,. \end{cases} \tag{18}$$

Clearly, the assertions "$\theta_i \in C_i^*$" for $i \in I_{\text{Reject}}$ correspond to the rejections made through Algorithm 1, so extra "free" information is provided by the assertions "$\theta_i \in C_i^*$" with $i \in I_{\text{Accept}}$, or by the assertion "$\theta_{m+1} \in C_{m+1}^*$" in case $|I_{\text{Reject}}| = m$. Note also that if $|I_{\text{Reject}}| < m$, the assertion "$\theta_{m+1} \in C_{m+1}^*$" is non-informative in that C_{m+1}^* then equals $\boldsymbol{\Theta}_{m+1}$. The proof of Theorem 1 is given in the appendix.

Theorem 1 The random regions $C_1^*, \ldots, C_{m+1}^*$ given by (18) simultaneously cover $\theta_1, \ldots, \theta_{m+1}$, respectively, with probability satisfying

$$\Pr\left[\theta_i \in C_i^* \text{ for all } 1 \le i \le m+1\right] \ge\ 1 - \alpha\,. \tag{19}$$

It follows from (19) that $C_1^*, \ldots, C_{m+1}^*$ constitute simultaneous $1-\alpha$ confidence regions for $\theta_1, \ldots, \theta_{m+1}$. Note from (18) that an informative assertion is: (a) made for each θ_i with $i \in I_{\text{Reject}}$ or with $i \in I_{\text{Accept}}$ and $w_i(I_{\text{Accept}}) > 0$ (unless $R_i \cup \{\theta_i\} = \boldsymbol{\Theta}_i$, as is the case with (8)); and (b) sometimes made also for θ_{m+1}, namely when $|I_{\text{Reject}}| = m$, i.e. when the assertions "$\theta_i \in R_i$" are made for all $\theta_1, \ldots, \theta_m$. Here "informative" means that an assertion is not just stating that a θ_i belongs to its range space $\boldsymbol{\Theta}_i$.

The implementation is simple. Given the quantities in (1)–(5) and the specified set of $w_i(I)$'s satisfying (13) for Algorithm 1, one first determines the index-sets I_{Reject} and I_{Accept} in (14), and then one determines the C_i^*'s from (18). In fact, as illustrated in Section 4.1, it is sometimes possible to determine I_{Reject} and I_{Accept} without numerical evaluation of $p_1, \ldots, p_m$.

3.2 Choice of $\boldsymbol{\theta_{m+1}}$ and $\boldsymbol{C_{m+1;1-\alpha}}$ to sharpen inferences about $\boldsymbol{\theta_1, \ldots, \theta_m}$

Choice 1 – for general $\boldsymbol{\theta_i}$'s and $\boldsymbol{R_i}$'s Recall from Section 1.1 that θ_i's, R_i's, and/or $C_{i;\gamma}$'s with distinct indexes $1 \le i \le m$ may be of different kinds/dimensions. In such a general situation, a possible choice for θ_{m+1} and $C_{m+1;1-\alpha}$ in (17) is as follows. Let $M = \{1, \ldots, m\}$, and set

$$\theta_{m+1} = (\theta_1, \ldots, \theta_m) \quad \text{and} \quad C_{m+1,1-\alpha} = C_{1,1-\alpha\, w_1(M)} \times \cdots \times C_{m,1-\alpha\, w_m(M)}\,, \tag{20}$$

where $\times$ stands for direct product, and the weights $w_i(M)$ are given by the specified set of $w_i(I)$'s used in Algorithm 1. That is, $C_{m+1;1-\alpha}$ is a rectangular region for the vector θ_{m+1} that consists of m Bonferroni-adjusted marginal confidence regions, so that (17) holds. Moreover, if $|I_{\text{Reject}}| = m$, then $S_1 \subset M$ is non-empty in Step $r = 1$ of Algorithm 1, and for each $i \in S_1$: $w_i(M) > 0$ and $p_i \le \alpha\, w_i(M)$, so $C_{i;1-\alpha\, w_i(M)} \subset R_i$ because of relation (4). This shows how (20) leads to sharpenings (of the form "$\theta_i \in C_{i;1-\alpha\, w_i(M)}$" with $C_{i;1-\alpha\, w_i(M)} \subset R_i$ for at least all $i \in S_1$) of the assertions "$\theta_i \in R_i$" made through $C_1^*, \ldots, C_m^*$ in (18) in case $|I_{\text{Reject}}| = m$. See Sections 4.1 and 4.2.2 for illustrations.

Choice 2 – for real $\boldsymbol{\theta_i}$'s in similar scales, and one-sided $\boldsymbol{R_i}$'s Other choices than (20) that concern $\theta_1, \ldots, \theta_m$ are possible in particular situations. Here is an example. Again, let $M = \{1, \ldots, m\}$, and suppose that for each $i \in M$, $\boldsymbol{\Theta}_i = \Re$, and the specified target region R_i and marginal-level γ confidence regions $C_{i;\gamma}$ for θ_i are of the form

$$R_i = (\theta_{i,0}\,, \infty) \quad \text{and} \quad C_{i;\gamma} = (L_{i,\gamma}, \infty)\,, \tag{21}$$

with $\theta_{i,0} \in (-\infty, \infty)$; cf. (6). Note that, for example, the intervals (11) and (12) are of the form $(L_{i,\gamma}, \infty)$. Now, set $\lambda = \min\{L_{i,1-\alpha} - \theta_{i,0};\ i \in M\}$. Then θ_{m+1} and $C_{m+1;1-\alpha}$ in (17) may be chosen as

$$\theta_{m+1} = (\theta_1, \ldots, \theta_m) \quad \text{and} \quad C_{m+1,1-\alpha} = (\theta_{1,0} + \lambda, \infty) \times \cdots \times (\theta_{m,0} + \lambda, \infty). \tag{22}$$

That is, $C_{m+1;1-\alpha}$ is a rectangular region for the vector θ_{m+1} that consists of m intervals $(\theta_{i,0} + \lambda, \infty)$. To see that (17) holds, let i_* denote the smallest (unknown) index $i \in M$ for which $\theta_i - \theta_{i,0}$ equals $\min(\theta_1 - \theta_{1,0}, \ldots, \theta_m - \theta_{m,0})$. Then: (a) $\Pr[L_{i_*;1-\alpha} - \theta_{i_*,0} < \theta_{i_*} - \theta_{i_*,0}] \geq 1 - \alpha$; (b) $\Pr[\lambda \leq L_{i_*,1-\alpha} - \theta_{i_*,0}] = 1$; so that (c) $\Pr[\lambda < \theta_i - \theta_{i,0} \text{ for all } i \in M] \geq 1 - \alpha$. Moreover, if $|I_{\text{Reject}}| = m$, then for each $i \in M$ there is a Step $r \geq 1$ in Algorithm 1 in which H_i is rejected, i.e. in which $i \in I_r$, $w_i(I_r) > 0$ and $p_i \leq \alpha w_i(I_r)$; so that $p_i \leq \alpha$, and thus $\theta_{i,0} \leq L_{i,1-\alpha}$ because of relation (4). That is, if $|I_{\text{Reject}}| = m$, then $\theta_{i,0} \leq L_{i,1-\alpha}$ for all $i \in M$, so $\lambda \geq 0$ in (22). This shows how (22) leads to sharpenings (of the form "$\theta_{i,0} + \lambda < \theta_i$" with $\lambda \geq 0$, for all $i \in M$) of the assertions "$\theta_{i,0} < \theta_i$" made through $C_1^*, \ldots, C_m^*$ in (18) in case $|I_{\text{Reject}}| = m$. See Section 4.1 for an illustration.

It is instructive to compare the sharpenings provided by (20) with those provided by (22) when the specified set of $w_i(I)$'s used in Algorithm 1 is given by (16) in the particular situation underlying (22) just described. Recall that (16) corresponds to the fixed-sequence MTP for $H_1, \ldots, H_m$, and suppose for simplicity that $\theta_{i,0} = 0$ in (21) and (22). It can then be verified that the sharpenings provided by (20) reduce to the single assertion that θ_1 belongs to $C_{1,1-\alpha} = (L_{1,1-\alpha}, \infty)$, because $w_1(M) = 1$ and $w_i(M) = 0$ for $2 \leq i \leq m$ in (20); whereas the sharpenings provided by (22) reduce to the assertions that θ_i belongs to (λ, ∞) for all $i \in M$, with $\lambda = \min\{L_{i,1-\alpha};\ i \in M\} \geq 0$. Thus, more assertions are made with (22), but the assertion about θ_1 with (20) is sharper in that $\lambda \leq L_{1,1-\alpha}$.

The fact that a common λ is used for the component intervals of $C_{m+1;1-\alpha}$ in (22) means that in contrast to (20), the choice (22) is reasonable only if $\theta_1, \ldots, \theta_m$ are in a common scale (or sufficiently similar scales), although the simultaneous coverage assertions are formally valid. For example, (22) is reasonable in a dose-response study where several dose are compared in a given sequence to a common control with respect to a response variable as described in Hsu and Berger (1999, Section 2). The simultaneous confidence intervals without multiplicity adjustment considered there essentially correspond to using the fixed-sequence weights (16) in Algorithm 1, and the marginal confidence regions (11) in (21), (22) and (18).

Variants of (20) and (22) can be obtained by restricting considerations to indexes in a pre-specified subset M^* of M of particular interest, i.e. to $C_{m+1;1-\alpha}$-inferences concerning the subvector θ_{m+1} of $(\theta_1, \ldots, \theta_m)$ with component-indexes in M^*. Then $C_{m+1;1-\alpha}$ in (20) can be replaced by the direct product of component regions $C_{i,1-\alpha\, w_i(M)/W}$, $i \in M^*$, where $W = \sum_{i \in M^*} w_i(M)$ is assumed to be positive; whereas $C_{m+1;1-\alpha}$ in (22) can be replaced by the direct product of component regions $(\theta_{i,0} + \lambda^*, \infty)$, $i \in M^*$, where $\lambda^* = \min\{L_{i,1-\alpha} - \theta_{i,0};\ i \in M^*\}$. Such assertions about θ_i's with $i \in M^*$ are at least as sharp as the corresponding assertions given by (20) and (22).

4 Illustrations

Two illustrations are now given using the weights $w_i(I)$ given by (15) for Holm's MTP – one where each $\theta_i \in \Re$, and one where each $\theta_i \in \Re^D$. Modifications required for other experimental settings and/or other MTPs based on weights $w_i(I)$ satisfying (13) should be evident.

4.1 Simultaneous confidence intervals aimed at showing that θ_i's are positive

Suppose that one is interested in simultaneous $1 - \alpha$ confidence intervals for quantities $\theta_1, \ldots, \theta_m$ in $\Re$ that: (a) are aimed at showing that θ_i's are >0; and (b) indicate by how much one missed the target assertion "$\theta_i > 0$" through a lower confidence bound for θ_i whenever this target assertion can-

not be made. As mentioned in connection with (6), an inequality $\theta_i > 0$ may correspond to superiority of one treatment relative to another with respect to a certain response characteristic, and the indexes $1 \leq i \leq m$ may then correspond to several response variables and/or several treatment comparisons. (Of course, any target assertion originally of the form (6) or (7) can be changed into a target assertion of positivity through substraction of $\theta_{i,0}$ and/or change of sign.)

Röhmel et al. (2006) considered a related multiple-testing problem where two treatments are compared and one is interested in showing non-inferiority for certain key variables and superiority for at least one of them. The motivating example mentioned by Röhmel et al. (2006) was a confirmatory clinical study where a drug A was to be compared to placebo P with respect to its pain-reducing effect, and where for ethical reasons an established pain-reducing drug was used as rescue medication (in case of too much pain for a subject – thus reducing the pain during the pain-assessment period for the subject). Because the amount of rescue medication given to a subject is an indirect measure of the subject's pain, the A versus P comparison was formulated as a multiple-testing non-inferiority/superiority problem of the kind just mentioned with two variables: a variable Y_1 reflecting pain, and the amount Y_2 of rescue medication given; assuming bivariate normality for Y_1 and Y_2 under each treatment. Röhmel et al. (2006) proposed a certain hierarchical MTP, and made comparisons with various previously proposed MTPs. In order to get directions right here, θ_i is defined as the $(\mathrm{P} - \mathrm{A})$-difference of true means of Y_i, so that $\theta_i > 0$ corresponds to A being superior to P. This example is now used to illustrate how simultaneous confidence regions (18) for θ_1 and θ_2 are obtained and how they behave.

Thus, suppose that in (1)–(5): $m = 2$, $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2 = \Re$, and R_i's and $C_{i;\gamma}$'s with $1 \leq i \leq m$ are of the form (21) with $\theta_{i,0} = 0$. The desired simultaneous confidence level $1 - \alpha$ is chosen to be $1 - 0.025$, because of the one-sided formulation of the problem.

Now, the following marginal-level $1 - \alpha/2$ and marginal-level $1 - \alpha$ ordinary one-sided two-sample t-intervals of the form (11) for θ_1 and θ_2 were reported by Röhmel et al. (2006):

$$\begin{aligned} C_{1,1-\alpha/2} &= (0.3141\,,\,\infty), \qquad & C_{2,1-\alpha/2} &= (-1.3252\,,\,\infty), \\ C_{1,1-\alpha} &= (0.5333\,,\,\infty), \qquad & C_{2,1-\alpha} &= (-1.0682\,,\,\infty). \end{aligned} \tag{23}$$

Comparing these intervals with the target regions $R_1 = R_2 = (0\,,\,\infty)$, it is evident from (10) that the corresponding p-values satisfy $p_1 \leq \alpha/2$ and $p_2 > \alpha$. It then follows from Algorithm 1 with weights $w_i(I)$ specified as (15) for Holm's MTP (assuming $v_1 = v_2 = 1/2$) that: $I_{\text{Reject}} = \{1\}$, $I_{\text{Accept}} = \{2\}$, and thus $|I_{\text{Reject}}| = |I_{\text{Accept}}| = 1 < m$. The informative simultaneous $1 - \alpha$ confidence assertions resulting from (18) based on (23) then are "$\theta_1 \in C_1^*$" = "$\theta_1 \in R_1$" and "$\theta_2 \in C_2^*$" = "$\theta_2 \in R_2 \cup C_{2,1-\alpha}$", i.e.

$$\text{“}\theta_1 > 0\text{”} \quad \text{and} \quad \text{“}\theta_2 > -1.0682\text{”}. \tag{24}$$

Röhmel et al. (2006) mentioned that the non-inferiority (lower) limits -1 and -2 for θ_1 and θ_2, respectively, were specified *a priori*. The simultaneous $1 - \alpha$ confidence assertions (24) thus enable us to: (a) make the inference that A is non-inferior to P with respect to θ_1 and θ_2 because (24) implies $\theta_1 > -1$ and $\theta_2 > -2$; (b) make the inference that A is superior to P with respect to θ_1 because (24) implies $\theta_1 > 0$; and (c) make the additional inference that $\theta_2 > -1.0682$. It is appealing that an inference can immediately be made from (24) about whether A is non-inferior to P with respect to θ_2 – for any alternative non-inferiority limits that one may be interested in.

Interestingly, Röhmel et al. (2006): (a) mentioned that a question that always arises in the context of multiple testing is the search for adequate confidence intervals or confidence sets; and (b) recommended (in view of the complexity of their hierarchical MTP, in case the correlation structure is unknown) to use a simple Bonferroni approach. Clearly, $C_{1,1-\alpha/2}$ and $C_{2,1-\alpha/2}$ in (23) constitute such simultaneous $1 - \alpha$ confidence intervals for θ_1 and θ_2, resulting in the assertions

$$\text{“}\theta_1 > 0.3141\text{”} \quad \text{and} \quad \text{“}\theta_2 > -1.3252\text{”}. \tag{25}$$

Note that the inference about θ_1 is sharper in (25) than in (24), whereas the inference about θ_2 is sharper in (24) than in (25). Intuitively, this sharper inference about θ_2 is possible by not making an inference sharper than necessary about θ_1 (as specified by the target assertion of positivity).

The m quantities θ_i are not in similar scales, so the choice (20) for θ_{m+1} and $C_{m+1;1-\alpha}$ seems more appropriate than (22). It is, however, informative to see how the regions (18) for θ_1 and θ_2 behave with the choices (20) and (22) if the lower bounds of the intervals for θ_2 in (23) are translated somewhat upwards. A translation by 1.1 leads to $C'_{2,1-\alpha/2} = (-0.2252\,,\,\infty)$ and $C'_{2,1-\alpha} = (0.0318\,,\,\infty)$, in which case (20) leads to "$\theta_1 > 0.3141$" and "$\theta_2 > 0$", whereas (22) leads to "$\theta_1 > 0.0318$" and "$\theta_2 > 0.0318$". A translation by 1.4 leads to $C''_{2,1-\alpha/2} = (0.0748\,,\,\infty)$ and $C''_{2,1-\alpha} = (0.3318\,,\,\infty)$, in which case (20) leads to "$\theta_1 > 0.3141$" and "$\theta_2 > 0.0748$", whereas (22) leads to "$\theta_1 > 0.3318$" and "$\theta_2 > 0.3318$".

Finally, it is should be noted that these developments do not require that Y_1 and Y_2 are bivariate normal – the t-intervals (23) require only marginal normality – and if appropriate, other assumptions might have been used for the marginal distributions of Y_1 and/or Y_2, e.g. a shift model in combination with non-parametric intervals (12). It should be evident how this application of Theorem 1 can be adapted to other experimental settings, other kinds of data and distributional assumptions, and/or other MTPs based on weights $w_i(I)$ satisfying (13).

4.2 Confidence regions corresponding to MTPs for sub-families F_i of H's

The θ_i's, R_i's and $C_{i;\gamma}$'s in (1)–(5) may be multi-dimensional. Each $C_{i;\gamma}$ may therefore correspond to a local (i.e. marginal) MTP for a sub-family F_i of null hypotheses H that concern the components of θ_i. This is now illustrated. First, a weighted version of the Holm-type MTP of Bauer et al. (1998, appendix) for sub-families $F_1, \ldots, F_m$ of H's is given in Section 4.2.1. This useful MTP was originally stated without weights for sub-families, but the version given here is a straightforward generalization. Then, in Section 4.2.2, a clinical-study application of this Holm-type MTP is considered where a local fixed-sequence MTP concerning θ_i-components is used within each F_i, and it is shown how the corresponding confidence regions (18) provide extra "free" information.

4.2.1 A weighted version of the Holm-type MTP of Bauer et al. (1998, appendix)

Consider a multiple-testing situation with n null hypotheses grouped into $m \geq 2$ sub-families $F_1, \ldots, F_m$, where F_i consists of $n_i \geq 1$ null hypotheses $H_{i,1}, \ldots, H_{i,n_i}$. Suppose that for each $i = 1, \ldots, m$, an MTP is given for F_i that locally (i.e. marginally within F_i) has multiplicity-adjusted (Westfall and Young, 1993) p-values $\tilde{p}_{i,1}, \ldots, \tilde{p}_{i,n_i}$ associated with $H_{i,1}, \ldots, H_{i,n_i}$, respectively, such that for any $0 < u < 1$, the $(\tilde{p}_{i,1}, \ldots, \tilde{p}_{i,n_i})$-based MTP for F_i that rejects $H_{i,j}$'s if and only if their adjusted p-values satisfy $\tilde{p}_{i,j} \leq u$ has multiple-level u. Note that such a local MTP for F_i does not necessarily have to belong to the Hommel et al. (2007, Section 2.2) class of MTPs based on underlying marginal p-values. Set $\tilde{p}_i = \max\{\tilde{p}_{i,1}, \ldots, \tilde{p}_{i,n_i}\}$, and note that this $\tilde{p}_i$ can be viewed as a p-value for the entire F_i in that for any $0 < u < 1$, the $(\tilde{p}_{i,1}, \ldots, \tilde{p}_{i,n_i})$-based MTP for F_i rejects all $H_{i,j}$'s in F_i if and only if $\tilde{p}_i \leq u$. Moreover, let $v_1, \ldots, v_m$ be positive weights summing up to 1 associated with the sub-families $F_1, \ldots, F_m$, respectively. The role of the $\tilde{p}_i$'s and v_i's here is similar to that of the p_i's and v_i's in Algorithm 1 with weights $w_i(I)$ given by (15) for Holm's MTP. Algorithm 2 is stated in terms of: the $\tilde{p}_{i,j}$'s and $\tilde{p}_i$'s just introduced, and the weights $w_i(I)$ given by (15) in terms of the $v_1, \ldots, v_m$ associated with $F_1, \ldots, F_m$.

Algorithm 2 Step 1: Set $I_1 = \{1, \ldots, m\}$. If there is an index $i_1 \in I_1$ for which $\tilde{p}_{i_1} \leq \alpha w_{i_1}(I_1)$, then reject the entire F_{i_1} and go to next step; otherwise, for each $i \in I_1$, reject all $H_{i,j}$'s within F_i for which $\tilde{p}_{i,j} \leq \alpha w_i(I_1)$, and stop.

Step $r < m$: Set $I_r = I_{r-1} - \{i_{r-1}\}$. If there is an index $i_r \in I_r$ for which $\tilde{p}_{i_r} \leq \alpha w_{i_r}(I_r)$, then reject the entire F_{i_r} and go to next step; otherwise, for each $i \in I_r$, reject all $H_{i,j}$'s within F_i for which $\tilde{p}_{i,j} \leq \alpha w_i(I_r)$, and stop.

Step m: Set $I_m = I_{m-1} - \{i_{m-1}\} \equiv \{i_m\}$. Reject all $H_{i_m,j}$'s within F_{i_m} for which $\tilde{p}_{i_m,j} \leq \alpha$, and stop.

It is understood in Step $1 \leq r < m$ that if there are more than one $i \in I_r$ for which $\tilde{p}_i \leq \alpha w_i(I_r)$, then i_r may be chosen arbitrarily as any of these i's. The set of rejected $H_{i,j}$'s is invariant under different such choices. Moreover it is understood that $H_{i,j}$'s that are not rejected through Algorithm 2 are accepted. It is appealing that when entire F_i's can no longer be rejected, it is still possible to reject null hypotheses within each of the remaining non-rejected F_i's. It can be verified that if each F_i consists of a single null hypothesis H_i, Algorithm 2 is equivalent to Algorithm 1 with weights (15).

The important result here is that the Holm-type MTP defined by Algorithm 2 for the family of all $n = n_1 + \cdots + n_m$ null hypotheses $H_{i,j}$ has multiple-level α. This can be shown through an extension of the argument sketched by Bauer et al. (1998, appendix). Power properties are, of course, highly dependent on how null hypotheses are grouped into F_i's , and on the choice of local MTPs for these F_i's.

4.2.2 Fixed-sequence local MTPs and confidence regions for θ_i-components

We now consider an interesting clinical-study application (Bauer, Brannath and Posch, 2001, Section 4.1) of the Holm-type MTP given by Algorithm 2, with $m = 2$ families F_1 and F_2, and local fixed-sequence MTPs within F_i's. Suppose D doses $\tau_1 < \cdots < \tau_D$ of a drug are to be compared to a control treatment τ_0 in two variables – an efficacy variable Y_E and a safety variable Y_S – with the aim of showing that certain doses τ_d are superior to τ_0 in Y_E and/or non-inferior to τ_0 in Y_S. Let

$$\begin{aligned} \theta_1 &= (\theta_{1,1}, \ldots, \theta_{1,D}) = (\mu_1^{Y_E} - \mu_0^{Y_E}, \ldots, \mu_D^{Y_E} - \mu_0^{Y_E}) \,, \\ \theta_2 &= (\theta_{2,1}, \ldots, \theta_{2,D}) = (\mu_1^{Y_S} - \mu_0^{Y_S}, \ldots, \mu_D^{Y_S} - \mu_0^{Y_S}) \,, \end{aligned} \tag{26}$$

where $\mu_d^{Y_E} - \mu_0^{Y_E}$ and $\mu_d^{Y_S} - \mu_0^{Y_S}$ are the true $(\tau_d - \tau_0)$-differences in Y_E and Y_S. Let F_1 consist of "efficacy" null hypotheses $H_{1,d}: \theta_{1,d} \leq 0$ (to identify doses τ_d with $\theta_{1,d} > 0$), and F_2 consist of "safety" null hypotheses $H_{2,d}: \theta_{2,d} \geq \delta$ (to identify doses τ_d with $\theta_{2,d} < \delta$, a given safety tolerance limit). Moreover, let v_1 and v_2 be positive, possibly unequal, weights associated with F_1 and F_2.

Now, assume that for each $1 \leq d \leq D$ and $0 < \gamma < 1$: (a) the marginal-level γ confidence interval $(L_{1,d;\gamma}, \infty)$ with $L_{1,d;\gamma} = \hat{\theta}_{1,d} - t_{f_{1,d},\gamma} s_{\hat{\theta}_{1,d}}$ is specified for $\theta_{1,d}$; and (b) the marginal-level γ confidence interval $(-\infty, U_{2,d;\gamma})$ with $U_{2,d;\gamma} = \hat{\theta}_{2,d} + t_{f_{2,d},\gamma} s_{\hat{\theta}_{2,d}}$ is specified for $\theta_{2,d}$. Here it is assumed that marginally for each index-pair (i, d), $\hat{\theta}_{i,d} \sim N(\theta_{i,d}, \sigma^2_{\hat{\theta}_{i,d}})$ and $s^2_{\hat{\theta}_{i,d}} / \sigma^2_{\hat{\theta}_{i,d}} \sim \chi^2_{(f_{i,d})} / f_{i,d}$ are independent; cf. Example 1. The underlying marginal t-tests for $H_{1,d}$ and $H_{2,d}$ that correspond to these confidence intervals have p-values satisfying

$$\begin{aligned} p_{1,d} &= (u \text{ such that } L_{1,d;1-u} = 0) \,, \\ p_{2,d} &= (u \text{ such that } U_{2,d;1-u} = \delta) \,. \end{aligned} \tag{27}$$

It is anticipated (but not assumed – the proposed procedure is valid anyhow) that $\theta_{1,1} \leq \cdots \leq \theta_{1,D}$ and $\theta_{2,1} \geq \cdots \geq \theta_{2,D}$, so the fixed-sequence testing is specified to be downward in dose within F_1, and upward in dose within F_2. In terms of (27), the local adjusted p-values introduced in Section 4.2.1 become

$$\begin{aligned} \tilde{p}_{1,d} &= \max\{p_{1,j};\ d \leq j \leq D\} \,, \qquad & \tilde{p}_1 &= \max\{p_{1,j};\ 1 \leq j \leq D\}, \\ \tilde{p}_{2,d} &= \max\{p_{2,j};\ 1 \leq j \leq d\}, \qquad & \tilde{p}_2 &= \max\{p_{2,j};\ 1 \leq j \leq D\}. \end{aligned} \tag{28}$$

The Holm-type MTP defined by Algorithm 2 based on (28) can then be used to make rejections in the total family of all $n = 2D$ null hypotheses in F_1 and F_2 at multiple-level α. As pointed out by Bauer et al. (2001, p. 609), this MTP can identify doses τ_d that: are effective and safe ($H_{1,d}$ and $H_{2,d}$ rejected), effective but possibly not safe ($H_{1,d}$ rejected but not $H_{2,d}$), and safe but possibly not effective ($H_{2,d}$ rejected but not $H_{1,d}$).

Extra "free" information can be obtained through (18) with quantities in (1)–(5) as follows: D-dimensional range-spaces $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2 = \Re^D$ for θ_1 and θ_2 in (26); D-dimensional rectangular target

regions $R_1 = (R_E)^D$ and $R_2 = (R_S)^D$ with $R_E = (0\,,\,\infty)$ and $R_S = (-\infty\,,\,\delta)$; null hypotheses $H_i : \theta_i \notin R_i$; marginal p-values p_i equal to $\tilde{p}_i$ in (28); and D-dimensional rectangular confidence regions $C_{i,\gamma}$ for θ_i based on Hsu-Berger-type confidence intervals for the components of θ_i that correspond to the fixed-sequence testing within F_i.

More precisely, the $C_{i,\gamma}$'s are defined as follows. Let $N_{1,\gamma}$ equal the number of indexes $1 \le d \le D$ for which $\tilde{p}_{1,d}$ in (28) is $\le 1-\gamma$, and let $N_{2,\gamma}$ equal the number of indexes $1 \le d \le D$ for which $\tilde{p}_{2,d}$ in (28) is $\le 1-\gamma$. That is, $N_{1,\gamma}$ equals the number of successive rejections of $H_{1,d}$'s downward in dose made by the fixed-sequence MTP used locally within F_1 with multiple level $1-\gamma$, and $N_{2,\gamma}$ equals the number of successive rejections of $H_{2,d}$'s upward in dose made by the fixed-sequence MTP used locally within F_2 with multiple level $1-\gamma$. The marginal-level γ rectangular confidence regions for θ_1 and θ_2 in (26) are given by

$$C_{1;\gamma} = \begin{cases} \Re^{D-1} \times (L_{1,D;\gamma}\,,\,\infty)\,, & \text{if} \quad N_{1,\gamma} = 0\,, \\ \Re^{D-N_{1,\gamma}-1} \times (L_{1,D-N_{1,\gamma};\gamma}\,,\,\infty) \times (R_E)^{N_{1,\gamma}}\,, & \text{if} \quad 0 < N_{1,\gamma} < D\,, \\ (L_{\min\,;\gamma},\,\infty)^D\,, & \text{if} \quad N_{1,\gamma} = D\,, \end{cases} \tag{29}$$

$$C_{2;\gamma} = \begin{cases} (-\infty\,,\,U_{2,1;\gamma}) \;\times\; \Re^{D-1}\,, & \text{if} \quad N_{2,\gamma} = 0\,, \\ (R_S)^{N_{2,\gamma}} \times (-\infty\,,\,U_{2,N_{2,\gamma}+1;\gamma}) \times \Re^{D-N_{2,\gamma}-1}\,, & \text{if} \quad 0 < N_{2,\gamma} < D\,, \\ (-\infty\,,\,U_{\max\,;\gamma})^D\,, & \text{if} \quad N_{2,\gamma} = D\,, \end{cases} \tag{30}$$

where $L_{\min\,;\gamma} = \min\{L_{1,d;\gamma};\ 1 \le d \le D\}$ and $U_{\max\,;\gamma} = \max\{U_{2,d;\gamma};\ 1 \le d \le D\}$. Here components equal to $\Re$ in (29) and (30) are non-informative about the corresponding components of θ_1 and θ_2 – they are included so that formally the $C_{i;\gamma}$'s become D-dimensional, as the θ_i's and R_i's. For example, with $N_{1,\gamma} = 0$ in (29), the assertion "$\theta_1 \in C_{1;\gamma}$" means "$L_{1,d;\gamma} < \theta_{1;D}$" whereas nothing informative is said about $\theta_{1;1}, \ldots, \theta_{1;D-1}$. It follows from Hsu and Berger (1999) that (29) covers θ_1, and (30) covers θ_2, with marginal probability $\ge\gamma$ each; so (3) is satisfied. Moreover, it can be verified that (2) and (4) are satisfied with p_i equal to $\tilde{p}_i$ in (28). For example, $C_{1;1-u} \subset R_1$ if and only if $L_{\min\,;1-u} \ge 0$; i.e. if and only if $p_{1,d}$ in (27) is $\le u$ for all $1 \le d \le D$; i.e. if and only if $\tilde{p}_1$ in (28) is $\le u$.

It can then be verified that with this choice of quantities in (1)–(5): (a) the simultaneous confidence regions C_1^* and C_2^* given by (18) for θ_1 and θ_2 imply the same rejections of $H_{i,d}$'s as the Holm-type MTP defined by Algorithm 2 based on the p-values (28); and (b) in case not all $n = 2D$ null hypotheses in F_1 and F_2 are rejected by this MTP (so that $I_{\text{Accept}} \ne \varnothing$ in (18)), the extra "free" information provided by C_1^* and/or C_2^* consists of the one-sided confidence bound for the $\theta_{i,d}$ that corresponds to the first non-rejected $H_{i,d}$ in the fixed sequence within each F_i that is not entirely rejected by the MTP. For example: (a) if nothing is rejected in F_1 and F_2 (so that $N_{1,1-\alpha w_1(M)} = 0$ and $N_{2,1-\alpha w_2(M)} = 0$), the extra "free" information consists in the assertions "$L_{1,D;1-\alpha w_1(M)} < \theta_{1;D}$" and "$\theta_{2;1} < U_{2,1;1-\alpha w_2(M)}$"; whereas (b) if the entire F_2 is rejected but nothing is rejected in F_1 (so that $N_{2,1-\alpha w_2(M)} = D$ and $N_{1,1-\alpha} = 0$), the extra "free" information consists in "$L_{1,D;1-\alpha} < \theta_{1;D}$". Here the weights $w_i(M)$ are given by (15) with $v_1 = v_2 = 1/2$ and $M = \{1\,,\,2\}$, so $w_1(M) = w_2(M) = 1/2$. In case all $n = 2D$ null hypotheses in F_1 and F_2 are rejected (so that $|I_{\text{Reject}}| = m$ in (18)), extra "free" information may be provided by C_{m+1}^* in (18) through the choice (20), i.e. through the two Bonferroni-type assertions "$\theta_1 \in C_{1;1-\alpha w_1(M)}$" and "$\theta_2 \in C_{2;1-\alpha w_2(M)}$", which mean "$L_{\min\,;1-\alpha w_1(M)} < \theta_{1,d}$" for all $1 \le d \le D$ and "$\theta_{2,d} < U_{\max\,;1-\alpha w_2(M)}$" for all $1 \le d \le D$.

It should be evident how this application of Theorem 1 and the Holm-type MTP given by Algorithm 2 can be adapted to other kinds of experimental settings, other kinds of data and distributional asumptions, and/or other local MTPs controlling their multiple levels within $F_1, \ldots, F_m$ that do not necessarily belong to the Hommel et al. (2007, Section 2.2) class of MTPs. A referee pointed out that if all local MTPs within $F_1, \ldots, F_m$ belong to this class (as they do in the clinical-study example just described), then it is possible to deduce this kind of results directly from (18) and Theorem 1, without

introducing multidimensional $C_{i,\gamma}$'s and locally adjusted p-values. Details about this are omitted here, but briefly, the idea is to construct a new MTP in the Hommel et al. (2007, Section 2.2) class, with appropriate rejections of $H_{i,j}$'s in $F_1, \ldots, F_m$, and with weights defined in terms of the $w_i(I)$'s and corresponding weights within F_i's that satisfy the required conditions corresponding to (13).

5 Concluding Comments

The simultaneous confidence regions (18) are such that if $|I_{\text{Reject}}| < m$, they do not provide more information about the θ_i's with $i \in I_{\text{Reject}}$ than the rejection assertions "$\theta_i \in R_i$" made by the underlying MTP given by Algorithm 1. This drawback seems however difficult to overcome if the MTP is α-exhaustive in the sense of Remark 2 in Section 2.2, which is the case for Holm's (1979) MTP with $w_i(I)$'s given by (15). For MTPs that are not α-exhaustive in this sense, it is sometimes possible to provide more information about θ_i's with $i \in I_{\text{Reject}}$ if $|I_{\text{Reject}}| < m$; see the last paragraph in this section.

This kind of drawback is present also with the simultaneous confidence bounds of Bofinger (1987) and Stefansson et al. (1988, Section 2) mentioned initially in Section 1, see the discussion in Hsu (1996, Section 3.1.1.2); and, of course, with the stepwise simultaneous confidence intervals without multiplicity adjustment of Berger and Hsu (1999), because the underlying fixed-sequence MTP is α-exhaustive.

The 5th International Conference on Multiple Comparison Procedures (MCP 2007) took place in Vienna, 9–11 July 2007. Interestingly, the organizers had put together a session where the author had a presentation about parts of the results in this article, and Klaus Strassburger had a presentation about results by him and Frank Bretz entitled Compatible Simultaneous Lower Confidence Bounds for the Holm Procedure and Other Bonferroni Based Closed Tests. The article by Hommel et al. (2007) had been published online some months earlier, and that article was referred to in both presentations. The two approaches underlying the simultaneous confidence regions, as well as their formulation, are quite different. However, as it appeared from Strassburger's presentation, the comparable results in terms of lower confidence bounds for real-valued $\theta_1, \ldots, \theta_m$ are similar, though not identical, for MTPs in the Hommel et al. (2007, Section 2.2) class that are α-exhaustive, e.g. for Holm's (1979) MTP; whereas for MTPs in this class that are not α-exhaustive, the approach by Strassburger and Bretz leads to confidence assertions for rejected H_i's that may be sharper than rejection assertions in case $|I_{\text{Reject}}| < m$. In practice, this may mean that at the planning stage of a confirmatory clinical study, one may have to choose between prespecifying in the protocol either: (a) an MTP that is not α-exhaustive for which confidence assertions are possible that may be sharper than rejection assertions for some rejected H_i's in case $|I_{\text{Reject}}| < m$; or (b) a corresponding α-exhaustive MTP with potentially more rejections and sharper confidence assertions for nonrejected H_i's, but without sharper confidence assertions than rejection assertions for rejected H_i's in case $|I_{\text{Reject}}| < m$, as illustrated by (24) versus (25); cf. also Remark 3 in Section 2.2. The proof of the main result (19) is an extension of Holm's (1979, pp. 69–70) short direct proof in that the idea is to show that a certain unobservable event with probability $\geq 1 - \alpha$ (the intersection event in the third row of (32)) is a subset of the event that no erroneous assertions are made; whereas the partitioning principle (Finner and Strassburger, 2002) seemed to be used in the other approach. If the MTP is not α-exhaustive, so that $\Pr\left(\bigcap_{i \in T^c}[\theta_i \in C_{i,1-\alpha\, w_i(T)}]\right)$ can be < 1, it is possible to extend the arguments in (32) to get assertions for rejected H_i's that may be sharper than rejection assertions in case $|I_{\text{Reject}}| < m$ – but this is outside the scope of this article. Of course, not many details are provided in 20-minute presentations, so it will be interesting to study and compare the two approaches when both are published. Anyhow, in view of the fact that simultaneous confidence regions corresponding to Holm's (1979) MTP had been lacking for a long time, the closeness in time of these two independent contributions was rather amazing.

Acknowledgements *The author is very grateful to two referees for their constructive comments and suggestions which considerably improved the article. In particular, one referee provided very detailed suggestions, including the present version of Algorithm 1, the alternative approach mentioned in the last paragraph of Section 4.2.2, and the present proof in the Appendix (which is more condensed and transparent than the original one).*

Appendix: Proof of Theorem 1

Suppose that the specified set of $w_i(I)$'s used in Algorithm 1 satisfy (13). Let T be the unknown, possibly empty, set of indexes $1 \leq i \leq m$ of true null hypotheses $H_i : \theta_i \notin R_i$. Moreover, let A denote the coverage event in (19), with complement

$$A^c = [\theta_i \notin C_i^* \text{ for some } 1 \leq i \leq m+1]\,. \tag{31}$$

The problem is to prove that $\Pr(A) \geq 1-\alpha$, or equivalently, that $\Pr(A^c) \leq \alpha$.

Proof in case $T = \varnothing$ Suppose that $T = \varnothing$, so that $\theta_i \in R_i$ for all $1 \leq i \leq m$**.** In view of (1) and (18), this implies $\theta_i \in R_i \subset C_i^*$ for $1 \leq i \leq m$. Therefore, A^c in (31) equals the event $[\theta_{m+1} \notin C_{m+1}^*]$, which according to (17) and (18) has probability $\leq \alpha$. This concludes the proof in case $T = \varnothing$.

Proof in case $T \neq \varnothing$ Suppose that $T \neq \varnothing$, and recall that $\theta_i \notin R_i$ for all $i \in T$. Then we have the following relations between events

$$\begin{aligned} A \equiv \cap_{i=1}^{m+1}[\theta_i \in C_i^*] &\supset \cap_{i\in T}[\theta_i \in C_{i,1-\alpha\, w_i(I_{\text{Accept}})}] \cap [T \subset I_{\text{Accept}}] \\ &\supset \cap_{i\in T}[\theta_i \in C_{i,1-\alpha\, w_i(T)}] \cap [T \subset I_{\text{Accept}}] \\ &= \cap_{i\in T}[\theta_i \in C_{i,1-\alpha\, w_i(T)}]\,. \end{aligned} \tag{32}$$

Here the $\supset$ in the first row follows from (18) and the relation $[T \subset I_{\text{Accept}}] \subset [\theta_{m+1} \in C_{m+1}^*]$. The $\supset$ in the second row of (32) follows from the fact that $T \subset I_{\text{Accept}}$, together with (13) and (5), implies that $C_{i,1-\alpha\, w_i(I_{\text{Accept}})} \supset C_{i,1-\alpha\, w_i(T)}$ for all $i \in T$. The equality in the third row of (32) follows from

$$\begin{aligned} \cap_{i\in T}[\theta_i \in C_{i,1-\alpha\, w_i(T)}] &\subset \cap_{i\in T}[C_{i,1-\alpha\, w_i(T)} \not\subset R_i] \\ &\subset \cap_{i\in T}[(p_i > \alpha\, w_i(T) \quad \text{and} \quad w_i(T) > 0\,) \text{ or } w_i(T) = 0] \\ &\subset [T \subset I_{\text{Accept}}]\,. \end{aligned} \tag{33}$$

The $\subset$ in the second row of (33) follows from (4). The $\subset$ in the third row follows from the fact that Algorithm 1 is equivalent to a closed-testing procedure based on the marginal level-α Bonferroni-type test for intersection hypotheses $H_I = \cap_{i\in I} H_i$ that rejects an H_I if and only if for some $i \in I$, $w_i(I) > 0$ and $p_i \leq \alpha\, w_i(I)$. More precisely, if the intersection event in the second row of (33) occurs, then H_T is not rejected by its marginal level-α Bonferroni-type test, so that no H_i with $i \in T$ is rejected by the closed-testing procedure, i.e. the event $[T \subset I_{\text{Accept}}]$ occurs. Finally, it follows from (3), (13), and Boole's inequality, that the intersection event in the third row of (32) has probability $\geq 1-\alpha$. This concludes the proof in case $T \neq \varnothing$.

Conflict of Interests Statement
The author has declared no conflict of interest.

References

Agresti, A. and Min, Y. (2001). On small-sample confidence intervals for parameters in discrete distributions. *Biometrics* **57**, 963–971.

Aitchinson, J. (1964). Confidence-region tests. *Journal of the Royal Statistical Society, Series B* **26**, 462–476.

Bauer, P., Brannath, W., and Posch, M. (2001). Multiple testing for identifying effective and safe treatments. *Biometrical Journal* **43**, 605–616.

Bauer, P., Röhmel, J., Maurer, W., and Hothorn, L. (1998). Testing strategies in multi-dose experiments including active control. *Statistics in Medicine* **17**, 2133–2146.

Berger, R. L. and Hsu, J. C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science* **11**, 283–319.

Bofinger, E. (1987). Stepdown procedures for comparison with a control. *Australian Journal of Statistics* **29**, 348–364.

Bofinger, E. (1992). Expanded confidence intervals, one-sided tests, and equivalence testing. *Journal of Biopharmaceutical Statistics* **2**, 181–188.

Finner, H. and Strassburger, K. (2002). The partitioning principle: a powerful tool in multiple comparisons. *The Annals of Statistics* **30**, 1194–1213.

Hayter, A. J. and Hsu, J. C. (1994). On the relationship between stepwise decision procedures and confidence sets. *Journal of the American Statistical Association* **89**, 128–136.

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

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65–70.

Hommel, G., Bretz, F., and Maurer, W. (2007). Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies. *Statistics in Medicine* **26**, 4063–4073.

Hsu, J. C. (1996). *Multiple Comparisons – Theory and Methods*. Chapman & Hall, London.

Hsu, J. C. and Berger, R. L. (1999). Stepwise confidence intervals without multiplicity adjustment for dose-response and toxicity studies. *Journal of the American Statistical Association* **94**, 468–482.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Fransisco.

Lehmann, E. L. and Romano, J. P. (2005). Generalizations of the familywise error rate. *Annals of Statistics* **33**, 1138–1154.

Röhmel, J., Gerlinger, C., Benda, N., and Läuter, J. (2006). On testing simultaneously non-inferiority in two multiple primary endpoints and superiority in at least one of them. *Biometrical Journal* **48**, 916–933.

Stefansson, G., Kim, W., and Hsu, J. C. (1988). On confidence sets in multiple comparisons. In: *Statistical Decision Theory and Related Topics IV* (Eds.: Gupta, S. S. and Berger, J. O.), Volume 2, pages 89–104. Springer-Verlag, New York.

Westfall, P. and Young, S. S. (1993). *Resampling-based multiple testing: examples and methods for p-value adjustment*. Wiley, New York.

