Alternative Confidence Regions for Bonferroni-Based Closed-Testing Procedures that are not Alpha-Exhaustive

Olivier Guilbaud*

AstraZeneca R&D, S-15185 Södertälje, Sweden

Received 22 May 2008, revised 14 May 2009, accepted 16 May 2009

This article complements the results in Guilbaud (*Biometrical Journal* 2008; **50**:678–692). Simultaneous confidence regions were derived in that article that correspond to any given multiple testing procedure (MTP) in a fairly large class of consonant closed-testing procedures based on marginal *p*-values and weighted Bonferroni tests for intersection hypotheses. This class includes Holm's MTP, the fixed-sequence MTP, gatekeeping MTPs, fallback MTPs, multi-stage fallback MTPs, and recently proposed MTPs specified through a graphical representation and associated rejection algorithm. More general confidence regions are proposed in this article. These regions are such that for certain underlying MTPs which are not alpha-exhaustive, they lead to confidence assertions that may be sharper than rejection assertions for some rejected null hypotheses H when not all Hs are rejected, which is not the case with the previously proposed regions. In fact, various alternative confidence regions may be available for such an underlying MTP. These results are shown through an extension of the previous direct arguments (without invoking the partitioning principle), and under the same general setup; so for instance, estimated quantities and marginal confidence regions are not restricted to be of any particular kinds/dimensions. The relation with corresponding confidence regions of Strassburger and Bretz (*Statistics in Medicine* 2008; **27**:4914–4927) is described. The results are illustrated with fallback and parallel-gatekeeping MTPs.

Key words: Fallback procedure; Graphical procedures; Holm procedure; Parallel-gatekeeping procedure; Simultaneous confidence regions.

Supporting Information for this article is available from the author or on the WWW under http://dx.doi.org/10.1002/bimj.200800113.

1 The Confidence Regions Proposed in Guilbaud (2008)

This article complements the results in Guilbaud (2008). The general setup described in Sections 1.1, 1.2, and 3 of that article is understood, and we refer to that article for additional details about assumptions, definitions, and notation, as well as illustrations. For convenience to the reader, this section provides a summary of the general situation considered and of certain key results.

We are interested in making simultaneous inferences about specified quantities, $\theta_1, \ldots, \theta_m$, based on marginal inferences about these quantities. In particular we are interested in simultaneous assertions of the form "$\theta_i \in R_i$" where $R_i \subset \Theta_i$ is a specified target region for θ_i, and if possible, additional simultaneous assertions about the θ_i s. Here Θ_i is the set of potential θ_i-values, assumed to be known; and R_i may correspond *e.g.* to a superiority, non-inferiority, or equivalence claim aimed at in a confirmatory clinical study.

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

Moreover, for each $i = 1,\ldots,m$: (i) p_i is an available marginal 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 desired level $0 < u < 1$ by rejecting H_i if and only if $p_i \leq u$; (ii) $C_{i;\gamma}$ is an available marginal level-γ confidence region for θ_i such that $\Pr[\theta_i \in C_{i;\gamma}] \geq \gamma$ for $0 < \gamma < 1$, and the family indexed by γ of $C_{i;\gamma}$ s is such that with probability one, $C_{i;\gamma'} \subset C_{i;\gamma''} \subset \Theta_i$ for $0 < \gamma' < \gamma'' < 1$, with the convention that $C_{i;\gamma} = \Theta_i$ if $\gamma = 1$; and (iii) for any given level $0 < u < 1$, the rejection event $[p_i \leq u]$ occurs if and only if $[C_{i;1-u} \subset R_i]$ occurs, *i.e.*

$$[p_i \leq u] = [C_{i;1-u} \subset R_i]. \tag{1}$$

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.

A multiple testing procedure (MTP) in the Hommel, Bretz, and Maurer (2007, Section 2.2) class of closed-testing procedures based on $p_1, \ldots, p_m$ is used to make rejections in the family of null hypotheses $H_1, \ldots, H_m$, with type-I family-wise error rate controlled in the strong sense to be $\leq \alpha$. Here $0 < \alpha < 1$ is assumed to be given. This class of MTPs is fairly large. It includes Holm's (1979) step-down MTP, the fixed-sequence MTP, gatekeeping MTPs, fallback MTPs, and multi-stage fallback MTPs; cf. Sections 2.3 and 3 for references and illustrations. It also includes the classes of graphical MTPs considered in Bretz *et al.* (2009) and Burman, Sonesson, and Guilbaud (2009), cf. Section 4.4.

An MTP in the Hommel *et al.* (2007, Section 2.2) class is defined in terms of given weights $w_1(I), \ldots, w_m(I)$, $\emptyset \neq I \subset \{1,\ldots,m\}$, that satisfy

$$\begin{aligned} &0 \leq w_i(I) \leq 1\ ,\ \textstyle\sum_{i=1}^m w_i(I) \leq 1, \text{ and} \\ &w_i(I) \leq w_i(J) \text{ for all } i, I, J \text{ with } i \in J \text{ and } J \subset I; \end{aligned} \tag{2}$$

and rejections of H_i s are made through the following algorithm with steps 0, 1, 2, ... The MTP is called α-exhaustive if $\sum_{i \in I} w_i(I) = 1$ for each non-empty $I \subset \{1,\ldots,m\}$. For example, Holm's (1979) MTP and the fixed-sequence MTP are α-exhaustive, whereas the MTPs considered in Section 3 are not. In the sequel, $M \equiv \{1,\ldots,m\}$.

Algorithm 1 Step 0 : Set $I_1 = M$.

Step $r \geq 1$: Set $S_r = \{i \in I_r; w_i(I_r) > 0$ and $p_i \leq \alpha w_i(I_r)\}$. If $S_r = \emptyset$ 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 index-sets I_{Reject} and $I_{\text{Accept}} = M - I_{\text{Reject}}$ of rejected and accepted H_i s are well-defined functions of $p_1, \ldots, p_m$, and $|I_{\text{Reject}}| + |I_{\text{Accept}}| = m$. It is important to note that $w_i(I)$ s with $i \notin I$ do not influence the MTP, *i.e.* such $w_i(I)$ s do not influence the behavior of the index-sets I_{Reject} and I_{Accept}.

Moreover, θ_{m+1} is any specified quantity of interest (typically a function of $\theta_1,\ldots,\theta_m$) with known set Θ_{m+1} of potential values, and $C_{m+1;1-\alpha}$ is any specified marginal confidence region for θ_{m+1} such that $\Pr[\theta_{m+1} \in C_{m+1;1-\alpha}] \geq 1-\alpha$. The main result in Guilbaud (2008, Theorem 1) was stated as follows in terms of the random regions $C_1^*, \ldots, C_{m+1}^*$ for $\theta_1,\ldots,\theta_{m+1}$ defined by

$$C_i^* = \begin{cases} R_i & \text{if } i \in I_{\text{Reject}}, \\ R_i \cup C_{i;1-\alpha\, w_i(I_{\text{Accept}})} & \text{if } i \in I_{\text{Accept}}, \\ C_{m+1;1-\alpha} & \text{if } i = m+1 \text{ and } |I_{\text{Reject}}| = m, \\ \Theta_{m+1} & \text{if } i = m+1 \text{ and } |I_{\text{Reject}}| < m. \end{cases} \tag{3}$$

Note that the assertions "$\theta_i \in C_i^*$" for $i \in I_{\text{Reject}}$ correspond to the rejections made through Algorithm 1, and hence 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 Θ_{m+1}.

Theorem 1. The random regions $C_1^*, \ldots, C_{m+1}^*$ given by Eq. (3) simultaneously cover $\theta_1, \ldots, \theta_{m+1}$, respectively, with probability satisfying

$$\Pr[\theta_i \in C_i^* \text{ for all } 1 \le i \le m+1] \ge 1 - \alpha. \tag{4}$$

It was shown in Guilbaud (2008, Section 3.2) how θ_{m+1} and $C_{m+1;1-\alpha}$ can be chosen to sharpen inferences about $\theta_1, \ldots, \theta_m$ in case $|I_{\text{Reject}}| = m$, *e.g.* through the choice $\theta_{m+1} = (\theta_1, \ldots, \theta_m)$ and the rectangular regions $C_{m+1;1-\alpha}$ in Guilbaud (2008, Eqs. (20) and (22)) that equal a direct product of component regions for the components $\theta_1, \ldots, \theta_m$. Illustrations concerning clinical studies were then given in Section 4 of that article, where $\theta_1, \ldots, \theta_m$, $R_1, \ldots, R_m$, and $C_{1;\gamma}, \ldots, C_{m;\gamma}$ were either one-dimensional or D-dimensional.

The regions (3) have the property that when $1 \le |I_{\text{Reject}}| < m$, the only assertion that can be made about a θ_i with $i \in I_{\text{Reject}}$ is the rejection assertion "$\theta_i \in R_i$", because the additional assertion "$\theta_{m+1} \in C_{m+1}^*$" then is non-informative. It was mentioned in Guilbaud (2008, Section 5) that this drawback seems difficult to circumvent if the underlying MTP is α-exhaustive, as is the case, *e.g.* with Holm's MTP and the fixed-sequence MTP. However, it was also mentioned there that if the MTP is not α-exhaustive, it is possible to get assertions for θ_i s with $i \in I_{\text{Reject}}$ that may be sharper than rejection assertions when $1 \le |I_{\text{Reject}}| < m$. The following sections concern this latter possibility.

2 More General Confidence Regions for $\theta_1, \ldots, \theta_m$

The main results about the simultaneous confidence regions proposed in this article are given in the following three subsections. The relation between these confidence regions and the corresponding confidence regions of Strassburger and Bretz (2008) is described in the last paragraph of Section 2.1. Illustrations are given in Section 3, and some additional results are given in Section 4.

2.1 The simultaneous confidence regions

In addition to the previous assumptions, it is assumed that θ_{m+1} is an arbitrary but given function $h(\theta_1, \ldots, \theta_m)$ of $\theta_1, \ldots, \theta_m$, so that $C_{m+1;1-\alpha}$ covers $\theta_{m+1} = h(\theta_1, \ldots, \theta_m)$ with probability $\ge 1 - \alpha$. All choices of θ_{m+1} considered in Guilbaud (2008, Section 3.2) to sharpen inferences about $\theta_1, \ldots, \theta_m$ are of this form, with θ_{m+1} equal to the vector (or a subvector of) $(\theta_1, \ldots, \theta_m)$; so for comparison purposes, it is assumed in the sequel that the same choice of $\theta_{m+1} = h(\theta_1, \ldots, \theta_m)$ and $C_{m+1;1-\alpha}$ is made in (3) and (4).

Moreover, for each $i = 1, \ldots, m$, let $C_{i;m+1;1-\alpha}$ be the projection of $C_{m+1;1-\alpha}$ on the set Θ_i of potential θ_i-values, given by

$$\begin{aligned} C_{i;m+1;1-\alpha} = \; & \{y \in \Theta_i; \text{there is a } (\theta_1, \ldots, \theta_{i-1}, y, \theta_{i+1}, \ldots, \theta_m) \in \Theta_1 \times \cdots \times \Theta_m \\ & \text{such that } h(\theta_1, \ldots, \theta_{i-1}, y, \theta_{i+1}, \ldots, \theta_m) \in C_{m+1;1-\alpha}\}. \end{aligned} \tag{5}$$

Thus, for example, if θ_{m+1} and $C_{m+1;1-\alpha}$ are of the form

$$\theta_{m+1} = (\theta_1, \ldots, \theta_m) \text{ and } C_{m+1,1-\alpha} = C_1 \times \cdots \times C_m, \tag{6}$$

where $C_i \subset \Theta_i$ is a component region for θ_i of the rectangular region $C_{m+1;1-\alpha}$, as in Guilbaud (2008, Eqs. (20) and (22)), then each $C_{i;m+1;1-\alpha}$ reduces to the component region C_i. Note from Eq. (5) that if the coverage event $[h(\theta_1, \ldots, \theta_m) \in C_{m+1;1-\alpha}]$ occurs, then each coverage event $[\theta_i \in C_{i;m+1;1-\alpha}]$ with $1 \le i \le m$ occurs, *i.e.*

$$[h(\theta_1, \ldots, \theta_m) \in C_{m+1;1-\alpha}] \subset \cap_{1 \le i \le m}[\theta_i \in C_{i;m+1;1-\alpha}]; \tag{7}$$

so in particular, the intersection event in the right-hand-side of (7) has probability $\ge 1 - \alpha$.

The main result in this article, Theorem 2, is stated in terms of the random regions $C_1^{**}, \ldots, C_{m+1}^{**}$ for $\theta_1, \ldots, \theta_{m+1}$ defined as follows in terms of the index-sets I_{Reject} and I_{Accept}, the weights $w_i(I)$

satisfying (2) used in Algorithm 1, and the random sets $C_{i;m+1;1-\alpha} \subset \Theta_i$, $1 \leq i \leq m$, given by Eq. (5). Recall from Section 1 that the $w_i(I)$ s in (2) may equal 0, and that for $1 \leq i \leq m$, by convention, $C_{i;\gamma} = \Theta_i$ if $\gamma = 1$. Now, for each $i = 1, \ldots, m+1$, let

$$C_i^{**} = \begin{cases} R_i & \text{if } i \in I_{\text{Reject}} \text{ and } |I_{\text{Reject}}| = m, \\ R_i \cap (C_{i;m+1;1-\alpha} \cup C_{i;1-\alpha \min\{w_i(J)\,;\,\emptyset \neq J \subset I_{\text{Accept}}\}}) & \text{if } i \in I_{\text{Reject}} \text{ and } |I_{\text{Reject}}| < m, \\ R_i \cup C_{i;1-\alpha\, w_i(I_{\text{Accept}})} & \text{if } i \in I_{\text{Accept}}, \\ C_{m+1;1-\alpha} & \text{if } i = m+1 \text{ and } |I_{\text{Reject}}| = m, \\ \Theta_{m+1} & \text{if } i = m+1 \text{ and } |I_{\text{Reject}}| < m. \end{cases} \tag{8}$$

Clearly: (i) if $|I_{\text{Reject}}| = m$, then Eq. (8) provides information about the θ_i s with $i \in I_{\text{Reject}}$ through the rejection assertions "$\theta_i \in R_i$" combined with the assertion "$\theta_{m+1} \in C_{m+1;1-\alpha}$", as Eq. (3) provides, whereas (ii) if $1 \leq |I_{\text{Reject}}| < m$, then Eq. (8) provides information about the θ_i s with $i \in I_{\text{Reject}}$ through the assertions "$\theta_i \in C_i^{**}$" where $C_i^{**} \subset R_i$, *i.e.* such assertions are at least as sharp as the corresponding rejection assertions "$\theta_i \in R_i$" that Eq. (3) provides.

The potential difference between the confidence assertions about $\theta_1, \ldots, \theta_m$ based on Eq. (8) and those based on Eq. (3) thus lies only in the assertions about θ_i s with $i \in I_{\text{Reject}}$ when $|I_{\text{Reject}}| < m$. It turns out that if the underlying MTP is α-exhaustive, such assertions based on Eq. (8) are always equal to the corresponding rejection assertions "$\theta_i \in R_i$" based on Eq. (3). This is shown in Section 2.2. The case of particular interest is therefore when the MTP is not α-exhaustive, because it is only then that assertions "$\theta_i \in C_i^{**}$" with $i \in I_{\text{Reject}}$ based on Eq. (8) can be stricly sharper than rejection assertions when $|I_{\text{Reject}}| < m$. This case is considered in Section 2.3.

Theorem 2. The random regions $C_1^{**}, \ldots, C_{m+1}^{**}$ given by Eq. (8) simultaneously cover $\theta_1, \ldots, \theta_{m+1}$, respectively, with probability satisfying

$$\Pr[\theta_i \in C_i^{**} \text{ for all } 1 \leq i \leq m+1] \geq 1 - \alpha. \tag{9}$$

The proof of inequality (9) is given in the Supporting Information for this article. It is an extension of the direct proof of inequality (4) given in Guilbaud (2008, Appendix).

The following particular choice of θ_{m+1} and $C_{m+1;1-\alpha}$ in (5), (7) and (8) is of interest in that it leads to a simplification of the confidence regions for $\theta_1, \ldots, \theta_m$ based on Eq. (8). Suppose for a moment that θ_{m+1} and $C_{m+1;1-\alpha}$ are chosen to be of the form (6) with Bonferroni-type component regions $C_i = C_{i;1-\alpha B_i}$, where $B_1, \ldots, B_m$ are arbitrary but given non-negative weights satisfying $\sum_{i=1}^m B_i \leq 1$. Then, $C_{i;m+1;1-\alpha}$ in (5), (7) and (8) equals $C_i = C_{i;1-\alpha B_i}$, and it can be verified (using $C_{i;\gamma'} \subset C_{i;\gamma''} \subset \Theta_i$ for $0 < \gamma' < \gamma'' \leq 1$, see Section 1) that the confidence regions for $\theta_1, \ldots, \theta_m$ based on Eq. (8) reduce to

$$C_i^{***} = \begin{cases} R_i \cap C_{i;1-\alpha w_i^{***}(I_{\text{Accept}})} & \text{if } i \in I_{\text{Reject}}, \\ R_i \cup C_{i;1-\alpha\, w_i(I_{\text{Accept}})} & \text{if } i \in I_{\text{Accept}}, \end{cases} \tag{10}$$

where $w_i^{***}(I_{\text{Accept}})$: (i) is equal to B_i if $I_{\text{Accept}} = \emptyset$, *i.e.* if $|I_{\text{Reject}}| = m$; and (ii) is equal to the minimum of B_i and $\min\{w_i(J)\,;\emptyset \neq J \subset I_{\text{Accept}}\}$ if $I_{\text{Accept}} \neq \emptyset$, *i.e.* if $|I_{\text{Reject}}| < m$. This means that the random regions $C_1^{***}, \ldots, C_m^{***}$ simultaneously cover $\theta_1, \ldots, \theta_m$, respectively, with probability $\geq 1 - \alpha$. These regions are used for the illustrations in Section 3, with certain underlying weights $B_1, \ldots, B_m$.

We end this section by describing, in terms of the present setup, the closely related confidence regions of Strassburger and Bretz (2008, Eq. (8)). Briefly: (i) they formulated their results in terms of real-valued θ_i s, and one-sided R_i s and $C_{i;\gamma}$ s of the form $R_i = (\theta_{i,0}, \infty)$ and $C_{i;\gamma} = (L_{i,\gamma}, \infty)$; and (ii) their confidence regions for $\theta_1, \ldots, \theta_m$ are of the form (10). These confidence regions for $\theta_1, \ldots, \theta_m$ are thus somewhat less general than those based on Eq. (8) in that $C_{m+1;1-\alpha}$ in (5), (7) and (8) is not limited to be rectangular or of Bonferroni type–see *e.g.* Guilbaud (2008, Eqs. (21) and (22)) for a situation where $C_{m+1;1-\alpha}$ is rectangular but not of Bonferroni type. Actually, this situation illustrates the fact that Hsu and Berger (1999, Section 2) intervals constitute a special case of Eq. (3), and thus of Eq. (8) (because the underlying fixed-sequence MTP is α-exhaustive, see

Section 2.2), whereas these intervals cannot be expressed as Eq. (10) because they are not based on Bonferroni-type intervals when $|I_{\mathrm{Reject}}| = m$.

2.2 The case when the MTP is α-exhaustive

Suppose that the MTP given by Algorithm 1 is α-exhaustive, as, *e.g.* Holm's (1979) MTP and the fixed-sequence MTP. Then, confidence assertions about $\theta_1, \ldots, \theta_m$ based on Eq. (8) are always equal to those based on Eq. (3). To show this, note first that the weights in (2) are such that $w_i(\{j\}) = 0$ if $i \neq j$. It follows that if $i \in I_{\mathrm{Reject}}$ and $|I_{\mathrm{Reject}}| < m$, then in Eq. (8), $\min\{w_i(J);\, \emptyset \neq J \subset I_{\mathrm{Accept}}\} = 0$, so that $C_i^{**} = R_i \cap (C_{i;m+1;1-\alpha} \cup \Theta_i) = R_i$, which is equal to $C_i^* = R_i$ in Eq. (3). This means that the confidence regions for $\theta_1, \ldots, \theta_m$ based on Eq. (8) always reduce to those based on Eq. (3).

2.3 The case when the MTP is not α-exhaustive

Suppose that the MTP given by Algorithm 1 is not α-exhaustive. Then typically (but see Section 4.3 for exceptions), it is possible to construct regions of the form (8) so that some confidence assertions "$\theta_i \in C_i^{**}$ " with $i \in I_{\mathrm{Reject}}$ can be strictly sharper than corresponding rejection assertions "$\theta_i \in R_i$" based on (3) in case $|I_{\mathrm{Reject}}| < m$.

As a simple but important illustration of this, consider the Bonferroni MTP that, in terms of given positive weights $v_1, \ldots, v_m$ summing up to 1, has $w_i(I)$ s in (2) given by $w_i(I) = v_i$, $1 \leq i \leq m$. Clearly, this MTP is not α-exhaustive. Now, choose θ_{m+1} and $C_{m+1;1-\alpha}$ to be of the form (6) with Bonferroni-type component regions $C_i = C_{i;1-\alpha v_i}$. It follows that $C_{i;m+1;1-\alpha}$ equals $C_{i;1-\alpha v_i}$ in Eqs. (5) and (8). Moreover, also $C_{i;1-\alpha \min\{w_i(J);\emptyset\neq J\subset I_{\mathrm{Accept}}\}}$ equals $C_{i;1-\alpha v_i}$ in Eq. (8), because $w_i(J) = v_i$ does not depend on $J \neq \emptyset$; and it follows from Eq. (1) that with this Bonferroni MTP, $i \in I_{\mathrm{Reject}}$ if and only if $C_{i;1-\alpha v_i} \subset R_i$. It can then be verified that in Eq. (8), $C_i^{**} = R_i \cap C_{i;1-\alpha v_i} = C_{i;1-\alpha v_i}$ if $i \in I_{\mathrm{Reject}}$ and $|I_{\mathrm{Reject}}| < m$. This means that when $1 \leq |I_{\mathrm{Reject}}| < m$: (i) all confidence assertions "$\theta_i \in C_i^{**}$" with $i \in I_{\mathrm{Reject}}$ can be strictly sharper than the corresponding rejection assertions "$\theta_i \in R_i$"; and (ii) these potentially sharper assertions consist of ordinary Bonferroni-adjusted marginal assertions. This is a nice property of the regions (8).

MTPs that are not α-exhaustive have $w_i(I)$ s often given with unnecessarily many 0 s. This is the case *e.g.* for the fallback MTP (Wiens and Dmitrienko, 2005), the multi-stage fallback MTP (Dmitrienko, Wiens, and Westfall, 2006b), and the Bonferroni parallel-gatekeeping MTP (Dmitrienko, Offen, and Westfall, 2003). Typically, the given $w_i(I)$ s for such an MTP are such that $w_i(\{j\}) = 0$ if $i \neq j$. As shown in Section 2.2, this relation implies that in Eq. (8), $C_i^{**} = R_i$ for $i \in I_{\mathrm{Reject}}$ when $|I_{\mathrm{Reject}}| < m$. This means that the confidence regions for $\theta_1, \ldots, \theta_m$ based on Eq. (8) always reduce to those based on Eq. (3).

However, a key point here is that $\min\{w_i(J);\, \emptyset \neq J \subset I_{\mathrm{Accept}}\}$ in Eq. (8) depends on $w_i(I)$ s with $i \notin I$, whereas such $w_i(I)$ s do not influence the MTP given by Algorithm 1, *i.e.* such $w_i(I)$ s do not influence the behavior of the index-sets I_{Reject} and I_{Accept}. We may thus try to modify $w_i(I)$ s with $i \notin I$ for I s for which $\sum_{k \in I} w_k(I) < 1$ in such a way that some confidence assertions about θ_i s with $i \in I_{\mathrm{Reject}}$ based on Eq. (8) may be strictly sharper than rejection assertions "$\theta_i \in R_i$" when $|I_{\mathrm{Reject}}| < m$. Of course, such modified $w_i(I)$ s must be pre-specified. The next section provides details about such modifications for two MTPs – a fallback MTP and a Bonferroni parallel-gatekeeping MTP.

The possibility of modifying $w_i(I)$ s with $i \notin I$ was not used by Strassburger and Bretz (2008). For example, the confidence assertions about θ_i s with $i \in I_{\mathrm{Reject}}$ for a fallback MTP proposed in their Section 3.3 cannot be strictly sharper than rejection assertions when $|I_{\mathrm{Reject}}| < m$, whereas it is shown in Section 3.1 that such sharper assertions are possible. In fact, it is shown there that one may choose between various alternative modifications of $w_i(I)$ s with $i \notin I$.

3 Illustrations with MTPs that are not α-exhaustive

To simplify the developments in this section, we consider only confidence regions of the form (10) with underlying weights $B_1, \ldots, B_m$ given by

$$B_i = w_i(M) \text{ for } i \in M \equiv \{1, \ldots, m\}. \tag{11}$$

The weights $B_1, \ldots, B_m$ are thus defined in terms of the $w_i(I)$ s with $i \in I$ in (2) that specify the MTP given by Algorithm 1. As shown in connection with Guilbaud (2008, Eqs. (20)), the choice (11) ensures that when $|I_{\text{Reject}}| = m$, $C_i^{***} = C_{i;1-\alpha B_i} \subset R_i$ (so that the assertion "$\theta_i \in C_i^{***}$" can be strictly sharper than the rejection assertion "$\theta_i \in R_i$") for at least all indexes i in the rejection set S_1 in Step $r = 1$ of Algorithm 1.

3.1 Fallback

Suppose that the sequence in which the null hypotheses $H_1, H_2, \ldots, H_m$ are stated is relevant and fixed, *e.g.* with the rejection assertion "$\theta_1 \in R_1$ " being of most importance, and the rejection assertion "$\theta_m \in R_m$" being of least importance. The fallback MTP is a generalization of the fixed-sequence MTP. It is defined in terms of *a priori* given levels $\alpha'_1, \alpha'_2, \ldots, \alpha'_m$ associated with $H_1, H_2, \ldots, H_m$, respectively, that satisfy $\alpha'_i \geq 0$ and $\sum_{i=1}^m \alpha'_i = \alpha$; *i.e.* these α'_i s constitute a Bonferroni-type split of α over the H_i s. Null hypotheses are then rejected in successive steps $r = 1, 2, \ldots, m$ as follows. Step 1: set $\alpha_1 = \alpha'_1$, and reject H_1 if and only if $p_1 \leq \alpha_1$; Step $r \geq 2$: set $\alpha_r = \alpha'_r$ if H_{r-1} was not rejected, and $\alpha_r = \alpha_{r-1} + \alpha'_r$ if H_{r-1} was rejected, and reject H_r if and only if $p_r \leq \alpha_r$. In contrast to the fixed-sequence MTP, it may thus be possible to reject a H_i even if there are preceding hypotheses that are not rejected. The fallback MTP reduces to the fixed-sequence MTP if $\alpha'_1 = 1$ and $\alpha'_2 = \cdots = \alpha'_m = 0$.

The fallback MTP just described can also be defined through Algorithm 1 with $w_i(I)$ s in (2) that depend on the levels $\alpha'_1, \alpha'_2, \ldots, \alpha'_m$, or equivalently, on the weights $w'_1, w'_2, \ldots, w'_m$ given by $\alpha'_i = w'_i\alpha$, $1 \leq i \leq m$. Expressions for the $w_i(I)$ s in terms of $\alpha'_1, \alpha'_2, \ldots, \alpha'_m$ were given by Wiens and Dmitrienko (2005, Appendix A). Here we view the fallback MTP as being specified by $w'_1, w'_2, \ldots, w'_m$. These weights are thus *a priori* given, and they satisfy $w'_i \geq 0$ and $\sum_{i=1}^m w'_i = 1$.

In the sequel we consider the special case with $m = 3$ null hypotheses H_1, H_2, H_3. Table 1 summarizes how the $w_i(I)$ s are defined in terms of the weights w'_1, w'_2, w'_3 in this case. This table corresponds to Wiens and Dmitrienko (2005, Table 1).

It is clear from Table 1 that the fallback MTP is not α-exhaustive, except in the special case with $w'_1 = 1$ and $w'_2 = w'_3 = 0$ that corresponds to the fixed-sequence MTP. It is also clear that,

Table 1 Weights w_i (I) for a fallback MTP with $m = 3$ null hypotheses H_1, H_2, H_3 based on given non-negative weights w'_1, w'_2, w'_3 summing up to 1. Modified values (subject to the restrictions (2)) in the marked cells do not influence the behavior of the index-sets I_{Reject} and I_{Accept} of the MTP.

I	w_1 (I)	w_2 (I)	w_3 (I)
$\{1,2,3\}$	w'_1	w'_2	w'_3
$\{1,2\}$	w'_1	w'_2	$\boxed{0}$
$\{1,3\}$	w'_1	0	$w'_2 + w'_3$
$\{1\}$	w'_1	$\boxed{0}$	$\boxed{0}$
$\{2,3\}$	0	$w'_1 + w'_2$	w'_3
$\{2\}$	$\boxed{0}$	$w'_1 + w'_2$	$\boxed{0}$
$\{3\}$	0	0	1

depending on the given values of w'_1, w'_2, w'_3, some (possibly all) of the 0 s in the five marked cells in Table 1 can be increased to positive values without violating (2) or modifying the MTP, *i.e.* without influencing the behavior of the index-sets I_{Reject} and I_{Accept}. To simplify developments and avoid certain trivial complications, it is supposed in the sequel that the weights w'_1, w'_2, w'_3 in Table 1 satisfy $w'_1 \geq w'_2 \geq w'_3 > 0$. See Hommel and Bretz (2008, p. 662) for a discussion about these weights.

We now show that it is possible to modify 0 s in the marked cells in Table 1 in such a way that some confidence assertions about θ_i s with $i \in I_{\mathrm{Reject}}$ based on Eqs. (10) and (11) can be strictly sharper than rejection assertions "$\theta_i \in R_i$ " when $1 \leq |I_{\mathrm{Reject}}| < 3$. Note from the first row of Table 1 that Eq. (11) means that the weights B_1, B_2, B_3 underlying Eq. (10) are equal to w'_1, w'_2, w'_3, respectively. Table 2 gives various alternative choices of modified $w_i(I)$ s, as well as corresponding $w_i^{***}(I)$ s, in terms of a parameter x satisfying $0 \leq x \leq w'_3$ whose value has to be pre-specified; and the last row of this table gives the value of $w_i^{***}(I_{\mathrm{Accept}})$ in Eq. (10) that follows from Eq. (11) if $I_{\mathrm{Accept}} = \emptyset$, *i.e.* if $|I_{\mathrm{Reject}}| = 3$.

For example, according to the definition of $w_i^{***}(I)$ given in connection with Eq. (10): (i) $w_3^{***}(\{1,2\})$ equals the minimum of $B_3 = w'_3$ and the minimum over $I = \{1,2\}, \{1\}, \{2\}$ of $w_3(I)$, so $w_3^{***}(\{1,2\}) = x$ because $w_3(\{2\}) = x$ and $x \leq w'_3$; (ii) $w_2^{***}(\{1\})$ equals the minimum of $B_2 = w'_2$ and the minimum over $I = \{1\}$ of $w_2(I)$ (*i.e.* a minimum over a single I-value), so $w_2^{***}(\{1\}) = w'_2$ because $w_2(\{1\}) = w'_2$; and (iii) $w_1^{***}(\{2\})$ equals the minimum of $B_1 = w'_1$ and the minimum over $I = \{2\}$ of $w_1(I)$ (*i.e.* a minimum over a single I-value), so $w_1^{***}(\{2\}) = w'_3 - x$ because $w_1(\{2\}) = w'_3 - x$ and $w'_3 - x \leq w'_3 \leq w'_1$. The $w_i(I)$ s and $w_i^{***}(I)$ s in Table 2 are to be used in Eq. (10) with $I = I_{\mathrm{Accept}}$.

The value of $x \in [0, w'_3]$ in Table 2 influences the properties of the resulting confidence regions (10), and hence a user has the possibility to choose this value accordingly. The behavior of these confidence regions is considered in Sections 3.1.1 and 3.1.2 under the specifications $x = 0$ and $x = w'_3$, respectively. It should be clear from the results in these two sections that (and how) an intermediate specification $0 < x < w'_3$ leads to an intermediate behavior. An example based on clinical study data and the specification $x = 0$ is given in Section 3.1.3.

3.1.1 Modified $w_i(I)$ s given by the pre-specification $x = 0$ in Table 2

Let us first consider what happens in this case when $I_{\mathrm{Reject}} = \{2,3\}$, *i.e.* when $I_{\mathrm{Accept}} = \{1\}$. Then, according to the previous description of how the fallback MTP successively rejects in steps in the sequence H_1, H_2, H_3, we have $p_1 > w'_1\alpha$, $p_2 \leq w'_2\alpha$, and $p_3 \leq (w'_2 + w'_3)\alpha$. In particular, it follows from

Table 2 Weights $w_i(I)$ and $w_i^{***}(I)$ to be used in Eq. (10) with $I = I_{\mathrm{Accept}}$ for a fallback MTP with $m = 3$ null hypotheses H_1, H_2, H_3 based on given weights $w'_1 \geq w'_2 \geq w'_3 > 0$ summing up to 1 and a given $x \in [0, w'_3]$. The value of x does not influence the behavior of the index-sets I_{Reject} and I_{Accept} of the MTP, but it does influence the behavior of the simultaneous confidence regions (10)

I	$w_1(I)$	$w_2(I)$	$w_3(I)$	$w_1^{***}(I)$	$w_2^{***}(I)$	$w_3^{***}(I)$
$\{1,2,3\}$	w'_1	w'_2	w'_3			
$\{1,2\}$	w'_1	w'_2	w'_3			x
$\{1,3\}$	w'_1	0	$w'_2 + w'_3$		0	
$\{1\}$	w'_1	w'_2	w'_3		w'_2	w'_3
$\{2, 3\}$	0	$w'_1 + w'_2$	w'_3	0		
$\{2\}$	$w'_3 - x$	$w'_1 + w'_2$	x	$w'_3 - x$		x
$\{3\}$	0	0	1	0	0	
$\emptyset$				w'_1	w'_2	w'_3

Eq. (1) that $C_{2;1-\alpha w'_2} \subset R_2$, and thus from Eq. (10) and Table 2 that the region C_2^{***} for θ_2 equals $R_2 \cap C_{2;1-\alpha w'_2} = C_{2;1-\alpha w'_2}$. This shows that the confidence assertion "$\theta_2 \in C_2^{***}$" can be strictly sharper than the rejection assertion "$\theta_2 \in R_2$". Moreover, if p_3 is not only $\leq (w'_2 + w'_3)\alpha$ but also $\leq w'_3\alpha$, then it follows from Eq. (1) that $C_{3;1-\alpha w'_3} \subset R_3$, and thus from Eq. (10) and Table 2 that the region C_3^{***} for θ_3 equals $R_3 \cap C_{3;1-\alpha w'_3} = C_{3;1-\alpha w'_3}$. This shows that the confidence assertion "$\theta_3 \in C_3^{***}$" can be strictly sharper than the rejection assertion "$\theta_3 \in R_3$", though only if $p_3 \leq w'_3\alpha$.

Next, let us consider what happens when $I_{\text{Reject}} = \{1,3\}$, *i.e.* when $I_{\text{Accept}} = \{2\}$. Then we have $p_1 \leq w'_1\alpha$, $p_2 > (w'_1 + w'_2)\alpha$, and $p_3 \leq w'_3\alpha$. If p_1 is not only $\leq w'_1\alpha$ but also $\leq w'_3\alpha$, then it follows from Eq. (1) that $C_{1;1-\alpha w'_3} \subset R_1$, and thus from Eq. (10) and Table 2 that the region C_1^{***} for θ_1 equals $R_1 \cap C_{1;1-\alpha w'_3} = C_{1;1-\alpha w'_3}$. This shows that the confidence assertion "$\theta_1 \in C_1^{***}$" can be strictly sharper than the rejection assertion "$\theta_1 \in R_1$", though only if $p_1 \leq w'_3\alpha$. Moreover, it follows from Table 2 that irrespective of how small p_3 is, the region C_3^{***} for θ_3 equals $R_3 \cap C_{3;1-\alpha 0} = R_3 \cap \Theta_3 = R_3$. This shows that the confidence assertion "$\theta_3 \in C_3^{***}$" equals the rejection assertion "$\theta_3 \in R_3$".

It can be verified by considering all possible sets I_{Reject}, that no other sets I_{Reject} with $1 \leq |I_{\text{Reject}}| < 3$ than $\{2,3\}$ and $\{1,3\}$ can lead to a confidence assertion "$\theta_i \in C_i^{***}$" with $i \in I_{\text{Reject}}$ that is strictly sharper than the corresponding rejection assertion "$\theta_i \in R_i$".

Finally, let us consider what happens when $I_{\text{Reject}} = \{1,2,3\}$, *i.e.* when $I_{\text{Accept}} = \emptyset$. Then we have $p_1 \leq w'_1\alpha$, $p_2 \leq (w'_1 + w'_2)\alpha$, and $p_3 \leq \alpha$, whereas the regions C_1^{***}, C_2^{***}, and C_3^{***} reduce to $R_1 \cap C_{1;1-\alpha w'_1}$, $R_2 \cap C_{2;1-\alpha w'_2}$, and $R_3 \cap C_{3;1-\alpha w'_3}$, respectively. It thus follows from Eq. (1) that: (i) the confidence assertion "$\theta_1 \in C_1^{***}$" can be strictly sharper than the rejection assertion "$\theta_1 \in R_1$"; and (ii) the confidence assertions "$\theta_2 \in C_2^{***}$" and "$\theta_3 \in C_3^{***}$" can be strictly sharper than the rejection assertions "$\theta_2 \in R_2$" and "$\theta_3 \in R_3$", though only if $p_2 \leq w'_2\alpha$ and $p_3 \leq w'_3\alpha$, respectively.

In summary, some confidence assertions about θ_i s with $i \in I_{\text{Reject}}$ based on Eq. (10) and Table 2 with $x = 0$ can be strictly sharper than the corresponding rejection assertions when I_{Reject} equals $\{1,3\}$, $\{2,3\}$, and $\{1,2,3\}$. In particular: (i) a confidence assertion about θ_1 strictly sharper than a rejection assertion is possible when $I_{\text{Reject}} = \{1,3\}$ and $I_{\text{Reject}} = \{1,2,3\}$; and (ii) a confidence assertion about θ_3 strictly sharper than a rejection assertion is possible when $I_{\text{Reject}} = \{2,3\}$ and $I_{\text{Reject}} = \{1,2,3\}$.

3.1.2 Modified $w_i(I)$ s given by the pre-specification $x = w'_3$ in Table 2

We do not give all details in this case, but compared with the previous results based on the specification $x = 0$, the following differences can be verified. A confidence assertion about θ_1 strictly sharper than a rejection assertion is now possible only when $I_{\text{Reject}} = \{1,2,3\}$, because the value of $w_1^{***}(\{2\})$ has decreased from w'_3 to 0. However, this is compensated in that a confidence assertion about θ_3 strictly sharper than a rejection assertion is now possible not only when $I_{\text{Reject}} = \{2,3\}$ and $I_{\text{Reject}} = \{1,2,3\}$, but also when $I_{\text{Reject}} = \{2\}$ and $I_{\text{Reject}} = \{1,2\}$, because the values of $w_3^{***}(\{2\})$ and $w_3^{***}(\{1,2\})$ have both increased from 0 to w'_3.

3.1.3 Example

Hartung *et al.* (2002) reported results from a study in which two doses of mitoxantrone were compared with placebo in multiple sclerosis patients with respect to five primary efficacy variables through a fixed-sequence MTP. Wiens and Dmitrienko (2005, Section 4) illustrated their results in terms of the (exploratory) lower-dose comparison, whereas we consider the (confirmatory) higher-dose comparison, because marginal confidence intervals for differences are available for that comparison. To illustrate the previous developments about Fallback MTPs with $m = 3$, we use only the first three primary efficacy variables. Briefly, these variables are, in relevant order: Y_1 = change from baseline of expanded disability status scale, Y_2 = change from baseline of ambulation index, and Y_3 = number of relapses treated with corticosteroids. For each i = 1, 2, 3, the comparison was in terms of the quantity $\theta_i = \Pr[Y'_i > Y''_i] - \Pr[Y'_i < Y''_i]$, where Y'_i and Y''_i denote independent random Y_i -variables from underlying patient populations treated with placebo and active dose, respectively; the aim was to show that $\theta_i > 0$, *i.e.* to reject the null hypothesis that $\theta_i \leq 0$. The following marginal two-sided 95% confidence intervals for θ_1, θ_2, θ_3 with endpoints of the form $\hat{\theta}_i \pm 1.96\,\mathrm{SE}_i$ were reported by Hartung *et al.* (2002, Table 2), respectively: (0.04, 0.44), (0.02, 0.40), and (0.18, 0.59). These are large-sample intervals based on the fact (Lachin, 1992, Section 2) that asymptotically, each $(\hat{\theta}_i - \theta_i)/\mathrm{SE}_i$ has a standard normal distribution $N(0,1)$. We now use these results to illustrate how the simultaneous confidence regions (10) for θ_1, θ_2, θ_3 based on a Fallback MTP and Table 2 with $x = 0$ are constructed, and how they behave.

For each $i = 1, 2, 3$, let: $\Theta_i = (-1, 1)$, $R_i = (0, 1)$, $p_i = \Phi(-\hat{\theta}_i/\mathrm{SE}_i)$, and $C_{i;\gamma} = (\hat{\theta}_i - z_\gamma \mathrm{SE}_i,\ 1)$; cf. Section 1. Here Φ denotes the standard normal distribution function, and $z_\gamma = \Phi^{-1}(\gamma)$; so asymptotically, relation (1) between the marginal p-value p_i for the one-sided $H_i:\ \theta_i \notin R_i$ *versus* $H_i^c : \theta_i \in R_i$ and the family indexed by γ of marginal one-sided confidence intervals $C_{i;\gamma}$ for θ_i is satisfied. The R_i s, the estimates $\hat{\theta}_i$ and SE_i (based on the reported two-sided intervals), and the corresponding p_i s are listed in Table 3 under the subheading Outcome Scenario 1. An alternative Outcome Scenario 2 is also given in Table 3 where the value of $\hat{\theta}_1$ (of $\hat{\theta}_2$) has been increased (decreased).

Now, suppose we had pre-specified that, based on these quantities, and with $\alpha = 0.025$, we would: (i) reject H_i s through the Fallback MTP given by Table 1 with weights $w_1' = 0.5$, $w_2' = 0.25$, $w_1' = 0.25$; and (ii) construct corresponding simultaneous $1-\alpha$ confidence regions (10) for θ_1, θ_2, θ_3 based on Table 2 with $x = 0$ as described in Section 3.1.1. This choice of w_1', w_2', w_3' is consistent (Hommel and Bretz, 2008, p. 662) with the given sequence H_1, H_2, H_3 in that for each I, the $w_i(I)$ s in Table 1 satisfy $w_i(I) \geq w_j(I)$ for any $i \leq j$ in I. Under Outcome Scenario 1, the Fallback MTP rejects all H_1, H_2, H_3, and hence $I_{\mathrm{Accept}} = \emptyset$, whereas under Outcome Scenario 2, only H_1 and H_3 are rejected, so $I_{\mathrm{Accept}} = \{2\}$. The value of the weights $w_i^{***}(I_{\mathrm{Accept}})$ and $w_i(I_{\mathrm{Accept}})$ in Eq. (10), as well as the resulting simultaneous $1-\alpha$ confidence intervals C_i^{***}, are listed in Table 3 for each outcome scenario.

Note that under Outcome Scenario 1, the confidence assertion "$\theta_2 \in C_2^{***}$ " is not sharper than the rejection assertion "$\theta_2 \in R_2$", because p_2 is not $\leq w_2'\alpha$; cf. the penultimate paragraph of Section 3.1.1. Note also that under Outcome Scenario 2, the confidence assertion "$\theta_3 \in C_3^{***}$" is not sharper than the rejection assertion "$\theta_3 \in R_3$" although p_3 is very small; cf. the second paragraph of Section 3.1.1. However, under each outcome scenario, there is at least one $i \in I_{\mathrm{Reject}}$ for which the confidence assertion "$\theta_i \in C_i^{***}$ " is strictly sharper than the corresponding rejection assertion "$\theta_i \in R_i$". Outcome Scenario 2 illustrates the fact that confidence assertions "$\theta_i \in C_i^{***}$ " with

Table 3 Values of quantities involved in simultaneous $1-\alpha$ confidence regions (10) for θ_1, θ_2, θ_3 based on Table 2 with $w_1' = 0.5$, $w_2' = 0.25$, $w_3' = 0.25$, $x = 0$, and $\alpha = 0.025$, under two outcome scenarios. The underlying fallback MTP based on w_i (I) s with $i \in I$ in Table 2 leads to $I_{\mathrm{Accept}} = \emptyset$ under Outcome Scenario 1, and to $I_{\mathrm{Accept}} = \{2\}$ under Outcome Scenario 2.

	Outcome Scenario 1			Outcome Scenario 2		
Quantity	$i = 1$	$i = 2$	$i = 3$	$i = 1$	$i = 2$	$i = 3$
R_i	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)
$\hat{\theta}_i$	0.240	0.210	0.385	0.300	0.190	0.385
SE_i	0.102	0.097	0.105	0.102	0.097	0.105
p_i	0.0093	0.0151	0.0001	0.0016	0.0250	0.0001
$w_i^{***}(I_{\mathrm{Accept}})$	0.5	0.25	0.25	0.25	–	0
$w_i(I_{\mathrm{Accept}})$	–	–	–	–	0.75	–
C_i^{***}	(0.011,1)	(0,1)	(0.124,1)	(0.045,1)	(−0.012,1)	(0,1)

$i \in I_{\text{Reject}}$ based on Eqs. (8) and (10) can be strictly sharper than rejection assertions also in case $|I_{\text{Reject}}| < m$, in contrast to those based on Eq. (3).

3.2 Bonferroni parallel gatekeeping

Bonferroni parallel gatekeeping was introduced by Dmitrienko *et al.* (2003), and further developed by Dmitrienko *et al.* (2006a) and Guilbaud (2007); see also Dmitrienko, Tamhane, and Wiens (2008). Here we consider the simple MTP used by Dmitrienko *et al.* (2003, Section 2.1). This is an MTP for $m = 4$ null hypotheses H_i grouped into two subfamilies, $F_1 = \{H_1, H_2\}$ and $F_2 = \{H_3, H_4\}$; where F_1 is a parallel gatekeeper for F_2 in that no H_i in F_2 can be rejected unless at least one H_i in F_1 is rejected. This MTP can be interpreted (Dmitrienko *et al.*, 2006a) as a two-step procedure with: (i) rejections in F_1 in Step 1 through a Bonferroni MTP with equal weights at local multiple level α; and (ii) rejections in F_2 in Step 2 through a Holm MTP with equal weights at local multiple level $(r_1/2)\alpha$. Here r_1 equals the number of rejections in Step 1; Step 2 is not performed if $r_1 = 0$.

The weights $w_i(I)$, $1 \leq i \leq 4$, of this MTP are given by the rows with $I \neq \emptyset$ in Table 4, where the values 0.5 in the two marked cells should be replaced by 0 s. These weights correspond to the ones in Dmitrienko *et al.* (2003, Table 1).

It is clear from Table 4 (with 0 s in the two marked cells) that this MTP is not α-exhaustive. It is also clear that the 0 s in the two marked cells can be increased to any positive values ≤ 0.5 without violating (2) or modifying the MTP, *i.e.* without influencing the behavior of the index-sets I_{Reject} and I_{Accept}.

We now show that with the modifications in the two marked cells (*i.e.* with 0 s increased to 0.5) in Table 4, some confidence assertions about θ_i s with $i \in I_{\text{Reject}}$ based on Eqs. (10) and (11) can be strictly sharper than rejection assertions "$\theta_i \in R_i$ " when $1 \leq |I_{\text{Reject}}| < 4$. Note from the first row of Table 4 that Eq. (11) means that the weights B_1, B_2, B_3, B_4 underlying Eq. (10) are equal to 0.5, 0.5, 0, 0, respectively. Table 4 gives the $w_i^{***}(I)$ s resulting from the modified $w_i(I)$ s, and its last row gives the value of $w_i^{***}(I_{\text{Accept}})$ in Eq. (10) that follows from Eq. (11) if $I_{\text{Accept}} = \emptyset$, *i.e.* if $|I_{\text{Reject}}| = 4$. For example, according to the definition of $w_i^{***}(I)$ given in connection with Eq. (10): (i) $w_2^{***}(\{1\})$ equals the minimum of $B_2 = 0.5$ and the minimum over $I = \{1\}$ of $w_2(I)$ (*i.e.* a minimum over a single I-value), so $w_2^{***}(\{1\}) = 0.5$ because $w_2(\{1\}) = 0.5$; and (ii) $w_3^{***}(\{1\})$ equals the minimum of $B_3 = 0$ and the minimum over $I = \{1\}$ of $w_3(I)$, so $w_3^{***}(\{1\}) = 0$. The $w_i(I)$ s and $w_i^{***}(I)$ s in Table 4 are to be used in Eq. (10) with $I = I_{\text{Accept}}$.

Let us first consider what happens when $I_{\text{Reject}} = \{2, 3, 4\}$, *i.e.* when $I_{\text{Accept}} = \{1\}$. Then, according to the previously mentioned two-step interpretation of the MTP, we have $p_1 > 0.5\alpha$, $p_2 \leq 0.5\alpha$, $p_3 \wedge p_4 \leq 0.25\alpha$, and $p_3 \vee p_4 \leq 0.5\alpha$. Here we use the notation $a \wedge b$ and $a \vee b$ for $\min(a, b)$ and $\max(a, b)$, respectively. In particular, it follows from Eq. (1) that $C_{2;1-\alpha 0.5} \subset R_2$, and thus from Eq. (10) and Table 4 that the region C_2^{***} for θ_2 equals $R_2 \cap C_{2;1-\alpha 0.5} = C_{2;1-\alpha 0.5}$. This shows that the confidence assertion "$\theta_2 \in C_2^{***}$ " can be strictly sharper than the rejection assertion "$\theta_2 \in R_2$". Moreover, it can be verified that the confidence assertions "$\theta_3 \in C_3^{***}$" and "$\theta_4 \in C_4^{***}$" are equal to the rejection assertions "$\theta_3 \in R_3$" and "$\theta_4 \in R_4$", because $w_3^{***}(I_{\text{Accept}}) = w_4^{***}(I_{\text{Accept}}) = 0$ in (10).

Similarly (by symmetry), when $I_{\text{Reject}} = \{1, 3, 4\}$, the confidence assertion "$\theta_1 \in C_1^{***}$" = "$\theta_1 \in C_{1;1-\alpha 0.5}$" can be strictly sharper than the rejection assertion "$\theta_1 \in R_1$", whereas the confidence assertions "$\theta_3 \in C_3^{***}$" and "$\theta_4 \in C_4^{***}$" are equal to the rejection assertions "$\theta_3 \in R_3$" and "$\theta_4 \in R_4$".

It can be verified by considering all possible sets I_{Reject}, that no other sets I_{Reject} with $1 \leq |I_{\text{Reject}}| < 4$ than $\{2, 3, 4\}$ and $\{1, 3, 4\}$ can lead to a confidence assertion "$\theta_i \in C_i^{***}$" with $i \in I_{\text{Reject}}$ that is strictly sharper than the corresponding rejection assertion "$\theta_i \in R_i$".

Finally, let us consider what happens when $I_{\text{Reject}} = \{1, 2, 3, 4\}$, *i.e.* when $I_{\text{Accept}} = \emptyset$. According to the two-step interpretation of the MTP, we have $p_1 \leq 0.5\alpha$, $p_2 \leq 0.5\alpha$, $p_3 \wedge p_4 \leq 0.5\alpha$, and

Table 4 Weights $w_i\,(I)$ and $w_i^{***}(I)$ to be used in (10) with $I = I_{\mathrm{Accept}}$ for a Bonferroni parallel-gatekeeping MTP with $m = 4$ null hypotheses in two subfamilies, $F_1 = \{H_1, H_2\}$ and $F_2 = \{H_3, H_4\}$. The specified value 0.5 in the two marked cells does not influence the behavior of the index-sets I_{Reject} and I_{Accept} of the MTP, but it does influence the behavior of the simultaneous confidence regions (10).

I	$w_1\,(I)$	$w_2\,(I)$	$w_3\,(I)$	$w_4\,(I)$	$w_1^{***}(I)$	$w_2^{***}(I)$	$w_3^{***}(I)$	$w_4^{***}(I)$
$\{1,2,3,4\}$	0.5	0.5	0	0				
$\{1,2,3\}$								
$\{1,2,4\}$	0.5	0.5	0	0			0	
$\{1,2\}$	0.5	0.5	0	0			0	0
$\{1,3,4\}$	0.5	0	0.25	0.25		0		
$\{1,3\}$	0.5	0	0.5	0		0		0
$\{1,4\}$	0.5	0	0	0.5		0	0	
$\{1\}$	0.5	0.5	0	0		0.5	0	0
$\{2,3,4\}$	0	0.5	0.25	0.25	0			
$\{2,3\}$	0	0.5	0.5	0	0			0
$\{2,4\}$	0	0.5	0	0.5	0		0	
$\{2\}$	0.5	0.5	0	0	0.5		0	0
$\{3,4\}$	0	0	0.5	0.5	0	0		
$\{3\}$	0	0	1	0	0	0		0
$\{4\}$	0	0	0	1	0	0	0	
$\emptyset$					0.5	0.5	0	0

$p_3 \vee p_4 \leq \alpha$. It can then be verified from Eq. (10) and Table 4 that: (i) the regions C_1^{***} and C_2^{***} reduce to the Bonferroni-type regions $R_1 \cap C_{1;1-\alpha0.5} = C_{1;1-\alpha0.5}$ and $R_2 \cap C_{2;1-\alpha0.5} = C_{2;1-\alpha0.5}$, so that the confidence assertions "$\theta_1 \in C_1^{***}$" and "$\theta_2 \in C_2^{***}$" can be strictly sharper than the corresponding rejection assertions "$\theta_1 \in R_1$" and "$\theta_2 \in R_2$", whereas (ii) the regions C_3^{***} and C_4^{***} reduce to $R_3 \cap C_{3;1-\alpha0} = R_3 \cap \Theta_3 = R_3$ and $R_4 \cap C_{4;1-\alpha0} = R_4 \cap \Theta_4 = R_4$, so that irrespective of how small p_3 and p_4 are, the confidence assertions "$\theta_3 \in C_3^{***}$" and "$\theta_4 \in C_4^{***}$" are equal to the rejection assertions "$\theta_3 \in R_3$" and "$\theta_4 \in R_4$".

In summary, confidence assertions about θ_i s with $i \in I_{\mathrm{Reject}} \cap \{1,2\}$ based on Eq. (10) and Table 4 can be strictly sharper than the corresponding rejection assertions when I_{Reject} equals $\{2,3,4\}$, $\{1,3,4\}$, and $\{1,2,3,4\}$, whereas confidence assertions about θ_i s with $i \in I_{\mathrm{Reject}} \cap \{3,4\}$ are always equal to the corresponding rejection assertions.

4 Concluding Comments and Additional Results

Some general comments concerning *a priori* specifications/choices are made in Sections 4.1 and 4.2. It is shown in Section 4.3 that there are MTPs which are not α-exhaustive and for which the confidence regions for $\theta_1, \ldots, \theta_m$ based on Eq. (8) are always equal to those based on Eq. (3). Finally, some remarks are made in Section 4.4 about graphical MTPs considered in Bretz *et al.* (2009) and in Burman *et al.* (2009).

4.1 Required pre-specifications given the MTP

A practical aspect with the proposed confidence regions for $\theta_1, \ldots, \theta_m$ is that one has to make various pre-specifications. In particular there is some flexibility in the choice of θ_{m+1} and $C_{m+1;1-\alpha}$ to be used to sharpen inferences about $\theta_1, \ldots, \theta_m$, and some possibilities in this context were mentioned in Guilbaud (2008, Section 3.2). If the MTP is not α-exhaustive, there is the additional flexibility in the choice of $w_i(I)$ s with $i \notin I$ to be used in (8) described in Section 2.3. The illustrations in Section 3 are based on confidence regions of the form (10) with the particular choice (11) of underlying weights $B_1, \ldots, B_m$. As shown in Section 3.1, even with this considerable restriction on the regions considered, there may be various alternative choices of $w_i(I)$ s with $i \notin I$ that are possible. This flexibility means that one has the opportunity to take the behavior of the corresponding alternative confidence regions into account when one makes the choice. Regions based on Eq. (10) may often be reasonable, but of course, depending on the situation considered, other regions may sometimes be preferable, *e.g.* regions (8) with θ_{m+1} and $C_{m+1;1-\alpha}$ of the form (6) where the components C_i of the rectangular region $C_{m+1;1-\alpha}$ are not of Bonferroni type, as in Guilbaud (2008, Eqs. (21) and (22)).

4.2 Choice between an MTP that is not α-exhaustive and an improved version that is

An MTP 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 (2) and are such that the resulting MTP becomes α-exhaustive (so $w'_i(I) = 0$ if $i \notin I$). An example of this is the Bonferroni MTP that can be improved to Holm's MTP based on the same weights.

One may therefore have a choice between using either: (i) an MTP that is not α-exhaustive for which confidence assertions are possible that may be strictly sharper than rejection assertions for some rejected H_i s in case not all H_i s are rejected; or (ii) a corresponding α-exhaustive MTP with potentially more rejections and sharper confidence assertions for nonrejected H_i s, but with confidence assertions equal to rejection assertions for rejected H_i s in case not all H_i s are rejected. In confirmatory clinical studies where rejection assertions are of particular importance, this kind of choice will probably often be in favor of the α-exhaustive MTP. However, other aspects than rejection power may sometimes be involved in such considerations. The appropriate choice may then not always be evident.

To illustrate such considerations, consider the parallel-gatekeeping MTP in Section 3.2. The weights $w'_i(I)$ of the corresponding improved MTP are obtained by keeping the original two 0 s in the marked cells of Table 4, and increasing the remaining single 0.5-value among the $w_i(I)$ s to 1 in each of the two rows with a marked cell. This improved MTP can be interpreted (Guilbaud, 2007, Section 5.1) as follows. If the two-step procedure mentioned in the first paragraph of Section 3.2 results in a rejection index-set equal to either $\{2,3,4\}$ or $\{1,3,4\}$, then one goes back in a third step to the single non-rejected H_i in $F_1 = \{H_1, H_2\}$, and reject this H_i if and only if $p_i \leq \alpha$. This three-step interpretation shows that Holm-type rejections are made in F_1, but only if the entire $F_2 = \{H_3, H_4\}$ is rejected in the second step. Thus, compared with the Holm-type rejections one would be able to make in F_1 if there were no F_2, the effect/cost of also considering potential rejections in F_2 lies in that possibly less is rejected in F_1. This dependence of rejections in F_1 on rejections in F_2 is quite natural and should, to the author's opinion, hardly undermine the credibility of a rejection in F_1 through the third step. This kind of dependence has been debated; see *e.g.* the discussion about the "independence condition" at the end of Dmitrienko and Tamhane (2007, Section 2), and about "Condition 1" in Hommel *et al.* (2007, Section 4).

Hence, for the sake of discussion, suppose that the dependence on rejections in F_2 is considered undesirable, and that one prefers to use the original two-step MTP that is not α-exhaustive. Then, instead of just pre-specifying that null hypotheses $H_i : \theta_i \notin R_i$ will be rejected through this two-step MTP to show that θ_i s belong to target regions R_i, one may, for example, pre-specify that confidence regions for $\theta_1, \ldots, \theta_4$ based on Eq. (10) and Table 4 will be constructed as described in Section 3.2. The resulting confidence assertions, including the implied rejection assertions "$\theta_i \in R_i$" for $i \in I_{\text{Reject}}$ from the MTP itself, will be simultaneously correct with confidence $\geq 1-\alpha$; and one then has the possibility of getting assertions that are strictly sharper than rejection assertions for the θ_i s with $i \in I_{\text{Reject}} \cap \{1,2\}$, possibly much sharper.

It should be observed, though, from the developments in Section 3.2 that the sharper assertions just mentioned about θ_i s with $i \in I_{\text{Reject}} \cap \{1,2\}$ are possible only if the entire F_2 is rejected, so also with this latter approach based on the two-step MTP there is a certain dependence of inferences about θ_1 and θ_2 on rejections in F_2.

4.3 A natural question concerning MTPs that are not α-exhaustive

In view of the developments in Sections 2.3 and 3, it is natural to ask whether it is possible for any MTP which is not α-exhaustive to modify its $w_i(I)$ s with $i \notin I$ in (2) in such a way that some confidence assertions about θ_i s with $i \in I_{\text{Reject}}$ based on (8) can be strictly sharper than rejections assertions when $|I_{\text{Reject}}| < m$. The answer to this question is no. That is, there are MTPs which are not α-exhaustive and for which such a modification is not possible.

To show this, note first that it follows from the arguments in Section 2.2 that such sharper confidence assertions about θ_i s with $i \in I_{\text{Reject}}$ are not possible if the $w_i(I)$ s with $i \in I$ in (2) are such that

$$w_i(\{i\}) = 1 \;\text{ for all }\; i \in M \equiv \{1,\ldots,m\} \tag{12}$$

because it then follows from (2) that $w_i(\{j\}) = 0$ if $i \neq j$. It thus only remains to show that there are MTPs satisfying (12) that are not α-exhaustive. A simple example is provided by the fallback MTP considered in the first two paragraphs in Section 3.1, with: $m \geq 3$, positive weights w'_i, and $w_i(\{i\})$ s modified to satisfy (12). In the special case given by Table 1, this modification means that $w_1(\{1\})$ and $w_2(\{2\})$ are increased to be equal to 1. Clearly, the resulting MTP has $w_i(I)$ s that satisfy (2) and (12), though this MTP is not α-exhaustive because $\sum_{i \in I} w_i(I) = 1 - w'_m < 1$ for $I = M - \{m\}$, see Wiens and Dmitrienko (2005, Appendix A).

4.4 Final remarks about weights $w_i(I)$ and graphical MTPs

One may sometimes have to derive weights $w_i(I)$, $i \in I$, to be able to apply (3), (8), or (10). For example, this may be the case if the MTP is specified through a graphical representation and associated rejection algorithm as in Bretz *et al.* (2009) or as in Burman *et al.* (2009). It should be noted that the graphical representations and approaches in these two articles are different, though underlying ideas are related. For such a graphical MTP, the weights $w_i(I)$, $i \in I$, can be obtained as described in, respectively: (i) remark (v) in the last paragraph of Bretz *et al.* (2009, p. 592); and (ii) the first paragraph of Burman *et al.* (2009, Section 3.3), *i.e.* $w_i(I)$ equals $\alpha_i(I^c)/\alpha$, $i \in I$, where $I^c = M - I$, and $\alpha_i(A)$ is defined in the first three paragraphs of Section 3.2 of that article in terms of a so called default graph for the MTP.

Briefly, a default graph consists of component sequences of H_i s, and of positive weights summing up to 1 associated with these sequences. Given a default graph, the determination of the weights $w_i(I)$ with $i \in I$ for any given $I \neq \emptyset$ is simple: (i) one first reduces the graph by dropping all H_i s with $i \in I^c$ from its component sequences (no reduction takes place if $I = M$), (ii) each H_i with $i \in I$ for which $w_i(I)$ is positive then occurs as the first hypothesis in one or more of the remaining reduced sequences, and (iii) the value of this positive $w_i(I)$ is simply equal to the sum of the initial weights of the remaining reduced sequences that have H_i as their first hypothesis. For example, it can be verified that any $w_i(I)$ with $i \in I$ in Table 1 can be obtained in this way from the three component sequences $\langle H_1, H_2, H_3\rangle$, $\langle H_2, H_3\rangle$, $\langle H_3\rangle$, with associated positive weights w'_1, w'_2, w'_3, respectively. (The default graph consists only of component sequences with positive weights; so in particular, if $w'_1 = 1$ and $w'_2 = w'_3 = 0$, then the default graph consists of the single component sequence $\langle H_1, H_2, H_3\rangle$ with associated weight 1.)

In view of the considerations in Section 2.3 it is also appealing that one can deduce immediately from a default graph whether the inequality $\sum_{i\in I} w_i(I)<1$ holds for a given I. This strict inequality holds if and only if I^c contains all indexes i of H_i s in at least one of the component sequences. For example, one can immediately see from the three component sequences with positive weights w'_1, w'_2, w'_3 just mentioned that $\sum_{i\in I} w_i(I)<1$ holds for $I^c=\{2,3\}$, $I^c=\{1,3\}$, and $I^c=\{3\}$, *i.e.* for $I=\{1\}$, $I=\{2\}$, and $I=\{1,2\}$, in accordance with Table 1. Burman *et al.* (2009) provided default graphs for various common and new MTPs, and described how others can be constructed.

Acknowledgements The main result in this article, Theorem 2, was originally inspired by the results by Klaus Strassburger and Frank Bretz presented at the 5th International Conference on Multiple Comparison Procedures (MCP 2007) in Vienna, 9–11, July 2007, and then sharpened to its present version after the first review period (during which Strassburger and Bretz (2008) was published) through very detailed comments from a referee. Among other things, this referee described how partitioning-based ideas (Finner and Strassburger, 2002) can be applied with the present setup. Details about this have been omitted. The author is very grateful to the referee, the associate editor, and the editor for their constructive comments and suggestions, which considerably improved the article.

Conflict of Interests Statement

The author has declared no conflict of interest.

References

Bretz, F., Maurer, W., Brannath, W. and Posch, M. (2009). A graphical approach to sequentially rejective multiple test procedures. *Statistics in Medicine* **28**, 586–604.

Burman, C.-F., Sonesson, C. and Guilbaud, O. (2009). A recycling framework for the construction of Bonferroni-based multiple tests. *Statistics in Medicine* **28**, 739–761.

Dmitrienko, A., Offen, W. and Westfall, P. (2003). Gatekeeping strategies for clinical trials that do not require all primary effects to be significant. *Statistics in Medicine* **22**, 2387–2400.

Dmitrienko, A. and Tamhane, A. C. (2007). Gatekeeping procedures with clinical trial applications. *Pharmaceutical Statistics* **6**, 171–180.

Dmitrienko, A., Tamhane, A. C., Wang, X. and Chen, X. (2006a). Stepwise gatekeeping procedures in clinical trial applications. *Biometrical Journal* **48**, 984–991.

Dmitrienko, A., Tamhane, A. C. and Wiens, B. L. (2008). General multistage gatekeeping procedures. *Biometrical Journal* **50**, 667–677.

Dmitrienko, A., Wiens, B. L. and Westfall, P. (2006b). Fallback tests in dose-response clinical trials. *Journal of Biopharmaceutical Statistics* **16**, 745–755.

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

Guilbaud, O. (2007). Bonferroni parallel gatekeeping – transparent generalizations, adjusted *p*-values, and short direct proofs. *Biometrical Journal* **49**, 917–927.

Guilbaud, O. (2008). Simultaneous confidence regions corresponding to Holm's step-down procedure and other closed-testing procedures. *Biometrical Journal* **50**, 678–692.

Hartung, H. P., Gonsette, R., König, N., Kwiecinski, H., Guseo, A., Morrissey, S. P., Krapf, H. and Zwingers, T. (2002). Mitoxantrone multiple sclerosis group study group. Mitoxantrone in progressive multiple sclerosis: a placebo-controlled, double-blind, randomised, multicentre trial. *Lancet* **360**, 2018–2025.

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

Hommel, G. and Bretz, F. (2008). Aesthetics and power considerations in multiple testing – a contradiction? *Biometrical Journal* **50**, 657–666.

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. 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.

Lachin, J. M. (1992). Some large-sample distribution-free estimators and tests for multivariate partially incomplete data from two populations. *Statistics in Medicine* **11**, 1151–1170.

Strassburger, K. and Bretz, F. (2008). Compatible simultaneous lower confidence bounds for the Holm procedure and other Bonferroni-based closed tests. *Statistics in Medicine* **27**, 4914–4927.

Wiens, B. L. and Dmitrienko, A. (2005). The fallback procedure for evaluating a single family of hypotheses. *Journal of Biopharmaceutical Statistics* **15**, 929–942.

