Sharper Confidence Intervals for Hochberg- and Hommel-Related Multiple Tests Based On an Extended Simes Inequality

Olivier GUILBAUD

This article concerns recent developments related to the now classical multiple-testing procedures (MTPs) of Holm, Hochberg, and Hommel based on marginal $p$-values. For a long time, the derivation of simultaneous confidence intervals (SCIs) corresponding to these MTPs was considered to be a difficult problem, but solutions were published in 2008 for Holm’s MTP, and in 2012 for Hochberg’s and Hommel’s MTPs. These SCIs turned out to be as simple and easily implemented as the MTPs themselves, and to be remarkably similar. However, they also turned out to have the property/limitation, shared with other powerful stepwise MTPs, that no confidence assertions sharper than rejection assertions are possible unless all null hypotheses are rejected. A possibility is then to construct related families of MTPs that do not have this limitation but are somewhat less powerful, so users may choose among various such trade-off MTPs. It is shown in this article how an extended Simes inequality can be used to construct Hochberg- and Hommel-related MTPs of this kind that: (i) are more powerful than corresponding trade-off MTPs proposed previously, and (ii) lead to SCIs that are sharper than the ones proposed previously. Corresponding Hommel-related MTPs and SCIs are considered for completeness and comparisons.

Key Words: Clinical trial; Closed-testing procedure; Confirmatory study; Simultaneous confidence intervals; Step-down testing procedure; Step-up testing procedure.

1. Introduction

This article concerns recent developments related to three now classical multiple-testing procedures (MTPs); namely Holm’s (1979) step-down MTP, Hochberg’s (1988) step-up MTP, and Hommel’s (1988) more powerful MTP that are based on marginal $p$-values $p_1,\ldots,p_m$ for the null hypotheses in a given family $\{H_1,\ldots,H_m\}$ of null hypotheses. These three MTPs control their Type-I family-wise error rate (FWER) in the strong sense, which nowadays is a regulatory requirement for confirmatory clinical trials aiming at supporting multiple claims.

Holm’s MTP is more powerful than the Bonferroni MTP, and as generally valid as this simple MTP. Hommel’s MTP is more powerful than Hochberg’s MTP, which in turn is more powerful than Holm’s MTP. Here “more powerful” is in the strong sense that for any outcome of $p_1,\ldots,p_m$, at least as much is rejected, possibly more. However, an important aspect of Hommel’s and
Hochberg’s MTPs in applications is that in contrast to Holm’s MTP, they are not generally valid. Hommel’s MTP is valid if the \( p \)-values \( p_1, \ldots, p_m \) are independent or positively dependent in such a way that the Simes (1986) inequality holds for any given subset of \( p_i \)'s. This condition on \( p \)-values is sufficient also for Hochberg’s MTP to be valid. Much research has been devoted to determine situations under which the Simes inequality holds; see for example, Sarkar and Chang (1997) and Sarkar (1998, 2008), with references.

For a long time, the derivation of simultaneous confidence intervals (SCIs) that correspond to these three classical MTPs was considered to be a difficult problem. Here, correspond means that the SCIs imply the same rejections as those made by the MTP, and thus prove extra information “for free.” A nice discussion can be found in Strassburger and Bretz (2008) of the state of the art before 2008 for Holm’s MTP, and of certain incorrect SCIs previously proposed for that MTP that seem to have been widely used in practice. Solutions were, however, obtained by Guilbaud (2008) and Strassburger and Bretz (2008) for Holm’s MTP, and by Guilbaud (2012) for Hochberg’s and Hommel’s MTPs.

The SCIs derived for these three MTPs turned out to be as simple and easily implemented as the MTPs themselves, and to be remarkably similar. However, these SCIs also turned out to have the property that no confidence assertions sharper than rejection assertions can be made unless all null hypotheses are rejected. This may be a disturbing limitation for practitioners, but it is not surprising—it is shared with SCIs for other powerful stepwise MTPs, including the fixed-sequence MTP (Hsu and Berger 1999, sec. 1), the step-down Dunnett MTP (Stefansson, Kim and Hsu 1988, sec. 3), and the step-up Dunnett MTP (Finner and Strassburger 2007, sec. 3).

It is then natural to try to construct families of MTPs related to Holm’s, Hochberg’s, and Hommel’s MTPs that do not have this limitation, at the cost of some loss in rejection power. A user may then choose among various such trade-off MTPs. Such families of MTPs were considered in Guilbaud (2012). The Holm-related MTPs considered there had previously been used in the gatekeeping context (Dmitrienko, Tamhane and Wiens 2008, sec. 3.1) where they were called truncated Holm MTPs. These can be viewed as MTPs between the Holm MTP and the Bonferroni MTP. The corresponding SCIs had briefly been discussed (though in other terms) in Strassburger and Bretz (2008, sec. 4.2), and they correspond to a special case of the Guilbaud (2009, Equation (10)) SCIs. In contrast, the Hochberg- and Hommel-related MTPs and the corresponding SCIs proposed in Guilbaud (2012) were new.

This article shows how alternative Hochberg- and Hommel-related MTPs can be constructed that: (i) are more powerful than the Hochberg- and Hommel-related MTPs proposed previously, and (ii) lead to SCIs that are sharper than the ones proposed previously. This improvement became possible through the derivation of an extended Simes inequality that is closely related to a recent result by Sarkar (2008, sec. 3). The results are formulated in terms of \( t \)-distributions, but are valid (at least approximately) more generally. A nice practical and theoretical aspect is that underlying correlations are only assumed to be nonnegative, as in Sarkar’s result. For completeness and comparison purposes, the Holm-related MTPs and corresponding SCIs just mentioned are also discussed—for simplicity under the same assumptions, though they are valid more generally.

The developments are in terms of one-sided null hypotheses, one-sided marginal tests/\( p \)-values, and one-sided SCIs. This was the case also for the developments in Guilbaud (2012) and Strassburger and Bretz (2008), as well as, for instance, in Hsu and Berger (1999, sec. 1) for the fixed-sequence MTP, in Stefansson, Kim and Hsu (1988, sec. 3) for the step-down Dunnett MTP, and in Finner and Strassburger (2007, sec. 3) for the step-up Dunnett MTP. The reason for not considering two-sided null hypotheses in this article is discussed in Section 10.1. It has to do with the fact that practitioners are typically interested in (and almost always make) additional decisions about the directions that led to the two-sided rejections, and the challenging problem of proving whether such additional directional decisions actually are valid with stepwise MTPs. The discussion in Section 10.1 is relevant also for MTPs that are more elaborated than the ones considered in this article; see for example, Dmitrienko, D’Agostino, and Huque (2013) for various elaborated MTPs that have stepwise components.

The article is organized as follows. The setup is described in Section 2, including the notation, the family of null hypotheses, and the assumptions. The formulation of an MTP as a closed-testing procedure (CTP) is described in Section 3. In view of their simplicity, the SCIs for the original Holm, Hochberg, and Hommel MTPs are introduced and discussed already in Section 4. The Liu (1996) class of CTPs is then described in Section 5, and the trade-off-type Holm-, Hochberg-, and Hommel-related MTPs are defined in Section 6 in terms of this class of CTPs. A simple algorithm is given in Section 7 that can be used to determine the null hypotheses rejected through these MTPs. The main results about the SCIs for these MTPs are given and discussed in Section 8. Section 9 provides illustrations based on the dose-finding trial example in Dmitrienko et al. (2009). Finally, some concluding comments and additional results are given in Section 10, including a discussion in Section 10.1 about why two-sided null hypotheses have not been considered, and a brief sketch in Section 10.2 of how the exact
inferences based on \(t\)-distributions that have been developed can be extended to approximate inferences in other situations through large-sample arguments.

The supplementary materials for this article consist of four appendices (A–D). Appendix A provides the extended Simes inequality (Theorem 1) and its proof. Appendix B provides a discussion of technical details and ideas underlying the proposed SCIs. In particular, the role of the extended Simes inequality is discussed in Section B.4. Appendix C provides proofs of miscellaneous assertions made in the developments. Appendix D provides a further discussion about the original Simes (1986) inequality and related results; including some historical details and an early result by Daniels (1945) that should be remembered, as well as a simple argument showing that in Simes’ proof based on independent \(p\)-values, one can drop the restrictive condition that \(p\)-values are uniformly distributed under their null hypotheses. Here “direct argument” means that no particular null hypotheses or test statistics underlying the \(p\)-values are introduced in the argument.

We refer to Guilbaud (2012) for more about SCIs for MTPs, and to Dmitrienko et al. (2009), Dmitrienko, D’Agostino, and Huque (2013), Dmitrienko and D’Agostino (2013), and Alosh, Bretz, and Huque (2013) for basic results, key issues, and recent advances in multiple testing in clinical drug development. It should also be mentioned that the pioneering article by Stefansson, Kim, and Hsu (1988) now has had its 25th birthday. This article was the first to show that SCIs corresponding to powerful stepwise MTPs were at all possible; see also the related further results and discussions in Hayter and Hsu (1994).

For convenience, no notational distinction is made between random quantities and the corresponding realizations in this article.

2. Setup, Notation, and Assumptions

The simultaneous inferences considered in this article concern \(m \geq 2\) specified real-valued quantities, \(\theta_1, \ldots, \theta_m\). Typically these unknown quantities \(\theta_i\) reflect comparisons of interest; for example, comparisons in a clinical trial in terms of differences in true means between two or more treatments; see the illustration in Section 9, and the extensions in Section 10.2. It is supposed that we are interested in making: (i) simultaneous assertions of the form

\[ \theta_i > \theta_{i,0}, \]  

and

\[ \theta_i < \theta_{i,0}, \]  

(1)

where \(\theta_{i,0}\) is a prespecified target boundary; and (ii) if possible, additional simultaneous confidence assertions about the \(\theta_i\)’s. To simplify the subsequent developments, though, all target assertions are supposed to be of the form “\(\theta_i > \theta_{i,0}\)” and “\(\theta_i < \theta_{i,0}\)” can be reformulated to be of the form “\(\theta_i > \theta_{i,0}\)” through an appropriate redefinition of \(\theta_i\) and \(\theta_{i,0}\). A target assertion “\(\theta_i > \theta_{i,0}\)” may, for example, correspond to a noninferiority or superiority assertion aimed at in a confirmatory clinical trial to support an intended claim about a treatment.

In the following, \(\alpha \in (0, 1/2)\) is a prespecified level, \(M\) denotes the index set \(\{1, 2, \ldots, m\}\), and for each subset \(I\) of \(M\), \(|I|\) denotes the number of elements in \(I\). We use the following standard notation. The normal distribution with mean \(\mu\) and variance \(b\) is denoted \(N(\mu, b)\), and the \(q\)-quantile and distribution function of the standard normal distribution \(N(0, 1)\) are denoted \(z_q\) and \(\Phi(\cdot)\), so that \(\Phi(z_q) = q\). The \(m\)-variate normal distribution with mean vector \(\mu\) and covariance matrix \(\Sigma\) is denoted \(N_m(\mu, \Sigma)\). The central \(t\)-distribution with \(\nu \geq 1\) degrees of freedom is denoted \(T_\nu\), its \(q\)-quantile is denoted \(t_{\nu, q}\), and its distribution function is denoted \(F_\nu(\cdot)\), so that \(F_\nu(t_{\nu, q}) = q\). Moreover, \(\chi^2_q\) denotes a chi-square distributed random variable with \(\nu \geq 1\) degrees of freedom.

The restriction \(\alpha < 1/2\) ensures that certain sign restrictions underlying the extended Simes inequality in Appendix A are fulfilled when this inequality is applied in the developments. Typically \(\alpha\) is chosen to be much smaller than 1/2, so this is not really a restriction in practice.

Let us now introduce, for \(i = 1, \ldots, m\), the null and alternative hypotheses, \(H_i\) and \(H_i^c\), about \(\theta_i \in (−\infty, \infty)\) given by

\[ H_i: \theta_i \leq \theta_{i,0}, \quad H_i^c: \theta_i > \theta_{i,0}, \]  

(2)

where \(\theta_{i,0} \in (−\infty, \infty)\) is a prespecified target value, see the discussion following (1). Note that it is the alternative hypothesis \(H_i^c\) that corresponds to the target assertion “\(\theta_i > \theta_{i,0}\)” aimed at.

The data we have at our disposal to make inferences about the quantities \(\theta_1, \ldots, \theta_m\) are supposed to be summarized in: (i) point estimators \(\hat{\theta}_1, \ldots, \hat{\theta}_m\) of these quantities; and (ii) standard errors \(s_{\hat{\theta}_1}, \ldots, s_{\hat{\theta}_m}\) of these point estimators, respectively. The following Assumption A about these random quantities leads to exact \(t\)-distribution-based inferences, including rejection decisions about \(H_i\)’s and SCIs for \(\theta_s\).

These distributional assumptions may seem rather strong, but they are in accordance with those used by Sarkar (2008, sec. 3) in his proof of the validity of the Simes inequality for \(t\)-distributions. Standard “large sample” arguments can, however, be invoked that show that approximate inferences can be made in situations where approximately normally distributed estimators \(\hat{\theta}_1, \ldots, \hat{\theta}_m\)
and associated standard errors are available, see Section 10.2.

Assumption A. (i) the data at our disposal for inferences about \( \theta_1, \ldots, \theta_m \) are summarized in the two independent statistics \((\hat{\theta}_1, \ldots, \hat{\theta}_m)\) and \( S > 0 \) that are such that

\[
\begin{pmatrix}
\hat{\theta}_1 \\
\vdots \\
\hat{\theta}_m
\end{pmatrix} \sim \mathcal{N}_m
\begin{pmatrix}
\theta_1 & \cdots & \theta_m \\
\vdots & \ddots & \vdots \\
\theta_m & \cdots & \theta_m
\end{pmatrix},
\sigma^2
\begin{pmatrix}
c_{1,1} & \cdots & c_{1,m} \\
\vdots & \ddots & \vdots \\
c_{m,1} & \cdots & c_{m,m}
\end{pmatrix}
\]

and

\[
S^2/\sigma^2 \sim \chi^2_v/\nu,
\]

where \( \sigma > 0 \) is an unknown scale factor; (ii) the empirical and theoretical standard error, \( \hat{s}_{\hat{\theta}} > 0 \) and \( \sigma_{\hat{\theta}} > 0 \), of the point estimator \( \hat{\theta} \) of \( \theta \) are of the form

\[
\hat{s}_{\hat{\theta}} = (S^2c_{ii})^{1/2}, \quad \sigma_{\hat{\theta}} = (\sigma^2c_{ii})^{1/2};
\]

and (iii) the \( (c_{ij}) \)-matrix in (3), the diagonal elements

\[c_{ii} > 0, \ldots, c_{m,m} > 0\]

are known and possibly different, and the off-diagonal elements

\[c_{ij}, i \neq j\]

are nonnegative, possibly unknown, and possibly different.

Remark 1. The fact that the off-diagonal elements \( c_{ij}, i \neq j \), are allowed to be different constitutes a considerable weakening of the earlier assumption (that off-diagonal elements are equal) used by Sarkar (1998, prop. 3.2) in his proof of the validity of the Simes inequality for \( t \)-distributions. Unbalanced designs are now covered by these assumptions, for example, many-to-one multiple comparisons based on samples of different size. See Sarkar (2008, sec. 3) for an interesting historical account of difficulties and advances in this context.

Assumption A implies that marginally for each \( i = 1, \ldots, m \): (i) \( \bar{\theta}_i \) and \( \hat{s}_{\hat{\theta}_i} \) are independent (observable) statistics such that \( \bar{\theta}_i \sim \mathcal{N}(\theta_i, \sigma^2) \) and \( \hat{s}_{\hat{\theta}_i}^{2} \sim \chi^2_v/\nu; \) (ii) for any given \( 0 < u < 1 \), the ordinary level-(1-\( u \)) confidence lower bound \( L_{i;1-u} \) for \( \theta_i \) based on the \( t \)-distributed ratio \( (\bar{\theta}_i - \theta_i)/s_{\hat{\theta}_i} \) is given by

\[
L_{i;1-u} = \bar{\theta}_i - t_{v,1-u} \hat{s}_{\hat{\theta}_i},
\]

so that

\[
Pr[L_{i;1-u} < \theta_i] = 1 - u;
\]

and (iii) the ordinary \( p \)-value \( p_i \) of the test of \( H_i \) versus \( H_i^c \) in (2) based on the \( t \)-statistic \( (\bar{\theta}_i - \theta_i, \bar{\theta}_i)/\hat{s}_{\hat{\theta}_i} \) is given by

\[
p_i = 1 - F_v[(\bar{\theta}_i - \theta_i, \bar{\theta}_i)/\hat{s}_{\hat{\theta}_i})
\]

\[= (u \text{ such that } L_{i;1-u} = \theta_i,0).\]

Finally, we note that under Assumption A, the \( m \) random quantities \( L_{i;(1-\alpha)/m}, i \in M \), constitute simultaneous \((1-\alpha)\) confidence lower bounds for \( \theta_1, \ldots, \theta_m \) in that

\[
Pr[L_{i;(1-\alpha)/m} < \theta_i \text{ for all } 1 \leq i \leq m] \geq 1 - \alpha.
\]

The multiplicity adjustment of the level of each lower bound in (8) is as if the bounds were independent. Relation (8) can be shown through an application of Slepian’s inequality in combination with Kimball’s inequality (Hsu 1996, theor. A.3.1 and cor. A.1.1) by conditioning on the random variable \( S \) in (3) and (4). The confidence bounds in (8) are sharper than the corresponding Bonferroni-type bounds \( L_{i;1-\alpha/m} \) that do not use the assumption that correlations between \( \bar{\theta}_i \)’s in (3) are nonnegative. As will be seen, these Bonferroni-type bounds, as well as the independence-type bounds in (8), occur in the expressions for the SCIs considered in Sections 4 and 8.

3. The Given MTP Viewed as a Closed-Testing Procedure (CTP)

The simultaneous \( 1-\alpha \) confidence assertions about \( \theta_1, \ldots, \theta_m \) to be considered correspond to a given MTP based on the marginal \( p \)-values (7) that rejects null hypotheses \( H_i: \theta_i \leq \theta_i,0 \) in the given family \( \{H_i: i \in M\} \) in favor of their alternatives \( H_i^c: \theta_i > \theta_i,0 \) at multiple-level \( \alpha \). Here multiple-level \( \alpha \) means that the Type-I FWER is strongly controlled to be at most \( \alpha \), that is, the probability of rejecting any true \( H_i \) in the family through the MTP is at most \( \alpha \), irrespective of the actual constellation of true and false \( H_i \)’s. The potential rejection assertions “\( \theta_i > \theta_i,0 \)” made through the MTP are thus equal to the target assertions aimed at.

It is assumed that the given MTP can be viewed as a CTP based on given local (i.e., marginal) level-\( \alpha \) tests for intersection hypotheses \( H_i = \bigcap_{c I \in H_i, \emptyset \neq I \subset M} \); and by definition, the CTP rejects any specific \( H_i \) if and only if each \( H_i \) with \( i \in I \subset M \) is rejected by its local/marginal level-\( \alpha \) test. Then: (i) the two complementary index sets

\[
I_{\text{Reject}} = \{i \in M: H_i \text{ is rejected by the given MTP/CTP},
\]

\[
I_{\text{Accept}} = M \setminus I_{\text{Reject}},
\]

are well-defined random quantities; (ii) their number of elements satisfy \( |I_{\text{Reject}}| + |I_{\text{Accept}}| = m \); and (iii) the probability is less than or equal to \( \alpha \) that \( I_{\text{Reject}} \) contains the index of any true \( H_i \), irrespective of the actual constellation of true and false \( H_i \)’s in the family. A simple algorithm is available to determine the index sets (9) for MTPs/CTPs considered in this article; see Section 7.

The index sets (9) and the rejection assertions “\( \theta_i > \theta_i,0 \)” , \( i \in I_{\text{Reject}} \), constitute the basis for the proposed simultaneous \( 1-\alpha \) confidence assertions about
\( \theta_1, \ldots, \theta_m \). The rejection assertions are always included in these simultaneous confidence assertions, so these latter assertions provide additional information “for free.”

4. The SCIs Corresponding to the Original Holm, Hochberg, and Hommel MTPs

The SCIs for these three classical MTPs are so simple and remarkably similar that we describe and discuss them already in this section. It is well known how these MTPs are defined in terms of the \( p \)-values \( p_1, p_2, \ldots, p_m \) given by (7), so it is easy to determine the index sets \( I_{\text{Reject}} \) and \( I_{\text{Accept}} \) in (9) for these MTPs; see for example, Dmitrienko, D'Agostino, and Huque (2013, sec. 5) and Dmitrienko and D’Agostino (2013, app. A.3). Algorithm 1 in Section 7 can also be used (with \( d_{\cdot, \cdot} \) ’s given by (19)) to determine these index sets.

The SCIs are defined in terms of: (i) the index sets \( I_{\text{Reject}} \) and \( I_{\text{Accept}} \), (ii) the target values \( \theta_{0,i} \) in (2), and (iii) the lower bounds \( L_{i,1-a/m} \) given by (5). We use the notation \( x \lor y \) for \( \max(x, y) \). Recall from (2) that the rejection of \( H_i \) in favor of \( H_F \) corresponds to the assertion “\( \theta_i \geq \theta_{0,i} \)” whereas the acceptance (i.e., nonrejection) of \( H_i \) corresponds to the noninformative assertion “\( \theta_i < -\infty \).”

Let \( I_{\text{Reject}} \) and \( I_{\text{Accept}} \) be the index sets (9) for Holm’s (1979) MTP based on the \( p \)-values (7), and consider the following intervals for \( \theta_i, i \in M \), respectively,

\[
\begin{align*}
\{ (\theta_{0,i} \lor L_{i,1-a/m}, \infty) \} & \text{ if } i \in I_{\text{Reject}} \text{ and } |I_{\text{Reject}}| = m, \\
\{ (\theta_{0,i}, \infty) \} & \text{ if } i \in I_{\text{Reject}} \text{ and } |I_{\text{Reject}}| < m, \\
\{ (L_{i,1-a/m}/|I_{\text{Accept}}|, \infty) \} & \text{ if } i \in I_{\text{Accept}}.
\end{align*}
\]

These intervals simultaneously cover \( \theta_1, \ldots, \theta_m \), respectively, with probability at least \( 1 - \alpha \). They were derived by Guilbaud (2008) and Strassburger and Bretz (2008) through different methods.

Let \( I_{\text{Reject}} \) and \( I_{\text{Accept}} \) be the index sets (9) for Hochberg’s (1988) MTP based on the \( p \)-values (7), and consider the following intervals for \( \theta_i, i \in M \), respectively,

\[
\begin{align*}
\{ (\theta_{0,i} \lor L_{i,1-a}/m, \infty) \} & \text{ if } i \in I_{\text{Reject}} \text{ and } |I_{\text{Reject}}| = m, \\
\{ (\theta_{0,i}, \infty) \} & \text{ if } i \in I_{\text{Reject}} \text{ and } |I_{\text{Reject}}| < m, \\
\{ (L_{i,1-a}/|I_{\text{Accept}}|, \infty) \} & \text{ if } i \in I_{\text{Accept}}.
\end{align*}
\]

These intervals simultaneously cover \( \theta_1, \ldots, \theta_m \), respectively, with probability at least \( 1 - \alpha \). They were derived by Guilbaud (2008); see eq. (58) and remark 9 in that article.

Let \( I_{\text{Reject}} \) and \( I_{\text{Accept}} \) be the index sets (9) for Hommel’s (1988) MTP based on the \( p \)-values (7), and consider the following intervals for \( \theta_i, i \in M \), respectively,

\[
\begin{align*}
\{ (\theta_{0,i} \lor L_{i,1-a}/m, \infty) \} & \text{ if } i \in I_{\text{Reject}} \text{ and } |I_{\text{Reject}}| = m, \\
\{ (\theta_{0,i}, \infty) \} & \text{ if } i \in I_{\text{Reject}} \text{ and } |I_{\text{Reject}}| < m, \\
\{ (L_{i,1-a}/|I_{\text{Accept}}|, \infty) \} & \text{ if } i \in I_{\text{Accept}}.
\end{align*}
\]

These intervals simultaneously cover \( \theta_1, \ldots, \theta_m \), respectively, with probability at least \( 1 - \alpha \). They were derived by Guilbaud (2012); see eq. (58’) and remark 11 in that article.

The expressions for the SCIs in (10)–(12) are very simple and remarkably similar. The only formal difference is that the Bonferroni-type lower bound \( L_{i,1-a/m} \) occurs in the first row in (10), whereas the independence-type lower bound \( L_{i,1-a}/m \) occurs in the first row of (11) and (12). Intuitively it is surprising that the same Bonferroni-type adjustment \( \alpha/|I_{\text{Accept}}| \) occurs in the third row of (10)–(12), because Hochberg’s and Hommel’s MTPs are not Bonferroni-based CTPs. The reason behind this common Bonferroni-type adjustment is actually far from evident; see Remark 3 in Section 8.

The SCIs in (10)–(12) have the following properties. For each \( i \in I_{\text{Reject}} \), the confidence assertion about \( \theta_i \) is at least as sharp as the rejection assertion “\( \theta_i \geq \theta_{0,i} \)” made through the MTP, possibly sharper—though the important point here is that such a sharper assertion is not possible in case \( |I_{\text{Reject}}| < m \). For each \( i \in I_{\text{Accept}} \), the confidence assertion about \( \theta_i \) is sharper than the noninformative acceptance assertion “\( \theta_i < -\infty \)” made through the MTP, but less sharp than the rejection assertion “\( \theta_i > \theta_{0,i} \)” thus indicating by how much one missed this target assertion.

It is interesting to note that in (10)–(12), the lower confidence bounds for \( \theta_i \)'s with \( i \in I_{\text{Accept}} \) are always at least as sharp as the corresponding Bonferroni bounds \( L_{i,1-a/m} \) in that

\[
L_{i,1-a/m} \leq L_{i,1-a}/|I_{\text{Accept}}| < \theta_{0,i} \quad \text{for all } i \in I_{\text{Accept}},
\]

where the inequality \( \leq \) is strict if at least one null hypothesis is rejected by the MTP. In case no null hypothesis is rejected by the MTP, the inequality \( \leq \) in (13) is an equality, and the \( m \) lower bounds in (10)–(12) are then all equal to the Bonferroni bounds \( L_{i,1-a/m} \). This could be used to construct lower confidence bounds that can beat Bonferroni bounds also in right neighborhoods of target values of interest by prespecifying \( \theta_i \)'s to be larger than these target values (instead of equal to them). This would obviously require various considerations regarding the choice of \( \theta_i \)'s, and we do not develop this idea in this article.

The essential difference between the SCIs (10)–(12) is of course that the three underlying MTPs may lead to different index sets (9). As is well known, the index sets
5. The Liu (1996) Class of CTPs Based on Ordered p-Values and a Critical Matrix

Consider the set of all p-values \( p_1, p_2, \ldots, p_m \) given by (7). For each nonempty \( I \subseteq M \), let \( p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(|I|)} \) denote the ordered values of the subset of \( p_i \)'s with index \( s \in I \). When \( I = M \), we sometimes simplify the \( p_{(r)} \)-notation to \( p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(|M|)} \).

Each CTP considered in this article for the family \( \{H_s : s \in M\} \) of null hypotheses given by (2) is based on local level-\( \alpha \) tests of intersection hypotheses \( H_I = \cap_{s \in I} H_s \) that have a rejection rule of the form

\[
P_{(s)} I \leq d_{(I, |I|) - r + 1} \quad \text{for some} \quad s \in \{1, 2, \ldots, |I|\}. \tag{15}
\]

Here the critical constants \( d_{(1)} \geq d_{(2)} \geq \cdots \geq d_{(|I|)} \) are prespecified and such that the test (15) of \( H_I \) has local (i.e., marginal) level \( \alpha \). There is no loss of generality in letting the critical constants \( d_{(r)} \) be nonincreasing in \( s \), because the ordered \( p \)-values \( p_{(s)} I \) in (15) are nondecreasing in \( s \).

The CTP based on the local level-\( \alpha \) tests (15) is determined by the prespecified critical constants. Such CTPs were considered by Liu (1996), who called the lower triangular matrix,

\[
\begin{pmatrix}
  d_{1,1} & d_{2,1} & d_{2,2} \\
  \vdots & \vdots & \vdots \\
  d_{r,1} & d_{r,2} & \cdots & d_{r,r} \\
  \vdots & \vdots & \ddots & \vdots \\
  d_{m,1} & d_{m,2} & \cdots & d_{m,r} & \cdots & d_{mm}
\end{pmatrix}, \tag{16}
\]

the critical matrix of the CTP. The local test (15) of \( H_I \) is thus based on the critical constants in row \( r = |I| \) of the matrix (16). It is important to note that these critical constants depend on \( I \) only through the number \(|I|\) of elements in \( I \). It is understood (without loss of generality) that

\[
d_{r,1} \geq d_{r,2} \geq \cdots \geq d_{r,r} \quad \text{for each row} \quad r = 1, \ldots, m. \tag{17}
\]

Moreover, for all CTPs considered in this article, the diagonal elements in (16) satisfy

\[
d_{1,1} \geq d_{2,2} \geq \cdots \geq d_{m,m}. \tag{18}
\]

The inequalities (18) are important for the subsequent developments in that they: (i) ensure the validity of Algorithm 1 described in Section 7 through which the index sets (9) can easily be determined; and (ii) lead to certain simplifications, notably the one discussed in Remark 3 in connection with (29). The inequalities (18) will, therefore, be verified in case they are not evident.

Many well-known MTPs can be viewed as CTPs based on (15)–(18); including the Bonferroni MTP, Holm’s (1979) MTP, Hochberg’s (1988) MTP, and Hommel’s (1988) MTP; see Bernhard, Klein, and Hommel (2004) for a review. The Bonferroni MTP has all its \( d_{r,s} \)'s equal to \( \alpha/m \), whereas the other three MTPs have \( d_{r,s} \) given for \( 1 \leq s \leq r \leq m \) by

\[
d_{r,s}^{\text{Holm}} = \frac{1}{r}, \\
\]

\[
d_{r,s}^{\text{Hochberg}} = \frac{1}{s}, \quad d_{r,s}^{\text{Hommel}} = \frac{\alpha - s + 1}{r}. \tag{19}
\]

Interestingly, it follows from (19) that the three MTPs have the same diagonal elements (18), and that these common diagonal elements are equal to

\[
\frac{\alpha}{1} \geq \frac{\alpha}{2} \geq \cdots \geq \frac{\alpha}{m}. \tag{20}
\]

We recognize these values as the critical values that occur in the step-down algorithm of Holm (1979), and in the step-up algorithm of Hochberg (1988). This relation between the critical matrix of Holm’s MTP, Hochberg’s MTP, and Hommel’s MTP was noted and discussed by Liu (1996).

6. The Holm(W), Hochberg(W), and Hommel(W) MTPs with Parameter

\[ 0 \leq W \leq 1 \]

We now introduce certain MTPs related to Holm’s (1979) MTP, Hochberg’s (1988) MTP, and Hommel’s
(1988) MTP that also belong to Liu’s (1996) class of CTPs, and define them in terms of their critical matrix. These MTPs depend on a parameter $0 \leq W \leq 1$ that specifies the closeness to the original Holm, Hochberg, and Hommel MTPs based on (19), with $W = 1$ corresponding to these original MTPs. The value of $0 \leq W \leq 1$ is to be chosen and prespecified by a user.

Let the Holm($W$) MTP, the Hochberg($W$) MTP, and the Hommel($W$) MTP, be the CTPs defined through (15)–(18) with critical constant $d_{r,s}$ given for $0 \leq W \leq 1$ and $1 \leq s \leq r \leq m$ by

$$d_{r,s}^{\text{Holm}(W)} = A_r(W) \frac{1}{r} - \frac{1}{s},$$

$$d_{r,s}^{\text{Hochberg}(W)} = B_r(W) \frac{1}{s} - \frac{1}{r},$$

$$d_{r,s}^{\text{Hommel}(W)} = B_r(W) \frac{r - s + 1}{r},$$

(21)

where $A_r(W)$ and $B_r(W)$ are the nondecreasing functions of $0 \leq W \leq 1$ given by

$$A_r(W) = \alpha \left[ 1 - \left( 1 - W \right) \left( m - r \right) / m \right],$$

$$B_r(W) = 1 - \frac{1 - \alpha}{\left[ 1 - W \right] \left( 1 - \alpha \right)^{1/m}} [1 - \alpha]^{r/m}. \quad (22)$$

It can be verified that as $W$ increases in the interval $[0, 1]$: (i) the value of $A_r(W)$ increases from $\alpha r / m$ to $\alpha$ if $1 \leq r < m$, and equals $\alpha$ if $r = m$; and (ii) the value of $B_r(W)$ increases from $1 - (1 - \alpha)^{r/m}$ to $\alpha$ if $1 \leq r < m$, and equals $\alpha$ if $r = m$. In particular, $A_r(W) \leq \alpha$, $B_r(W) \leq \alpha$, and $A_r(1) = B_r(1) = \alpha$.

The critical constants $d_{r,s}$ in (21) satisfy the important inequalities (18) for any given $0 \leq W \leq 1$. For the Holm($W$)-related $d_{r,s}$’s this follows from the fact that $A_r(W)/r$ equals $\alpha \left[ W/r + (1 - W)/m \right]$, which is nonincreasing as $r$ increases. For the Hochberg($W$)-, and Hommel($W$)-related $d_{r,s}$, this follows from the fact shown in Appendix C.1 that the ratio $B_r(W)/r$ is nonincreasing as $r$ increases.

It is clear from (22) that with $W = 1$, the critical constants in (21) equal the corresponding ones in (19). This means that the Holm(1), Hochberg(1), and Hommel(1) MTPs are equal to the original Holm (1979), Hochberg (1988), and Hommel (1988), MTPs, respectively.

It is also clear from (22) that, with $W = 0$, the Holm($W$) MTP has $d_{r,s}$’s that are all equal to $\alpha / m$, as the Bonferroni MTP. This means that the Holm(0) MTP is equal to the Bonferroni MTP.

The rejection index sets $I_{\text{Reject}}$ for the Holm($W$), Hochberg($W$), and Hommel($W$) MTPs are nondecreasing as $W$ increases, because $A_r(W)$ and $B_r(W)$ in (21) and (22) are nondecreasing as $W$ increases. It follows in particular that for any given $0 \leq W \leq 1$, these MTPs reject, respectively: (i) at most as much as the Holm (1979) MTP, the Hochberg (1988) MTP, and the Hommel (1988) MTP; and (ii) at least as much as the Holm(0) MTP, the Hochberg(0) MTP, and the Hommel(0) MTP.

It is shown in Appendix C.2 that for any given $0 < W < 1$, the index sets (9) of the Holm($W$), Hochberg($W$), and Hommel($W$) MTPs satisfy

$$I_{\text{Holm}(W)} \supset I_{\text{Hochberg}(W)} \supset I_{\text{Hommel}(W)},$$

(23)

This generalizes the relations (14). It follows in particular that the Holm($W$), Hochberg($W$), and Hommel($W$) MTPs reject at least as much as the Holm(0) MTP, that is, as the Bonferroni MTP.

Remark 2. Guilbaud (2012, sec. 4.3) considered another version of the Hochberg($W$) and Hommel($W$) MTPs defined for $W_{a,m} \leq W \leq 1$ where the lower bound $W_{a,m}$ equals $1 - [\alpha/m]/[1 - (1 - \alpha)^{1/m}]$ and is such that $0 < W_{a,m} < 1$. The $d_{r,s}$’s of these MTPs are defined by substituting

$$B_r^*(W) = \alpha - (1 - W)(m - r)[1 - (1 - \alpha)^{1/m}] \quad (24)$$

for $B_r(W)$ in the last two rows of (21). In the sequel we will refer to these MTPs based on (24) as the Hochberg*(W) and Hommel*(W) MTPs.

It is clear from (22) and (24) that $B_r^*(1) = B_r(1) = \alpha$ for all $1 \leq r < m$. It follows that the Hochberg*(1) and Hommel*(1) MTPs are equal to the original Hochberg and Hommel MTPs, respectively, as the Hochberg(1) MTP and the Hommel(1) MTPs. Moreover, it is shown in Appendix C.3 that for $W_{a,m} < W < 1$,

$$B_r^*(W) < B_r(W) < \alpha, \quad \text{if } 1 \leq r < m,$n

$$B_r^*(W) = B_r(W) = \alpha, \quad \text{if } r = m. \quad (25)$$

It follows that for $W_{a,m} < W < 1$, the Hochberg($W$), and Hommel($W$) MTPs are sharper than the Hochberg*(W) and Hommel*(W) MTPs, respectively, in that at least as much is rejected, possibly more. Another advantage of the Hochberg($W$) and Hommel($W$) MTPs is that they are defined also for $0 \leq W < W_{a,m}$. It is understood in the sequel that the Hochberg($W$) MTP and the Hommel($W$) MTP are the MTPs based on (21) and (22) that are defined for all $0 \leq W \leq 1$.

7. Algorithm to Determine the Index Sets $I_{\text{Reject}}$ and $I_{\text{Accept}}$ in (9)

The following simple algorithm can be used to determine the index sets (9) for MTPs/CTPs that can be
represented in terms of a critical matrix (16) satisfying (17) and (18). In particular, the index sets (9) for the Holm(W), Hochberg(W), and Hommel(W) MTPs with \( d_{r,s} \)'s given by (21) can be determined through this algorithm for any given \( 0 \leq W \leq 1 \).

**Algorithm 1** (Klein (1998)). Assuming that the inequalities (18) hold: (i) determine the index-set

\[
K = \{ r \in M; p_{(m-s+1)} \geq d_{r,s} \} \text{ for all } s = 1, \ldots, r; \]

(ii) if \( K = \emptyset \), then set \( I_{\text{Reject}} \) equal to \( M \); and (iii) if \( K \neq \emptyset \), then let \( k = \max(K) \), and set \( I_{\text{Reject}} \) equal to \( \{ i \in M; p_i \leq d_{k,k} \} \).

This is a corrected version of the corresponding algorithm in Guilbaud (2012, p. 327). It originates from Klein’s (1998) unpublished diploma thesis, but its published version mentioned as Klein’s theorem in Bernhard, Klein, and Hommel (2004, p. 8) contained a recently detected index error that unfortunately was reproduced in Guilbaud (2012); see the corrections in Bernhard, Klein and Hommel (2013) and Guilbaud (2013). Only the numerical illustrations in Guilbaud (2012, sec. 7) were affected by this index error, and a corrected version of that section was provided in the Supporting Information for Guilbaud (2013).

8. The SCIs Corresponding to the Holm(W), Hochberg(W), and Hommel(W) MTPs

The SCIs for \( \theta_1, \ldots, \theta_m \) considered here are defined in terms of: (i) the target values \( \theta_{i,0} \) in (2); (ii) the lower bounds \( L_{i,1−u} \) in (5); (iii) the functions \( A_r(W) \) and \( B_r(W) \) given by (22); and (iv) the index-sets \( I_{\text{Reject}} \) and \( I_{\text{Accept}} \) in (9) that can be determined through Algorithm 1 with \( d_{r,s} \)'s given by (21) for the Holm(W), Hochberg(W), and Hommel(W) MTPs. Some technical details and underlying ideas are discussed in Appendix B; but briefly, these SCIs are special cases of the SCIs in Liu (1996) class, which in turn follow from the more general SCIs in Liu (2012, eq. (19)). Recall that the value of \( 0 \leq W \leq 1 \) is to be prespecified by a user, and that \( W = 1 \) corresponds to the original Holm (1979), Hochberg (1988), and Hommel (1988) MTPs.

Let \( I_{\text{Reject}} \) and \( I_{\text{Accept}} \) be the index sets (9) determined through Algorithm 1 with \( d_{r,s} = d_{r,s}^{\text{Holm}(W)} \) in (21) for the Holm(W) MTP based on the p-values (7), and consider the following intervals for \( \theta_i, i \in M \), respectively,

\[
(\theta_{i,0} \lor L_{i,1−(1−\alpha)/m}, \infty), \quad \text{if } i \in I_{\text{Reject}} \text{ and } |I_{\text{Reject}}| = m, \\
(\theta_{i,0} \lor L_{i,1−(1−W)m}, \infty), \quad \text{if } i \in I_{\text{Reject}} \text{ and } |I_{\text{Reject}}| < m, \\
(L_{i,1−B_r(W)}, \infty), \quad \text{if } i \in I_{\text{Accept}}, \\
\text{with } r = |I_{\text{Accept}}|. 
\]  

These intervals simultaneously cover \( \theta_1, \ldots, \theta_m \), respectively, with probability at least \( 1 − \alpha \). These SCIs are not new. For \( W = 1 \), they equal the SCIs in (10) derived by Guilbaud (2008) and Strassburger and Bretz (2008). For \( 0 \leq W < 1 \), they were briefly discussed (though in other terms) in Strassburger and Bretz (2008, Section 4.2), and they can be viewed as a special case of the Guibaud (2009, eq. (10)) SCIs with weights \( B_i = 1/m \).

Let \( I_{\text{Reject}} \) and \( I_{\text{Accept}} \) be the index sets (9) determined through Algorithm 1 with \( d_{r,s} = d_{r,s}^{\text{Hochberg}(W)} \) in (21) for the Hochberg(W) MTP based on the p-values (7), and consider the following intervals for \( \theta_i, i \in M \), respectively,

\[
(\theta_{i,0} \lor L_{i,1−(1−\alpha)/m}, \infty), \quad \text{if } i \in I_{\text{Reject}} \text{ and } |I_{\text{Reject}}| = m, \\
(\theta_{i,0} \lor L_{i,1−(1−W)m}, \infty), \quad \text{if } i \in I_{\text{Reject}} \text{ and } |I_{\text{Reject}}| < m, \\
(L_{i,1−B_r(W)}, \infty), \quad \text{if } i \in I_{\text{Accept}}, \\
\text{with } r = |I_{\text{Accept}}|. 
\]  

These intervals simultaneously cover \( \theta_1, \ldots, \theta_m \), respectively, with probability at least \( 1 − \alpha \). For \( W = 1 \), they equal the SCIs in (12) derived by Guilbaud (2012)
for Hommel’s (1988) MTP; whereas for $0 \leq W < 1$, they are new.

For $W = 1$, the SCIs (26)–(28) have the properties discussed in Section 4. For $0 \leq W < 1$, the SCIs have the following properties. For each $i \in I_{\text{Reject}}$, the confidence assertion about $\theta_i$ is at least as sharp as the rejection assertion “$\theta_i > \theta_{i,0}$” made through the MTP, possibly sharper—the important point here is that such a sharper assertion is possible even in case $|I_{\text{Reject}}| < m$. For each $i \in I_{\text{Accept}}$, the confidence assertion about $\theta_i$ is sharper than the noninformative acceptance assertion “$\theta_i > -\infty$” made through the MTP but less sharp than the rejection assertion “$\theta_i > \theta_{i,0}$,” thus indicating by how much one missed this target assertion.

For $W = 0$, the SCIs (26) reduce to the Bonferroni bounds $L_{1,1} = \alpha/m$ for all $1 \leq i \leq m$, irrespective of the outcome of the index sets $I_{\text{Reject}}$ and $I_{\text{Accept}}$, because $A_r(0)/r = \alpha/m$. This reduction result is quite appealing because the Holm(0) MTP equals the Bonferroni MTP.

Remark 3. The confidence intervals in the third row of (10)–(12) and (26)–(28) have a simple relation to the diagonal elements $d_{r,r}$ of the critical matrix (16) defining each underlying MTP. It can be verified from (19) and (21) that each of these confidence intervals is of the simple form

$$(L_{1,1} - d_{r,r}, \infty), \quad \text{if } i \in I_{\text{Accept}}, \quad (with \ r = |I_{\text{Accept}}|).$$

Remark 4. Let $0 \leq W' < W'' \leq 1$ be given, and let $H(W)$ denote for a moment any one of the Holm($W$), Hochberg($W$), and Hommel($W$) MTPs. Then $H(W'')$ can be said to be better than $H(W')$ in that it always rejects at least as much, possibly more; see Section 6. In contrast, one cannot say that the $H(W'')$-based SCIs in (26)–(28) are better than the $H(W')$-based SCIs, nor vice versa. Here “are better than” is in the strong sense “are contained within” with probability one. Rather, one can say that the $H(W'')$-based SCIs and the $H(W')$-based SCIs correspond to different tradeoffs between the aim of rejecting as many $H_i$’s as possible and the aim of getting sharper confidence assertions about $\theta_i$’s associated with rejected $H_i$’s. Note, for example, that in case $H(W')$ rejects some $H_i$’s and $H(W'')$ does not reject all $H_i$’s, then the $H(W')$-based lower confidence bound in the second row of (26)–(28) can be larger than the corresponding $H(W'')$-based lower confidence bound; see the illustration in Section 9.2.

Remark 5. For any given $0 \leq W \leq 1$, the SCIs (28) for the Hommel($W$) MTP are at least as sharp, possibly sharper than, the SCIs (27) for the Hochberg($W$) MTP, which in turn are at least as sharp as, possibly sharper than, the SCIs (26) for the Holm($W$) MTP. This can be verified using the relations (23), the inequality $1 - (1 - \alpha)^{1/m} > \alpha/m$, and the fact shown in Appendix C.2 that the difference $B_r(W) - A_r(W)$ given by (22) is nonnegative. It should be remembered in this context that the Holm($W$)-based inferences (rejections and SCIs) are valid under weaker assumptions than the Hochberg($W$)- and Hommel($W$)-based inferences.

Remark 6. The SCIs (27) and (28) are at least as sharp as, possibly sharper (in case $W < 1$) than, the SCIs in Guilbaud (2012, eqs. (58) and (58')) that correspond to the Hochberg($W$) and Hommel($W$) MTP discussed in Remark 2. The latter SCIs are given for any given $W_{a,m} \leq W \leq 1$ by the expressions (27) and (28) with: (i) $B_r(W)$ in the third row replaced by $B_r^*(W)$ in (24); and (ii) index sets $I_{\text{Reject}}$ and $I_{\text{Accept}}$ for the Hochberg($W$) and Hommel($W$) MTP obtained through Algorithm 1 and the $d_{r,s}$’s described in Remark 2. The improvement over the Hochberg($W$)- and Hommel($W$)-based inferences is a consequence of the relations (25) between $B_r^*(W)$ and $B_r(W)$. These relations, and the role in this context of the extended Simes inequality derived in Appendix A, are discussed in Appendix B.4; see also the illustration of the improvement in Section 9.3.

9. Illustrations Based on the Dose-Finding Trial Example in Dmitrienko et al. (2009)

We use the same illustration as in Guilbaud (2012, sec. 7) based on the dose-finding trial example in Dmitrienko et al. (2009). We can then make some comparisons with previous results. The trial is a parallel-group study where patients with dyslipidemia are treated either with placebo (labeled $D_0$), or with one of four doses (labeled $D_1$, $D_2$, $D_3$, $D_4$), in increasing dose order) of the drug under investigation. The study is balanced in that the size of each treatment group is 77. The inferences of interest concern the four ($D_i - D_0$)-differences, $i = 1,2,3,4$, in true means, $\theta_1$, $\theta_2$, $\theta_3$, $\theta_4$, with respect to increase in high-density lipoprotein cholesterol after 12 weeks of treatment in patients. These differences are estimated by the corresponding four ($D_i - D_0$)-differences in sample means, $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$, $\hat{\theta}_4$, with standard errors, $s_{\hat{\theta}_1}, s_{\hat{\theta}_2}, s_{\hat{\theta}_3}, s_{\hat{\theta}_4}$, based on a pooled estimator $S^2$ of an underlying common
variance $\sigma^2$. This pooled $S^2$ is based on the five within-
group variances.

The situation is according to the setup in Section 2, with $m = 4$, $M = \{1, 2, 3, 4\}$, $\nu = 5$ ($77 - 1$) = 380 
degrees of freedom, constants $c_{i,j}$ in (3) given by the 
balanced design (i.e., $c_{i,j} = 2/77$, and $c_{i,j} = 1/77$ for 
$i \neq j$), and null hypotheses $H_1$, $H_2$, $H_3$, $H_4$ given by (2) 
with each $\theta_{i,0} = 0$. The specified value of $\alpha$ is 0.025.

9.1 Outcome Scenario

The outcome scenario considered here is the one la-
beled Scenario 1 in Dmitrienko et al. (2009), and Out-
come Scenario 1 in Guilbaud (2012). In this outcome 
scenario, the observed $(D_i - D_0)$-differences, $\theta_1$, $\theta_2$, $\theta_3$, $\theta_4$, 
are equal to 2.90, 3.14, 3.56, 3.81, and their standard 
errors, $s_{\theta_1}$, $s_{\theta_2}$, $s_{\theta_3}$, $s_{\theta_4}$, are each equal to 1.44. The ob-
served difference $\hat{\theta}_i$ is thus positive for $1 \leq i \leq 4$, and 
increases as $i$ increases. The marginal $p$-values $p_1$, $p_2$, $p_3$, $p_4$, based on these observed differences and standard 
errors are equal to 0.0224, 0.0149, 0.0069, 0.0042; these values differ slightly from those for Scenario 1 in 
Dmitrienko et al. (2009, tab. 2.1) due to the rounding in 
the observed values of differences and standard errors 
just given.

9.2 Inferences Based on (26)–(28) for Some Values of $0 \leq W \leq 1$

Table 1 provides the rejection index set for the 
Holm($W$), Hochberg($W$), and Hommel($W$) MTPs, as 
well as the corresponding simultaneous $1 - \alpha$ confidence 
lower bounds for $\theta_1$, $\theta_2$, $\theta_3$, $\theta_4$ in (26)–(28), in the 
outcome scenario just described. With these data, the results 
for the Hommel($W$) MTP turn out to be equal to those for 
the Hochberg($W$) MTP, so they are not repeated in the 
table. Recall that the $W$-value is to be prespecified by a 
user.

First, consider the results in the left part of Table 1. 
With $W = 1$ (original Holm MTP), only $H_3$ and $H_4$ 
are rejected, so no confidence assertion sharper than a 
rejection assertion “$\theta_i > 0$” is made about $\theta_1$ and $\theta_4$. 
Note also that as $W$ decreases in the seven rows with 
$I_{\text{Reject}} = \{3, 4\}$, the assertions about $\theta_1$ and $\theta_2$ become 
less sharp. Moreover, as $W$ further decreases in the four 
rows with $I_{\text{Reject}} = \{4\}$: (i) the assertions about $\theta_1$, $\theta_2$, and 
$\theta_3$ become less sharp; whereas (ii) the assertion about 
$\theta_4$ becomes sharper, even though not all hypotheses are 
rejected. In particular, with $W = 0$ (Bonferroni MTP), the 
assertion about $\theta_4$ is sharper than the assertions made 
with larger $W$-values, including the one made with the 
original Holm MTP. These results are in accordance with 
the discussion in Remark 4.

Next, consider the results in the right part of Ta-
ble 1. With $W = 1$ (original Hochberg MTP), all four 
hypotheses are rejected, and a confidence assertion about 
$\theta_4$ is made that is sharper than a rejection assertion. In-
terestingly, an equally sharp assertion about $\theta_4$ is made 
with $W = 0.9$. Otherwise, similar comments can be made 
about the monotonic behavior of rows as $W$ decreases as 
those made about the left part of Table 1. Again, these 
results are in accordance with Remark 4.

The fact that the SCIs in the right part of Table 1 are 
at least as sharp, often sharper, than those in the left part 
is in accordance with the discussion in Remark 5.

9.3 Improvement Over the Hochberg- and 
Hommel-Related Inferences in Guilbaud (2012)

As discussed in Remarks 2 and 6, for any given 
$W_{a,m} \leq W \leq 1$: (i) the Hochberg($W$) and Hommel($W$) 
MTPs reject at least as much as, possibly more (in 
case $W < 1$) than, the Hochberg*$W$ and Hommel*$W$
MTPs in Guilbaud (2012, sec. 4.3); and (ii) the SCIs (27) and (28) are at least as sharp, possibly sharper (in case \( W<1 \)) than, the SCIs in Guilbaud (2012, eqs. (58) and (58')) that correspond to the Hochberg\(^*(W)\) and Hommel\(^*(W)\) MTPs. Recall also that the Hochberg\((W)\) and Hommel\((W)\) MTPs and associated SCIs (27) and (28) have the further advantage that they are defined also for \( 0 \leq W < W_{\alpha,m} \).

Table 2 illustrates the improvement in the outcome scenario described in Section 9.1. With these data, the results for the Hommel\((W)\) and Hommel\(^*(W)\) MTPs turn out to be equal to those for the Hochberg\((W)\) and Hochberg\(^*(W)\) MTPs, respectively, so they are not repeated in the table. We restrict considerations to \( W\)-values equal/close to 0.86, 0.79, 0.33, 0.32; because for these four values, the Hochberg\((W)\) MTP rejects more than the Hochberg\(^*(W)\) MTP.

The improvement in the left part (compared to the right part) of Table 2 may be small, except when the Hochberg\((W)\) reject more than the Hochberg\(^*(W)\). Anyhow, whenever there is a difference, small or large, it is always in favor of the Hochberg\((W)\) MTP and its associated SCIs. This illustrates the fact discussed in Remarks 2 and 6 that the Hochberg\((W)\) and Hommel\((W)\) MTPs and their associated SCIs are preferable to the Hochberg\(^*(W)\) and Hommel\(^*(W)\) MTPs and their associated SCIs.

**10. Concluding Comments and Additional Results**

If the aim in a particular situation is to reject as many null hypotheses as possible, then it may be reasonable to restrict considerations to the original Holm, Hochberg, and Hommel MTPs and their associated SCIs given by (10)–(12). This may be the case in confirmatory clinical trials where target assertions may correspond, for example, to noninferiority and/or superiority assertions aimed at to support intended claims about treatments. The confidence assertions associated with nonrejected hypotheses that can then be made may be of some value in that they indicate by how much one missed the target assertions aimed at. Otherwise, if one is interested in the possibility of making assertions sharper than rejection assertions also when not all hypotheses are rejected, then one may consider trade-off procedures such as the Holm\((W)\), Hochberg\((W)\), and Hommel\((W)\) MTPs with \( 0 \leq W <1 \) and their associated SCIs given by (26) and (28). The choice of \( 0 \leq W <1 \) to be used then depends on the anticipated behavior of the MTPs and their SCIs. A further discussion of these aspects is outside the scope of this article.

We emphasize that the main results in this article are based on the setup and assumptions made in Section 2. The assumption that correlations between \( \hat{\theta}_i \)'s in (3) are nonnegative is important for the Hochberg\((W)\)- and Hommel\((W)\)-related inferences. If this assumption is questionable, then one may consider the Holm\((W)\)-related inferences instead, because these are valid without this assumption.

We end this article with: (i) a discussion about why we have not considered two-sided null hypotheses in this article, and (ii) a brief description of how large sample arguments can be invoked that show that approximate inferences can be made under rather weak distributional assumptions in many situations.

**10.1 About One-Sided Directional Decisions Supplementing Two-Sided Rejections**

An MTP may of course be applied to a family of two-sided null-hypotheses under appropriate assumptions. For example, under the setup of Section 2 (with two-sided null-hypotheses and \( p\)-values instead of one-sided), Holm’s (1979) MTP is valid without any assumption about \( \hat{\theta}_i \)-correlations, whereas the assumption about
nonnegative $\hat{\theta}_i$-correlations does not seem to be sufficient for the validity of the Simes inequality that underlies the Hochberg (1988) and Hommel (1988) MTPs; see Sarkar (2008, theor. 3.1 part (ii)) for a stronger sufficient condition on correlations.

However, even if an MTP for a family of two-sided null-hypotheses is valid, it is seldom sufficient to make decisions just about rejections. Practitioners are typically interested in, and almost always make, additional decisions about the directions that led to the two-sided rejections, based, for example, on point estimates. Two kinds of errors are then possible: (i) incorrect rejections of true two-sided null hypotheses, so called Type I errors; and/or (ii) correct rejections of false two-sided null hypotheses but directional decisions in the wrong directions, so called Type III errors.

Few theoretical results are, however, available about whether powerful stepwise-like MTPs with multiple-level $\alpha$ also control the probability of making any Type-I or Type-III error to be less than or equal to $\alpha$. Positive results have been obtained under rather restrictive independence or conditional independence assumptions about test statistics (Shaffer 1980; Liu 1997; Finner 1999; Hochberg, Liu, and Parmat 2000, Section 3; Sarkar, Sen, and Finner 2004), for instance for Holm’s and Hochberg’s MTPs based on two-sided marginal $p$-values. Unfortunately, even simple dependence structures are not covered by these results—a prominent example being the many-to-one multiple comparison problem (Finner 1999, sec. 4). In fact, even in the independence case, a proof seems to be lacking for Hommel’s MTP based on two-sided $p$-values, though simulations indicate that additional directional decisions are valid (Hochberg, Liu, and Parmat 2000, sec. 3).

Thus, in many (not to say most) situations of practical interest, a proof is lacking of the validity of one-sided directional decisions supplementing two-sided rejections for Holm’s, Hochberg’s, and Hommel’s MTPs based on two-sided $p$-values, as well as for many other powerful stepwise-like MTPs. In their concluding remarks, Sarkar, Sen, and Finner (2004, p. 96) stated that while it is believed that positive results might hold for certain types of dependent test statistics, the directional decision problem still remains to be one of the most challenging problems in multiple testing. For these reasons, the developments in this article have been in terms of one-sided null hypotheses.

The validity of additional directional decisions may of course be investigated empirically in a given situation encountered in practice, for example, through simulations for parameter constellations in a neighborhood of the anticipated constellation. However, as long as proofs are lacking, doubts may potentially be raised about the validity, for example, by regulatory agencies.

Anyhow, in view of the current state of knowledge in this context, it is the author’s opinion that practitioners should be (made) aware of this validity problem, and of its potential consequences. It seems reasonable to require that if one plans to use a stepwise MTP (or a more elaborated MTP with stepwise components) for a family of two-sided null hypotheses in a confirmatory study, and two-sided rejections may be supplemented by one-sided directional decisions (as will almost always be the case in interpretations of the results), then one should explicitly state already at the planning stage of the study whether a proof of the validity of such additional one-sided directional decisions is available, or whether this validity is supported only empirically; and give appropriate references.

The article by Westfall, Bretz, and Tobias (2013) appeared during the revision of the present article. This is an extensive investigation of the directional error problem for stepwise and closed-testing MTPs, mainly using numerical and simulation methods. Briefly, for certain standard situations of practical interest to biopharmaceutical research, including many-to-one and other pair-wise comparisons based on the $t$-distribution and possibly small and/or unbalanced sample sizes, they found no cases of excess directional error rates. The article by Westfall, Bretz, and Tobias could therefore be a suitable reference for practitioners to support the validity of additional one-sided directional decisions in such standard situations. On the other hand, the authors also provided new counterexamples where directional error rates are uncontrolled based on linear models and CTPs for null hypotheses about linear functions of regression parameters, so the directional decision problem in multiple testing is indeed a challenging one.

Finally, it should be mentioned in this context that Holm (1979, p. 68) suggested that one could consider the larger family of $2m$ one-sided null hypotheses of the form $H_1^\circ : \theta_i \leq \theta_{i,0}$ and $H_m^\circ : \theta_i \geq \theta_{i,0}$ and apply his step-down MTP with one-sided $p$-values to this larger family to make valid directional decisions. Guilbaud (2012, sec. 8.2) showed how Holm’s suggestion can be adapted to the Holm($W$) MTPs and how corresponding two-sided SCIs for $\theta_1, \ldots, \theta_m$ can be constructed. A simple variant of this idea, applicable to the Hochberg($W$) and Hommel($W$) MTPs, would be to: (i) consider the two families $\{H_1^\circ, \ldots, H_m^\circ\}$ and $\{H_1^\circ, \ldots, H_m^\circ\}$; and (ii) combine the one-sided inferences that can be made in each family through a Bonferroni adjustment over the two families (assigning half of the desired maximum error rate to each family) to get valid directional decisions and corresponding two-sided SCI for $\theta_1, \ldots, \theta_m$. If the correlation structure of the $\theta_i$’s is not in accordance with Assumption A in the family $\{H_1^\circ, \ldots, H_m^\circ\}$, then one can try to reformulate and exchange the one-sided null hypotheses within
some pairs of one-sided null hypotheses, and accordingly redefine the $\hat{\theta}_i$'s, $\hat{\theta}_a$'s, and $\theta_{i,a}$'s, so that the resulting correlation structure of the $\hat{\theta}_i$ is appropriate. This is always possible if $m = 2$, and is sometimes possible if $m > 2$. In the situation considered in Section 9, no such reformulation and exchange of one-sided null hypotheses within pairs is necessary. We omit the details.

10.2 Inferences Based on “Large Sample” Approximations

The rather strong Assumption A in Section 2 leads to exact inferences about $\theta_1, \ldots, \theta_m$ based on the $t$-distribution. Standard “large sample” arguments can, however, be invoked to show that corresponding approximate inferences can be made in various situations.

The data available for inferences are then supposed to be summarized in the estimators $\hat{\theta}_1, \ldots, \hat{\theta}_m$ with associated standard errors $s_1, \ldots, s_m$. Letting $n$ denote the sample size (or some other suitable information measure), it is assumed that as $n \to \infty$: (i) the distribution of $\sqrt{n}(\hat{\theta}_1 - \theta_1, \ldots, \hat{\theta}_m - \theta_m)^T$ tends to an $m$-variate normal distribution with mean $(0, \ldots, 0)^T$ and covariance matrix $(\sigma_{i,j})$ having positive diagonal elements $\sigma_{i,i}$ and nonnegative off-diagonal elements; and (ii) each standardized difference $\sqrt{n}(s_i - \sqrt{\sigma_{i,i}})$ tends in probability to zero. This holds in many common situations.

Then, for large values of $n$: (i) the vector $(\hat{\theta}_1, \ldots, \hat{\theta}_m)^T$ is approximately distributed according to an $m$-variate normal distribution with mean $(\theta_1, \ldots, \theta_m)^T$ and covariance matrix having diagonal elements $(s_1, \ldots, s_m)$ and nonnegative off-diagonal elements; (ii) $L_{z_{1-u}} = \hat{\theta}_1 - z_{1-u}s_1^2$ is such that $Pr[L_{i:1-u} < \theta_i]$ is approximately equal to $1 - u$; (iii) $p_i \equiv 1 - \Phi((\hat{\theta}_i - \theta_{i,a})/s_i^2)$ can be used to test the null hypothesis $H_i$ in (2) at an approximate level $0 < u < 1$ by rejecting $H_i$ if and only if $p_i \leq u$; and (iv) with these normal-approximation-based $L_{i:1-u}$'s and $p_i$'s, the Holm($W$), Hochberg($W$), and Hommel($W$) MTPs defined through (21) have multiple-level approximately equal to $\alpha$, and their associated SCIs in (26)–(28) have simultaneous confidence level approximately equal to $1 - \alpha$.

As a simple illustration of these approximate inferences, consider the situation described in Section 9, with the modification/complication that: (i) sample sizes are not equal; and (ii) the within-group variability is anticipated to differ so much between groups that we are reluctant to pool sample variances $S_i^2$. Such a situation was discussed by Dunnett and Tamhane (1991, p. 994) who mentioned the possibility of using Holm’s (1979) MTP based on the (Group$_i$ − Group$_0$)-differences $\hat{\theta}_i$ in sample means and standard errors $s_i$ of the form $(S_i^2/n_i + S_0^2/n_0)^{1/2}$. It can be verified from the two preceding paragraphs that with these $\hat{\theta}_i$'s and standard errors $s_i$, and if the ratios $n_i/n$ tend to given positive proportions as $n \equiv n_0 + \cdots + n_m$ tends to infinity and the normal-approximation-based $L_{i:1-u}$'s and $p_i$'s are used, then for any given $0 \leq W \leq 1$, the Holm($W$), Hochberg($W$), and Hommel($W$) MTPs, as well as their associated SCIs in (26)–(28), are approximately valid for large values of $n$.

Suppose that in the preceding situation we are interested in group comparisons of, say, survival (or death) rates $P_i(t)$ after a given time $t$ on treatment instead of means. The comparisons may then be in terms of, for example, differences, ratios, or odds ratios; with $\theta_i$'s specified for all $1 \leq i \leq m$ either as, respectively, $\hat{\theta}_i = P_i(t) - P_0(t)$, $\hat{\theta}_i = \log(P_i(t)/P_0(t))$, or $\hat{\theta}_i = \log([P_i(t)/Q_i(t)]/[P_0(t)/Q_0(t)])$ where $Q_i(t) \equiv 1 - P_i(t)$. Let $\hat{P}_i(t)$ denote the Kaplan–Meier estimator of $P_i(t)$, and let $\hat{\alpha}_i^2(t)$ denote the associated Greenwood variance estimator. Approximate inferences about the $\theta_i$'s may then be obtained by substituting $\hat{P}_i(t)$ for $P_i(t)$ in the $\theta_i$-definitions; and (ii) standard errors $s_i$ in terms of $\hat{\alpha}_i^2(t)$ obtained (if necessary) through the “delta-method.” For example, if $\theta_i$'s are specified as rate differences, then each $s_i^2$ is simply equal to $\hat{\alpha}_i^2(t) + \hat{\alpha}_j^2(t)$; whereas if $\theta_i$'s are specified as log-odds-ratios as just described, then according to the delta-method, each $s_i^2$ equals $\hat{\alpha}_i^2(t)/[\hat{P}_i(t)/Q_i(t)]^2 + \hat{\alpha}_j^2(t)/[\hat{P}_0(t)/Q_0(t)]^2$ where $Q_i(t) \equiv 1 - \hat{P}_i(t)$. We omit technical details.

Acknowledgments

Parts of the research reported in this article were presented in an invited talk at the International Conference on Simultaneous Inference that took place in Hannover, Germany, 24–26 September 2013. The author is grateful to the referees for their comments.

Supplementary Materials

The supplementary materials for this article consist of four appendices (A-D). See the third paragraph from the end of Section 1 for a description of the contents of the appendices.

[Received November 2012. Revised December 2013]

References

Alosh, M., Bretz, F., and Huque, M. (2013), “Advanced Multiplicity Adjustment Methods in Clinical Trials,” Statistics in Medicine, DOI: 10.1002/sim.5974. [125]

Bernhard, G., Klein, M., and Hommel, G. (2004), “Global and Multiple Test Procedures Using Ordered p-Values—A Review,” Statistical Papers, 45, 1–14. [128,130]

Bernhard, G., Klein, M., and Hommel, G. (2013), “Erratum: Global and Multiple Test Procedures Using Ordered p-Values—A Review,” Statistical Papers, DOI: 10.1007/bf02778266. [130]
Daniels, H. E. (1945), “The Statistical Theory of the Strength of Bundles of Threads,” *Proceedings of the Royal Statistical Society, Series A*, 183, 405–435. DOI: 10.1098/rspa.1945.0011. [125]

Dmitrienko, A., Bretz, F., Westfall, P. H., Troendle, J., Wiens, B. L., Tamhane, A. C., and Hsu, J. C. (2009), “Multiple Testing Methodology,” in *Multiple Testing Problems in Pharmaceutical Statistics*, eds. A. Dmitrienko, A. C. Tamhane, and F. Bretz, Boca Raton, FL: Chapman & Hall/CRC, pp. 35–98. [124,125,131,132]

Dmitrienko, A., and D’Agostino, R. B. (2013), “Traditional Multiplicity Adjustment Methods in Clinical Trials,” *Statistics in Medicine*, DOI: 10.1002/sim.5990. [125,127]

Dmitrienko, A., Tamhane, A. C., and Wiens, B. L. (2009), “Multiple Testing Adjustment Methods in Clinical Trials,” *Statistics in Medicine*, DOI: 10.1002/sim.5990. [124,125,127]

Dmitrienko, A., Tamhane, A. C., and Wiens, B. L. (2008), “General Multistage Gatekeeping Procedures,” *Biometrical Journal*, 50, 667–677. [124]

Dunnett, C. W., and Tamhane, A. C. (1991), “Step-Down Multiple Tests for Comparing Treatments With a Control in Unbalanced One-Way Layouts,” *Statistics in Medicine*, 10, 939–947. [135]

Finner, H. (1999), “Stepwise Multiple Test Procedures and Control of Directional Errors,” *The Annals of Statistics*, 27, 274–289. [134]

Finner, H., and Strassburger, K. (2007), “Step-Up Related Simultaneous Confidence Intervals for MCC and MCB,” *Biometrical Journal*, 49, 40–51. [124]

Guilbaud, O. (2008), “Simultaneous Confidence Regions Corresponding to Holm’s Step-Down Procedure and Other Closed-Testing Procedures,” *Biometrical Journal*, 50, 678–692. [124,127,130]

——— (2009), “Alternative Confidence Regions for Bonferroni-Based Closed-Testing Procedures that are not Alpha-Exhaustive,” *Biometrical Journal*, 51, 721–735. [124,130]

——— (2012), “Simultaneous Confidence Regions for Closed Tests, Including Holm-, Hochberg-, and Hommel-Related Procedures,” *Biometrical Journal*, 54, 317–342. [124,125,127,129,130,131,132,134]

——— (2013), “Corrections: Simultaneous Confidence Regions for Closed Tests, Including Holm-, Hochberg-, and Hommel-Related Procedures,” *Biometrical Journal*, 55, 634–635. [130]

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. [125]

Hochberg, Y. (1988), “A Sharper Bonferroni Procedure for Multiple Significance Testing,” *Biometrika*, 75, 800–802. [123,127,128,129,130,134]

Hochberg, Y., Liu, W., and Parmat, Y. (2000), “On the Directional Decision Problem,” in *Perspectives in Statistical Sciences*, eds. A. Basu, J. Gosh, P. K. Sen, and B. Sinha, India: Oxford University Press, pp. 179–190. [134]

Holm, S. (1979), “A Simple Sequentially Rejective Multiple Test Procedure,” *Scandinavian Journal of Statistics*, 6, 65–70. [123,127,128,129,130,133,134,135]

Hommel, G. (1988), “A Stagewise Rejective Multiple Test Procedure Based on a Modified Bonferroni Test,” *Biometrika*, 75, 383–386. [123,127,128,129,130,134]

Hsu, J. C. (1996), *Multiple Comparisons—Theory and Methods*, London: Chapman & Hall. [126]

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. [124]

Klein, M. (1998), “Multiple Testprozeduren: Step-Down und Step-Up,” unpublished diploma thesis, University Mainz, Mainz, Germany. [In German, unpublished, but available from Markus Zervas: markus.zervas@t-online.de] [130]

Liu, W. (1996), “Multiple Tests of a Non-Hierarchical Finite Family of Hypotheses,” *Journal of the Royal Statistical Society*, Series B, 58, 455–461. [124,128,130,131]

——— (1997), “Control of Directional Errors With Step-Up Multiple tests,” *Statistics and Probability Letters*, 31, 239–242. [134]

Sarkar, S. K. (1998), “Some Probability Inequalities for Ordered MTP2 Random Variables: A Proof of the Simes Conjecture,” *The Annals of Statistics*, 26, 494–504. [124,126]

——— (2008), “On the Simes Inequality and its Generalization,” in *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen (Vol. 1)*, eds. N. Balakrishnan, E. Pena, and M. J. Silvapulle, Beachwood, Ohio: IMS Collections, 231–242, DOI: 10.1214/193940370000000167. [124,125,126,134]

Sarkar, S. K., and Chang, C. K. (1997), “Simes’ Method for Multiple Hypothesis Testing With Positively Dependent Test Statistics,” *Journal of the American Statistical Association*, 92, 1601–1608. [124]

Sarkar, S. K. Sen, P. K., and Finner, H. (2004), “On Two Results in Multiple Testing,” in *Recent Developments in Multiple Comparison Procedures* (Vol. 47), eds. Y. Benjamini, F. Bretz, and S. K. Sarkar, Beachwood, Ohio: IMS Lecture Notes—Monograph Series, pp. 89–99, DOI: 10.1214/lms/1196285628. [134]

Simes, R. J. (1986), “An Improved Bonferroni Procedure for Multiple Tests of Significance,” *Biometrika*, 73, 751–754. [124,125]

Shaffer, J. P. (1980), “Control of Directional Errors With Stagewise Multiple Test Procedures,” *The Annals of Statistics*, 8, 1342–1348. [134]

Stefansson, G., Kim, W. C., and Hsu, J. C. (1988), “On Confidence Sets in Multiple Comparisons,” in *Statistical Decision Theory and Related Topics IV* (Vol. 2), eds. S. S. Gupta and J. O. Berger, New York: Springer-Verlag, pp. 89–104. [124,125]

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. [124,127,130]

Westfall, P. H., Bretz, F., and Tobias, R. D. (2013), “Directional Error Rates of Closed Testing Procedures,” *Statistics in Biopharmaceutical Research*, 6, 239–258. [124]

About the Author

Olivier Guilbaud, Sk¨ondalsv¨agen 113A 4tr, SE-12868, Sk¨onidal, Sweden (E-mail: olivier.jm.guilbaud@gmail.com). The author is retired - future updates of the contact information will be available through the membership directory of the International Statistical Institute (ISI).