

A general multistage procedure for k -out-of- n gatekeeping

Dong Xi^{a,*†} and Ajit C. Tamhane^b

We generalize a multistage procedure for parallel gatekeeping to what we refer to as k -out-of- n gatekeeping in which at least k out of n hypotheses ($1 \leq k \leq n$) in a gatekeeper family must be rejected in order to test the hypotheses in the following family. This gatekeeping restriction arises in certain types of clinical trials; for example, in rheumatoid arthritis trials, it is required that efficacy be shown on at least three of the four primary endpoints. We provide a unified theory of multistage procedures for arbitrary k , with $k = 1$ corresponding to parallel gatekeeping and $k = n$ to serial gatekeeping. The theory provides an insight into the construction of truncated separable multistage procedures using the closure method. Explicit formulae for calculating the adjusted p -values are given. The proposed procedure is simpler to apply for this particular problem using a stepwise algorithm than the mixture procedure and the graphical procedure with memory using entangled graphs. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: closed procedure; error rate function; familywise error rate; Hochberg procedure; Holm procedure; Hommel procedure; separable procedures; truncated procedures

1. Introduction

Research on gatekeeping procedures has flourished over the past decade as they are increasingly used in clinical trials involving tests of multiple null hypotheses, which can be hierarchically grouped into $m \geq 2$ ordered families, F_1, \dots, F_m . Hierarchical ordering is induced by the so-called gatekeeping restrictions; for example, efficacy on secondary endpoints cannot be tested unless efficacy on primary endpoints is demonstrated, or superiority cannot be tested unless noninferiority is established. In general, hypotheses in families F_j for $j > i$ are retained (accepted) without tests if a specified gatekeeping restriction on the rejection of hypotheses in family F_i is not met. Family F_i is called a gatekeeper for family F_{i+1} and by extension to all families F_j for $j > i$.

The earliest works [1, 2] dealt with the serial gatekeeping problem in which the gatekeeping condition was that all the hypotheses in the gatekeeper family are rejected. Dmitrienko *et al.* [3] considered the parallel gatekeeping problem in which the gatekeeping condition was that at least one hypothesis in the gatekeeper family is rejected. Dmitrienko *et al.* [4] generalized the serial and parallel gatekeeping problems to the tree gatekeeping problem. These authors proposed procedures on the basis of the closure method of Marcus *et al.* [5] and which used weighted Bonferroni tests for the intersection hypotheses. Dmitrienko *et al.* [6–8] proposed mixture gatekeeping procedures on the basis of the closure method, which generalized the scope of the tree gatekeeping procedures. Bretz *et al.* [9, 10] proposed a graphical approach to gatekeeping. Recently, Maurer and Bretz [11] have developed a graphical approach with memory that deals with general logical restrictions of the type handled by mixture procedures.

Dmitrienko *et al.* [6–8] noted that tree gatekeeping cannot be used to specify the k -out-of- n gatekeeping restriction for general k . As an example, consider the Food and Drug Administration Guidance for Industry on rheumatoid arthritis [12], which states that ‘... trial results were considered to support a

^aIIS Statistical Methodology, Novartis Pharmaceuticals Corporation, One Health Plaza, East Hanover, NJ 07936, U.S.A.

^bDepartment of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, U.S.A.

*Correspondence to: Dong Xi, IIS Statistical Methodology, Novartis Pharmaceuticals Corporation, One Health Plaza, East Hanover, NJ 07936, U.S.A.

†E-mail: dong.xi@novartis.com

conclusion of effectiveness when statistical evidence of efficacy was shown for at least three of the four measures . . . ? There is an additional requirement that this condition must be met in at least four of the six monthly visits, but development of a procedure that meets this second condition is not considered in the present paper.

Dmitrienko *et al.* [6–8] used restriction functions to deal with general types of logical restrictions in mixture procedures. However, mixture procedures, being closed multiple test procedures (MTPs), do not have obvious stepwise shortcuts. Maurer and Bretz [11] gave an example of the graphical approach for $k = 3$ and $n = 4$. It involves taking a convex combination of four graphs (called entangled graphs). If one restricts to standard p -value-based MTPs of Holm [13], Hochberg [14], and Hommel [15], the proposed procedure is easier to describe and perform than the mixture procedure and the entangled graphical procedure with memory. We have also extended the methods to the fallback MTP [16, 17] and the Dunnett MTP [18], but we do not give those extensions here for lack of space.

The MTPs used to test individual families F_i are called component MTPs, which together constitute the multistage gatekeeping procedure. Maurer *et al.* [1] showed that the familywise error rate (FWER) is strongly controlled (1) for the serial gatekeeping problem by using a multistage MTP that tests each family F_i using a component MTP at level α and proceeds to test F_{i+1} iff all hypotheses in F_i are rejected ($i = 1, \dots, m - 1$).

For parallel gatekeeping, Dmitrienko *et al.* [3] proposed a closed MTP, which used the weighted Bonferroni test for all intersection hypotheses and accounted for the parallel gatekeeping restriction by assigning suitable weights to the hypotheses in each intersection. Dmitrienko *et al.* [19] showed that this closed MTP has a multistage shortcut, which uses the Bonferroni MTPs for the first $m - 1$ families and the Holm MTP for the last family. Guilbaud [20] showed that this multistage procedure can be derived directly (without recourse to the closure principle) and the Holm MTP can be replaced by any FWER controlling MTP in the last family. Dmitrienko *et al.* [21] showed that any separable MTP, which is more powerful than the Bonferroni MTP, can be used as a component MTP in the first $m - 1$ families. We generalize this multistage procedure in the present paper.

The outline of the paper is as follows. Section 2 extends the definition of separability introduced in [21] for parallel gatekeeping ($k = 1$) to arbitrary k ; we refer to it as k -separability. Section 3 derives k -separable truncated stepwise MTPs called k -truncated MTPs. Section 4 states the generalized k -out-of- n gatekeeping MTP. Section 5 derives adjusted p -values for this MTP. Section 6 applies the procedures to the rheumatoid arthritis example. Section 7 gives a discussion of the results. The Appendix includes all the proofs.

2. Error rate function and separability

From now to Section 3, we will consider a single family F of n hypotheses, H_1, \dots, H_n , with p -values p_1, \dots, p_n . We shall return to the multifamily gatekeeping setting in Section 4. Throughout, we shall assume that the p_i are uniformly distributed over $[0, 1]$ under the respective null hypotheses H_i ($i = 1, \dots, n$). For the Hochberg and the Hommel procedures, we will assume independence or positive regression dependence among the p_i 's to ensure that the Simes test [22] remains conservative [23, 24]. Let $p_{(1)} \leq \dots \leq p_{(n)}$ denote the ordered p -values and let $H_{(1)}, \dots, H_{(n)}$ denote the corresponding hypotheses.

We require any MTP to strongly control [25] FWER; that is, for any specified α ,

$$\text{FWER} = \sup_{H(I)} P\{\text{Reject at least one true } H_i, i \in I | H(I)\} \leq \alpha \quad (1)$$

where the supremum is taken over the subset of the parameter space where $H(I) = \bigcap_{i \in I} H_i$ holds for all intersection hypotheses and $I \subseteq \{1, \dots, n\}$ is an arbitrary nonempty index set.

Dmitrienko *et al.* [21] introduced the error rate function to construct multistage MTPs for parallel gatekeeping. The error rate function of an MTP with nominal FWER α is defined as

$$e(I | \alpha) = \sup_{H(I)} P \left[\bigcup_{i \in I} \{\text{Reject } H_i\} \mid H(I) \right]$$

It is the maximum probability of at least one false rejection of H_i , $i \in I$ when using an MTP controlling FWER at level α . In addition, $e(\emptyset | \alpha) = 0$, and we can set $e(N | \alpha) = \alpha$.

In general, the error rate function is difficult to calculate exactly especially when the test statistics are correlated. Therefore, we use an easily computable upper bound instead but continue to regard it as the true error rate function of the MTP and still denote it by $e(I | \alpha)$. For the Bonferroni MTP, this upper bound equals $e(I | \alpha) = (|I|/n)\alpha$.

For parallel gatekeeping, Dmitrienko *et al.* [21] defined an α -level MTP to be separable if its error rate function satisfies

$$e(I | \alpha) < \alpha \quad \text{if } |I| < n, \quad e(I | \alpha) = \alpha \quad \text{if } |I| = n \quad (2)$$

where $|I|$ is the cardinality of the index set I . Obviously, the Bonferroni MTP is separable, but it is known from [21] that the Holm, Hochberg, and Hommel stepwise MTPs are not separable.

For the k -out-of- n gatekeeping problem, we generalize the definition (2) as follows. An α -level MTP is said to be k -separable if its error rate function satisfies

$$e(I | \alpha) < \alpha \quad \text{if } |I| \leq n - k, \quad e(I | \alpha) = \alpha \quad \text{if } |I| > n - k \quad (3)$$

In the sequel, we will see that the k -separability condition ensures that a positive significance level $\leq \alpha$ can be carried over to the following family iff at least k hypotheses are rejected; that is, no more than $n - k$ hypotheses is accepted, in the gatekeeper family. For $k = 1$, this definition reduces to the separability definition (2), whereas for $k = n$, this definition is vacuous, which agrees with the fact that separability is not required for serial gatekeeping because full α is carried over to the next family iff all n hypotheses are rejected.

3. k -Truncated separable multiple test procedures

3.1. Closed multiple test procedures

The Holm, Hochberg, and Hommel stepwise MTPs are more powerful than the Bonferroni MTP, but they cannot be used to test gatekeeper families (except for serial gatekeepers) because they are not k -separable for any $k < n$. For parallel gatekeeping, Dmitrienko *et al.* [21] proposed truncated versions of these stepwise MTPs to make them separable and still be more powerful than the Bonferroni MTP. These truncated MTPs use convex combinations of the critical constants of the original stepwise MTPs and the Bonferroni MTP. For example, for a prespecified truncation parameter $\gamma \in [0, 1)$, the truncated Holm MTP tests and rejects $H_{(i)}$ for $i = 1, 2, \dots, n$, as long as

$$p_{(i)} \leq \left(\frac{\gamma}{n - i + 1} + \frac{1 - \gamma}{n} \right) \alpha.$$

It stops at the first step for which this inequality is violated and accepts all the remaining hypotheses. The Hochberg and Hommel MTPs can be similarly modified.

In order to satisfy the k -separability condition (3), we need to use a closed MTP, which tests each intersection hypothesis $H(I)$ at local level $\alpha(I)$ such that $\alpha(I) < \alpha$ if $|I| \leq n - k$ and $\alpha(I) = \alpha$ if $|I| > n - k$. Any function $\alpha(I)$, which satisfies this condition, can be used to construct a k -truncated separable MTP. It is convenient to use the following function parameterized by a truncation parameter $\gamma \in [0, 1)$:

$$\alpha(I) = \begin{cases} \left[\gamma + \frac{(1-\gamma)m}{n-k+1} \right] \alpha & \text{if } 1 \leq m \leq n - k \\ \alpha & \text{if } m > n - k \end{cases} \quad (4)$$

where we have denoted $|I| = m$. The reason for using this parametric function is that it can be used to construct truncated stepwise MTPs such as the truncated Holm MTP given earlier, which depends on a single parameter γ . When $\gamma = 1$, this function gives the untruncated MTP, and when $\gamma = 0$, it tests $H(I)$ for $|I| \leq n - k$ at a truncated level $|I|/(n - k + 1) \times \alpha$. We will give some comments on the choice of γ later in this section.

The error rate function of the closed MTP using this form of $\alpha(I)$ is given by

$$e(I | \alpha) = \begin{cases} 0 & \text{if } m = 0 \\ \alpha(I) & \text{if } m > 0 \end{cases} \quad (5)$$

It is easy to see that $e(I | \alpha)$ satisfies the k -separability condition.

Dmitrienko *et al.* [21] showed that if a gatekeeper family is tested using an MTP at level α with error rate function $e(I | \alpha)$ and if A is the index set of the accepted hypotheses, then the significance level transferred to the next family equals $\alpha - e(A | \alpha)$. Thus, if $|A| > n - k$, that is, if less than k hypotheses are rejected, then $\alpha - e(A | \alpha) = 0$, and so testing stops with acceptance of all the remaining hypotheses thus satisfying the k -out-of- n gatekeeping condition.

For parallel gatekeeping ($k = 1$), one must use a separable MTP to test all hypotheses in a gatekeeper family, whereas for serial gatekeeping ($k = n$), a non-separable MTP can be used. For $1 < k < n$, we need to switch from an untruncated MTP to a k -truncated MTP after k hypotheses are rejected in the case of a step-down MTP or from a k -truncated MTP to an untruncated MTP after $n - k$ hypotheses are accepted in the case of a step-up MTP. In the following sections, we will present k -truncated versions of the Holm, Hochberg, and Hommel MTPs; we give their derivations in the Appendix.

The choice of γ for the primary family is somewhat subjective. It is easy to check that $e(I | \alpha)$ defined in (5) is an increasing function of γ , and therefore, the significance level transferred to the secondary family is a decreasing function of γ . Thus, increasing γ gains power for the primary family at the expense of the secondary family. To strike a balance, we usually use $\gamma = 0.5$, but a larger γ may be used if more rejections are desired in the primary family.

The choice of k is of course dictated by clinical considerations rather than statistical considerations. We note in passing, however, that $\alpha(I)$ is an increasing function of k , and so the number of rejected primary hypotheses will increase with k .

3.2. k -Truncated Holm multiple test procedure

The k -truncated Holm MTP operates as follows. Begin testing with $H_{(1)}$. In general, if all $H_{(j)}$ for $j < i$ are rejected, then test $H_{(i)}$ and reject it if

$$p_{(i)} \leq \begin{cases} \frac{\alpha}{n-i+1} & \text{if } 1 \leq i \leq k \\ \left(\frac{\gamma}{n-i+1} + \frac{1-\gamma}{n-k+1}\right)\alpha & \text{if } k < i \leq n \end{cases} \quad (6)$$

and proceed to the next step. Otherwise, accept $H_{(i)}$ and all the remaining hypotheses. Thus, use the untruncated Holm MTP until k hypotheses are rejected, and then switch to the k -truncated Holm MTP.

Theorem 1

The k -truncated Holm MTP is an exact shortcut to the closed MTP, which rejects each intersection hypothesis $H(I)$ if

$$p_i \leq \alpha(I)/m \text{ for at least one } i \in I \quad (7)$$

where $\alpha(I)$ is given by (4). Because this closed MTP tests each $H(I)$ at level $\alpha(I)$ using the Bonferroni test, (5) gives the exact error rate function of the k -truncated Holm MTP.

Let A denote the index set of the accepted hypotheses and let $r = n - |A|$ denote the number of rejected hypotheses. Then from (5), we see that the significance level transferred to the next family is given by

$$\alpha - e(A | \alpha) = \begin{cases} \alpha & \text{if } r = n \\ \frac{r-k+1}{n-k+1}(1-\gamma)\alpha & \text{if } k \leq r < n \\ 0 & \text{if } r < k \end{cases} \quad (8)$$

Note that if all n hypotheses are rejected, then the full α level is transferred to the next family, whereas if less than n but greater than k hypotheses are rejected, then the α level transferred to the next family is proportional to the the number of hypotheses rejected in excess of $k - 1$ with $(1 - \gamma)\alpha/(n - k + 1)$ level transferred for each such rejected hypothesis. This comes from the Bonferroni part of the truncated Holm MTP. When all n hypotheses are rejected, there is an additional increase of $\gamma\alpha$ level coming from the Holm part.

Bretz *et al.* [9] showed that the Holm MTP can be represented graphically so that the significance level assigned to any rejected hypothesis is transferred to unrejected hypotheses within the same family. If all hypotheses in the primary family are rejected, then the full significance level at which the last hypothesis is tested and rejected is transferred to the secondary family. Clearly, if less than k hypotheses are rejected, then no α level is transferred to the next family.

3.3. Li's truncated Holm multiple test procedure

Li [26] proposed a general truncated Holm MTP for parallel gatekeeping, which uses n prespecified non-negative constants, $\varepsilon_1, \dots, \varepsilon_n$, such that $\sum_{i=1}^n \varepsilon_i = \alpha$ and $\varepsilon_0 = 0$. In general, Li's MTP tests and rejects $H_{(i)}$ if all $H_{(j)}$ for $j < i$ are rejected and

$$p_{(i)} \leq \frac{\alpha - \varepsilon_0 - \varepsilon_1 - \dots - \varepsilon_{i-1}}{n - i + 1} \quad (i = 1, \dots, n).$$

If $H_{(i)}$ is rejected, then ε_i is transferred to the next family, and the remaining level is used to test $H_{(i+1)}$ ($i = 1, \dots, n - 1$). If $H_{(i+1)}$ is accepted, then all the remaining hypotheses are also accepted, and the significance level equals to $\varepsilon_1 + \dots + \varepsilon_i$ is transferred to test the next family. If all n hypotheses are rejected, then the total significance level transferred to the next family equals $\sum_{i=1}^n \varepsilon_i = \alpha$. If no hypothesis is rejected, then, of course, no significance level is transferred to the next family, thus satisfying the parallel gatekeeping condition. Note that if $\varepsilon_i = 0$ for $i < k$ and $\varepsilon_i > 0$ for $i \geq k$, then this MTP satisfies the k -out-of- n gatekeeping condition.

Li's truncated Holm MTP is also an exact shortcut to a closed MTP that uses the Bonferroni test as local test of intersection hypotheses. To ensure consonance and thus a shortcut, the monotonicity condition of [27] requires that

$$\frac{\alpha}{n} \leq \frac{\alpha - \varepsilon_1}{n - 1} \leq \dots \leq \frac{\alpha - \varepsilon_1 - \dots - \varepsilon_{n-1}}{1}.$$

The choice of $\varepsilon_1, \dots, \varepsilon_n$ subject to the aforementioned constraint is more difficult than the choice of a single γ . For the k -out-of- n gatekeeping condition, the two choices can be made equal if we choose

$$\varepsilon_1 = \dots = \varepsilon_{k-1} = 0, \varepsilon_k = \dots = \varepsilon_{n-1} = \frac{(1 - \gamma)\alpha}{n - k + 1}, \varepsilon_n = \gamma\alpha + \frac{(1 - \gamma)\alpha}{n - k + 1}.$$

In the special case $k = n - 1$, we see that $\varepsilon_{n-1} = (1 - \gamma)\alpha/2$ and $\varepsilon_n = (1 + \gamma)\alpha/2$. Thus, for the first $n - 1$ steps, we use the untruncated Holm MTP, and only in the last step, we use the truncated Holm MTP with the critical constant $(1 + \gamma)\alpha/2$ instead of α for comparing with $p_{(n)}$.

3.4. k -Truncated Hochberg multiple test procedure

The k -truncated Hochberg MTP uses the same critical constants as the k -truncated Holm MTP (6) but operates in the reverse order as follows. Begin testing with $H_{(n)}$. In general, if all $H_{(j)}$ for $j > i$ are accepted, then test $H_{(i)}$ and reject it if

$$p_{(i)} \leq \begin{cases} \frac{\alpha}{n-i+1} & \text{if } 1 \leq i \leq k \\ \left(\frac{\gamma}{n-i+1} + \frac{1-\gamma}{n-k+1} \right) \alpha & \text{if } k < i \leq n \end{cases} \quad (9)$$

If $H_{(i)}$ is rejected, then reject all the remaining hypotheses; otherwise, proceed to the next step. Thus, until $n - k$ hypotheses are accepted, we apply the k -truncated Hochberg MTP, and then we switch to the untruncated Hochberg MTP.

Theorem 2

The k -truncated Hochberg MTP is an exact shortcut to the closed MTP, which rejects each intersection hypothesis $H(I)$ if, for at least one $i \in I$,

$$p_{(i)} \leq \begin{cases} \left(\frac{\gamma}{m-i+1} + \frac{1-\gamma}{n-k+1} \right) \alpha & \text{if } 1 \leq i \leq m, m \leq n - k \\ \left(\frac{\gamma}{m-i+1} + \frac{1-\gamma}{n-k+1} \right) \alpha & \text{if } m + 1 - (n - k) \leq i \leq m, m > n - k \\ \frac{\alpha}{m-i+1} & \text{if } 1 \leq i < m + 1 - (n - k), m > n - k \end{cases} \quad (10)$$

This k -truncated Hochberg MTP is more conservative than the closed MTP, which tests each $H(I)$ at level $\alpha(I)$ using critical constants $\alpha(I)/i$ ($i = 1, \dots, m$), where $\alpha(I)$ is given by (4). Hence, (5) is an upper bound on its exact error rate function.

3.5. *k*-Truncated Hommel multiple test procedure

For parallel gatekeeping, Brechenmacher *et al.* [28] proposed a truncated Hommel MTP. The following is its generalization to a *k*-truncated Hommel MTP.

- Step 1. Accept $H_{(n)}$ if

$$p_{(n)} > \left(\gamma + \frac{1 - \gamma}{n - k + 1} \right) \alpha;$$

otherwise, reject all hypotheses and stop.

- Step $i = 2, \dots, n - k$. Accept $H_{(n-i+1)}$ and go to the next step if

$$p_{(n-i+j)} > \left(\frac{j\gamma}{i} + \frac{1 - \gamma}{n - k + 1} \right) \alpha \quad (j = 1, \dots, i);$$

otherwise, stop and reject any hypothesis $H_{(j)}$ satisfying

$$p_{(j)} \leq \left(\frac{\gamma}{i - 1} + \frac{1 - \gamma}{n - k + 1} \right) \alpha.$$

- Step $i = n - k + 1, \dots, n$. Accept $H_{(n-i+1)}$ and go to the next step if

$$p_{(n-i+j)} > \frac{j\alpha}{i} \quad (j = 1, \dots, i);$$

otherwise, stop and reject any hypothesis $H_{(j)}$ satisfying

$$p_{(j)} \leq \frac{\alpha}{i - 1}.$$

Theorem 3

The *k*-truncated Hommel MTP is an exact shortcut to the closed MTP, which rejects each intersection hypothesis $H(I)$ if, for at least one $i \in I$,

$$p_{(i)} \leq \begin{cases} \left(\frac{i\gamma}{m} + \frac{1 - \gamma}{n - k + 1} \right) \alpha & \text{if } 1 \leq m \leq n - k \\ \frac{i\alpha}{m} & \text{if } m > n - k \end{cases} \quad (11)$$

This *k*-truncated Hommel MTP is more conservative than the closed MTP, which tests each $H(I)$ at level $\alpha(I)$ using the Simes test, where $\alpha(I)$ is given by (4). Hence, (5) is an upper bound on its exact error rate function.

Simple algebra shows that, for the same *k* and γ , the critical constants used to compare $p_{(i)}$ with are ordered from the largest to the smallest, respectively, in the tests (11), (10), and (7) of the intersection hypothesis $H(I)$. Hence, the truncated Hommel MTP is more powerful than the truncated Hochberg MTP, which in turn is more powerful than the truncated Holm MTP.

4. Generalized multistage multiple test procedure for *k*-out-of-*n* gatekeeping

In this section, we apply the *k*-truncated MTPs derived in the previous section to construct a generalization of the multistage MTP proposed in [21] for parallel gatekeeping to *k*-out-of-*n* gatekeeping. Suppose that the hypotheses, H_1, \dots, H_n , are grouped into *m*-ordered families, F_1, \dots, F_m . Let $F_i = \{H_j, j \in N_i\}$ consist of n_i hypotheses where $N_1 = \{1, \dots, n_1\}$ and

$$N_i = \{n_1 + \dots + n_{i-1} + 1, \dots, n_1 + \dots + n_i\} \quad (i = 2, \dots, m).$$

Suppose further that, for each F_i ($i = 1, \dots, m - 1$), a k_i -out-of- n_i gatekeeping condition is specified where $1 \leq k_i \leq n_i$. We require an MTP that satisfies the strong FWER control requirement (1) for the overall family $F = \bigcup_{i=1}^m F_i$ subject to the specified gatekeeping conditions.

We propose the following multistage MTP. Let $\alpha_1 = \alpha$. At the i th stage ($i = 1, \dots, m - 1$), test the hypotheses in family F_i using any k_i -separable MTP \mathcal{P}_i at level α_i . Let $A_i \subseteq N_i$ be the index set of the accepted hypotheses $H_j \in F_i$. Set

$$\alpha_{i+1} = \alpha_i - e_i(A_i | \alpha_i) \quad (i = 1, \dots, m - 1) \tag{12}$$

where $e_i(A_i | \alpha_i)$ is the error rate function of \mathcal{P}_i . For $i = 1, \dots, m - 1$, if $\alpha_{i+1} > 0$, then proceed to test family F_{i+1} at level α_{i+1} using a k_{i+1} -separable MTP \mathcal{P}_{i+1} . For F_m , use any non-separable MTP \mathcal{P}_m at level α_m if $\alpha_m > 0$. If $\alpha_{i+1} = 0$, then stop testing and accept all the remaining hypotheses. Because \mathcal{P}_i is k_i -separable, $\alpha_{i+1} = 0$ iff $|A_i| > n_i - k_i$, that is, iff less than k_i hypotheses in family F_i are rejected.

Theorem 4

The generalized multistage MTP stated earlier satisfies the strong FWER control requirement (1).

5. Adjusted p -values

For simplicity, we consider two families, $F_1 = \{H_1, \dots, H_{n_1}\}$ and $F_2 = \{H_{n_1+1}, \dots, H_{n_1+n_2}\}$, with F_1 being a k_1 -out-of- n_1 gatekeeper for F_2 . First, let us focus on the adjusted p -values for primary hypotheses. For k_1 -truncated Holm MTP and Hochberg MTP, the adjusted p -values are obtained by simply extending their corresponding formulae for the untruncated case; see, for example, [29, p. 68]. Thus, the adjusted p -values $\tilde{p}_{(i)}$ for hypotheses $H_{(i)}$ ($i = 1, \dots, n_1$) for the k_1 -truncated Holm MTP with the truncation parameter γ are given by

$$\tilde{p}_{(i)} = \begin{cases} \min\{1, n_1 p_{(1)}\} & \text{if } i = 1 \\ \min\{1, \max\{\tilde{p}_{(i-1)}, (n_1 - i + 1)p_{(i)}\}\} & \text{if } i = 2, \dots, k_1 - 1 \\ \min\left\{1, \max\left\{\tilde{p}_{(i-1)}, p_{(i)} / \left(\frac{\gamma}{n_1 - i + 1} + \frac{1 - \gamma}{n_1 - k_1 + 1}\right)\right\}\right\} & \text{if } i = k_1, \dots, n_1 \end{cases} \tag{13}$$

Similarly, the adjusted p -values for the k_1 -truncated Hochberg MTP are given by

$$\tilde{p}_{(i)} = \begin{cases} \min\left\{1, p_{(n_1)} / \left(\frac{\gamma}{n_1 - i + 1} + \frac{1 - \gamma}{n_1 - k_1 + 1}\right)\right\} & \text{if } i = n_1 \\ \min\left\{\tilde{p}_{(i+1)}, p_{(i)} / \left(\frac{\gamma}{n_1 - i + 1} + \frac{1 - \gamma}{n_1 - k_1 + 1}\right)\right\} & \text{if } i = n_1 - 1, \dots, k_1 \\ \min\{\tilde{p}_{(i+1)}, (n_1 - i + 1)p_{(i)}\} & \text{if } i = k_1 - 1, \dots, 1 \end{cases} \tag{14}$$

Wright [30] gave an algorithm for computing the adjusted p -values for the untruncated Hommel MTP. However, that algorithm seems difficult to extend to the k -truncated Hommel MTP, so we computed the adjusted p -values by numerical search by a method analogous to the one given in [21].

Next, we derive expressions for the adjusted p -values \tilde{p}_j of secondary hypotheses H_j ($j = n_1 + 1, \dots, n_1 + n_2$) given the adjusted p -values \tilde{p}_i of primary hypotheses H_i ($i = 1, \dots, n_1$). First, consider the simplest case where the primary family is a serial gatekeeper ($k_1 = n_1$). In that case, we use untruncated MTPs in both the primary and the secondary families. In fact, because all primary hypotheses must be rejected to pass the primary gate, the adjusted p -value for the primary gate equals $p^* = \max_{i=1, \dots, n_1} \tilde{p}_{(i)}$. In particular, for the untruncated Hochberg and the Hommel MTPs, $p^* = \tilde{p}_{(n_1)} = p_{(n_1)}$. To define the adjusted p -values for the secondary hypotheses, let \bar{p}_j denote the adjusted p -value of a secondary hypothesis H_j , adjusted for the MTP \mathcal{P}_2 used within F_2 but not adjusted for primary gatekeeping. Then, we have

$$\tilde{p}_j = \max(p^*, \bar{p}_j), \quad (j = n_1 + 1, \dots, n_1 + n_2).$$

Next, consider the case $1 \leq k_1 < n_1$. Let $r_1 = \#\{\tilde{p}_i \leq \alpha \mid i = 1, \dots, n_1\}$ denote the number of rejected hypotheses in F_1 . We begin with the observation that if $r_1 < k_1$, then all secondary hypotheses H_j are accepted without a test, and if $r_1 \geq k_1$, then H_j is rejected iff $\bar{p}_j \leq \alpha_2$, which, by substituting from (8), is equivalent to

$$\frac{(n_1 - k_1 + 1)\bar{p}_j}{(1 - \gamma)(r_1 - k_1 + 1)} \leq \alpha.$$

Note that \tilde{p}_j is the smallest α that results in r_1 rejections in F_1 and satisfies the aforementioned inequality. Determination of this smallest α requires careful checking of the conditions that α must satisfy. This is performed in the following theorem.

Theorem 5

If $1 \leq k_1 < n_1$, then the adjusted p -values \tilde{p}_j of the secondary hypotheses H_j ($j = n_1 + 1, \dots, n_1 + n_2$) for different ranges of the values of \bar{p}_j are given by the following expressions:

$$\tilde{p}_j = \tilde{p}_{(k_1)} \text{ if } \bar{p}_j \leq \frac{(1-\gamma)\tilde{p}_{(k_1)}}{n_1 - k_1 + 1} \tag{15}$$

If

$$\frac{(1-\gamma)\tilde{p}_{(k_1)}}{n_1 - k_1 + 1} < \bar{p}_j < \frac{(1-\gamma)(n_1 - k_1)\tilde{p}_{(n_1-1)}}{n_1 - k_1 + 1}$$

then define r_1 as the smallest value of r ($r = k_1, \dots, n_1 - 2$) that satisfies the inequality

$$\frac{(1-\gamma)(r - k_1 + 1)\tilde{p}_{(r)}}{n_1 - k_1 + 1} \leq \bar{p}_j < \frac{(1-\gamma)(r - k_1 + 2)\tilde{p}_{(r+1)}}{n_1 - k_1 + 1} \tag{16}$$

Then \tilde{p}_j is given by

$$\tilde{p}_j = \begin{cases} \frac{(n_1 - k_1 + 1)\bar{p}_j}{(1-\gamma)(r_1 - k_1 + 1)} & \text{if } \frac{(1-\gamma)(r_1 - k_1 + 1)\tilde{p}_{(r_1)}}{n_1 - k_1 + 1} \leq \bar{p}_j < \frac{(1-\gamma)(r_1 - k_1 + 1)\tilde{p}_{(r_1+1)}}{n_1 - k_1 + 1} \\ \tilde{p}_{(r_1+1)} & \text{if } \frac{(1-\gamma)(r_1 - k_1 + 1)\tilde{p}_{(r_1+1)}}{n_1 - k_1 + 1} \leq \bar{p}_j < \frac{(1-\gamma)(r_1 - k_1 + 2)\tilde{p}_{(r_1+1)}}{n_1 - k_1 + 1} \end{cases} \tag{17}$$

Otherwise, if

$$\bar{p}_j \geq \frac{(1-\gamma)(n_1 - k_1)\tilde{p}_{(n_1-1)}}{n_1 - k_1 + 1}$$

then

$$\tilde{p}_j = \begin{cases} \frac{(n_1 - k_1 + 1)\bar{p}_j}{(1-\gamma)(n_1 - k_1)} & \text{if } \frac{(1-\gamma)(n_1 - k_1)\tilde{p}_{(n_1-1)}}{n_1 - k_1 + 1} \leq \bar{p}_j < \frac{(1-\gamma)(n_1 - k_1)\tilde{p}_{(n_1)}}{n_1 - k_1 + 1} \\ \tilde{p}_{(n_1)} & \text{if } \frac{(1-\gamma)(n_1 - k_1)\tilde{p}_{(n_1)}}{n_1 - k_1 + 1} \leq \bar{p}_j < \tilde{p}_{(n_1)} \\ \bar{p}_j & \text{if } \bar{p}_j \geq \tilde{p}_{(n_1)} \end{cases} \tag{18}$$

Remark 1

Contrary to what one might expect, the aforementioned expressions do not simplify for parallel gatekeeping because one must begin by checking (15) for $k_1 = 1$. Dmitrienko *et al.* [21] suggested a numerical search method in this case to find the smallest α to reject each H_j ($j = n_1 + 1, \dots, n_1 + n_2$). The closed formulae given here offer an easier method of calculation. If $k_1 = n_1 - 1$, then the \tilde{p}_j are given by (18).

6. Example

We return to the rheumatoid arthritis trial example given in Section 1. There are four primary endpoints: physician global assessment (H_1), patient global assessment (H_2), swollen joint count (H_3), and painful joint count (H_4). Aaltonen *et al.* [31] mentioned multiple secondary endpoints, but we will consider only one: American College of Rheumatology 50% improvement at 6 months (H_5). Thus, $F_1 = \{H_1, H_2, H_3, H_4\}$ and $F_2 = \{H_5\}$. As stated in Food and Drug Administration guidance [12], we will assume that at least three out of four primary endpoints must demonstrate efficacy; that is, $k_1 = 3$. Suppose that the raw p -values for the five hypotheses are as follows:

$$p_1 = 0.01, p_2 = 0.02, p_3 = 0.024, p_4 = 0.04, p_5 = 0.01,$$

and $\alpha = 0.05$.

We will compare truncated Holm (T-HOLM), truncated Hochberg (T-HOCH), and truncated Hommel (T-HOML) procedures where all truncated MTPs use $\gamma = 0.5$ and $k_1 = 3$. In this setting, Li's method

is the same as the T-HOLM. The critical constants for each hypothesis for the T-HOLM and T-HOCH procedures are given in Table I. The critical matrix for T-HOML is as given in Table II. Because we used the untruncated MTP to test H_1, H_2, H_3 , the critical constant of the respective truncated MTP differs only for H_4 . This critical constant is

$$\left(\frac{0.5}{4-4+1} + \frac{0.5}{4-3+1} \right) 0.05 = 0.0375,$$

whereas for the untruncated MTPs, it is 0.05.

First, we apply T-HOLM to F_1 . In Step 1, H_1 is rejected because $p_1 = 0.01 \leq 0.0125$, but H_2 cannot be rejected because $p_2 = 0.02 > 0.0167$. So H_3 and H_4 are also accepted. No significance level will be transferred to F_2 because only H_1 is rejected by T-HOLM, and thus, H_5 is also accepted without test.

Next, we apply T-HOCH to F_1 . In Step 1, H_4 is accepted because $p_4 = 0.04 > 0.0375$, but H_3 is rejected because $p_3 = 0.024 \leq 0.025$. So H_2 and H_1 are also rejected. Because T-HOCH rejects H_3, H_2 , and H_1 , using (12) and (5), we calculate the significance level transferred to F_2 as

$$\alpha_2 = \alpha - \left[\gamma + \frac{(1-\gamma)|A_1|}{n_1 - k_1 + 1} \right] \alpha = 0.05 - \left[0.5 + \frac{0.5 \times 1}{4} \right] (0.05) = 0.0125.$$

Because $p_5 = 0.01 \leq \alpha_2 = 0.0125$, T-HOCH rejects H_5 .

Finally, we apply T-HOML to F_1 . In Step 1, H_4 is accepted because $p_4 = 0.04 > 0.0375$, but it rejects all the remaining hypotheses with p -values less than 0.0375 because $p_3 = 0.024 \leq 0.025$. So it rejects H_3, H_2 , and H_1 . Because T-HOML makes the same rejections as T-HOCH, it also rejects H_5 .

Table III shows the adjusted p -values of all hypotheses for the three MTPs calculated using the formulae (13) and (14) for the primary hypotheses and Theorem 5 for the secondary hypotheses. Note that they agree with the rejection decisions obtained by stepwise applications of these MTPs at $\alpha = 0.05$. In addition, adjusted p -values of T-HOML are less than or equal to those of T-HOCH, although they lead to the same rejections. Figure 1 shows \tilde{p}_5 as a function of \bar{p}_5 for this example. Note that the three MTPs share the same figure because they differ only in the adjusted p -values for the primary hypotheses.

In summary, T-HOLM is not able to win on the 3-out-of-4 win criterion on the primary endpoints and hence does not go on to test the secondary endpoint. T-HOCH and T-HOML win on the 3-out-of-4 win

Table I. Critical constants for truncated Holm and truncated Hochberg.

MTP	H_1	H_2	H_3	H_4
T-HOLM	0.0125	0.0167	0.025	0.0375
T-HOCH	0.0125	0.0167	0.025	0.0375

MTP, multiple test procedure; T-HOLM, truncated Holm; T-HOCH, truncated Hochberg.

Table II. Critical matrix for truncated Hommel.

$$C_{\text{T-HOML}} = \begin{pmatrix} 0.0375 & & & & \\ 0.0375 & 0.0250 & & & \\ 0.0500 & 0.0333 & 0.0167 & & \\ 0.0500 & 0.0375 & 0.0250 & 0.0125 & \end{pmatrix}$$

T-HOML, truncated Hommel.

Table III. Adjusted p -values.

MTP	H_1	H_2	H_3	H_4	H_5
T-HOLM	0.040*	0.060	0.060	0.060	0.060
T-HOCH	0.040*	0.048*	0.048*	0.054	0.048*
T-HOML	0.032*	0.040*	0.048*	0.054	0.048*

MTP, multiple test procedure; T-HOLM, truncated Holm; T-HOCH, truncated Hochberg; T-HOML, truncated Hommel.

* The adjusted p -values significant at the 0.05 level are marked with an asterisk.

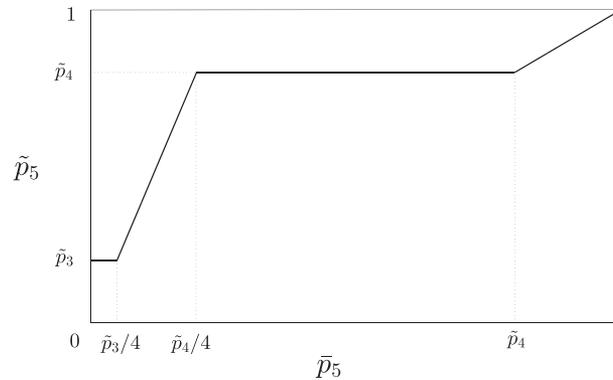


Figure 1. \tilde{p}_5 as a function of \bar{p}_5 .

criterion on the primary endpoints, and both declare a significant result on the secondary endpoint. Note that the adjusted p -values for T-HOLM, T-HOCH, and T-HOMM are in non-increasing order for each hypothesis corresponding to decreasing order of their powers noted previously.

7. Discussion

In this paper, we have extended the separability concept introduced in [21] for parallel gatekeeping to k -separability for $1 \leq k \leq n$. Using this concept, we have shown how to construct k -separable MTPs and incorporate them into a general multistage MTP thus providing a unified framework for any k -out-of- n gatekeeping problem ranging from parallel gatekeeping ($k = 1$) to serial gatekeeping ($k = n$). A novel feature of a k -separable MTP is that it is a hybrid MTP, which switches between a k -truncated MTP and an untruncated MTP. We have applied this general methodology to derive k -separable versions of the Holm, Hochberg, and Hommel MTPs. We have also derived explicit expressions for the adjusted p -values of the secondary hypotheses when the primary family is a k -out-of- n gatekeeper. Such expressions were not available before even for $k = 1$.

It is to be noted that k -out-of- n gatekeeping does not cover the most general types of logical restrictions in which testability of a secondary hypothesis depends upon whether certain individual primary hypotheses are rejected or not. Mixture procedures handle such logical restrictions via the so-called restriction functions, but the resulting closed procedures do not have multistage shortcuts.

Appendix

In this appendix, we give derivations of the k -truncated MTPs from Section 3 and proofs of their FWER control. We first review a result used in the sequel that characterizes the conditions under which closed MTPs have a step-down or a step-up shortcut [32].

As in Section 2, without loss of generality, we will denote any nonempty subset I of the index set $N = \{1, 2, \dots, n\}$ by $\{1, 2, \dots, m\}$. Let $p_{(1)} \leq \dots \leq p_{(m)}$ denote the ordered p -values and let $H_{(1)}, \dots, H_{(m)}$ denote the corresponding hypotheses. Let $H(I) = \bigcap_{i \in I} H_i$. Suppose that there exist critical constants $c_{m1} \geq \dots \geq c_{mm}$ such that

$$P \{p_{(1)} > c_{mm}\alpha, \dots, p_{(m)} > c_{m1}\alpha \mid H(I)\} \geq 1 - \alpha \tag{A.1}$$

Then the α -level local test of $H(I)$ is reject $H(I)$ if

$$p_{(m-i+1)} \leq c_{mi}\alpha \text{ for at least one } i = 1, \dots, m \tag{A.2}$$

We can determine the critical constants c_{mi} by assuming that $p_{(1)} \leq \dots \leq p_{(m)}$ are order statistics from a uniform $[0, 1]$ distribution under $H(I)$.

Let C denote an $n \times n$ lower-triangular matrix with entries c_{mi} ($i = 1, \dots, m, m = 1, \dots, n$). Liu [32] showed that the closure MTP based on (A.1) has a step-down shortcut if the row entries of C are equal and a step-up shortcut if the column entries are equal. For a step-down MTP, the critical constants for comparing the ordered p -values are given by the first column (c_{11}, \dots, c_{n1}) of C and for a step-up

Next, we show that the k -truncated Hommel MTP stated in Section 3.5 is a shortcut to the closed MTP (11). The proof is analogous to the one for the untruncated Hommel MTP [32]. Suppose that the k -truncated Hommel MTP stops at Step i ($1 \leq i \leq n - k + 1$) and rejects a hypothesis $H_{(j)}$ with

$$p_{(j)} \leq \left(\frac{\gamma}{i-1} + \frac{1-\gamma}{n-k+1} \right) \alpha,$$

where $j \in \{1, \dots, n - i + 1\}$. Then, it accepts $H_{(j+1)}, \dots, H_{(n)}$. It is easy to see that $\bigcap_{\ell=n-i+2}^n H_{(\ell)}$ is accepted by the closed procedure at the $(i-1)$ th step. For $\ell \in \{j+1, \dots, n-i+1\}$, $\left(\bigcap_{j=n-i+3}^n H_{(j)} \right) \cap H_{(\ell)}$ will be accepted because $p_{(\ell)} \geq \left(\frac{\gamma}{i-1} + \frac{1-\gamma}{n-k+1} \right) \alpha$. Thus, $H_{(j+1)}, \dots, H_{(n)}$ are accepted by the closed procedure. In contrast, each $H_{(J)} = \bigcap_{(\ell) \in J} H_{(\ell)}$, with $J = \{(i_1), \dots, (i_r)\}$, $\{i_1, \dots, i_{r_0}\} \subseteq \{1, \dots, j\}$, and $\{i_{r_0}, \dots, i_r\} \subseteq \{j+1, \dots, n\}$ ($1 \leq r_0 \leq r$), will be rejected because

$$\max_{1 \leq q \leq r_0} (p_{(i_q)}) \leq p_{(j)} \leq \left(\frac{\gamma}{i-1} + \frac{1-\gamma}{n-k+1} \right) \alpha \leq \left(\frac{\gamma}{r-r_0} + \frac{1-\gamma}{n-k+1} \right) \alpha$$

by noting that $r - r_0 \leq i - 1$. So individual hypotheses $H_{(1)}, \dots, H_{(j)}$ will be rejected. Thus, the two methods give the same rejections when the k -truncated Hommel MTP stops at the i th step, $i = 1, \dots, n - k + 1$. The same result can be shown when $i = n - k + 2, \dots, n$. \square

Proof of Theorem 4

The proof of Proposition 4.1 from [21] applies here without any change because it does not utilize the fact that the multistage MTP is for parallel gatekeeping ($k = 1$). It only depends on the definitions of the FWER and the error rate function, which are the same for k -out-of- n gatekeeping. \square

Proof of Theorem 5

First, we prove (15). Consider increasing α beginning at 0. As long as $\alpha < \tilde{p}_{(k_1)}$, no secondary hypothesis can be rejected because $r < k_1$. The smallest α at which both $r = k_1$ and H_j are rejected, that is,

$$\bar{p}_j \leq \alpha_2 = \frac{(1-\gamma)\alpha}{n_1 - k_1 + 1},$$

is $\alpha = \tilde{p}_{(k_1)}$, and so $\tilde{p}_j = \tilde{p}_{(k_1)}$.

Next, let us prove (17). Let r_1 denote the smallest r ($k_1 \leq r \leq n_1 - 2$), which simultaneously satisfies $\tilde{p}_{(r)} \leq \alpha < \tilde{p}_{(r+1)}$ (so that r primary hypotheses are rejected) and

$$\bar{p}_j \leq \alpha_2 = \frac{(1-\gamma)(r - k_1 + 1)\alpha}{n_1 - k_1 + 1}.$$

Combination of these two inequalities is

$$\frac{(1-\gamma)(r - k_1 + 1)\tilde{p}_{(r)}}{n_1 - k_1 + 1} \leq \bar{p}_j < \frac{(1-\gamma)(r - k_1 + 1)\tilde{p}_{(r+1)}}{n_1 - k_1 + 1} \tag{A.7}$$

which is (16). Substituting $r = r_1$ in the aforementioned inequality and solving for α give the first equation in (17).

If we increase α from $\tilde{p}_{(r_1)}$ to $\tilde{p}_{(r_1+1)}$ so that $r_1 \rightarrow r_1 + 1$, then the inequality (A.7) becomes

$$\frac{(1-\gamma)(r_1 - k_1 + 2)\tilde{p}_{(r_1+1)}}{n_1 - k_1 + 1} \leq \bar{p}_j < \frac{(1-\gamma)(r_1 - k_1 + 2)\tilde{p}_{(r_1+2)}}{n_1 - k_1 + 1} \tag{A.8}$$

We need to determine \tilde{p}_j when \bar{p}_j is between the upper limit of (A.7) and the lower limit of (A.8), that is, when

$$\frac{(1-\gamma)(r_1 - k_1 + 1)\tilde{p}_{(r_1+1)}}{n_1 - k_1 + 1} \leq \bar{p}_j < \frac{(1-\gamma)(r_1 - k_1 + 2)\tilde{p}_{(r_1+1)}}{n_1 - k_1 + 1}.$$

In this case, the number of rejected primary hypotheses equals $r_1 + 1$, and so the condition for rejection of H_j is

$$\bar{p}_j \leq \alpha_2 = \frac{(1 - \gamma)(r_1 - k_1 + 2)\alpha}{n_1 - k_1 + 1},$$

and the smallest α for which this inequality is satisfied is $\tilde{p}_j = \tilde{p}_{(r_1+1)}$, which is the second equation in (17).

To prove (18), we set $r_1 = n_1 - 1$. The proofs of the first two equations are similar to those given earlier of the two equations in (17) except that the upper limit of the range of \bar{p}_j is $\tilde{p}_{(n_1)}$ in the second equation. This is because if $\bar{p}_j \geq \tilde{p}_{(n_1)}$, then by setting $\alpha = \bar{p}_j$, we get $r = n_1$ because $\tilde{p}_{(n_1)} \leq \bar{p}_j = \alpha$. Because all n_1 primary hypotheses are rejected, we have $\alpha_2 = \alpha = \bar{p}_j$, which is the smallest value of α at which H_j is rejected. This proves the last equation of (18). \square

Acknowledgements

We thank the editor and two referees for very useful comments that helped to improve the article. This paper forms a part of the first author's doctoral dissertation under the second author's supervision in the Department of Statistics at Northwestern University, U.S.A.

References

1. Maurer W, Hothorn L, Lehmacher E. Multiple comparisons in drug clinical trials and preclinical assays: a priori ordered hypotheses. In *Biometrie in der Chemisch-in-Pharmazeutischen Industrie*, Vol. 6, Vollman J (ed.). Stuttgart: Fischer-Verlag, 1995; 3–18.
2. Westfall PH, Krishen A. Optimally weighted, fixed-sequence, and gatekeeping multiple testing procedures. *Journal of Statistical Planning and Inference* 2001; **99**:25–40.
3. Dmitrienko A, Offen WW, Westfall PH. Gatekeeping strategies for clinical trials that do not require all primary effects to be significant. *Statistics in Medicine* 2003; **22**:2387–2400. DOI: 10.1002/sim.1526.
4. Dmitrienko A, Wiens BL, Tamhane AC, Wang X. Tree-structured-structured gatekeeping tests in clinical trials with hierarchically ordered multiple objectives. *Statistics in Medicine* 2007; **26**:2465–2478. DOI: 10.1002/sim.2716.
5. Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 1976; **63**:655–660.
6. Dmitrienko A, Kordzakhia G, Tamhane AC. Multistage and mixture parallel gatekeeping procedures for clinical trials. *Journal of Pharmaceutical Statistics* 2011; **21**:726–747. DOI: 10.1080/10543406.2011.551333.
7. Dmitrienko A, Tamhane AC. Mixtures of multiple testing procedures for gatekeeping applications in clinical trial applications. *Statistics in Medicine* 2011; **30**:1473–1488. DOI: 10.1002/sim.2716.
8. Dmitrienko A, Tamhane AC. General theory of mixture procedures for gatekeeping. *Biometrical Journal* 2013; **55**(3):402–419. DOI: 10.1002/bimj.201100258.
9. Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Statistics in Medicine* 2009; **28**:586–604. DOI: 10.1002/sim.3495.
10. Bretz F, Posch M, Glimm E, Klinglmueller F, Maurer W, Rohmeyer K. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biometrical Journal* 2011; **53**:894–913. DOI: 10.1002/bimj.201000239.
11. Maurer W, Bretz F. Memory and other properties of multiple test procedures generated by entangled graphs. *Statistics in Medicine* 2013; **32**:1739–1753. DOI: 10.1002/sim.5711.
12. Food and Drug Administration. *Guidance for Industry—Clinical Development Programs for Drugs, Devices, and Biological Products for the Treatment of Rheumatoid Arthritis (RA)*, U.S. Department of Health and Human Services, 1999.
13. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 1979; **6**:65–70.
14. Hochberg Y. A sharper Bonferroni procedure for multiple significance testing. *Biometrika* 1988; **75**:800–802.
15. Hommel G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* 1988; **75**:383–386.
16. Wiens BL. A fixed-sequence Bonferroni procedure for testing multiple endpoints. *Pharmaceutical Statistics* 2003; **2**:211–215. DOI: 10.1002/pst.64.
17. Wiens BL, Dmitrienko A. The fallback procedure for evaluating a single family of hypotheses. *Journal of Biopharmaceutical Statistics* 2005; **15**:929–942. DOI: 10.1080/10543400500265660.
18. Dunnett CW. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* 1995; **50**:1096–1121.
19. Dmitrienko A, Tamhane AC, Wang X, Chen X. Stepwise gatekeeping procedures in clinical trial applications. *Biometrical Journal* 2006; **48**:984–991.
20. Guilbaud O. Bonferroni parallel gatekeeping—transparent generalizations, adjusted p -values, and short direct proofs. *Biometrical Journal* 2007; **49**:917–927. DOI: 10.1002/bimj.200610361.
21. Dmitrienko A, Tamhane AC, Wiens BL. General multistage gatekeeping procedures. *Biometrical Journal* 2008; **50**:667–677. DOI: 10.1002/bimj.200710464.

22. Simes RJ. An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 1986; **73**:751–754.
23. Sarkar SK, Chang CK. The Simes method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Association* 1997; **92**:1601–1608. DOI: 10.1080/01621459.1997.10473682.
24. Sarkar SK. Probability inequalities for ordered MTP2 random variables: a proof of the Simes conjecture. *The Annals of Statistics* 1998; **26**:494–504. DOI: 10.1214/aos/1028144846.
25. Hochberg Y, Tamhane AC. *Multiple Comparison Procedures*. Wileys: New York, 1987.
26. Li J. A simple and flexible Holm gatekeeping procedure. *Biometrical Journal* 2011; **53**:797–809. DOI: 10.1002/bimj.201000040.
27. Hommel G, Bretz F, Maurer W. Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies. *Statistics in Medicine* 2007; **26**:4063–4073. DOI: 10.1002/sim.2873.
28. Brechenmacher T, Xu J, Dmitrienko A, Tamhane AC. A mixture gatekeeping procedure based on the Hommel test for clinical trial applications. *Journal of Pharmaceutical Statistics* 2011; **21**:748–767. DOI: 10.1080/10543406.2011.551334.
29. Dmitrienko A, Tamhane AC, Bretz F. *Multiple Testing Problems in Pharmaceutical Statistics*. Taylor & Francis/CRC Press: Boca Raton, Florida, 2009.
30. Wright SP. Adjusted p-values for simultaneous inference. *Biometrics* 1992; **48**:1005–1013.
31. Aaltonen KJ, Virkki LM, Malmivaara A, Kontinen YT, Nordström DC, Blom M. Systematic review and meta-analysis of the efficacy and safety of existing TNF blocking agents in treatment of rheumatoid arthritis. *PLoS ONE* 2012; **7**(1):e30275. DOI: 10.1371/journal.pone.0030275.
32. Liu W. Multiple tests of a non-hierarchical finite family of hypotheses. *Journal of Royal Statistical Society, Series B* 1996; **58**:455–461.