A Mixture Gatekeeping Procedure Based on the Hommel Test for Clinical Trial Applications

Thomas Brechenmacher (Dainippon Sumitomo Pharma Co., Ltd.)
Jane Xu (Sunovion Pharmaceuticals Inc.)
Alex Dmitrienko (Eli Lilly and Company)
Ajit C. Tamhane (Northwestern University)

December 9, 2010

Abstract

When conducting clinical trials with hierarchically ordered objectives, it is essential to use multiplicity adjustment methods that control the familywise error rate in the strong sense while taking into account the logical relations among the null hypotheses. This paper proposes a gatekeeping procedure based on the Hommel (1988) test which offers power advantages compared to other p-value based tests proposed in the literature. A general description of the procedure is given and details are presented on how it can be applied to complex clinical trial designs. Two clinical trial examples are given to illustrate the methodology developed in the paper.

Keywords: Multiple testing; Gatekeeping procedure; Hommel procedure; Closed testing; Clinical trials; Hierarchical objectives.

1 Introduction

Gatekeeping procedures are often used to address multiplicity issues in clinical trials with hierarchically ordered objectives, e.g., comparisons of multiple doses with placebo on the primary endpoint followed by the same comparisons on the secondary endpoint. Testing multiple null hypotheses may result in substantial inflation of the familywise error rate (FWER), i.e. the probability of erroneously rejecting at least one true null hypothesis. It is essential to ensure that a multiple testing procedure (MTP) controls the FWER in the strong sense (Hochberg and Tamhane, 1987), i.e. under any configuration of the true and false null hypotheses, at a preassigned level of significance α.

Over the past decade, many gatekeeping procedures (Dmitrienko and Tamhane, 2009) have been proposed to address the problem of FWER control when the null hypotheses are hierarchically ordered with logical relations. The null hypotheses are grouped into families and tested in a sequential manner beginning with the most important family, e.g., the family of hypotheses on the primary endpoint. Inferences in each family depend on the acceptance or rejection of null hypotheses in the earlier families. Conversely, acceptance or rejection of a null hypothesis in a given family may not depend on the results in later families — the so-called *independence condition*.

Different types of logical gatekeeping constraints among the families of hypotheses have been studied including *serial gatekeeping* (Maurer et al. 1995, Bauer et al. 1998, Westfall and Krishen 2001), *parallel gatekeeping* (Maurer et al. 1995, Dmitrienko, Offen and Westfall 2003) and their generalization referred to as *tree-structured gatekeeping* (Dmitrienko et al. 2007, 2008). Dmitrienko and Tamhane (2010a,b) recently proposed a general framework for constructing gatekeeping procedures, called the mixture method. This approach enables clinical trial sponsors to set up highly flexible decision rules involving complex logical restrictions between the null hypotheses while controlling the FWER in the strong sense.

To motivate the mixture gatekeeping procedures, consider the problem of testing the effect of a treatment at multiple dose levels versus a common control with respect to multiple endpoints. The endpoints are ordered by the importance of clinical objectives beginning with the primary endpoint. The null hypotheses corresponding to different endpoints are grouped into separate families.

Figure 1 illustrates this hypothesis testing problem using an example from a trial in patients with schizophrenia. The aim of the trial was to compare three doses (L: Low dose; M: Medium dose; H: High dose) of a novel atypical antipsychotic to a control for three ordered endpoints. In Figure 1, logical restrictions are represented by arrows. In this example, if H_1 fails to be rejected, H_4 and H_7 should be automatically accepted. This is done to reflect the fact that if the low dose cannot be shown to be superior to the control for the primary endpoint, there is no value in further testing

the efficacy of that dose for the two secondary endpoints. The trial's sponsor was interested in defining a gatekeeping procedure to help enrich the product label by including information on the significant treatment effects for the primary as well as both the secondary endpoints.

[Insert Figure 1 here]

This paper applies the mixture method using the Hommel (1988) MTP as a component procedure. The resulting mixture gatekeeping procedure will be referred to as the *Hommel-based gatekeeping procedure*. The Hommel MTP is more powerful than other p-value based MTPs proposed in the literature, e.g., Bonferroni, Holm (1979) and Hochberg (1988) MTPs. Thus the new procedure proposed in this paper presents an obvious advantage over the existing p-value based gatekeeping procedures.

The paper is organized as follows: Section 2 presents the statement of the problem and gives a brief desciption of the general method for constructing mixture gatekeeping procedures. In Section 3, the Hommel-based gatekeeping procedure is defined for parallel gatekeeping and its strong control of the FWER is established. This procedure is generalized in Section 3.3 to tree-structured gatekeeping problems. In Section 4, two clinical trial examples are presented to illustrate the procedure. Finally, Section 5 offers brief conclusions. The proofs of theoretical results are given in the Appendix.

2 Mixture gatekeeping procedures

Consider a problem involving $n \geq 2$ null hypotheses $H_1, \ldots, H_n$ that are grouped into m families with n_i null hypotheses in the ith family ($n_1 + \ldots + n_m = n$). Specifically, let $F_i = \{H_j,\ j \in N_i\}$, $i = 1, \ldots, m$, $m \geq 2$, where

$$N_1 = \{1, \ldots, n_1\}, N_i = \{n_1 + \ldots + n_{i-1} + 1, \ldots, n_1 + \ldots + n_i\},\ i = 2, \ldots, m.$$

Further consider closed testing procedures (Marcus, Peritz and Gabriel, 1976) $\mathcal{P}_i$, $i = 1, \ldots, m$, known as *component procedures*, that provide *local FWER control*, i.e., $\mathcal{P}_i$ controls the FWER in the strong sense at any preassigned level α within F_i. Our goal is to build a mixture of the component procedures $\mathcal{P}_i$, $i = 1, \ldots, m$, which will be denoted by $\mathcal{P}$, using the methodology proposed in Dmitrienko and Tamhane (2010a,b). This gatekeeping procedure $\mathcal{P}$ will be a closed testing procedure with strong *global FWER control* within the combined family of null hypotheses $F = F_1 \cup \ldots \cup F_m$. Global FWER control helps the trial's sponsor justify inclusion of all significant findings with respect to the primary and secondary objectives in the product label.

Before we give the method of constructing mixture procedures, we need to introduce the key concepts of the error rate function and separability.

2.1 Error rate function and separability

To simplify the notation, in this section we will consider the problem of testing a single family of null hypotheses $F = \{H_1, \ldots, H_n\}$. Denote the marginal (raw) p-values, obtained from appropriate statistical tests, associated with these null hypotheses by $p_1, \ldots, p_n$.

Consider the index set $N = \{1, \ldots, n\}$ and a subset $I \subseteq N$. Further, consider an MTP for testing the null hypotheses $H_1, \ldots, H_n$ which controls the FWER in the strong sense within this family at level α. The error rate function of this MTP is defined as the maximum probability of making at least one Type I error when the null hypotheses H_i, $i \in I$, are true and the remaining null hypotheses are false (Dmitrienko, Tamhane and Wiens, 2008). The error rate function is denoted by $e(I|\alpha)$ and formally defined as

$$e(I|\alpha) = \sup P\left\{\bigcup_{i\in I}(\text{Reject } H_i)\middle| H(I)\right\},$$

where the supremum of the probability is computed over the entire parameter space corresponding to the intersection hypothesis $H(I) = \cap_{i\in I} H_i$.

Generally, an easily computable upper bound on $e(I|\alpha)$ is used since an exact expression is difficult to obtain. We will treat the upper bound itself as the actual error rate function and denote it by the same notation $e(I|\alpha)$. The following natural conditions are generally satisfied by the error rate functions of standard MTPs (or they are easy to enforce if violated).

- $e(\emptyset|\alpha) = 0$, i.e., no Type I error is committed when all null hypotheses are false.
- $e(I|\alpha) \leq e(J|\alpha)$ if $I \subseteq J \subseteq N$, i.e., the type I error rate is a monotone function of the set of true null hypotheses.
- $e(N|\alpha) = \alpha$, i.e., the Type I error rate is maximized when all null hypotheses are true and equals the nominal significance level α.

The error rate function is used to define the minimum fraction of α remaining after testing the null hypotheses in a given family and which can be carried over to the next family. To ensure that the fraction of α carried over to the next family is positive when some null hypotheses are rejected in the current family, the MTP used to test the family needs to be *separable*. An MTP is said to be separable if its error rate function is strictly less than (separates from) α when not all null hypotheses are true, i.e., $e(I|\alpha) < \alpha$ if I is a proper subset of N.

We are now ready to introduce mixture procedures.

2.2 Mixture gatekeeping procedures

Consider the setup introduced at the beginning of this section and further assume that the component procedures $\mathcal{P}_1, \ldots, \mathcal{P}_{m-1}$ are separable. Each $\mathcal{P}_i$ is a closed procedure and thus there exists a *local test* of every non-empty intersection hypothesis $H(I_i) = \cap_{j \in I_i} H_j$ where $I_i \subseteq N_i$. Let $p_i(I_i)$ denote the *local p-value* for testing $H(I_i)$ using $\mathcal{P}_i$.

Consider the combined family of null hypotheses $F = F_1 \cup \ldots \cup F_m$. The mixture procedure is based on the closure principle (Marcus et al. 1976) to control the FWER for this family at a preassigned level α. Thus it consists of α-level local tests for all non-empty intersection hypotheses $H(I) = \cap_{j \in I} H_j$, where $I \subseteq N$ and $N = \{1, \ldots, n\}$.

Suppose the intersection hypothesis $H(I)$ contains null hypotheses from k families with $1 \leq k \leq m$, i.e.,

$$I = I_{i_1} \cup \cdots \cup I_{i_k}, \quad \text{where } I_{i_j} \subseteq N_{i_j},\ j = 1, \ldots, k,$$

and $I_{i_1}, \ldots, I_{i_k}$ are all non-empty. To simplify the notation, we will denote $I_{i_1}, \ldots, I_{i_k}$ by $I_1, \ldots, I_k$ even though the indices $\{i_1, \ldots, i_k\}$ may not be sequential starting with 1. This notation will be adopted throughout this paper and will not be mentioned again.

The local α-level test of $H(I)$ is based on a function of $p_1(I_1), \ldots, p_k(I_k)$, called the *mixing function*, denoted by $\phi_I(p_1(I_1), \ldots, p_k(I_k))$. The range of this function is the interval $[0, 1]$ and it satisfies the following properties:

- Property 1: If $k = 1$ then $\phi_I(p_1(I_1), \ldots, p_k(I_k)) = p_1(I_1)$ and if $k > 1$ then $\phi_I(p_1(I_1), \ldots, p_k(I_k)) \leq p_1(I_1)$.
- Property 2: $P_{H(I)}\{\phi_I(p_1(I_1), \ldots, p_k(I_k)) \leq \alpha\} \leq \alpha$.

Then the local test of $H(I)$ is defined as reject $H(I)$ if

$$\phi_I(p_1(I_1), \ldots, p_k(I_k)) \leq \alpha. \tag{1}$$

Note that if $I_i \subset N_i$ for $i = 1, \ldots, j-1$ and $I_j = N_j$ for some $j \leq k$ then the above rejection rule does not depend on $p_{j+1}(I_{j+1}), \ldots, p_k(I_k)$. Property 2 guarantees that every such local test is an α-level test. Therefore $\mathcal{P}$ controls the FWER at level α because it is a closed procedure.

We will focus on a particular mixing function, called the *the Bonferroni mixing function*, which has the following general form for any specified $I = I_1 \cup \cdots \cup I_k$:

$$\phi_I(p_1(I_1), \ldots, p_k(I_k)) = \min\left(\frac{p_1(I_1)}{c_1(I|\alpha)}, \ldots, \frac{p_k(I_k)}{c_k(I|\alpha)}\right), \tag{2}$$

where the $c_i(I|\alpha)$ are the coefficients defined recursively as follows: Let $e_i(I_i|\alpha)$ be the error rate function of $\mathcal{P}_i$, $f_i(I_i|\alpha) = e_i(I_i|\alpha)/\alpha$ and $c_1(I|\alpha) = 1$. Then

$$c_i(I|\alpha) = c_{i-1}(I|\alpha)[1 - f_{i-1}(I_{i-1}|\alpha)] = \prod_{j=1}^{i-1}[1 - f_j(I_j|\alpha)],\ i = 2, \ldots, k. \tag{3}$$

The weight $c_i(I|\alpha)$ reflects the relative importance of family F_i within the intersection hypothesis $H(I)$. The hypotheses from the first family in the intersection hypothesis are the most important hypotheses within $H(I)$, so the weight on I_1 is set to $c_1(I|\alpha) = 1$. The importance of the other families is determined by computing the fraction of α carried over from the preceding families within $H(I)$. Specifically, recalling the definition of the error rate function, we see that $1 - f_i(I_i|\alpha)$ is the minimum fraction of α available for testing the null hypotheses in F_{i+1} when the null hypotheses H_j, $j \in I_i$, are accepted. If I_i is a proper subset of the index set N_i, the separability condition defined in Section 2.1 ensures that the weight $c_{i+1}(I|\alpha)$ associated to I_{i+1} is greater than zero.

Note that if all the error rate functions $e_i(I_i|\alpha)$ are proportional to α then all the $f_i(I_i|\alpha)$ as well as the $c_i(I|\alpha)$ defined in (3) are independent of α and so may be denoted simply by $f_i(I_i)$ and $c_i(I)$, respectively. Hence $\phi(I)$ defined in (2) is independent of α. Then the p-value, $p(I)$, corresponding to test (1) of $H(I)$, simply equals $\phi(I)$. Otherwise, $p(I)$ must be obtained numerically by finding the smallest α that satisfies the inequality (1).

Since the mixture procedure is a closed procedure, it rejects any individual hypothesis H_j iff all intersection hypotheses $H(I)$ with $j \in I$ are rejected by their respective α-level local tests. Therefore the adjusted p-value for each H_j in F can be defined as the maximum over the local p-values for the intersection hypotheses containing this null hypothesis, i.e.,

$$\widetilde{p}_j = \max_{I:j\in I} p(I). \tag{4}$$

The null hypothesis H_j is rejected by the mixture gatekeeping procedure iff $\widetilde{p}_j \leq \alpha$. Since this procedure is based on α-level local tests for all intersection hypotheses in F, the closure principle guarantees that the mixture gatekeeping procedure controls the global FWER in the strong sense at level α.

3 Hommel-based gatekeeping procedure

We will use the general method for constructing a mixture procedure to obtain the Hommel-based gatekeeping procedure, i.e., a mixture procedure consisting of

Hommel-type procedures as component procedures. However, the basic Hommel procedure is not separable (see the proof of Proposition 1 in the Appendix). Therefore, we use its truncated version for the first $m-1$ component procedures which we discuss first.

3.1 Truncated Hommel MTP

As in Section 2.1, for notation convenience, we will consider a single family of null hypotheses $F = \{H_1, \ldots, H_n\}$ with the corresponding marginal (raw) p-values, $p_1, \ldots, p_n$. Recall that the Hommel MTP is derived by applying the closure principle in which all intersection hypotheses $H(I) = \cap_{i \in I} H_i$ for $I \subseteq N$ are tested using the Simes (1986) test. Let k denote the number of elements in I and let $p_{(1)} \leq \cdots \leq p_{(k)}$ denote the ordered raw p-values associated with the null hypotheses H_i, $i \in I$. Then the α-level Simes test rejects $H(I)$ iff

$$p_{(i)} \leq \frac{i\alpha}{k} \text{ for at least one } i = 1, \ldots, k.$$

The truncated Hommel MTP employs convex combinations of the critical constants of the original Hommel MTP and the critical constants of the Bonferroni MTP in the closed procedure. Thus it rejects $H(I)$ at level α iff

$$p_{(i)} \leq \left(\frac{i\gamma}{k} + \frac{1-\gamma}{n} \right) \alpha \text{ for at least one } i = 1, \ldots, k, \tag{5}$$

where the *truncation fraction* γ is between 0 and 1. The local p-value for $H(I)$ is given by

$$p(I) = \min_{i=1,\ldots,k} \left[\frac{p_{(i)}}{i\gamma/k + (1-\gamma)/n} \right]. \tag{6}$$

In Proposition 1 of the Appendix, we show that the truncated Hommel MTP satisfies the separability condition when $0 \leq \gamma < 1$.

For $\gamma = 0$ this MTP reduces to the Bonferroni MTP while for $\gamma = 1$ it reduces to the regular Hommel MTP. Note that truncating the Hommel MTP incurs some power loss, which is smaller for higher values of γ. As is the case for the regular Hommel MTP, there is a stepwise shortcut to this truncated Hommel MTP which can be applied in a step-up manner as shown in the Appendix.

The truncated Hommel MTP controls the FWER in the strong sense under the same conditions as those necessary for the validity of the Simes test, i.e., when the test statistics are independent (Simes, 1986) or positively dependent (Sarkar and Chang, 1997 and Sarkar, 1998). The test statistics are said to be positively dependent if their joint distribution is multivariate totally positive of order two (Karlin 1968).

For example, the multivariate t-distribution with a common positive correlation satisfies this condition. Thus both the regular and truncated Hommel MTPs control the FWER in problems involving comparisons of multiple doses of a treatment to a common control in trials with a balanced design, i.e., all treatment groups have the same sample size which may be different than the control group sample size.

Upper bound on the error rate function

The following easy-to-compute upper bound on the error rate function of the truncated Hommel MTP is derived in Proposition 1 given in the Appendix:

$$e(I|\alpha,\gamma) = (\gamma + (1-\gamma)|I|/n)\alpha \text{ if } |I| > 0 \tag{7}$$

and $e(I|\alpha,\gamma) = 0$ if $|I| = 0$ where $|I|$ is the cardinality of the index set I. It is easy to see that $e(I|\alpha,\gamma)$ satisfies the conditions listed in section 2.1.

Note that $e(I|\alpha,\gamma)$ is proportional to α and hence $f(I|\alpha,\gamma) = e(I|\alpha,\gamma)/\alpha$ is independent of α; we therefore denote it by $f(I|\gamma)$. As noted above, if these upper bounds are used for the component truncated Hommel MTPs for $F_1,\ldots,F_{m-1}$ then we can define the p-value, $p(I)$, for every intersection hypothesis $H(I)$, to be simply equal to $\phi(I)$ given by (2) since all the $c_i(I|\alpha)$ are independent of α (and hence are denoted by $c_i(I)$).

An exact expression for the error rate function of the truncated Hommel MTP under the assumption of independent test statistics is given in the Appendix. We also explain there why we did not use it in the final implementation of the Hommel-based gatekeeping procedure.

3.2 Hommel-based gatekeeping procedure

The Hommel-based gatekeeping procedure uses the truncated Hommel MTP with the truncation fraction $0 \le \gamma_i < 1$ as the component MTP $\mathcal{P}_i$ $(i = 1,\ldots,m-1)$ and the regular Hommel MTP as $\mathcal{P}_m$. In clinical trial applications, the truncation fractions γ_i need to be prespecified at the planning stage. A high value of γ_i increases the power of $\mathcal{P}_i$ but penalizes the powers of $\mathcal{P}_j$ for $j > i$ since their weights $c_j(I)$ (here we are suppressing the dependence of $c_j(I)$ on $\gamma_1,\ldots,\gamma_{j-1}$ for notational convenience) are reduced. Appropriate values for the truncation fractions that maximize a desired power function can be determined by simulations based on the trial assumptions; see Section 4 for examples of power functions that can be used for selecting the truncation fractions. The component procedure $\mathcal{P}_m$ is the last one in the sequence and thus it does not need to be separable. Hence the regular Hommel MTP can be used instead of the truncated Hommel MTP in the last family to improve the overall power.

Specifically, for any index set $I = I_1 \cup \cdots \cup I_k$, the local p-values for the truncated Hommel MTP are given by

$$p_i(I_i) = \min_{j \in I_i} \left[\frac{p_{(j)}}{j\gamma_i/|I_i| + (1-\gamma_i)/n_i} \right], i = 1, \ldots, k$$

where $\gamma_i = 1$ for the regular Hommel MTP which is used for the last family F_m. Further, using the upper bound on the error rate function (7), we can write

$$f_i(I_i|\gamma_i) = \frac{e(I_i|\alpha, \gamma_i)}{\alpha} = \left(\gamma_i + \frac{(1-\gamma_i)|I_i|}{n_i} \right), i = 1, \ldots, k. \tag{8}$$

As noted before, the fractions $f_i(I_i|\gamma_i)$ are independent of α. Finally the coefficients $c_i(I)$ used to weigh the local p-values are given by $c_1(I) = 1$ and

$$c_i(I) = c_{i-1}(I)[1 - f_{i-1}(I_{i-1}|\gamma_{i-1})], i = 2, \ldots, k,$$

Finally, the p-value, $p(I)$, for testing $H(I)$ is obtained as follows.

- Case 1. If I_i is a proper subset of the index set N_i for all $i = 1, \ldots, k$,

$$p(I) = \phi(I) = \min \left(\frac{p_1(I_1)}{c_1(I)}, \ldots, \frac{p_k(I_k)}{c_k(I)} \right).$$

- Case 2. If I_i is a proper subset of N_i for $i = 1, \ldots, j-1$ and $I_j = N_j$ for some $j \leq k$,

$$p(I) = \phi(I) = \min \left(\frac{p_1(I_1)}{c_1(I)}, \ldots, \frac{p_j(I_j)}{c_j(I)} \right).$$

3.3 Logical restrictions

An important feature of gatekeeping procedures is that logical restrictions can be specified to account for clinically relevant logical relationships among the null hypotheses of interest. In this paper, logical restrictions will be formulated in terms of serial and parallel rejection sets that were defined in the context of tree gatekeeping procedures (Dmitrienko et al., 2007, 2008). Logical restrictions based on serial and parallel rejection sets can be extended to more general monotone logical restrictions (Dmitrienko and Tamhane, 2010) but this approach will not be pursued here.

For each null hypothesis H_j in F_i, $i = 2, \ldots, m$, we define the serial and parallel rejection sets, denoted by S_j and P_j, at least one of which is non-empty and which contain indices of the null hypotheses from $F_1, \ldots, F_{i-1}$. The null hypothesis H_j is said to be *testable* if all null hypotheses in S_j are rejected and at least one null

hypothesis in P_j is rejected; otherwise H_j is said to be *non-testable* and is accepted without a test.

We now show how to account for testability of hypotheses as specified through their serial and parallel rejection sets. Let $H(I)$, $I \subseteq N$, be any intersection hypothesis and consider the non-empty index sets associated with $F_1, \ldots, F_k$, i.e.,

$$I = I_1 \cup \ldots \cup I_k, \quad \text{where } I_i \subseteq N_i,\ i = 1, \ldots, k.$$

The index sets $I_1, \ldots, I_k$ need to be modified to account for the logical restrictions. The restricted index set I_i^* is defined as the subset of I_i that consists of only the testable null hypotheses. It is obtained by removing any null hypothesis H_j, $j \in I_i$ if at least one hypothesis from S_j is included in I or all hypotheses from P_j are included in I.

Going back to the schizophrenia trial example introduced in Section 1, the logical relationships among the nine null hypotheses of no treatment effect are defined based on serial rejection sets as

- Family F_1: S_1, S_2 and S_3 are empty.
- Family F_2: $S_4 = \{1\}$, $S_5 = \{2\}$ and $S_6 = \{3\}$.
- Family F_3: $S_7 = \{1, 4\}$, $S_8 = \{2, 5\}$ and $S_9 = \{3, 6\}$.

To illustrate the definition of restricted index sets, we consider the intersection hypothesis $H_1 \cap H_4 \cap H_5 \cap H_8$, i.e., $H(I)$ with $I = \{1, 4, 5, 8\}$. Here $I_1 = \{1\}$, $I_2 = \{4, 5\}$, $I_3 = \{8\}$. To obtain the restricted index sets, we remove the null hypotheses that are not testable according to the logical relationships. First, H_4 is not testable since the null hypothesis H_1 belonging to S_4 is included in I; thus $I_2^* = \{5\}$. Similarly, H_8 is not testable since H_5 belonging to S_8 is included in I; thus $I_3^* = \emptyset$.

Logical restrictions can be applied to the Hommel-based gatekeeping procedure by slightly modifying the definition of the local tests for the intersection hypotheses in F as follows:

- Case 1. If I_i^* is a proper subset of N_i for all $i = 1, \ldots, k$, the intersection hypothesis $H(I)$ is rejected if

$$\phi(I) = \min\left(\frac{p_1(I_1)}{c_1(I)}, \frac{p_2(I_2^*)}{c_2(I)}, \ldots, \frac{p_k(I_k^*)}{c_k(I)}\right) \leq \alpha.$$

- Case 2. If I_i^* is a proper subset of N_i for all $i = 1, \ldots, j-1$ and $I_j^* = N_j$ for some $j = 1, \ldots, k$, then the intersection hypothesis $H(I)$ is rejected if

$$\phi(I) = \min\left(\frac{p_1(I_1)}{c_1(I)}, \frac{p_2(I_2^*)}{c_2(I)}, \ldots, \frac{p_j(I_j^*)}{c_j(I)}\right) \leq \alpha.$$

- Case 3. If I_i^* is empty for some $i = 1, \ldots, k$, then the term $p_i(I_i^*)/c_i(I)(I)$ is deleted from the formulas given above.

Note that the $c_i(I)$, $i = 1, \ldots, k$ are not affected by the logical restrictions.

As shown in Dmitrienko and Tamhane (2010), if the component procedures are consonant (Gabriel, 1969) in the first $m-1$ families then this procedure is in conformance with the logical restrictions, i.e., a null hypothesis H_j cannot be rejected if at least one null hypothesis in S_j or all null hypotheses in P_j are accepted. Neither the Hommel MTP nor the truncated Hommel MTP is consonant. Therefore, to ensure conformance with the logical restrictions, the Hommel-based gatekeeping procedure needs to be enforced as follows. Let $\widetilde{p}_j$ be the adjusted p-value for the null hypothesis H_j in F obtained using the Hommel-based gatekeeping procedure defined above. Modify the adjusted p-values as

$$p_j^* = \max\left(\widetilde{p}_j,\ \max_{r\in S_j} p_r^*,\ \min_{r\in P_j} p_r^*\right).$$

This modification ensures that the adjusted p-value for H_j is $> \alpha$ if one or more null hypotheses in S_j or all null hypotheses in P_j have adjusted p-values $> \alpha$ and thus H_j cannot be rejected. In the schizophrenia trial example, the adjusted p-value for H_7 is re-defined as $p_7^* = \max(\widetilde{p}_7,\ p_1^*,\ p_4^*)$ since $S_7 = \{1, 4\}$ and thus H_7 cannot be rejected if either H_1 or H_4 is accepted.

4 Examples

In this section, we give two examples. The first example considers a clinical trial with tree gatekeeping restrictions involving parallel rejection sets. This example is used to compare the Hommel-based gatekeeping procedure to the Bonferroni-based gatekeeping procedure. The second example considers a real clinical trial with tree gatekeeping restrictions involving serial rejection sets. This example is used to illustrate how to determine the truncation fractions.

4.1 Hypertension trial

This example is from Dmitrienko et al. (2007). Consider a clinical trial in patients with hypertension in which the aim is to compare a new anti-hypertension treatment to an active control with regard to four endpoints:

- Primary endpoint (P): Mean reduction in systolic blood pressure.
- Two secondary endpoints (S1 and S2): Mean reduction in diastolic blood pressure and proportion of patients with controlled systolic/diastolic blood pressure.

- Tertiary endpoint (T): Average blood pressure based on ambulatory blood pressure monitoring.

Although the primary comparison of this trial is non-inferiority, superiority is also tested conditionally upon establishing non-inferiority for each endpoint. The resulting eight null hypotheses, denoted by $H_1, \ldots, H_8$, are grouped into four families. The families of null hypotheses, logical relationships and raw two-sided p-values are shown in Figure 2.

[Insert Figure 2 here]

As can be seen from this figure, the index sets associated with the four families are

$$N_1 = \{1\},\ N_2 = \{2, 3, 4\},\ N_3 = \{5, 6, 7\},\ N_4 = \{8\},$$

and thus $n_1 = 1$, $n_2 = 3$, $n_3 = 3$ and $n_4 = 1$. Also, only parallel rejection sets are defined in this example (a serial rejection set consisting of only one null hypothesis is equivalent to a parallel rejection set). These sets are defined as follows:

- Family F_1: P_1 is empty.
- Family F_2: $P_2 = P_3 = P_4 = \{1\}$.
- Family F_3: $P_5 = \{2\}$, $P_6 = \{2, 4\}$ and $P_7 = \{4\}$.
- Family F_4: $P_8 = \{6\}$.

The Hommel-based gatekeeping procedure is constructed based on a mixture of the truncated Hommel MTPs in F_1, F_2 and F_3 and the regular Hommel MTP in F_4. The truncation fractions $0 \leq \gamma_i < 1$, $i = 2, 3$, must be pre-specified. Note that since H_1 is the only null hypothesis in F_1, H_1 is simply rejected iff $p_1 \leq \alpha$ and thus γ_1 does not need to be specified. In this example we set $\gamma_2 = \gamma_3 = 0.9$. In general, the choice of these values can be investigated through simulations as explained in the second example.

To compute the adjusted p-values for the eight null hypotheses, local p-values for the $2^8 - 1 = 255$ intersection hypotheses need to be obtained based on the algorithm described in Sections 3. As an example, consider the intersection hypothesis $H_2 \cap H_6 \cap H_7 \cap H_8$, i.e., $H(I)$ with $I = \{2, 6, 7, 8\}$. The index sets associated with F_1 through F_4 are given by

$$I_1 = \emptyset,\ I_2 = \{2\},\ I_3 = \{6, 7\},\ I_4 = \{8\}.$$

Here H_6 contained in I_3 belongs to P_8 and so $I_4^* = \emptyset$. The other I_i^* are the same as the I_i. From the definition of the local p-values in Section 3.3, the local p-value for

the intersection hypothesis $H(I)$ is given by

$$p(I) = \min\left(\frac{p_2(I_2^*)}{c_2(I)}, \frac{p_3(I_3^*)}{c_3(I)}\right),$$

where $p_2(I_2^*)$ and $p_3(I_3^*)$ are the local p-values for the truncated Hommel MTP defined in (6):

$$\begin{aligned} p_2(I_2^*) &= \frac{p_2}{\gamma_2/|I_2^*| + (1-\gamma_2)/n_2} = \frac{0.008}{0.9/1 + 0.1/3} = 0.009, \\ p_3(I_3^*) &= \min\left(\frac{p_6}{\gamma_3/|I_3^*| + (1-\gamma_3)/n_3}, \frac{p_7}{2\gamma_3/|I_3^*| + (1-\gamma_3)/n_3}\right) \\ &= \min\left(\frac{0.010}{0.9/2 + 0.1/3}, \frac{0.302}{0.9 + 0.1/3}\right) = 0.021. \end{aligned}$$

Further, $c_2(I) = 1$ and

$$c_3(I) = c_2(I)\left[1 - \left(\gamma_2 + \frac{(1-\gamma_2)|I_2|}{n_2}\right)\right] = 1 - \left(0.9 + \frac{0.1}{3}\right) = 0.067.$$

Thus the local p-value for $H(I)$ is equal to $\min(0.009, 0.021/0.067) = 0.009$.

Once the local p-values for all intersection hypotheses have been computed, the adjusted p-value for each null hypothesis is obtained from (4). Since the truncated Hommel MTP is not consonant, a final modification needs to be performed as described in Section 3.3 to arrive at the adjusted p-values:

$$\begin{gathered} p_1^* = \widetilde{p}_1,\ p_2^* = \max\left(\widetilde{p}_2,\ p_1^*\right),\ p_3^* = \max\left(\widetilde{p}_3,\ p_1^*\right),\ p_4^* = \max\left(\widetilde{p}_4,\ p_1^*\right), \\ p_5^* = \max\left(\widetilde{p}_5,\ p_2^*\right),\ p_6^* = \max\left(\widetilde{p}_6,\ \min\left(p_2^*,\ p_4^*\right)\right),\ p_7^* = \max\left(\widetilde{p}_7,\ p_4^*\right), \\ p_8^* = \max\left(\widetilde{p}_8,\ p_6^*\right). \end{gathered}$$

A similar algorithm can be applied to define the Bonferroni-based gatekeeping procedure, i.e., a mixture of the Bonferroni MTPs in F_1, F_2 and F_3 and the Holm (1979) MTP in F_4.

The adjusted p-values for the eight null hypotheses for the Hommel-based and Bonferroni-based gatekeeping procedures are given in Table 1. As can be seen from this table, the Hommel-based gatekeeping procedure rejects five null hypotheses at the two-sided 0.05 level. In particular, this procedure establishes superiority of the new treatment to the active control with respect to Endpoint P and non-inferiority for Endpoints S1, S2 and T. The Bonferroni-based gatekeeping procedure rejects only four null hypotheses; in particular, it fails to establish non-inferiority for Endpoint S2. This example demonstrates the power advantage of the Hommel-based gatekeeping

procedure compared to the gatekeeping procedure based on the Bonferroni MTP. Note that the Hommel-based gatekeeping procedure is uniformly more powerful than gatekeeping procedures constructed based on MTPs that are less powerful than the Hommel MTP, e.g., Holm and Hochberg.

[Insert Table 1 here]

4.2 Schizophrenia trial

Let us return to the schizophrenia trial example introduced in Section 1. This trial was designed to evaluate the safety and efficacy of three doses of a new treatment compared to placebo in regard to three endpoints:

- Primary endpoint (P): Mean change from baseline in the Positive and Negative Syndrome Scale (PANSS) total score at Week 6.
- First key secondary endpoint (S1): Mean change from baseline in the Clinical Global Impression-Severity (CGI-S) score at Week 6.
- Second key secondary endpoint (S2): Mean change from baseline in the PANSS total score at Day 4.

The logical relationships among the resulting nine null hypotheses of no treatment effect are displayed in Figure 1 and explained in Section 3.3.

The truncation fractions $0 \leq \gamma_i < 1$, $i = 1, 2$, determine the balance of power in the three families of null hypotheses. Consider for example γ_1. This parameter influences power of the hypothesis tests in F_1 (Endpoint P tests) and F_2 (Endpoint S1 tests). The larger the value of γ_1, the higher the power of the Endpoint P tests and the lower the power of the Endpoint S1 tests. Similar considerations apply to γ_2, which determines the power balance between F_2 (Endpoint S1 tests) and F_3 (Endpoint S2 tests). In order to select optimum values for the truncation fractions, simulations can be performed based on the study design assumptions to maximize a pre-defined power function. In this example, several power functions for various scenarios were investigated. For the sake of simplicity, we only present here the simulation results for two power functions under the effect sizes and correlation assumptions shown in Tables 2 and 3:

- Power function 1: Probability to reject at least two null hypotheses in F_1 and at least one in F_2.
- Power function 2: Probability to reject at least two null hypotheses in F_1, at least two in F_2 and at least one in F_3.

[Insert Tables 2 and 3 here]

Table 4 presents the power simulation results based on 100,000 simulations for both power functions and for $\gamma_1, \gamma_2 = 0(0.1)0.9$ assuming a multivariate normal distribution for the test statistics and a sample size of 120 patients per arm. As shown in this table, the first power function is maximized for $\gamma_1 = 0.5$ and $\gamma_2 = 0.9$ whereas the second for $\gamma_1 = 0$ and $\gamma_2 = 0.2$. Note that $\gamma_1 = 0$ implies the use of the Bonferroni MTP in F_1. In this clinical trial, the first power function was considered to be a better fit to the trial objectives and thus $\gamma_1 = 0.5$ and $\gamma_2 = 0.9$ were chosen.

[Insert Table 4 here]

Table 5 lists the raw two-sided p-values and adjusted p-values computed using the Hommel-based gatekeeping procedure with the logical restrictions displayed in Figure 1. It follows from Table 5 that the procedure rejects the null hypotheses H_2 and H_5 at the two-sided 0.05 level (Dose M was significantly more effective than placebo for both Endpoints P and S1).

[Insert Table 5 here]

5 Conclusions

The Hommel-based gatekeeping procedure introduced in this paper is a closed testing procedure where p-values for each intersection hypothesis $H(I)$ are computed in two steps. First, the intersection hypothesis is decomposed into its sub-intersections, $H(I_i)$, where each $H(I_i)$ consists of the hypotheses that belong to both $H(I)$ and F_i. Next a local p-value, $p(I_i)$, is computed for each $H(I_i)$ using the truncated Hommel MTP (except for $H(I_m)$ which uses the regular Hommel MTP). In the second step, an overall p-value for the intersection hypothesis is computed using the Bonferroni mixing function which serves as a bridge between the different families to define a global closed testing procedure. Logical relations between the hypotheses are taken into account through the tree-gatekeeping approach which enables us to identify non-testable hypotheses in each $H(I_i)$ which are then eliminated from that intersection hypothesis before computing its local p-value.

The separability condition defined in Section 2.1 ensures that some part of α, if any remaining, can always be carried over to the next family. Since the Hommel MTP is not separable, we use the truncated Hommel MTP which is separable. The truncation fraction defined in this manner plays an important role in the algorithm. Locally within a family, a high value of the truncation fraction provides the best power. However, it reduces the power in all subsequent families. In order to choose optimum values for the truncation fractions, simulations based on the trial assumptions should be performed at the planning stage.

Serial and parallel logical restrictions can be implemented in the Hommel-based gatekeeping procedure. This enables clinical trial sponsors to set up complex hierarchical structures among clinical endpoints or clinical objectives in general. Further, well-defined logical restrictions can enhance the power of the procedure.

In this paper an efficient multiplicity adjustment procedure is defined for clinical trials with hierarchical objectives which controls the global FWER in the strong sense under the same conditions as required by the Hommel MTP. This method, termed the Hommel-based gatekeeping procedure, yields more power compared to gatekeeping procedures derived from other p-value based MTPs proposed in the literature, e.g., the Bonferroni, Holm and Hochberg.

SAS code for implementing the Hommel-based gatekeeping procedure can be downloaded from `www.multxpert.com`.

Acknowledgment. The authors are grateful to Prof. Sanat Sarkar for providing a shorter proof of Proposition 2.

References

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

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

[3] Dmitrienko, A., Tamhane, A.C., (2009). Gatekeeping procedures in clinical trials. *Multiple Testing Problems in Pharmaceutical Statistics*. Dmitrienko, A., Tamhane, A.C., Bretz, F. (editors). Chapman and Hall/CRC Press, New York.

[4] Dmitrienko, A., Tamhane, A.C. (2010a). Mixtures of multiple testing procedures for gatekeeping applications in clinical trials. *Statistics in Medicine*. To appear.

[5] Dmitrienko, A., Tamhane, A.C. (2010b). Theory of mixture gatekeeping procedures for clinical trials. In preparation.

[6] Dmitrienko, A., Tamhane, A., Liu, L. (2008). Mixtures of multiple testing procedures with gatekeeping applications. Northwestern University. Department of Industrial Engineering and Management Sciences. Working Paper 08-04. Available at `http://www.iems.northwestern.edu/research/papers.html`.

[7] Dmitrienko, A., Tamhane, A.C., Liu, L., Wiens, B.L. (2008). A note on tree gatekeeping procedures in clinical trials. *Statistics in Medicine*. 27, 3446–3451.

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

[9] Dmitrienko, A., Wiens, B.L., Tamhane, A.C., Wang X. (2007). Tree-structured gatekeeping tests in clinical trials with hierarchically ordered multiple objectives. *Statistics in Medicine*. 26, 2465-2478.

[10] Gabriel, K.R. (1969). Simultaneous test procedures—Some theory of multiple comparisons. *Annals of Mathematical Statistics*. 40, 224–250.

[11] Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*. 75, 800–802.

[12] Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Multiple Scandanavian Journal of Statistics*. 6, 65–70.

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

[14] Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*. 75, 383–386.

[15] Karlin, S. (1968). *Total Positivity*. Stanford University Press, Stanford.

[16] Marcus, R. Peritz, E., Gabriel, K.R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*. 63, 655–660.

[17] Maurer, W., Hothorn, L. A., Lehmacher, W. (1995). Multiple comparisons in drug clinical trials and preclinical assays: a priori ordered hypotheses. *Biometrie in der Chemisch-in-Pharmazeutischen Industrie*. 6. Vollman, J. (editor). Fischer-Verlag, Stuttgart, 3–18.

[18] Sarkar, S. (2010). Personal communication.

[19] Sarkar, S., Chang, C.K. (1997). Simes' method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Association*. 92, 1601–1608.

[20] Sarkar, S.K. (1998). Some probability inequalities for censored MTP2 random variables: A proof of the Simes conjecture. *The Annals of Statistics*. 26, 494–504.

[21] Simes, R.J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*. 63, 655–660.

[22] Westfall, P. H., Krishen, A. (2001). Optimally weighted, fixed sequence, and gatekeeping multiple testing procedures. *Journal of Statistical Planning and Inference*. 99, 25–40.

Appendix

Separability of the truncated Hommel MTP

Proposition 1 *The regular Hommel MTP is not separable, but the truncated Hommel MTP is separable for any* $0 \leq \gamma < 1$ *if the test statistics are positive dependent.*

Proof. Consider $n \geq 2$ hypotheses, $H_1, \ldots, H_n$. To see that the regular Hommel MTP is not separable, consider the configuration $I = \{n\}$, i.e., only H_n is true, and $H_1, \ldots, H_{n-1}$ are infinitely false so that $p_1, \ldots, p_{n-1} \to 0$ and $p_n = p_{(n)}$. Therefore the probability that H_n is rejected, i.e., $e(I|\alpha)$ with $I = \{n\}$, equals α. Thus $e(I|\alpha) = \alpha$ even though I is a proper subset of N, so the separability condition is violated.

To show the separability of the truncated Hommel MTP, suppose that $X_1, \ldots, X_n$ are continuous random variables with common c.d.f. $F(x)$. Let $X_{(1)} < \ldots < X_{(n)}$ denote their ordered values. Further let $U_{(1)} < \ldots < U_{(n)}$ denote the ordered values of the random variables $U_i = 1 - F(X_i)$, $i = 1, \ldots, n$. Since the U_i's are uniform $[0, 1]$ random variables, from Simes (1986), Sarkar and Chang (1997) and Sarkar (1998) it follows that for any $0 \leq y \leq 1$,

$$P\left(\bigcup_{i=1}^{n}\left\{U_{(i)} \leq \frac{iy}{n}\right\}\right) \leq y$$

with an equality if the random variables $X_1, \ldots, X_n$ are independent and an inequality if they are positively dependent.

Let $F = \{H_1, \ldots, H_n\}$ and $N = \{1, \ldots, n\}$. Select any nonempty $I \subset N$ and let $H(I) = \cap_{i \in I} H_i$. Further, let $|I| = k$. Using the simplified notation in which the indexes in I are numbered sequentially $1, \ldots, k$, let $p_{(1)} < \ldots < p_{(k)}$ denote the ordered raw p-values associated with H_i, $i \in I$. By the closure principle, an upper bound on the error rate function of the truncated Hommel MTP is given by

$$e(I|\alpha, \gamma) = P_{H(I)}\left(\bigcup_{i=1}^{k}\left\{p_{(i)} \leq \left(\frac{i\gamma}{k} + \frac{1-\gamma}{n}\right)\alpha\right\}\right).$$

Since $i \geq 1$, we have

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

and thus

$$e(I|\alpha, \gamma) \leq P_{H(I)}\left(\bigcup_{i=1}^{k}\left\{p_{(i)} \leq \frac{i(\gamma + (1-\gamma)k/n)\alpha}{k}\right\}\right).$$

Since the test statistics are assumed to be positive dependent, the Simes inequality can be applied with $U_{(i)} = p_{(i)}$, $i = 1, \ldots, k$, and thus the right-hand side of this inequality is no greater than $(\gamma + (1-\gamma)k/n)\alpha$. Since $k < n$ and $0 \leq \gamma < 1$, it follows that $e(I|\alpha, \gamma) < \alpha$, which implies that the truncated Hommel MTP is separable.

Generalized Simes Identity and Exact Error Rate Function of the Truncated Hommel Procedure

Proposition 2 *Let $U_1, \ldots, U_n$ be i.i.d. uniform $[0,1]$ random variables (r.v.'s) and $U_{(1)} < \ldots < U_{(n)}$ be their ordered values. Further let $\alpha, \beta > 0$ be fixed constants such that $\alpha + \beta < 1$. Then*

$$P\left(U_{(i)} > \frac{i\alpha}{n} + \beta,\ i = 1, \ldots, n\right) = (1-\beta)^{n-1}[1-(\alpha+\beta)]. \tag{9}$$

For $\beta = 0$, this reduces to the Simes (1986) identity.

Proof: Our original proof used the induction method of Simes (1986). The following proof is due to Sarkar (2010). Note that an alternative way of writing the Simes identity is as follows: Let $X_{(1)} < \cdots < X_{(n)}$ be order statistics from a continuous c.d.f. $F(x)$. Then

$$P(X_{(i)} > c_i, i = 1, \ldots, n) = 1 - \alpha, \tag{10}$$

where $F(c_i) = i\alpha/n$. Now for $X_{(i)}$ we will substitute $U_{(i)}$ conditioned on $U_{(1)} > \beta$. In that case, the $U_{(i)}$'s are order statistics from a uniform distribution on $[\beta, 1]$ with c.d.f. $F(u) = (u-\beta)/(1-\beta)$. Thus

$$F(i\alpha/n + \beta) = \frac{i\alpha}{n(1-\beta)}. \tag{11}$$

Also $P(U_{(1)} > \beta)) = (1-\beta)^n$. Thus we can write

$$\begin{aligned}
& P\left(U_{(i)} > \frac{i\alpha}{n} + \beta,\ i = 1, \ldots, n\right) \\
= \ & (1-\beta)^n P\left(U_{(i)} > \frac{i\alpha}{n} + \beta,\ i = 1, \ldots, n \middle| U_{(1)} > \beta\right) \\
= \ & (1-\beta)^n \left(1 - \frac{\alpha}{1-\beta}\right) \quad \text{(using (10) and (11))} \\
= \ & (1-\beta)^n[1-(\alpha+\beta)],
\end{aligned}$$

which is the desired result.

Under the intersection hypothesis $H(I) = \cap_{i=1}^k H_i$ where $k = |I|$, the exact error rate function of the truncated Hommel procedure equals

$$e(I|\gamma) = 1 - P_{h(I)}\left(\cap_{i=1}^k \left\{p_{(i)} > \left[\frac{i\gamma}{k} + \frac{1-\gamma}{n}\right]\right\}\right).$$

Substitute $n = k = |I|$, $\alpha = \gamma\alpha$ and $\beta = (1-\gamma)\alpha/n$ in (9) which gives the following exact expression for the error rate function of the truncated Hommel MTP:

$$e(I|\alpha,\gamma) = 1 - \left[1 - \frac{(1-\gamma)\alpha}{n}\right]^{|I|-1}\left[1 - \left(\gamma\alpha + \frac{(1-\gamma)\alpha}{n}\right)\right] \text{ if } |I| > 0$$

and $e(I|\alpha,\gamma) = 0$ if $|I| = 0$.

We did not use this exact expression because of the following reasons.

- This expression assumes independence whereas (7) is general.
- This expression is not proportional to α; as a result, the corresponding $f(I|\alpha,\gamma) = e(I|\alpha,\gamma)/\alpha$ depends on α. As can be seen from (2), this makes the mixing function $\phi(I)$ depend on α and so the local p-value to test $H(I)$ cannot be found by simply equating it to $\phi(I)$ and must be obtained numerically by solving the inequality (1).
- We compared powers using this exact expression and the upper bound (7) and found the power gains using the former to be quite marginal.

Step-up algorithm for the truncated Hommel MTP

Consider null hypotheses $H_1, \ldots, H_n$ and their associated ordered raw p-values $p_{(1)} < \ldots < p_{(n)}$. The truncated Hommel procedure can be applied using the following algorithm for $0 < \gamma \leq 1$. For $\gamma = 0$, the usual Bonferroni procedure is used.

- Step 1. Accept $H_{(n)}$ and go to the next step if

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

  Otherwise reject all hypotheses and stop.

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

$$p_{(n-k+i)} > \left(\frac{i\gamma}{k} + \frac{1-\gamma}{n}\right)\alpha \text{ for } i = 1, \ldots, k.$$

  Otherwise stop and reject all remaining hypotheses $H_{(i)}$, $i = 1, \ldots, n-k+1$, statisfying

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

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

$$p_{(i)} > \left(\frac{i\gamma}{n} + \frac{1-\gamma}{n}\right)\alpha \text{ for } i = 1, \ldots, n;$$

otherwise reject it if

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

Figure 1. Example of a clinical trial with three doses and three endpoints.

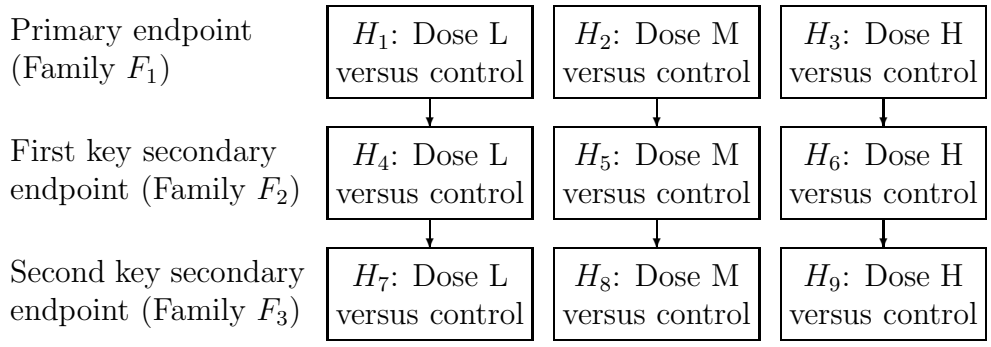

Figure 2. Gatekeeping structure in the hypertension trial example.

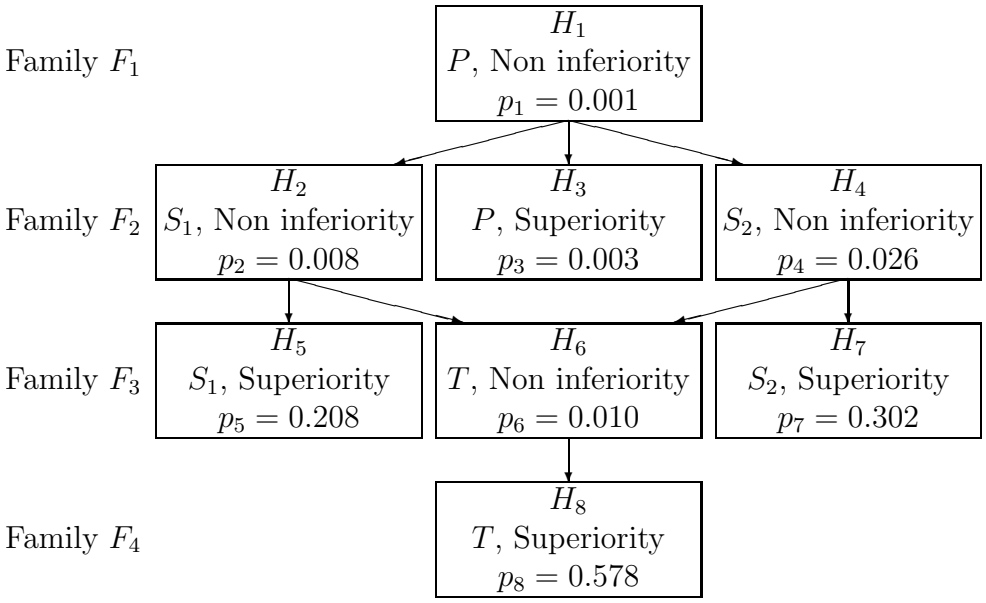

Table 1. Adjusted p-values in the hypertension trial example using the Hommel-based and Bonferroni-based gatekeeping procedures

Family	Null hypothesis	Raw p-value	Adjusted p-value	
			Hommel-based†	Bonferroni-based
F_1	H_1	0.001	0.001*	0.001*
F_2	H_2	0.008	0.017*	0.024*
	H_3	0.003	0.009*	0.009*
	H_4	0.026	0.028*	0.078
F_3	H_5	0.208	0.324	0.624
	H_6	0.010	0.030*	0.045*
	H_7	0.302	0.324	0.906
F_4	H_8	0.578	0.578	0.867

*The asterisk identifies the adjusted p-values that are significant at the 0.05 level.
†The Hommel-based gatekeeping procedure uses $\gamma_2 = \gamma_3 = 0.9$.

Table 2. Effect sizes (standardized mean difference) in the schizophrenia trial example.

Endpoint	Low dose versus placebo	Medium dose versus placebo	High dose versus placebo
P	0.3	0.4	0.7
S1	0.2	0.3	0.5
S2	0.1	0.2	0.3

Table 3. Correlation coefficients ρ between test statistics for the same dose in the schizophrenia trial example. (Correlation coefficients between test statistics for different doses were set to $\rho/2$.)

Endpoint	P	S1	S2
P	1	0.8	0.4
S1	0.8	1	0.3
S2	0.4	0.3	1

Table 4. Simulated powers (%) of the Hommel-based gatekeeping procedure for two power functions in the schizophrenia trial example as functions of the truncation fractions γ_1 and γ_2. (The highest power entry for each power function is shown in a box.)

Power function 1[†]										
	γ_2									
γ_1	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.0	77.4	77.5	77.7	77.8	77.9	78.0	78.1	78.2	78.3	78.4
0.1	77.6	77.7	77.9	78.0	78.1	78.3	78.4	78.5	78.6	78.7
0.2	77.8	78.0	78.1	78.2	78.4	78.5	78.6	78.7	78.9	79.0
0.3	78.0	78.1	78.3	78.4	78.5	78.7	78.8	78.9	79.0	79.2
0.4	78.1	78.3	78.4	78.5	78.7	78.8	78.9	79.1	79.2	79.3
0.5	78.2	78.3	78.5	78.6	78.7	78.9	79.0	79.1	79.2	**79.4**
0.6	78.1	78.3	78.4	78.5	78.7	78.8	79.0	79.1	79.2	79.3
0.7	77.8	77.9	78.1	78.3	78.4	78.5	78.7	78.8	78.9	79.0
0.8	77.1	77.3	77.4	77.6	77.7	77.8	78.0	78.1	78.2	78.4
0.9	75.6	75.7	75.8	76.0	76.1	76.3	76.4	76.6	76.7	76.8
Power function 2[‡]										
	γ_2									
γ_1	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.0	27.3	27.6	**27.7**	27.6	27.5	27.0	26.3	25.5	24.3	22.3
0.1	27.0	27.2	27.4	27.4	27.1	26.8	26.3	25.4	24.3	22.4
0.2	26.8	27.0	27.1	27.0	26.9	26.6	26.0	25.3	24.3	22.5
0.3	26.5	26.7	26.8	26.7	26.5	26.3	25.8	25.1	24.1	22.5
0.4	26.3	26.4	26.5	26.4	26.2	25.9	25.5	24.8	23.9	22.4
0.5	26.0	26.1	26.1	26.0	25.9	25.6	25.2	24.6	23.7	22.3
0.6	25.8	25.8	25.8	25.7	25.6	25.3	24.9	24.3	23.5	22.2
0.7	25.5	25.5	25.5	25.4	25.2	25.0	24.6	24.1	23.4	22.1
0.8	25.3	25.3	25.2	25.1	24.9	24.7	24.3	23.8	23.1	21.9
0.9	25.0	25.0	25.0	24.8	24.6	24.4	24.1	23.6	23.0	21.9

[†]Power function 1: Reject at least two hypotheses in F_1 and at least one in F_2.
[‡]Power function 2: Reject at least two hypotheses in F_1, at least two in F_2 and at least one in F_3.

Table 5. Adjusted p-values in the schizophrenia trial example obtained from the Hommel-based gatekeeping procedure with the truncation fractions $\gamma_1 = 0.5$ and $\gamma_2 = 0.9$.

Family	Null hypothesis	Raw p-value	Adjusted p-value
F_1	H_1	0.394	0.591
	H_2	0.011	0.034*
	H_3	0.163	0.391
F_2	H_4	0.365	0.591
	H_5	0.005	0.034*
	H_6	0.169	0.543
F_3	H_7	0.241	0.591
	H_8	0.296	0.591
	H_9	0.263	0.591

The asterisk identifies the adjusted p-values that are significant at the 0.05 level.