Journal of Biopharmaceutical Statistics, 21: 726–747, 2011

ISSN: 1054-3406 print/1520-5711 online
DOI: 10.1080/10543406.2011.551333

MULTISTAGE AND MIXTURE PARALLEL GATEKEEPING PROCEDURES IN CLINICAL TRIALS

Alex Dmitrienko[1], **George Kordzakhia**[2], **and Ajit C. Tamhane**[3]
[1]*Eli Lilly and Company, Indianapolis, Indiana, USA*
[2]*Food and Drug Administration, Silver Spring, Maryland, USA*
[3]*Northwestern University, Evanston, Illinois, USA*

Gatekeeping procedures have been developed to solve multiplicity problems arising in clinical trials with hierarchical objectives where the null hypotheses that address these objectives are grouped into ordered families. A general method for constructing multistage parallel gatekeeping procedures was proposed by Dmitrienko et al. (2008). The objective of this paper is to study two related classes of parallel gatekeeping procedures. Restricting to two-family hypothesis testing problems, we first use the mixture method developed in Dmitrienko and Tamhane (2011) to define a class of parallel gatekeeping procedures derived using the closure principle that can be more powerful than multistage gatekeeping procedures. Second, we show that power of multistage gatekeeping procedures can also be improved by using α-exhaustive tests for the component procedures. Extensions of these results for multiple families are stated. Illustrative examples from clinical trials are given.

Key Words: Closure principle; Familywise error rate; Multiple comparisons; Mixture procedure; Parallel gatekeeping.

1. INTRODUCTION

Analysis of trials with multiple objectives has attracted much attention in the clinical trial literature. Recent developments in this area include a new class of testing methods, known as gatekeeping methods, for hypothesis testing problems with multiple families of null hypotheses. This methodology enables clinical trial sponsors to build multiple testing procedures that take into account several sources of multiplicity, e.g., multiplicity induced by multiple primary and secondary endpoints, multiple dose-control tests, multiple patient populations, etc. A key property of these procedures is that they control the global familywise error rate (FWER) in the strong sense (Hochberg and Tamhane, 1987) across multiple families and thus help clinical trial sponsors enrich product labels by including relevant secondary objectives. Dmitrienko and Tamhane (2009) gave a review of the literature on the subject.

Received September 15, 2010; Accepted December 10, 2010
Address correspondence to Alex Dmitrienko, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, Indiana 46285, USA; E-mail: dmitrienko_alex@lilly.com

To help define key concepts in gatekeeping procedures, consider a clinical trial with hierarchically ordered null hypotheses of no treatment effect corresponding to specified multiple objectives. To simplify the ideas and notation, we restrict to the case of two families of null hypotheses (labeled as primary and secondary). The primary family serves as a *gatekeeper* for the secondary family. We focus on procedures that satisfy the *parallel gatekeeping condition* introduced in Dmitrienko et al. (2003). This condition states that the null hypotheses in the secondary family can be tested if and only if at least one null hypothesis in the primary family is rejected; otherwise all null hypotheses in the secondary family are accepted without tests.

A general method for constructing parallel gatekeeping procedures was proposed in Dmitrienko et al. (2008). Using this method, one can define a broad class of gatekeeping procedures that use powerful procedures such as the Hochberg, Hommel, and Dunnett procedures as component procedures to test the primary and secondary null hypotheses. In hypothesis testing problems with two families, these gatekeeping procedures have a simple two-stage testing structure (more generally a multistage structure) that streamlines their implementation and interpretation. We refer to them as *two-stage parallel gatekeeping procedures*.

In this paper we introduce two classes of parallel gatekeeping procedures that provide a power advantage over multistage gatekeeping procedures. The first class is defined using the mixture method developed in Dmitrienko and Tamhane (2011), which uses a direct application of the closure principle (Marcus et al., 1976). These parallel gatekeeping procedures will be termed *mixture parallel gatekeeping procedures*. It is important to note that multistage gatekeeping procedures proposed in Dmitrienko et al. (2008) were derived using a method that was not explicitly based on the closure principle. It is shown in this paper that in two-family problems, mixture gatekeeping procedures can be more powerful than two-stage gatekeeping procedures. At a conceptual level, the relationship between the multistage and mixture gatekeeping frameworks is similar to that between the Hochberg and Hommel procedures (Hommel, 1988,?). The Hochberg procedure was derived using a direct argument without an explicit reference to the closure principle and has a simple stepwise structure. The Hommel procedure was defined as a closed testing procedure by specifying tests for all intersection hypotheses in the closed family. This procedure is uniformly more powerful than the Hochberg procedure but lacks the simple stepwise structure.

Two-stage gatekeeping procedures satisfy the independence condition, which requires that the inferences on the primary null hypotheses be independent of the inferences on the secondary null hypotheses. This condition is generally required in hypothesis testing problems involving primary and secondary endpoints and helps to ensure that the inferences on the primary endpoints are not influenced by those on the secondary endpoints. However, it is less relevant and can be dropped in other hypothesis testing problems arising in clinical trials. We show that if this condition is relaxed then the gatekeeping procedure can be made more powerful. This is accomplished by employing α-exhaustive (Grechanovsky and Hochberg, 1999) local tests for all intersection hypotheses in the closed family when the two-stage procedure is expressed in its equivalent form as a closed procedure. The resulting procedure is based on a three-stage algorithm where the third stage retests the null hypotheses that were not rejected at the first stage using a more powerful procedure if all the secondary null hypotheses are rejected. Clearly, this procedure does not satisfy the independence condition but it is uniformly more powerful than the two-stage gatekeeping procedure satisfying the condition.

The paper is organized as follows. Section 2 reviews the two-stage and mixture parallel gatekeeping procedures and shows that they are equivalent if the multiple testing procedure used in the primary family is consonant; if this condition is relaxed then the mixture procedure can be made more powerful. Section 3 shows that the power of two-stage gatekeeping procedures can be enhanced by employing α-exhaustive local tests in the closed family for the intersection hypotheses, and the resulting closed procedure has a simple three-stage structure with retesting. Examples are given in both these sections to illustrate the new parallel gatekeeping procedures. Section 4 briefly outlines extensions to general hypothesis testing problems with an arbitrary number of families and Section 5 discusses software implementation of two- and three-stage gatekeeping procedures. The proofs of the propositions are given in the appendix.

2. TWO-STAGE AND MIXTURE PARALLEL GATEKEEPING PROCEDURES

Consider a multiple testing problem arising in a clinical trial with n null hypotheses denoted by H_i, $i = 1, \ldots, n$, which are grouped into a primary family F_1 of n_1 null hypotheses and a secondary family F_2 of n_2 null hypotheses ($n_1 + n_2 = n$). Denote the two families by

$$F_1 = \{H_i, i \in N_1\}, \quad F_2 = \{H_i, i \in N_2\}$$

where N_1 and N_2 are the index sets for the null hypotheses included in the two families, respectively, i.e.,

$$N_1 = \{1, \ldots, n_1\}, \quad N_2 = \{n_1 + 1, \ldots, n_1 + n_2\}$$

Let $N = N_1 \cup N_2 = \{1, \ldots, n\}$. As noted before, F_1 is a parallel gatekeeper for F_2. We require a gatekeeping procedure that satisfies the parallel gatekeeping condition and controls the global FWER in the strong sense at a prespecified α level. In other words, the probability of rejecting any true null hypothesis must be $\leq \alpha$ for all possible combinations of the true and false null hypotheses in the two families.

In this section we define the two-stage and mixture parallel gatekeeping procedures that are built from predefined multiple testing procedures for testing the primary and secondary null hypotheses. These procedures are denoted by $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively, and we refer to them as the *primary and secondary component procedures*. For example, we may test the primary null hypotheses using the Bonferroni procedure and, if the parallel gatekeeping condition is satisfied, use the Holm (1979) procedure in the secondary family. We assume that both component procedures are closed (Marcus et al., 1976) and thus provide strong local control of FWER within the corresponding family.

2.1. Two-Stage Gatekeeping Procedure

The two-stage gatekeeping procedure developed in Dmitrienko et al. (2008) is built around the concept of the *error rate function*. The error rate function of the primary component $\mathscr{P}_1$, denoted by $e_1(I_1 \mid \alpha)$, where $I_1 \subseteq N_1$ is the index set of true hypotheses, is defined as the probability of incorrectly rejecting at least one true null hypothesis H_i, $i \in I_1$. The error rate function is assumed to be monotone, i.e., $e_1(I_1 \mid \alpha) \leq e_1(J_1 \mid \alpha)$ if $I_1 \subseteq J_1 \subseteq N_1$ and, in addition, $e_1(I_1 \mid \alpha) = \alpha$ if $I_1 = N_1$.

Generally the exact error rate function is difficult to evaluate for most procedures, so we use a simple upper bound instead. For convenience, we refer to the upper bound itself as the error rate function and use the same notation $e_1(I_1 \mid \alpha)$ for it. For example, the upper bound on the error rate function of the Bonferroni procedure that tests each H_i, $i \in N_1$, at level α/n_1, is

$$e_1(I_1 \mid \alpha) = \alpha |I_1| / n_1$$

where $|I_1|$ denotes the number of elements in the index set I_1.

The portion of the α that can be carried over from F_1 to F_2 depends on the set of primary null hypotheses accepted by $\mathscr{P}_1$ and it is quantified via the error rate function of the primary component. For this portion to be positive when $\mathscr{P}_1$ rejects at least one primary null hypothesis, it is required that $\mathscr{P}_1$ must be *separable*, i.e., $e_1(I_1 \mid \alpha) < \alpha$ for all proper subsets I_1 of N_1. The Bonferroni procedure is clearly separable; however, the standard stepwise procedures such as in Holm (1979), Hommel (1988), and Hommel (1988) are not separable. Dmitrienko et al. (2008) showed that these procedures can be made separable by forming their *truncated* versions that use convex combinations of the critical constants of the original procedures with those of the Bonferroni procedure. The truncated procedures are less powerful than the original procedures but are more powerful than the Bonferroni procedure.

As shown in Dmitrienko et al. (2008), an upper bound on the error rate function of the truncated Holm procedure is given by

$$e_1(I_1 \mid \alpha) = \left(\gamma + (1 - \gamma) \frac{|I_1|}{n_1} \right) \alpha \tag{1}$$

if I_1 is nonempty and 0 otherwise. Here γ is the truncation fraction with $0 \leq \gamma \leq 1$. Note that these truncated procedures are separable for $\gamma < 1$. As shown in Brechenmacher et al. (2011), the same upper bound can be used for the truncated Hochberg and Hommel procedures. This bound is generally conservative and can be improved if additional assumptions on the joint distribution of the hypothesis test statistics can be made; e.g., a sharper bound can be derived under the independence assumption.

In the following two-stage procedure, $\mathscr{P}_1$ is assumed to be separable and may be chosen to be the Bonferroni procedure or one of the more powerful truncated procedures.

- Stage 1. The primary null hypotheses are tested using $\mathscr{P}_1$ at level $\alpha_1 = \alpha$. Let $A_1 \subseteq N_1$ be the index set of these null hypotheses accepted by $\mathscr{P}_1$.
- Stage 2. If at least one primary null hypothesis is rejected, i.e., if $A_1 \subset N_1$, the secondary null hypotheses are tested using $\mathscr{P}_2$ at level $\alpha_2 = \alpha_1 - e_1(A_1 \mid \alpha_1)$.

Dmitrienko et al. (2008) proved that this general two-stage gatekeeping procedure controls the global FWER at the α level. Note that the secondary null hypotheses are not tested if no hypotheses are rejected in the primary family and thus this two-stage procedure satisfies the parallel gatekeeping condition. Further, by construction, the inferences in the primary family do not depend on the inferences in the secondary family; thus, the independence condition is satisfied. This procedure is illustrated later in Example 1.

A key advantage of the two-stage gatekeeping procedure is that it is transparent and clearly demonstrates the process of performing multiplicity adjustments. However, this procedure is not the most powerful available. It is shown later that for certain types of primary component procedures the power of the two-stage procedure can be improved without sacrificing the parallel gatekeeping and independence properties or global FWER control.

2.2. Mixture Gatekeeping Procedure

Dmitrienko and Tamhane (2011) proposed a mixture approach for combining the component procedures $\mathscr{P}_1$ and $\mathscr{P}_2$ to construct a parallel gatekeeping procedure that strongly controls the global FWER. As already mentioned, this approach is based on the closure principle, which requires local α-level tests of all intersection hypotheses $H(I) = \bigcap_{i \in I} H_i$, where I is a nonempty subset of N. Selecting any intersection hypothesis $H(I)$, we partition it as $H(I) = H(I_1) \cap H(I_2)$, where

$$H(I_j) = \bigcap_{i \in I_j} H_i, \quad I_j \subseteq N_j, \quad j = 1, 2$$

and at least one of the index sets, I_1 and I_2, is nonempty. Let $p_j(I_j)$ be the local p value for the intersection hypothesis $H(I_j)$ using the component procedure $\mathscr{P}_j$, $j = 1, 2$. We assume that $p_j(I_j)$ provides a local α-level test of $H(I_j)$ for any α, i.e., under $H(I)$,

$$P(\text{Reject } H(I_j)) = P(p_j(I_j) \leq \alpha) \leq \alpha \tag{2}$$

Since $\mathscr{P}_j$ itself is a closed procedure, any null hypothesis $H_i \in F_j$ is rejected by $\mathscr{P}_j$ at level α if and only if $p_j(I_j) \leq \alpha$ for all index sets I_j such that I_j includes the index i, $j = 1, 2$.

We make a simplifying assumption that the error rate function $e_1(I_1 \mid \alpha)$ of $\mathscr{P}_1$ is proportional to α. As seen from Eq. (1), this assumption is satisfied by the Bonferroni and truncated Holm, Hochberg, and Hommel procedures. Under this assumption, the mixture gatekeeping procedure defines the local p value, $p(I)$, for the intersection hypothesis $H(I)$ as a function of $p_1(I_1)$ and $p_2(I_2)$ as follows.

- Case 1 (the intersection hypothesis includes only the primary null hypotheses, i.e., $I = I_1$ and I_2 is empty). The local p value for the mixture gatekeeping procedure equals the local p value of $\mathscr{P}_1$, i.e., $p(I) = p_1(I_1)$.
- Case 2 (the intersection hypothesis includes only the secondary null hypotheses, i.e., $I = I_2$ and I_1 is empty). The local p value for the mixture gatekeeping procedure equals the local p value of $\mathscr{P}_2$, i.e., $p(I) = p_2(I_2)$.
- Case 3 (the intersection hypothesis includes both primary and secondary null hypotheses, i.e., $I = I_1 \cup I_2$ and both I_1 and I_2 are nonempty). The local p value

for the mixture gatekeeping procedure is given by the following formula that combines the local p values $p_1(I_1)$ and $p_2(I_2)$:

$$p(I) = \min\left(p_1(I_1), \frac{p_2(I_2)}{1 - f_1(I_1)}\right) \tag{3}$$

where $f_1(I_1) = e_1(I_1 \mid \alpha)/\alpha$. The min function in the preceding formula for $p(I)$ is referred to as the *Bonferroni mixing function* in Dmitrienko and Tamhane (2011).

This procedure is illustrated in Example 1 later in this section.

Note that since $e_1(I_1 \mid \alpha)$ is assumed to be proportional to α, the fraction $f_1(I_1)$ is independent of α. This allows us to define the local p value, $p(I)$, simply as a function of $p_1(I_1)$ and $p_2(I_2)$ as in Eq. (3). If $f_1(I_1)$ were a function of α then $p(I)$ would need to be calculated using the general definition of adjusted p values involving an iterative calculation; see Dmitrienko and Tamhane (2011, section 2.1) for more details.

The following proposition gives the conditions under which the mixture procedure controls the FWER.

Proposition 1. *The mixture parallel gatekeeping procedure controls the global FWER at level α if $\mathcal{P}_1$ is any separable FWER-controlling procedure and $\mathcal{P}_2$ is any FWER-controlling procedure.*

The proof is given in the appendix.

2.3. Relationship Between the Two-Stage and Mixture Gatekeeping Procedures

Although the two-stage and the mixture procedures are defined using two seemingly different methods, they are in fact closely related to each other. The relationship between the two procedures is described in Propositions 2 and 3. Proposition 2 shows that the two-stage procedure based on the component procedures $\mathcal{P}_1$ and $\mathcal{P}_2$ is equivalent to the mixture procedure based on the same two component procedures if $\mathcal{P}_1$ is a consonant procedure (Gabriel, 1969), i.e., if $\mathcal{P}_1$ rejects an intersection hypothesis $H(I_1)$, $I_1 \subseteq N_1$, then it rejects at least one null hypothesis $H_i \in I_1$. Examples of consonant separable procedures include the truncated versions of the Holm, Hochberg and fallback procedures. On the other hand, the truncated Hommel procedure is not consonant (Westfall et al., 1999, section 2.5.4).

A key assumption made in Proposition 2 is that the primary component procedure is consonant. Proposition 3 shows that if $\mathcal{P}_1$ is nonconsonant, the mixture gatekeeping procedure provides a uniform power advantage over the two-stage gatekeeping procedure. In particular, the mixture procedure rejects a primary null hypothesis if and only if it is rejected by the two-stage procedure; however, the mixture procedure may reject more secondary null hypotheses than the two-stage procedure.

Proposition 2. *For any separable and consonant FWER-controlling primary component procedure $\mathcal{P}_1$ and FWER-controlling secondary component procedure $\mathcal{P}_2$,* *the two-stage parallel gatekeeping procedure is equivalent to the mixture parallel gatekeeping procedure.*

The proof is given in the appendix.

Example 1 Consider a clinical trial conducted to evaluate the efficacy of a new treatment compared to a placebo on two primary and two secondary endpoints. Denote the two primary null hypotheses by H_1 and H_2 and the two secondary null hypotheses by H_3 and H_4. The primary family serves as a parallel gatekeeper for the secondary family. We will construct a mixture gatekeeping procedure with the following component procedures:

- Truncated Hochberg procedure ($\mathcal{P}_1$): A truncated version of the Hochberg procedure with a prespecified truncation parameter ($0 \le \gamma < 1$), which is separable, needs to be used in the primary family for a positive α to be carried over to the secondary family if at least one primary null hypothesis is rejected.
- Hochberg procedure ($\mathcal{P}_2$): Note that the regular Hochberg procedure, which is not separable, can be used in the secondary family since it is the last family in the sequence.

Let p_i denote the p value for testing the null hypothesis H_i, $i = 1, \ldots, 4$. Let $p_{(1)} < p_{(2)}$ denote the ordered p values in the primary family and $p_{(3)} < p_{(4)}$ denote the ordered p values in the secondary family.

Assuming that the truncated and regular Hochberg procedures control the FWER within the primary and secondary families, e.g., the joint distribution of the hypothesis test statistics within each family is multivariate totally positive of order two (Sarkar, 1998; Sarkar and Chang, 1997), the mixture gatekeeping procedure is defined as follows. We first need to compute the local p values $p_1(I_1)$ and $p_2(I_2)$ for the primary and secondary component procedures, respectively, where $I_1 \subseteq N_1 = \{1, 2\}$ and $I_2 \subseteq N_2 = \{3, 4\}$. Beginning with the local p values for the primary component procedure, consider the index set $I_1 = \{1, 2\}$. Note that $\mathcal{P}_1$ rejects the intersection hypothesis $H(I_1) = H_1 \cap H_2$ if

$$p_{(1)} < (\gamma/2 + (1-\gamma)/2)\alpha \quad \text{or} \quad p_{(2)} < (\gamma + (1-\gamma)/2)\alpha$$

The local p value for $H(I_1)$, i.e., $p_1(I_1)$, is defined as the smallest α for which either of these two inequalities is satisfied and thus

$$p_1(I_1) = \min\left(\frac{p_{(1)}}{\gamma/2 + (1-\gamma)/2}, \frac{p_{(2)}}{\gamma + (1-\gamma)/2}\right)$$

Further, considering $I_1 = \{1\}$ and $I_2 = \{2\}$, note that $\mathcal{P}_1$ rejects H_1 if $p_1 < (\gamma + (1-\gamma)/2)\alpha$ and H_2 if $p_2 < (\gamma + (1-\gamma)/2)\alpha$. Therefore,

$$p_1(I_1) = \frac{p_i}{\gamma + (1-\gamma)/2} \quad \text{if } I_1 = \{i\}, \; i = 1, 2$$

Similarly, the local p values for the secondary component procedure are given by

$$p_2(I_2) = \begin{cases} \min(2p_{(3)}, p_{(4)}) & \text{if } I_2 = \{3, 4\} \\ p_3 & \text{if } I_2 = \{3\} \\ p_4 & \text{if } I_2 = \{4\} \end{cases}$$

The second step is to obtain the local $p$values $p(I)$, where $I \subseteq N = \{1, 2, 3, 4\}$, for the mixture gatekeeping procedure. This computation makes use of the error rate function for the primary component procedure (truncated Hochberg procedure), which in this particular case has the following simple form:

$$e_1(I_1 \mid \alpha) = \begin{cases} \alpha & \text{if } I_1 = \{1, 2\} \\ (\gamma + (1 - \gamma)/2)\alpha & \text{if } I_1 = \{1\} \text{ or } I_1 = \{2\} \\ 0 & \text{if } I_1 = \emptyset \end{cases}$$

The local p values are displayed in Table 1. Note that $p(I)$ for the mixture gatekeeping procedure equals $p_1(I_1)$ for $\mathscr{P}_1$ when $I = \{1, 2, 3, 4\}$, $\{1, 2, 3\}$, $\{1, 2, 4\}$, and $\{1, 2\}$ since $I_1 = N_1$ in these cases and thus $f_1(I_1) = 1$.

The mixture gatekeeping procedure rejects a null hypothesis H_i if $p(I) \leq \alpha$ for all I such that $i \in I$. Applying this rule to each of the four null hypotheses and noting that the truncated Hochberg procedure is consonant, it is easy to verify that, as stated in Proposition 1, the mixture gatekeeping procedure has the following two-stage structure:

- Stage 1. The null hypotheses in the primary family are tested using the truncated Hochberg procedure at the full α level.

Table 1 Local p values for the mixture gatekeeping procedure in Example 1

Index set			Local p value
I	I_1	I_2	$p(I)$
$\{1, 2, 3, 4\}$	$\{1, 2\}$	$\{3, 4\}$	$p_1(I_1)$
$\{1, 2, 3\}$	$\{1, 2\}$	$\{3\}$	$p_1(I_1)$
$\{1, 2, 4\}$	$\{1, 2\}$	$\{4\}$	$p_1(I_1)$
$\{1, 2\}$	$\{1, 2\}$	$\emptyset$	$p_1(I_1)$
$\{1, 3, 4\}$	$\{1\}$	$\{3, 4\}$	$\min(p_1(I_1), 2p_2(I_2)/(1 - \gamma))$
$\{1, 3\}$	$\{1\}$	$\{3\}$	$\min(p_1(I_1), 2p_2(I_2)/(1 - \gamma))$
$\{1, 4\}$	$\{1\}$	$\{4\}$	$\min(p_1(I_1), 2p_2(I_2)/(1 - \gamma))$
$\{1\}$	$\{1\}$	$\emptyset$	$p_1(I_1)$
$\{2, 3, 4\}$	$\{2\}$	$\{3, 4\}$	$\min(p_1(I_1), 2p_2(I_2)/(1 - \gamma))$
$\{2, 3\}$	$\{2\}$	$\{3\}$	$\min(p_1(I_1), 2p_2(I_2)/(1 - \gamma))$
$\{2, 4\}$	$\{2\}$	$\{4\}$	$\min(p_1(I_1), 2p_2(I_2)/(1 - \gamma))$
$\{2\}$	$\{2\}$	$\emptyset$	$p_1(I_1)$
$\{3, 4\}$	$\emptyset$	$\{3, 4\}$	$p_2(I_2)$
$\{3\}$	$\emptyset$	$\{3\}$	$p_2(I_2)$
$\{4\}$	$\emptyset$	$\{4\}$	$p_2(I_2)$

- Stage 2. If at least one primary null hypothesis is rejected, the null hypotheses in the secondary family are tested using the Hochberg procedure at level α_2, where $\alpha_2 = \alpha$ if both primary null hypotheses are rejected and $\alpha_2 = (1-\gamma)\alpha/2$ if only one primary null hypothesis is rejected.

To illustrate this two-stage gatekeeping procedure, we assume that the one-sided raw p values for the four null hypotheses are as follows:

$$p_1 = 0.0110, \quad p_2 = 0.0193, \quad p_3 = 0.0042, \quad p_4 = 0.0057$$

Here $p_1 < p_2$ are the ordered p values in the primary family and $p_3 < p_4$ are the ordered p values in the secondary family. We further assume a one-sided $\alpha = 0.025$.

The truncated Hochberg procedure with $\gamma = \frac{1}{2}$ will be used in the primary family at level $\alpha_1 = \alpha$ and the regular Hochberg procedure will be used in the secondary family at level α_2 computed from the error rate function (1):

- Stage 1. The truncated Hochberg procedure fails to reject the null hypothesis H_2 because $p_2 = 0.0193 > \gamma\alpha_1 + (1-\gamma)\alpha_1/2 = 0.01875$. However, it rejects H_1 since $p_1 = 0.0110 < \alpha_1/2 = 0.0125$.
- Stage 2. Since one primary null hypothesis is rejected in Stage 1, the two-stage gatekeeping procedure passes the parallel gatekeeper. Using

$$\alpha_2 = \alpha_1 - e_1(A_1 \mid \alpha_1) = \alpha - \left(\gamma + \frac{(1-\gamma)}{2}\right)\alpha = 0.00625$$

  the Hochberg procedure rejects both secondary null hypotheses since the larger secondary p value, $p_4 = 0.0057 < \alpha_2 = 0.00625$.

To summarize, the two-stage gatekeeping procedure rejects one primary and two secondary null hypotheses in this example.

Proposition 3. *For any separable and nonconsonant FWER-controlling $\mathcal{P}_1$ and general FWER-controlling $\mathcal{P}_2$, the mixture parallel gatekeeping procedure is uniformly more powerful than the two-stage parallel gatekeeping procedure, i.e., the former rejects as many and potentially more null hypotheses than the latter.*

The proof of Proposition 3 follows directly from Part 1 of the proof of Proposition 2 and is omitted.

Proposition 3 is illustrated in the following example.

Example 2 Consider a clinical trial with $F_1 = \{H_1, H_2, H_3, H_4\}$ and $F_2 = \{H_5\}$. The raw one-sided p values for the five null hypotheses are shown in Table 2. To construct the two-stage and mixture gatekeeping procedures, we use the truncated Hommel procedure (which is nonconsonant) with $\gamma = \frac{3}{4}$ as $\mathcal{P}_1$ and the regular Hommel procedure as $\mathcal{P}_2$. The adjusted p values produced by the two procedures are displayed in Table 2 (the adjusted p values for the two-stage gatekeeping procedure are computed using the direct-calculation algorithm introduced in Dmitrienko et al. (2008) and the adjusted p values for the mixture gatekeeping procedure are found using the algorithm defined in section 2.2). Using

Table 2 Raw and adjusted p values in Example 2

			Adjusted p value	
Family	Null hypothesis	Raw p value	Two-stage procedure	Mixture procedure
Primary	H_1	0.0053	0.0210*	0.0210*
	H_2	0.0126	0.0276	0.0276
	H_3	0.0131	0.0276	0.0276
	H_4	0.0224	0.0276	0.0276
Secondary	H_5	0.0022	0.0276	0.0233*

*Significant at the one-sided 0.025 level.

one-sided $\alpha = 0.025$, the two-stage procedure rejects H_1 and then proceeds to test H_5, which cannot be rejected. By contrast, the mixture procedure rejects H_1 as well as H_5. This demonstrates the improved power of the mixture gatekeeping procedure compared to the two-stage gatekeeping procedure.

It is clear from the proof of Proposition 2 that the mixture procedure is equivalent to $\mathscr{P}_1$ within the primary family, i.e., the mixture procedure rejects any primary null hypothesis if and only if $\mathscr{P}_1$ rejects that null hypothesis. This implies that the inferences in F_1 are not affected by the rejection or acceptance of secondary null hypotheses and thus the independence condition is satisfied even if $\mathscr{P}_1$ is nonconsonant. This statement is formulated as Proposition 4.

Proposition 4. *For any separable FWER-controlling $\mathscr{P}_1$ and FWER-controlling $\mathscr{P}_2$, the mixture parallel gatekeeping procedure satisfies the independence condition.*

The proof of Proposition 4 is omitted.

Although choosing $\mathscr{P}_1$ to be nonconsonant allows the mixture procedure to gain power, there is a risk of violating the all-too-important parallel gatekeeping condition. The following example illustrates this phenomenon.

Example 3 Using a setting very similar to the one used in Example 2, consider a two-family hypothesis testing problem with $F_1 = \{H_1, H_2, H_3\}$ and $F_2 = \{H_4\}$. The raw one-sided p values for these null hypotheses along with their adjusted p values are displayed in Table 3. The adjusted p values are computed using a mixture gatekeeping procedure based on the truncated Hommel procedure with $\gamma = \frac{3}{4}$ in the primary family and the regular Hommel procedure in the secondary family. It follows from Table 3 that none of the primary null hypotheses can be rejected at one-sided $\alpha = 0.025$; however, the mixture gatekeeping procedure still rejects the secondary null hypothesis. This clearly violates the parallel gatekeeping condition and is due to the fact that the truncated Hommel procedure is nonconsonant.

To address this problem, the adjusted p values in the secondary family need to be modified to enforce the parallel gatekeeping condition. This can be accomplished using the readjustment algorithm suggested by Kordzakhia et al. (2008). The

Table 3 Raw and adjusted p values in Example 3

Family	Null hypothesis	Raw p value	Adjusted p value (mixture procedure)
Primary	H_1	0.0125	0.0262
	H_2	0.0143	0.0262
	H_3	0.0218	0.0262
Secondary	H_4	0.0010	0.0245*

*Significant at the one-sided 0.025 level. Note that after the readjustment algorithm is applied, the adjusted p value for H_4 is set to 0.0262 and thus it satisfies the parallel gatekeeping condition.

adjusted p values for the secondary null hypotheses are readjusted to

$$\tilde{p}_i' = \max\left(\tilde{p}_i, \min_{j \in N_1} \tilde{p}_j\right), \quad i \in N_2$$

where $\tilde{p}_i$ is the adjusted p value for the null hypothesis H_i produced by the mixture gatekeeping procedure. In this example, the parallel gatekeeping condition is enforced by setting the adjusted p value for H_4 to 0.0262 and thus making it nonsignificant at $\alpha = 0.025$.

It is natural to ask whether or not the power advantage due to using a nonconsonant primary component in the mixture gatekeeping procedure is lost if the p values need to be readjusted. Example 2 shows that this is generally not the case. Specifically, the re adjusted p value for the null hypothesis H_5 is equal to the original adjusted p value and thus even after readjustment the mixture gatekeeping procedure rejects more null hypotheses than the two-stage gatekeeping procedure.

3. α-EXHAUSTIVE THREE-STAGE PARALLEL GATEKEEPING PROCEDURES

An important feature of the two-stage gatekeeping procedure introduced in section 2 is that it satisfies the independence condition. The closed representation of the two-stage gatekeeping procedure introduced in section 2 provides an insight into this property. As shown later, the independence is achieved by testing some of the intersection hypotheses in the closed family at a level that is less than α. Multiple testing procedures of this kind are known as non-α-exhaustive procedures and, as demonstrated by Grechanovsky and Hochberg (1999), one can build uniformly more powerful procedures by forcing the size of all local tests in the closed family to be exactly α.

The use of α-exhaustive gatekeeping procedures based on the Bonferroni procedure has been discussed by Dmitrienko et al. (2005, chapter 2), Guilbaud (2007), Bretz et al. (2009), and Burman et al. (2009). These gatekeeping procedures do not satisfy the independence condition and cannot be expressed as multistage procedures that test families sequentially from the first one to the last one. However, Guilbaud (2007) proved that an alternative multistage representation exists. This representation involves retesting; i.e., the families are first tested sequentially and then, if certain additional conditions are met, the families are retested in a reverse order using more powerful procedures than the ones used originally. It is shown in this section that the power of the two-stage gatekeeping procedure can be improved uniformly by constructing an α-exhaustive mixture gatekeeping procedure and, further, this mixture gatekeeping procedure is actually based on a multistage algorithm with retesting.

Consider the two-family hypothesis testing problem studied in section 2 but suppose that the independence condition is not applicable because of the nature of the multiple objectives addressed in the trial. Selecting two component procedures and assuming that $\mathscr{P}_1$ is consonant, the two-stage gatekeeping procedure is equivalent to the mixture gatekeeping procedure derived from the same two components. Note that the decision rules used in this closed procedure tests all intersection hypotheses $H(I)$ at level α with the exception of local tests for $H(I)$ with $I = I_1 \subset N_1$. In this case, the local p value is given by $p(I) = p_1(I_1)$ and the level of the associated test is strictly less than α. This follows from the fact that $\mathscr{P}_1$ is a separable procedure, and so under $H(I)$,

$$P(p(I) \leq \alpha) = P(p_1(I_1) \leq \alpha) < \alpha$$

if $I_1 \subset N_1$. Given this, it is easy to uniformly improve the power of the mixture gatekeeping procedure without compromising global FWER control. This is achieved by increasing the size of the local tests for $H(I)$ with $I = I_1 \subset N_1$ to α.

To define the α-exhaustive mixture gatekeeping procedure, let $\mathscr{P}_1^*$ denote an α-exhaustive version of $\mathscr{P}_1$. It is also a closed procedure with the local p values for the intersection hypotheses in the closed family denoted by $p_1^*(I_1)$, $I_1 \subseteq N_1$. For example, an α-exhaustive version of the Bonferroni procedure is the Holm procedure and an α-exhaustive version of any truncated procedure is the regular version of that procedure. To illustrate, consider the mixture gatekeeping procedure defined in Example 1 of section 2. The primary component procedure $\mathscr{P}_1$ in this example is the truncated Hochberg procedure. An α-exhaustive version of $\mathscr{P}_1$ is the regular Hochberg procedure and thus the local p values for all the intersection hypotheses are given by

$$p_1^*(I_1) = \begin{cases} \min(2p_{(1)}, p_{(2)}) & \text{if } I_1 = \{1, 2\} \\ p_1 & \text{if } I_1 = \{1\} \\ p_2 & \text{if } I_1 = \{2\} \end{cases}$$

It is easy to see that $p_1^*(I_1) \leq p_1(I_1)$ for all I_1 and the regular Hochberg procedure provides a uniform power advantage over the truncated Hochberg procedure. In general, the local p values for $\mathscr{P}_1^*$ are chosen to ensure that the local test for each intersection hypothesis $H(I_1)$ is α-level and thus the α-exhaustive procedure $\mathscr{P}_1^*$ is uniformly more powerful than the original procedure $\mathscr{P}_1$.

As in section 2, we define the α-exhaustive mixture gatekeeping procedure by specifying a local p value for each intersection hypothesis in the closed family. We assume for simplicity that the error rate function of $\mathscr{P}_1$ is proportional to α. Select an arbitrary nonempty index set $I \subseteq N$ and let $I_1 = I \cap N_1$ and $I_2 = I \cap N_2$. The local

p value for the intersection hypothesis $H(I)$ is defined as follows:

- Case 1 (the intersection hypothesis includes only the primary null hypotheses, i.e., $I = I_1$ and I_2 is empty). The local p value is defined as $p(I) = p_1^*(I_1)$.
- Case 2 (the intersection hypothesis includes only the secondary null hypotheses, i.e., $I = I_2$ and I_1 is empty). The local p value is defined as $p(I) = p_2(I_2)$.
- Case 3 (the intersection hypothesis includes both primary and secondary null hypotheses, i.e., $I = I_1 \cup I_2$ and I_1 and I_2 are both nonempty). The local p value is defined as

$$p(I) = \min\left(p_1(I_1), \frac{p_2(I_2)}{1 - f_1(I_1)}\right)$$

where $f_1(I_1) = e_1(I_1 \mid \alpha)/\alpha$.

It is easy to verify that the resulting mixture gatekeeping procedure is α-exhaustive and still controls the global FWER in the strong sense.

As an illustration, the α-exhaustive mixture gatekeeping procedure in Example 1 is based on the local p values displayed in Table 1 with the exception of the local p values for the intersection hypotheses $H(I)$ with $I = \{1, 2\}$, $\{1\}$ and $\{2\}$. For these intersection hypotheses, $p(I) = p_1^*(I)$.

The α-exhaustive mixture gatekeeping procedure defined earlier admits a useful three-stage representation with retesting. The main difference between this three-stage procedure and the two-stage gatekeeping procedure is that to construct an α-exhaustive procedure, one needs to return to the primary family if all secondary null hypotheses are rejected. The first two stages are the same as before, but the following third stage is added.

- Stage 3. If all secondary null hypotheses are rejected but some primary null hypotheses are accepted, then retest those accepted primary null hypotheses using the α-exhaustive component procedure $\mathscr{P}_1^*$ at level $\alpha_3 = \alpha$.

As was noted earlier, a key feature of the three-stage gatekeeping procedure is that the primary null hypotheses may be tested twice. They are tested in Stage 1 using $\mathscr{P}_1$ at the full α level and, further, if all secondary null hypotheses are rejected in Stage 2, the primary null hypotheses are retested using the α-exhaustive version $\mathscr{P}_1^*$ of $\mathscr{P}_1$ again at the full α level. The three-stage procedure rejects at least as many null hypotheses as the two-stage procedure ($\mathscr{P}_1^*$ will reject the null hypotheses rejected by $\mathscr{P}_1$ and possibly some that were accepted by $\mathscr{P}_1$) and is hence more attractive than the two-stage procedure if the independence condition is not imposed.

The three-stage procedure provides a generalization of Bonferroni-based chain procedures introduced in Bretz et al. (2009). It is shown in section 4 that the method defined earlier can be used to construct α-exhaustive multistage gatekeeping procedures based on more powerful component procedures, e.g., truncated Hommel procedure, truncated Hochberg procedure, etc.

The following proposition is an extension of a theorem in Guilbaud (2007, section 5) that deals with a three-stage representation of α-exhaustive Bonferroni-based gatekeeping procedures.

Proposition 5. *For any separable and consonant FWER-controlling $\mathcal{P}_1$ and a general FWER-controlling $\mathcal{P}_2$, the α-exhaustive mixture parallel gatekeeping procedure is equivalent to the three-stage parallel gatekeeping procedure.*

Example 1 Revisited We return to Example 1 from section 2 to illustrate the three-stage gatekeeping procedure with retesting and compare it with the two-stage gatekeeping procedure. The first two stages in the three-stage procedure are identical to those in the two-stage procedure. Recall that the two-stage procedure rejected one primary null hypothesis in Stage 1 and two secondary null hypotheses in Stage 2 and, since both secondary null hypotheses are rejected, the primary null hypotheses H_1 and H_2 are retested in Stage 3:

- Stage 3. The primary null hypotheses were tested using the truncated Hochberg procedure in Stage 1 and thus H_1 and H_2 are retested in Stage 3 using the regular Hochberg procedure at level $\alpha_3 = \alpha$. Since $p_2 = 0.0193 < \alpha_3 = 0.025$, the Hochberg procedure rejects H_2 and hence also H_1 (which was, of course, rejected by $\mathcal{P}_1$). Note that H_2 was not rejected by the two-stage procedure.

The three-stage gatekeeping procedure gains power by retesting the primary null hypotheses at the full α level and, in this example, this translates into an additional rejected null hypothesis in the primary family.

4. GATEKEEPING PROCEDURES FOR MULTIFAMILY HYPOTHESIS TESTING PROBLEMS

The parallel gatekeeping procedures introduced in sections 2 and 3 were formulated for testing problems with two families of null hypotheses. In this section we briefly outline how the new methods can be generalized to testing problems with an arbitrary number of families.

Consider a clinical trial with n null hypotheses grouped into $m \geq 2$ families. The ith family of null hypotheses is defined as $F_i = \{H_j : j \in N_i\}$, where the index sets $N_1, \dots, N_m$ are defined as

$$N_1 = \{1, \dots, n_1\}, \quad N_i = \{n_1 + \cdots + n_{i-1} + 1, \dots, n_1 + \cdots + n_i\}, \quad i = 2, \dots, m$$

and $n_1 + \cdots + n_m = n$. The families are tested sequentially beginning with F_1, and F_i is a parallel gatekeeper for F_{i+1} $(i = 1, \dots, m-1)$.

We begin with an extension of the mixture parallel gatekeeping procedure defined in section 2. Consider the problem of constructing a mixture of the component procedures, $\mathcal{P}_1, \dots, \mathcal{P}_m$, used in these m families. It is assumed that each component procedure is closed and provides local FWER control within the corresponding family and $\mathcal{P}_1, \dots, \mathcal{P}_{m-1}$ are separable. Further, let $e_i(I_i \mid \alpha)$ denote the error rate function of $\mathcal{P}_i$. We assume, as in section 2, that the error rate functions are proportional to α so that the fractions $f_i(I_i) = e_i(I_i \mid \alpha)/\alpha$ are independent of α, $i = 1, \dots, m-1$.

We now define a general mixture gatekeeping procedure for testing the null hypotheses in the combined family $F = F_1 \cup \cdots \cup F_m$ using the closure principle. Consider an arbitrary nonempty index set $I \subseteq N$, where $N = \{1, \dots, n\}$, and let

$I_i = I \cap N_i$, $i = 1, \ldots, m$. The local p value for the intersection hypothesis $H(I)$ is defined as follows:

- Case 1 ($I = I_i$ for some $i = 1, \ldots, m$). The local p value for $H(I)$ is defined as $p(I) = p_i(I_i)$, where $p_i(I_i)$ is the local p value for $H(I_i)$ using the component procedure $\mathcal{P}_i$.
- Case 2 ($I = I_{i_1} \cup \cdots \cup I_{i_s}$, where $I_{i_1}, \ldots, I_{i_s}$ are nonempty and $s \geq 2$). For notational simplicity and without loss of generality, relabel the index sets so that $I_{i_1} = I_1, \ldots, I_{i_s} = I_s$. Then the local p value for $H(I)$ is defined as

$$p(I) = \min\left(\frac{p_1(I_1)}{b_1}, \frac{p_2(I_2)}{b_2}, \ldots, \frac{p_s(I_s)}{b_s}\right)$$

  where b_i, $i = 1, \ldots, s$, are defined as before.

The resulting gatekeeping procedure rejects the null hypothesis H_i if and only if $p(I) \leq \alpha$ for all index sets I containing the index i. As in Proposition 1, it can be shown that each local p value defines an α-level local test and thus, by the closure principle, this gatekeeping procedure controls the global FWER strongly at level α. Properties of general mixture gatekeeping procedures will be studied in a separate paper.

Further, an extension of the α-exhaustive multistage gatekeeping procedure introduced in section 3 to hypothesis testing problems with $m \geq 3$ families of null hypotheses can be constructed along the lines of the multistage algorithm proposed in Guilbaud (2007, section 5.3) for Bonferroni-based gatekeeping procedures. Let $\mathcal{P}_i^*$ denote an α-exhaustive version of the component procedure $\mathcal{P}_i$, $i = 1, \ldots, m$ (note that $\mathcal{P}_m^* = \mathcal{P}_m$). The general α-exhaustive multistage gatekeeping procedure with retesting is as follows:

- Stage 1. Test the null hypotheses in F_1 using $\mathcal{P}_1$ at level $\alpha_1 = \alpha$. Let A_1 denote the index set of the null hypotheses accepted by $\mathcal{P}_1$.
- Stage $i = 2, \ldots, m$. If at least one null hypothesis is rejected in F_{i-1} then test the null hypotheses in F_i using $\mathcal{P}_i$ at level $\alpha_i = \alpha_{i-1} - e_{i-1}(A_{i-1} \mid \alpha_{i-1})$. Let A_i denote the index set of the null hypotheses accepted by $\mathcal{P}_i$.
- Stage $m + 1$. If all null hypotheses are rejected in F_m then retest the null hypotheses in F_{m-1} using the α-exhaustive procedure $\mathcal{P}_{m-1}^*$ at level α_{m-1}.
- Stage $i = m + j$ ($j = 2, \ldots, m - 1$). If all null hypotheses are rejected in F_{m-j+1}, the null hypotheses in F_{m-j} are retested using the α-exhaustive component procedure $\mathcal{P}_{m-j}^*$ at level α_{m-j}.

Using arguments similar to those utilized in the proof of Proposition 5, it can be shown that this multistage gatekeeping procedure is equivalent to an α-exhaustive version of the general mixture gatekeeping procedure. This α-exhaustive procedure is a closed procedure based on the local p values defined as follows:

- Case 1 ($I = I_i$ for some $i = 1, \ldots, m$). The local p value for $H(I)$ is defined as $p(I) = p_i^*(I_i)$, where $p_i^*(I_i)$ is the local p value for $H(I_i)$ using the component procedure $\mathcal{P}_i^*$.

- Case 2 ($I = I_{i_1} \cup \cdots \cup I_{i_s}$, where $I_{i_1}, \ldots, I_{i_s}$ are nonempty and $s \geq 2$). Assuming again that $I_{i_1} = I_1, \ldots, I_{i_s} = I_s$, the local p value for $H(I)$ is defined as

$$p(I) = \min\left(\frac{p_1(I_1)}{b_1}, \frac{p_2(I_2)}{b_2}, \ldots, \frac{p_s^*(I_s)}{b_s}\right)$$

where $b_1 = 1$ and $b_i = b_{i-1}[1 - f_{i-1}(I_{i-1})]$, $i = 2, \ldots, s$.

The equivalence implies that the multistage parallel gatekeeping procedure defined earlier is also α-exhaustive and controls the global FWER over all m families in the strong sense at α.

5. SOFTWARE IMPLEMENTATION

Multistage parallel gatekeeping procedures with and without retesting can be implemented using the R package developed by Alex Dmitrienko, Eric Nantz, and Gautier Paux (Multxpert package). For more information on the Multxpert package, visit the Multiplicity Expert web site at www.multxpert.com.

The ParGateAdjP function included in this package computes adjusted p values and generates decision rules for general multistage gatekeeping procedures defined in section 4. As an illustration, consider the two-family hypothesis testing problem in Example 1. The raw p values in the primary and secondary families are specified as follows:

```
# Primary family
rawp1<-c(0.0110,0.0193)
label1<-"Primary endpoints"
# Secondary family
rawp2<-c(0.0042,0.0057)
label2<-"Secondary endpoints"
```

The second step is to define the primary and secondary component procedures, e.g., the truncated Hochberg procedure (truncation parameter $\gamma = 0.5$) and regular Hochberg procedure (truncation parameter $\gamma = 1$):

```
# Primary family
family1<-list(label=label1, rawp=rawp1, proc="Hochberg",
    procpar=0.5)
# Secondary family
family2<-list(label=label2, rawp=rawp2, proc="Hochberg",
    procpar=1)
# List of gatekeeping procedure parameters
gateproc<-list(family1, family2)
```

To compute the adjusted p values for the two-stage procedure, the independence parameter in the ParGateAdjP function is set to TRUE (the independence condition is imposed):

```
pargateadjp(gateproc, independence=TRUE)
```

The resulting adjusted p values are given by

$$\tilde{p}_1 = 0.0220, \quad \tilde{p}_2 = 0.0257, \quad \tilde{p}_3 = 0.0228, \quad \tilde{p}_4 = 0.0228$$

To implement the three-stage procedure with retesting, the independence parameter is set to FALSE:

```
pargateadjp(gateproc, independence=FALSE)
```

The three-stage procedure produces the following adjusted p values:

$$\tilde{p}_1 = 0.0220, \quad \tilde{p}_2 = 0.0228, \quad \tilde{p}_3 = 0.0228, \quad \tilde{p}_4 = 0.0228$$

These adjusted p values are uniformly smaller than those produced by the two-stage procedure, which illustrates the higher power of the three-stage procedure.

The ParGateAdjP function can also generate decision rules for multistage gatekeeping procedures. To obtain decision rules for the three-stage procedure, the global FWER (alpha) needs to be specified and the printDecisionRules parameter needs to be set to TRUE:

```
pargateadjp(gateproc, independence=FALSE, alpha=0.025,
  printDecisionRules=TRUE)
```

This function call produces the following output, which shows the individual steps in the underlying algorithm:

```
Family 1 (Primary endpoints) is tested using Hochberg procedure
(truncation parameter=0.5) at alpha1=0.025.
Null hypothesis 1 is rejected.
Null hypothesis 2 is accepted.

One or more null hypotheses are rejected in Family 1 and the
parallel gatekeeping procedure passes this family. Based on
the error rate function of Hochberg procedure (truncation
parameter=0.5), alpha2=0.0062 is carried over to Family 2.

Family 2 (Secondary endpoints) is tested using Hochberg
procedure (truncation parameter=1) at alpha2=0.0062.
Null hypothesis 3 is rejected.
Null hypothesis 4 is rejected.

All null hypotheses are rejected in Family 2 and the parallel
gatekeeping procedure passes this family. Retesting begins and
alpha3=0.025 is carried over to Family 1.

Family 1 (Primary endpoints) is retested using Hochberg
procedure (truncation parameter=1) at alpha3=0.025.
Null hypothesis 1 is rejected.
Null hypothesis 2 is rejected.
```

APPENDIX

Proof of Proposition 1

By the closure principle, the mixture gatekeeping procedure controls the global FWER at level α if each local test is an α-level test. We now verify that the local p values defined earlier give α-level tests of the intersection hypothesis $H(I) = H(I_1) \cap H(I_2)$. Assume that $H(I)$ is true and hence both $H(I_1)$ and $H(I_2)$ are true. In Case 1 or Case 2, we have $p(I) = p_1(I_1)$ or $p(I) = p_2(I_2)$, respectively, and thus the test for $H(I)$ is an α-level test. In Case 3, by the definition of $p(I)$ and the Bonferroni inequality,

$$\begin{aligned} P(p(I) \leq \alpha) &= P(p_1(I_1) \leq \alpha \text{ or } p_2(I_2) \leq \alpha(1 - f_1(I_1)) \\ &\leq P(p_1(I_1) \leq \alpha) + P(p_2(I_2) \leq \alpha(1 - f_1(I_1)) \end{aligned}$$

Since $H(I_1)$ is true, it follows from the definition of the error rate function that

$$P(p_1(I_1) \leq \alpha) \leq \alpha f_1(I_1)$$

Further, since $H(I_2)$ is true, we have

$$P(p_2(I_2) \leq \alpha(1 - f_1(I_1)) \leq \alpha(1 - f_1(I_1))$$

Adding the two inequalities we get the desired result that $P(p(I) \leq \alpha) \leq \alpha$ and thus the mixture gatekeeping procedure controls FWER $\leq \alpha$.

Proof of Proposition 2

The proof consists of two parts. We show in Part 1 that the mixture gatekeeping procedure rejects every null hypothesis rejected by the two-stage gatekeeping procedure. Further, assuming that $\mathcal{P}_1$ is consonant, it is demonstrated in Part 2 that any null hypothesis rejected by the mixture gatekeeping procedure is also rejected by the two-stage gatekeeping procedure.

Part 1. Suppose that the two-stage procedure rejects a primary null hypothesis H_i. Therefore $\mathcal{P}_1$ rejects H_i at level α and, since $\mathcal{P}_1$ is a closed procedure, $p_1(I_1) \leq \alpha$ for all $I_1 \subseteq N_1$ such that $i \in I_1$. Now select any index set $I \subseteq N$ that contains i and let $I_1 = I \cap N_1$ and $I_2 = I \cap N_2$. Consider the following two cases:

- Case 1 ($I_2 = \emptyset$). In this case we have $p(I) = p_1(I_1) \leq \alpha$.
- Case 2 ($I_2 \neq \emptyset$). In this case we have

$$p(I) = \min\left(p_1(I_1), \frac{p_2(I_2)}{1 - f_1(I_1)}\right) \leq p_1(I_1) \leq \alpha$$

In both cases, $p(I) \leq \alpha$ for any index set I containing i, which implies that the mixture procedure rejects the primary null hypothesis H_i.

Next suppose that the two-stage procedure rejects a secondary null hypothesis H_j, which means that $\mathcal{P}_2$ rejects H_j at level $\alpha_2 = \alpha - e_1(A_1 \mid \alpha)$. Select any $I \subseteq N$ such that $j \in I$ and let $I_1 = I \cap N_1$ and $I_2 = I \cap N_2$. Also let $R_1 = N_1 \backslash A_1$ be the index set of the rejected null hypotheses. Consider the following two cases:

- Case 1 ($I_1 \cap R_1 \neq \emptyset$). Since I_1 includes indices of null hypotheses rejected by $\mathscr{P}_1$ at level α, we have $p_1(I_1) \leq \alpha$. Thus we conclude that

$$p(I) = \min\left(p_1(I_1), \frac{p_2(I_2)}{1 - f_1(I_1)}\right) \leq p_1(I_1) \leq \alpha$$

- Case 2 ($I_1 \cap R_1 = \emptyset$). Here $I_1 \subseteq A_1$. By the monotonicity of $e_1(A_1 \mid \alpha)$, we have $f_1(I_1) \leq f_1(A_1)$. Since $\mathscr{P}_2$ rejects H_j at level $\alpha_2 = \alpha - e_1(A_1 \mid \alpha)$ and $j \in I_2$, we have

$$p_2(I_2) \leq \alpha - e_1(A_1 \mid \alpha) = \alpha(1 - f_1(A_1)) \leq \alpha(1 - f_1(I_1))$$

  and so

$$p(I) = \min\left(p_1(I_1), \frac{p_2(I_2)}{1 - f_1(I_1)}\right) \leq \frac{p_2(I_2)}{1 - f_1(I_1)} \leq \alpha$$

Since in both cases $p(I) \leq \alpha$ if I includes j, the mixture procedure rejects the secondary null hypothesis H_j.

Part 2. Suppose that the mixture procedure rejects a primary null hypothesis H_i. In other words, $p(I) \leq \alpha$ for all $I \subseteq N$ which contain i. It immediately follows that $p_1(I_1) \leq \alpha$ for any $I_1 \subseteq N_1$ if $i \in I_1$. Hence, $\mathscr{P}_1$ rejects H_i at level α and so H_i is rejected by the two-stage procedure.

Next suppose that the mixture procedure rejects a secondary null hypothesis H_j. Consider any index set $I = A_1 \cup I_2$, where I_2 is an arbitrary subset of N_2 such that $j \in I_2$. If $p_1(A_1) \leq \alpha$, then by the consonance property, $\mathscr{P}_1$ would reject at least one primary null hypothesis H_i, $i \in A_1$. However, all null hypotheses in A_1 are accepted, which implies that $p_1(A_1) > \alpha$. On the other hand, the mixture procedure rejects H_j and thus $p(I) \leq \alpha$. Therefore we must have

$$\frac{p_2(I_2)}{1 - f_1(A_1)} \leq \alpha$$

This implies that

$$p_2(I_2) \leq \alpha(1 - f_1(A_1)) = \alpha - e_1(A_1 \mid \alpha)$$

Since this is true for any $I_2 \subseteq N_2$ with $j \in I_2$, we conclude that $\mathscr{P}_2$ rejects H_j at level $\alpha_2 = \alpha - e_1(A_1 \mid \alpha)$, which implies that the secondary null hypothesis H_j is rejected by the two-stage procedure. The proof of Proposition 2 is complete.

Proof of Proposition 5

Using the same logic as in the proof of Proposition 2, it is easy to see that identical decision rules are used by the two procedures for the secondary null hypotheses and thus it is sufficient to focus on the primary null hypotheses.

Part 1. Suppose that the three-stage procedure with retesting rejects a primary null hypothesis H_i, and consider the following two cases:

- Case 1 (H_i is rejected in Stage 1). In this case, $p_1(I_1) \leq \alpha$ for all index sets $I_1 \subseteq N_1$ such that $i \in I_1$. Select any index set $I = I_1 \cup I_2$ with $I_2 \subseteq N_2$ and note that

$$p(I) = \min\left(p_1(I_1), \frac{p_2(I_2)}{1 - f_1(I_1)}\right) \leq p_1(I_1) \leq \alpha$$

- Case 2 (H_i is rejected in Stage 3). In this case, $\mathscr{P}_2$ rejects all secondary null hypotheses at level $\alpha_2 = \alpha - e_1(A_1 \mid \alpha)$ (note that A_1 is the index set of the primary null hypotheses accepted by $\mathscr{P}_1$ but not necessarily by $\mathscr{P}_1^*$) and $\mathscr{P}_1^*$ rejects H_i at level α. Select any index set I which contains i. Consider three subcases:
  - Case 2A ($I = I_1 \subseteq N_1$). Recall that $p(I) = p_1^*(I_1)$ since $I \subseteq N_1$ and, further, $p_1^*(I_1) \leq \alpha$ for any $I_1 \subseteq N_1$ with $i \in I_1$ since H_i is rejected by $\mathscr{P}_1^*$ at level α in Stage 3.
  - Case 2B ($I = I_1 \cup I_2$, $I_1, I_2 \neq \emptyset$, $I_1 \cap R_1 \neq \emptyset$). In this case, $p_1(I_1) \leq \alpha$ since all primary null hypotheses H_j with $j \in R_1$ are rejected by $\mathscr{P}_1$ at level α. This implies that, for any $I_2 \subseteq N_2$,

$$p(I) = \min\left(p_1(I_1), \frac{p_2(I_2)}{1 - f_1(I_1)}\right) \leq p_1(I_1) \leq \alpha$$

  - Case 2C ($I = I_1 \cup I_2$, $I_1, I_2 \neq \emptyset$, $I_1 \cap R_1 = \emptyset$). In this case, $I_1 \subseteq A_1$. By the monotonicity of $e_1(A_1 \mid \alpha)$, we have $f_1(I_1) \leq f_1(A_1)$. Since $\mathscr{P}_2$ rejects all secondary null hypotheses H_j at level $\alpha_2 = \alpha - e_1(A_1 \mid \alpha)$, we have, for any $I_2 \subseteq N_2$,

$$p_2(I_2) \leq \alpha - e_1(A_1 \mid \alpha) = \alpha(1 - f_1(A_1)) \leq \alpha(1 - f_1(I_1))$$

  and thus

$$p(I) = \min\left(p_1(I_1), \frac{p_2(I_2)}{1 - f_1(I_1)}\right) \leq \frac{p_2(I_2)}{1 - f_1(I_1)} \leq \alpha$$

In all cases just considered, $p(I) \leq \alpha$ for any index set I containing i and thus the α-exhaustive mixture procedure rejects the primary null hypothesis H_i.

Part 2. Suppose that the α-exhaustive mixture procedure rejects a primary null hypothesis H_i. This implies that $p(I) \leq \alpha$ for any index set I with $i \in I$. Given this, consider the following two cases:

- Case 1 ($p_1(I_1) \leq \alpha$ for all index sets $I_1 \subseteq N_1$ containing i). In this case, H_i is rejected by $\mathscr{P}_1$ at level α in Stage 1.
- Case 2 ($p_1(I_1) > \alpha$ for some index sets $I_1 \subseteq N_1$ containing i). In this case, H_i is not rejected by $\mathscr{P}_1$ at α and thus $i \in A_1$. Now consider any index set $I = A_1 \cup I_2$,

where $I_1 = A_1$ and I_2 is an arbitrary subset of N_2. Recall that $i \in I$ and thus the intersection hypothesis $H(I)$ is rejected by the α-exhaustive mixture procedure, i.e.,

$$p(I) = \min\left(p_1(I_1), \frac{p_2(I_2)}{1 - f_1(I_1)}\right) \leq \alpha$$

However, due to the assumption that $\mathcal{P}_1$ is consonant, we have $p_1(I_1) = p_1(A_1) > \alpha$. This implies that

$$p_2(I_2) \leq \alpha(1 - f_1(I_1)) = \alpha(1 - f_1(A_1)) = \alpha - e_1(A_1 \mid \alpha)$$

for any $I_2 \subseteq N_2$ and thus all secondary null hypotheses are rejected by $\mathcal{P}_2$ at level $\alpha_2 = \alpha - e_1(A_1 \mid \alpha)$ in Stage 2. Further, consider any index set $I_1 \subseteq N_1$ containing i. Since $i \in I_1$, the intersection hypothesis $H(I_1)$ is rejected by the α-exhaustive mixture procedure and thus $p_1(I_1) \leq \alpha$. On the other hand, $p(I) = p_1^*(I_1)$, where $I_1 = I$. We conclude that $p_1^*(I_1) \leq \alpha$ for any $I_1 \subseteq N_1$ and thus H_i is rejected by the α-exhaustive primary component procedure $\mathcal{P}_1^*$ at α after all secondary null hypotheses are rejected by $\mathcal{P}_2$ (this rejection occurs in Stage 3).

These two cases demonstrate that the primary null hypothesis H_i is rejected by the three-stage procedure. The proof of Proposition 5 is complete.

DISCLAIMER

Views expressed in this paper are authors' personal views and not necessarily those of the U.S. Food and Drug Administration.

REFERENCES

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

Brechenmacher, T., Xu, J., Dmitrienko, A., Tamhane, A.C. (2011). A mixture gatekeeping procedure based on the Hommel test for clinical trial applications. *Journal of Biopharmaceutical Statistics* 21:748–767.

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

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

Dmitrienko, A., Tamhane, A. C. (2011). Mixtures of multiple testing procedures for gatekeeping applications in clinical trials. To appear in *Statistics in Medicine*, in press.

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.

Dmitrienko, A., Molenberghs, G., Chuang-Stein, C., Offen, W. (2005). *Analysis of Clinical Trials Using SAS: A Practical Guide*. Cary, NC: SAS Press, pp. 67–127.

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

Gabriel, K. R. (1969). Simultaneous test procedures – some theory of multiple comparisons. *Annals of Mathematical Statistics* 40:224–250.

Grechanovsky, E., Hochberg, Y. (1999). Closed procedures are better and often admit a shortcut. *Journal of Statistical Planning and Inference* 76:79–91.

Guilbaud, O. (2007). Bonferroni parallel gatekeeping—transparent generalizations, adjusted *p* values and short direct proofs. *Biometrical Journal* 49:217–227.

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

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

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

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

Kordzakhia, G., Dinh, P., Bai, S., Lawrence, J., Yang, P. (2008). Bonferroni-based tree-structured gatekeeping testing procedures. Unpublished manuscript.

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

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.

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

Westfall, P. H., Tobias, R. D., Rom, D., Wolfinger, R. D., Hochberg, Y. (1999). *Multiple Comparisons and Multiple Tests Using the SAS System*. Cary, NC: SAS Press.