# General Theory of Mixture Procedures for Gatekeeping

Alex Dmitrienko (Quintiles)
Ajit C. Tamhane (Northwestern University)

**Abstract**

The paper introduces a general approach to constructing mixture-based gatekeeping procedures in multiplicity problems with two or more families of hypotheses. Mixture procedures serve as extensions of and overcome limitations of some previous gatekeeping approaches such as parallel gatekeeping and tree-structured gatekeeping. This paper offers a general theory of mixture procedures constructed from nonparametric (p-value based) or parametric (normal theory based) procedures and studies their properties. It is also shown that the mixture procedure for parallel gatekeeping is equivalent to to the multistage gatekeeping procedure. A clinical trial example is used to illustrate the mixture approach and implementation of mixture procedures.

## 1 Introduction

Gatekeeping procedures address the problems of testing logically related hypotheses that are grouped into hierarchically ordered families. Such problems arise in clinical trials involving multiple endpoints, noninferiority/superiority tests, multiple doses, etc. This has been an active area of research in the last decade. Much of this work deals with serial gatekeeping (Maurer, Hothorn and Lehmacher, 1995; Bauer *et al.*, 1998; Westfall and Krishen, 2001) and parallel gatekeeping (Dmitrienko, Offen and Westfall, 2003; Dmitrienko, Tamhane and Wiens, 2008), and their generalization to tree-structured gatekeeping (Dmitrienko *et al.*, 2007; Dmitrienko *et al.*, 2008a). However, this generalization does not cover all types of logical restrictions that are employed in clinical trial applications. Furthermore, the tree-structured gatekeeping approach is designed to use only the Bonferroni procedure for testing the intersection hypotheses within the closed testing framework (Marcus, Peritz and Gabriel 1976), so there is a potential for improving its power.

Dmitrienko *et al.* (2008b) proposed a general multistage (or stepwise) parallel gatekeeping procedure that uses more powerful (than Bonferroni) component multiple testing procedures (MTPs) for the individual families. Dmitrienko and Tamhane (2011) introduced an alternative procedure, called the mixture procedure, based on the closure principle. This procedure not only allows the use of more powerful component MTPs but it can also deal with more general logical restrictions than the tree-structured gatekeeping approach allows. In Dmitrienko and Tamhane (2011) we restricted to only two families of hypotheses and the exposition of the ideas was mainly through examples. In this paper we extend the mixture procedures to arbitrary number of families; in addition we provide their general theory and study their properties. Several examples are given to illustrate the mixture procedures.

The outline of the paper is as follows. Section 2 introduces some background and notation. Section 3 reviews the multistage parallel gatekeeping procedures. Section 4 defines the basic mixture procedure for parallel gatekeeping and shows that it is equivalent to the multistage procedure. Section 5 extends the mixture framework to general gatekeeping restrictions through restriction functions. Section 6 introduces a clinical trial example to illustrate mixture procedures with general gatekeeping restrictions. Calculations for this example were done by using R programs available at http://multxpert.com/. Finally, Section 7 gives concluding remarks. Proofs of the theoretical results are given in the Appendix.

# 2 Background and notation

Consider the problem of testing $n \geq 2$ hypotheses, $H_1, \ldots, H_n$, which are grouped into $m \geq 2$ ordered families, $F_j = \{H_i : i \in N_j\}$, where $N_1 = \{1, \ldots, n_1\}, N_j = \{n_1+\ldots+n_{j-1}+1, \ldots, n_1+\ldots+n_j\}$ are the index sets of the hypotheses ($j = 2, \ldots, m$) and $\sum_{j=1}^m n_j = n$. Let $F = \bigcup_{j=1}^m F_j$ be the overall family and $N = \bigcup_{j=1}^m N_j = \{1, \ldots, n\}$ be the corresponding index set of all hypotheses. In the simple setting of serial or parallel gatekeeping, family $F_j$ serves as a gatekeeper for family $F_{j+1}$ ($j = 1, \ldots, m-1$). If the gatekeeper $F_j$ is not passed then all hypotheses in families $F_k$ for $k > j$ are regarded as *non-testable*, i.e., they are accepted without tests, while if the gatekeeper $F_j$ is passed then the hypotheses in family $F_{j+1}$ are regarded as *testable*, i.e., they are tested to decide whether to accept or reject them. In *serial gatekeeping*, the condition for passing the gatekeeper is rejection of all hypotheses in the gatekeeper, while in *parallel gatekeeping*, the condition is rejection of at least one hypothesis in the gatekeeper. Since the testability of the hypotheses in later families is conditioned on rejection of the hypotheses in earlier families, the hypotheses are said to have *logical restrictions*.

We assume that all procedures considered in this paper satisfy the *strong family-*

*wise error rate (FWER) control* requirement (Hochberg and Tamhane 1987):

$$\text{FWER} = P(\text{Reject at least one true hypothesis } H_i,\ i \in N) \leq \alpha \tag{1}$$

for a specified $\alpha$ for any combination of the true and false hypotheses in the overall family $F$.

# 3 Multistage procedures for parallel gatekeeping

Dmitrienko *et al.* (2008b) proposed a method for constructing multistage parallel gatekeeping procedures based on the concepts of error rate functions and separable procedures. To define these two concepts, consider a single family $F = \{H_1, \ldots, H_n\}$ with $n \geq 2$ hypotheses and for any nonempty subset $I \subseteq N = \{1, \ldots, n\}$, let $H(I) = \bigcap_{i \in I} H_i$ denote an intersection hypothesis. Then the *error rate function* of an MTP, denoted by $\mathcal{P}$, for a fixed level $\alpha$, is defined as

$$e(I|\alpha) = \sup P\{\text{Reject at least one } H_i,\ i \in I | H(I)\}, \tag{2}$$

where the supremum is taken over the parameter space in which the $H_i,\ i \in I$ are true and the $H_i,\ i \notin I$ are false. Thus $e(I|\alpha)$ is the FWER of $\mathcal{P}$ viewed as a function of $I$ for fixed $\alpha$. The error rate function can always be monotonized so that $e(I|\alpha) \leq e(J|\alpha)$ for $I \subseteq J$ if it is not already monotonic. In addition, $e(\emptyset|\alpha) = 0$ and we can always set $e(N|\alpha) = \alpha$.

Usually, an exact expression for $e(I|\alpha)$ is not available or depends upon unknown correlations among the test statistics. So we use an easy-to-compute upper bound and treat it as the true error rate function. For example, for the Bonferroni procedure, we have $e(I|\alpha) = (|I|/n)\alpha$ where $|I|$ is the cardinality of index set $I$.

A procedure is said to be *separable* if its error rate function satisfies that $e(I|\alpha) < \alpha$ for all proper subsets $I \subset N$. The Bonferroni procedure is clearly separable and the single-step Dunnett procedure (Dunnett, 1955) can also be shown to be separable. On the other hand, most $p$-value based stepwise procedures, e.g., Holm (1979), Hochberg (1988), Hommel (1988), and fallback (Wiens, 2003; Wiens and Dmitrienko, 2005), are not separable. Similarly, the stepwise Dunnett procedures (Naik, 1975; Dunnett and Tamhane, 1992) are not separable.

It is clear from the definition of separable procedures that, when viewed as closed procedures (Marcus *et al.*, 1976), they do not test all intersection hypotheses at full $\alpha$, i.e., they are not $\alpha$-exhaustive. Therefore they are conservative with respect to the corresponding non-separable procedures. Non-separable procedures can be thought of as "greedy" procedures in the sense that a positive significance level is carried forward to the next family only when all hypotheses in a given family are

rejected. Thus, non-separable procedures can be used only for serial gatekeeping. Separable procedures, on the other hand, support more flexible rules for transferring significance levels along a sequence of families. Specifically, separable procedures allow a positive significance level to be carried forward to the next family provided at least one hypothesis is rejected in the current family, which is consistent with the parallel gatekeeping approach considered in this paper.

Non-separable procedures can be made separable by applying them using convex combinations of their critical constants with those of the corresponding separable procedures. For example, the critical constants of the Holm, Hochberg, Hommel or fallback procedures can be combined with those of the Bonferroni procedure and the critical constants of the stepwise Dunnett procedures can be combined with those of the single-step Dunnett procedure. We refer to such hybrid procedures as *truncated procedures*. They are less powerful than the original non-separable procedures but are more powerful than the corresponding separable procedures (Bonferroni or Dunnett). As an example, let $p_1, \ldots, p_n$ denote the raw $p$-values of the hypotheses $H_1, \ldots, H_n$. Further let $p_{(1)} \leq \cdots \leq p_{(n)}$ denote their ordered values and let $H_{(1)}, \ldots, H_{(n)}$ denote the respective ordered hypotheses. Fix $\gamma \in [0, 1]$, called the *truncation fraction*. Then the *truncated Holm procedure* rejects the hypothesis $H_{(i)}$ if and only if (iff)

$$p_{(j)} \leq \left[\frac{\gamma}{n-j+1} + \frac{1-\gamma}{n}\right] \alpha \text{ for } j = 1, \ldots, i.$$

Note that for $\gamma = 0$ this procedure reduces to the Bonferroni procedure and for $\gamma = 1$, it reduces to the original Holm procedure. It was shown in Dmitrienko *et al.* (2008b) that an upper bound on the error rate function of this procedure is given by

$$e(I|\alpha) = \begin{cases} 0 & \text{if } |I| = 0, \\ (\gamma + (1-\gamma)|I|/n)\alpha & \text{if } |I| > 0. \end{cases} \tag{3}$$

It is easy to see that this procedure is separable iff $\gamma < 1$.

A brief discussion about the choice of $\gamma$ is in order. As $\gamma$ is increased for a given family, the power for that family increases often at the expense of the power for the next family. Therefore a proper balance must be struck between these two powers. An optimal choice of $\gamma$ can be made if a clinically relevant optimality criterion is specified as illustrated in Brechenmacher *et al.* (2011), Millen and Dmitrienko (2011) and Dmitrienko *et al.* (2011b).

It should be noted that there are many possible approaches to constructing truncated procedures, e.g., instead of truncating critical points as we have done above, one could truncate $\alpha$ at which the closure procedure tests each intersection hypothesis. We adopted the present approach because it gives simple testing algorithms and simple formulas for error rate functions.

Returning to the gatekeeping problem of testing $n$ hypotheses grouped into $m$ families, the method for constructing multistage parallel gatekeeping procedures proposed in Dmitrienko *et al.* (2008b) assumes that component MTPs $\mathcal{P}_1, \ldots, \mathcal{P}_m$ are specified for the families $F_1, \ldots, F_m$, respectively, such that $\mathcal{P}_1, \ldots, \mathcal{P}_{m-1}$ are separable while $\mathcal{P}_m$ may be non-separable. Denote the error rate function of $\mathcal{P}_j$ by $e_j(I_j|\alpha)$ where $I_j \subseteq N_j$ $(j = 1, \ldots, m)$. The multistage procedure operates as follows.

- Step 1: Test the hypotheses in $F_1$ at level $\alpha_1 = \alpha$ using $\mathcal{P}_1$. Let $A_1 \subseteq N_1$ be the index set of the accepted hypotheses and let $\alpha_2 = \alpha_1 - e_1(A_1|\alpha_1)$. If $A_1 = N_1$, i.e., if all hypotheses in $F_1$ are accepted then $e_1(A_1|\alpha_1) = \alpha_1$ and so $\alpha_2 = 0$, in which case stop testing and accept all remaining hypotheses. Otherwise go to Step 2.

- Step $j$ $(2 \leq j \leq m-1)$: Test the hypotheses in $F_j$ at level $\alpha_j$ using $\mathcal{P}_j$. Let

$$\alpha_{j+1} = \alpha_j - e_j(A_j|\alpha_j),\ (2 \leq j \leq m-1), \tag{4}$$

  where $A_j \subseteq N_j$ is the index set of the accepted hypotheses in $F_j$. If all hypotheses in $F_j$ are accepted using $\mathcal{P}_j$, i.e., if $A_j = N_j$, then $e_j(A_j|\alpha_j) = \alpha_j$ and so $\alpha_{j+1} = 0$, in which case stop testing and accept all hypotheses in $F_k$ for $k > j$. Otherwise go to Step $j + 1$.

- Step $m$: Test the hypotheses in $F_m$ at level $\alpha_m$ using $\mathcal{P}_m$ and stop testing.

**Example 1:** Suppose that $\mathcal{P}_1, \ldots, \mathcal{P}_{m-1}$ are Bonferroni procedures and $\mathcal{P}_m$ is any non-separable procedure. In Dmitrienko *et al.* (2008b) we showed that

$$\alpha_{k+1} = \alpha \prod_{j=1}^{k} \left( \frac{r_j}{n_j} \right),\ k = 1, \ldots, m-1,$$

where $r_j$ is the number of rejected hypotheses from $F_j$. Note that if $r_j = 0$ then $\alpha_k = 0$ for all $k > j$ and so all hypotheses in $F_k$ for $k > j$ are non-testable thus satisfying the parallel gatekeeping condition.

More generally, let $e_j(I_j|\alpha)$ be the error rate function of $\mathcal{P}_j$. Further let $f_j(I_j|\alpha) = e_j(I_j|\alpha)/\alpha$ denote the error rate fraction (the error rate function expressed as a fraction of $\alpha$). Then we have $\alpha_1 = \alpha$ and

$$\alpha_{k+1} = \alpha \prod_{j=1}^{k} [1 - f_j(I_j|\alpha)]\ (1 \leq k \leq m-1). \tag{5}$$

We will use this equation in the sequel. ■

Equation (4) uses what we call as the "use it or lose it" principle. In other words, if the $\alpha$ is not used to reject a hypothesis then it is lost, while if the hypothesis is rejected then it can be carried forward to test subsequent hypotheses in the sequence. This same principle underlies the fixed-sequence (Maurer *et al.*, 1995), fallback (Wiens, 2003; Wiens and Dmitrienko, 2005) and chain (Millen and Dmitrienko, 2011) procedures in problems with a priori ordered hypotheses.

Serial and parallel gatekeeping paradigms were extended in tree-structured gatekeeping (Dmitrienko *et al.*, 2007) by defining serial and parallel rejection sets for each hypothesis $H_i$ (denoted by $R_i^S$ and $R_i^P$, respectively). Specifically, for any hypothesis $H_i \in F_j$ $(j > 1)$, $R_i^S$ and $R_i^P$ are the sets of hypotheses belonging to $F_k$ for $k < j$ such that $H_i$ is testable iff all hypotheses in $R_i^S$ and at least one hypothesis in $R_i^P$ are rejected. Clearly, if $R_i^S = F_{j-1}$ and $R_i^P = \emptyset$ for all $H_i \in F_j$ $(j > 1)$ then we have serial gatekeeping, while if $R_i^P = F_{j-1}$ and $R_i^S = \emptyset$ for all $H_i \in F_j$, $j > 1$ then we have parallel gatekeeping.

# 4 Mixture procedures for parallel gatekeeping

In this section we define a very general method for constructing multiple testing procedures, termed the *mixture method*. As shown in this paper, the mixture method can be applied to build a broad class of gatekeeping procedures for problems with several families of hypotheses. This includes gatekeeping procedures developed in recent publications as well as some new gatekeeping procedures.

Denote the closure of the family $F_j$ by $\overline{F}_j$, being the set of all non-empty intersections of the hypotheses in $F_j$. Let $\mathcal{P}_1, \ldots, \mathcal{P}_m$ denote the component MTPs for families $F_1, \ldots, F_m$, respectively. We assume that each $\mathcal{P}_j$ is a closed procedure which controls the FWER within $F_j$ at any preassigned level $\alpha$ and $\mathcal{P}_1, \ldots, \mathcal{P}_{m-1}$ are separable. Here $\mathcal{P}_j$ being closed means that there exist $\alpha$-level tests (called the *local tests*) for all intersection hypotheses $H(I_j) \in \overline{F}_j$ such that $\mathcal{P}_j$ rejects any hypothesis $H_i \in F_j$ at level $\alpha$ iff the local tests reject all $H(I_j) \in \overline{F}_j$ at level $\alpha$ for all $I_j$ containing $i$. Let $p_j(I_j)$ denote the $p$-value (called the *local p-value*) associated with the local test of the intersection hypothesis $H(I_j)$. For example, if $\mathcal{P}_j$ is the Holm procedure then the local test of each $H(I_j)$ is the Bonferroni test and

$$p_j(I_j) = |I_j| \min_{i \in I_j}(p_i). \tag{6}$$

A mixture procedure composed of component procedures $\mathcal{P}_1, \ldots, \mathcal{P}_m$ and denoted by $\mathcal{P}$ is a closed procedure for testing all hypotheses in $F = \bigcup_{j=1}^m F_j$. Let $\overline{F}$ denote the closed family of non-empty intersection hypotheses in all $m$ families. Since $\mathcal{P}$ is a closed procedure, we need to define the local tests for all $H(I) \in \overline{F}$.

First we define the notion of a *mixing function*. Consider any index set $I$ consisting of the union of $s$ ($1 \leq s \leq m$) nonempty subsets, $I_{j_1}, \ldots, I_{j_s}$ ($1 \leq j_1 < \cdots < j_s \leq m$). To avoid double subscripts and keep the notation simple, we will relabel these subsets as $I_1, \ldots, I_s$ keeping in mind that they are not necessarily the first $s$ consecutive index subsets. Let $H(I) = \bigcap_{j=1}^{s} H(I_j)$. A mixing function $\phi_I(p_1(I_1), \ldots, p_s(I_s))$ is a function on the interval $[0, 1]$ which satisfies the following properties.

- Property 1: For $s = 1$, $\phi_I(p_1(I_1)) = p_1(I_1)$ and for $s > 1$, $\phi_I(p_1(I_1), \ldots, p_s(I_s)) \leq p_1(I_1)$.
- Property 2: $P\{\phi_I(p_1(I_1), \ldots, p_s(I_s)) \leq \alpha\} \leq \alpha$.

It can be shown that Property 1 ensures independence for $F_1$ (i.e., inferences in $F_1$ are independent of those in $F_j$ for $j > 1$) which is a desirable property for a gatekeeping procedure. Property 2 ensures that the local tests of intersection hypotheses are of level $\leq \alpha$. In what follows, it is assumed that all probability expressions are evaluated under appropriate null hypotheses (e.g., in case of Property 2, under $H(I) = H(I_1) \bigcap \cdots \bigcap H(I_s)$), which will not be shown notationally.

Define the local test of the intersection hypothesis $H(I)$ as reject $H(I)$ if

$$\phi_I(p_1(I_1), \ldots, p_s(I_s)) \leq \alpha. \tag{7}$$

By Property 2 it follows that $\mathcal{P}$ controls the FWER at level $\alpha$ because it is a closed procedure.

A particular class of mixing functions that we will study in this paper has the following general form. For any specified $I = I_1 \bigcup \cdots \bigcup I_s$,

$$\phi_I(p_1(I_1), \ldots, p_s(I_s)) = \min\left(\frac{p_1(I_1)}{c_1(I|\alpha)}, \ldots, \frac{p_s(I_s)}{c_s(I|\alpha)}\right), \tag{8}$$

where the $c_j(I|\alpha)$ are the coefficients determined to satisfy Property 2 subject to $1 = c_1(I|\alpha) \geq \cdots \geq c_s(I|\alpha) \geq 0$. Property 1 is obviously satisfied by this mixing function. Monotonicity is imposed on the coefficients so that greater weights are assigned to the $p$-values for the intersection hypotheses from the earlier families than to those from the later families. This helps account for the hierarchical structure of the problem in the sense that power of the tests of the more important hypotheses (included in the earlier families) are improved at the expense of power of the tests of the less important hypotheses (included in the later families). Using this mixing function, the local test (7) simplifies to

$$p_j(I_j) \leq \alpha c_j(I|\alpha) \text{ for at least one } j = 1, \ldots, s. \tag{9}$$

Different mixing functions differ in their choice of the coefficients $c_j(I|\alpha)$. We will consider two mixing functions: Bonferroni and parametric. The Bonferroni mixing function ignores correlations among the test statistics while the parametric mixing function takes them into account by assuming a joint distribution for the test statistics.

**Remark 1:** The $c_j(I|\alpha)$ that satisfy Property 2 always exist but are not unique. A trivial choice is $c_1(I|\alpha) = 1$ and $c_j(I|\alpha) = 0$ for $j = 2, \ldots, s$. ■

## 4.1 Bonferroni mixing function

The Bonferroni mixing function uses the coefficients defined recursively as follows: Let $c_1(I|\alpha) = 1$ and

$$c_j(I|\alpha) = c_{j-1}(I|\alpha)[1 - f_{j-1}(I_{j-1}|\alpha)] = \prod_{k=1}^{j-1}[1 - f_k(I_k|\alpha)],\ j = 2, \ldots, s, \tag{10}$$

where $I = I_1 \bigcup \cdots \bigcup I_s$ and $f_k(I_k|\alpha) = e_k(I_k|\alpha)/\alpha$. Substituting this formula for $c_j(I|\alpha)$ in (9), we get that the intersection hypothesis $H(I)$ is rejected at level $\alpha$ iff

$$p_j(I_j) \leq \alpha \prod_{k=1}^{j-1}[1 - f_k(I_k|\alpha)] \text{ for at least one } j = 1, \ldots, s. \tag{11}$$

For the sake of simplicity, henceforth we will assume that the error rate functions $e_j(I_j|\alpha)$ of all the component procedures $\mathcal{P}_j$ are proportional to $\alpha$. Therefore the error rate fractions $f_j(I_j|\alpha)$ and hence the $c_j(I_j|\alpha)$ are independent of $\alpha$ and we denote them simply by $f_j(I_j)$ and $c_j(I_j)$, respectively. Note that the error rate function (3) satisfies this property. This error rate function can be used as an upper bound on the exact error rate function of not only the truncated Holm procedure, but also the truncated Hochberg and truncated Hommel procedures (see Brechenmacher *et al.*, 2011).

The main consequence of this simplifying assumption is that we can define the local $p$-value of any $H(I)$ simply by equating it to the corresponding mixing function, i.e.,

$$p(I) = \phi_I(p_1(I_1), \ldots, p_s(I_s)) = \min\left(\frac{p_1(I_1)}{c_1(I)}, \ldots, \frac{p_s(I_s)}{c_s(I)}\right),$$

since this equals the smallest $\alpha$ at which $H(I)$ can be rejected using the local test (7). It should be noted that the local $p$-values are well-defined also in the general case where the $e_j(I_j|\alpha)$ are not proportional to $\alpha$, but the calculation of $p(I)$ is more involved in that case and must be done numerically as explained in Remark 3.

Once the $p(I)$-values are evaluated for all $H(I)$ then using the closure principle, the adjusted $p$-value for any individual hypothesis $H_i$ can be defined as $\widetilde{p}_i = \max_{I \ni i} p(I)$,

where the maximum is taken over all index sets $I$ that contain $i$. We can reject $H_i$ at level $\alpha$ if $\widetilde{p}_i \leq \alpha$.

**Remark 2:** As an example of an error rate function which is not proportional to $\alpha$, consider again the truncated Holm procedure but suppose that the test statistics and hence their $p$-values are independent. Then its exact error rate function is given by

$$e(I|\alpha) = 1 - \left[1 - \left(\frac{\gamma}{|I|} + \frac{(1-\gamma)}{n}\right)\alpha\right]^{|I|}.$$

The upper bound given in (3) is the Bonferroni upper bound on this expression. For another example, Brechenmacher *et al.* (2011) have given the following exact expression for the error rate function of the truncated Hommel procedure under the same assumption of independent test statistics:

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

which is not proportional to $\alpha$. ■

**Remark 3:** If the $c_j(I|\alpha)$ are functions of $\alpha$, the local test (9) is still well-defined as it can be applied for any fixed $\alpha$. To compute $p(I)$, we can numerically solve for the smallest $\alpha$ that satisfies the inequality

$$\min\left(\frac{p_1(I_1)}{c_1(I|\alpha)}, \ldots, \frac{p_s(I_s)}{c_s(I|\alpha)}\right) \leq \alpha.$$

A smallest $\alpha$ satisfying this inequality always exists since the test rejects if $\alpha = 1$ and accepts if $\alpha = 0$. However, one cannot interpret that smallest $\alpha$ as $p(I)$ unless the test is $\alpha$-consistent (Roth, 1999; Lehmann and Romano, 2005), i.e., the indicator function of the test (which equals 1 if the test rejects and 0 if it accepts) is nondecreasing in $\alpha$. It is immediately evident that the local test (9) will be $\alpha$-consistent if $\alpha c_j(I|\alpha)$ is nondecreasing in $\alpha$ for all $j$. In particular, this is true if the $c_j(I|\alpha)$ are independent of $\alpha$ for all $j$ and $I$. ■

We next state the main result of this paper.

**Proposition 1** *If the component procedures $\mathcal{P}_1, \ldots, \mathcal{P}_{m-1}$ are separable then the local test (7) of each intersection hypothesis $H(I)$ which uses the Bonferroni mixing function defined by (8) and (10) is an $\alpha$-level test, i.e., it satisfies Property 2. Hence by the closure property, the mixture procedure $\mathcal{P}$ strongly controls the FWER at level $\alpha$.*

Although not needed for the above proposition, the subsequent propositions make the assumption that the error rate functions of the component MTPs are proportional

to $\alpha$, so that we can equate $p(I)$ for any intersection hypothesis $H(I)$, where $I = I_1 \bigcup \cdots \bigcup I_s$, to the corresponding mixing function $\phi_I(p_1(I_1), \ldots, p_s(I_s))$.

To state the following proposition, we need the concept of consonance introduced by Gabriel (1969). An MTP is said to be consonant if whenever it rejects an intersection hypothesis $H(I) = \bigcap_{i \in I} H_i$, it rejects at least one component hypothesis $H_i$, $i \in I$.

**Proposition 2** *If the component procedures $\mathcal{P}_1, \ldots, \mathcal{P}_{m-1}$ are separable and consonant then the mixture procedure $\mathcal{P}$ based on the Bonferroni mixing function is equivalent to the multistage procedure defined in Section 3 in that both procedures make the same decisions, i.e., they both reject the same hypotheses.*

If any one or more of the component MTPs is allowed to be non-consonant then the equivalence claimed in this proposition may no longer hold as the following example shows. It also shows that the mixture procedure can be more powerful than the multistage procedure in that case.

**Example 2:** This example, taken from Dmitrienko, Kordzakhia and Tamhane (2011), is designed to illustrate an important property of mixture procedures. Consider two families, $F_1 = \{H_1, H_2, H_3, H_4\}$ and $F_2 = \{H_5\}$. The first family serves as a parallel gatekeeper for the other family. The raw $p$-values for the five null hypotheses are $p_1 = 0.0053$, $p_2 = 0.0126$, $p_3 = 0.0131$, $p_4 = 0.0224$ and $p_5 = 0.0022$. To construct the two-stage and mixture gatekeeping procedures, we will use the truncated Hommel procedure (which is non-consonant since the untruncated Hommel procedure is non-consonant; see Westfall *et al.*, 1999, Section 2.5.4) with $\gamma = 3/4$ as the primary component procedure $\mathcal{P}_1$ and the regular Hommel procedure as the secondary component procedure $\mathcal{P}_2$. The adjusted $p$-values for the two-stage procedure are computed by the direct-calculation algorithm introduced in Dmitrienko *et al.* (2008b) and the adjusted $p$-values for the mixture procedure are found by the method described in Section 4.1. Using a one-sided $\alpha = 0.025$, it is shown in Dmitrienko, Kordzakhia and Tamhane (2011) that the adjusted $p$-values produced by the two-stage procedure for the five null hypotheses are 0.0210, 0.0276, 0.0276, 0.0276, 0.0276, respectively. This means that the two-stage procedure rejects $H_1$ and thus passes the parallel gatekeeper. However, the only null hypothesis in the second family cannot be rejected. The mixture procedure provides an advantage over the two-stage procedure and rejects two null hypotheses in this problem. In particular, the adjusted $p$-values for the five null hypotheses are given by 0.0210, 0.0276, 0.0276, 0.0276, 0.0233, respectively, and thus the mixture procedure rejects both $H_1$ and $H_5$. ■

## 4.2 Parametric mixing function

As before, assume that $H(I) = \bigcap_{i \in I} H_i$ is true where $I = I_1 \bigcup \cdots \bigcup I_s$. To obtain an $\alpha$-level test of each $H(I)$ we must have under $H(I)$,

$$P\left\{\phi_I(p_1(I_1), \ldots, p_s(I_s)) \leq \alpha\right\} = P\left\{p_j(I_j) \leq \alpha c_j(I|\alpha) \text{ for at least one } j = 1, \ldots, s\right\} \leq \alpha. \tag{12}$$

The coefficients $c_j(I|\alpha)$ are calculated as follows starting with $c_1(I|\alpha) = 1$. Having recursively computed $c_k(I|\alpha)$ for $k = 1, \ldots, j-1$, solve for $c_j(I|\alpha)$ from

$$P\left\{p_1(I_1) \leq \alpha \text{ or } \cdots \text{ or } p_{j-1}(I_{j-1}) \leq \alpha c_{j-1}(I|\alpha) \text{ or } p_j(N_j) \leq \alpha c_j(I|\alpha)\right\} = \alpha \tag{13}$$

for $j = 2, \ldots, s$. Notice that in the last term of the probability expression we use $p_j(N_j)$ and not $p_j(I_j)$. The reason for this will become clear in Proposition 3.

To solve equation (13), one needs to know the joint distribution of the $p_j(I_j)$ values which is often difficult to specify. It is easier to specify the joint distribution of the associated test statistics. For the sake of illustration, suppose that $t$-statistics, denoted by $t_i$, are used to test the hypotheses $H_i$ $(1 \leq i \leq n)$ and union-intersection statistics $t_j(I_j) = \max_{i \in I_j} t_i$ are used to test the intersection hypotheses $H(I_j) = \bigcap_{i \in I_j} H_i$. Further suppose that $t_1, \ldots, t_n$ have an $n$-variate $t$-distribution with $\nu$ degrees of freedom (d.f.) and correlation matrix $R = \{\rho_{ij}\}$ under the overall intersection hypothesis $H(N) = \bigcap_{i \in N} H_i$.

Let $R(N_j)$ denote the submatrix of $R$ corresponding to the joint distribution of the $t_i$ for $i \in N_j$ and let $t^*(\alpha|\nu, n_j, R(N_j))$ be the upper $\alpha$ critical constant of the distribution of $t_j(N_j) = \max_{i \in N_j} t_i$. Note that in order for the procedure $\mathcal{P}_j$ to be separable $(1 \leq j \leq m-1)$, it must use this common critical constant to test all intersection hypotheses $H(I_j) = \bigcap_{i \in I_j} H_i$ so that if $I_j \subset N_j$ then

$$\begin{aligned} e_j(I_j|\alpha) &= P\{t_j(I_j) \geq t^*(\alpha|\nu, n_j, R(N_j))\} \\ &< P\{t_j(N_j) \geq t^*(\alpha|\nu, n_j, R(N_j))\} \\ &= e_j(N_j|\alpha) = \alpha. \end{aligned}$$

The Dunnett (1955) procedure is an example of such a separable procedure. On the other hand, $\mathcal{P}_m$ can be non-separable, e.g., the step-down Dunnett procedure of Naik (1975) which uses the critical constant $t^*(\alpha|\nu, |I_m|, R(I_m))$ (where $R(I_m)$ is the submatrix of $R(N_m)$ corresponding to the correlation matrix of the $t_i,\ i \in I_m$) to test the intersection hypothesis $H(I_m)$. To keep the exposition simple, we will assume that $\mathcal{P}_m$ is also separable and uses a common critical constant $t^*(\alpha|\nu, n_m, R(N_m))$ to test all intersection hypotheses $H(I_m)$ for $I_m \subseteq N_m$.

The coefficients $c_j(I|\alpha)$ are calculated recursively starting with $c_1(I|\alpha) = 1$. Next we calculate $c_2(I|\alpha)$ from the equation

$$P\left\{t_1(I_1) \geq t^*(\alpha|\nu, n_1, R(N_1)) \text{ or } t_2(N_2) \geq t^*(\alpha c_2(I|\alpha)|\nu, n_2, R(N_2))\right\} = \alpha.$$

Note that the evaluation of this probability requires the knowledge of not only the correlation matrices $R(N_1)$ and $R(N_2)$, but also the cross-correlation matrix between the $t_i,\ i \in I_1$ and $t_j,\ j \in N_2$. In general, having recursively computed $c_2(I|\alpha), \ldots, c_{j-1}(I|\alpha)$, we calculate $c_j(I|\alpha)$ from the equation

$$P\{t_1(I_1) \geq t^*(\alpha|\nu, n_1, R(N_1)) \text{ or } \cdots \text{ or } t_{j-1}(I_{j-1}) \geq t^*(\alpha c_{j-1}(I|\alpha)|\nu, n_{j-1}, R(N_{j-1}))$$
$$\text{or } t_j(N_j) \geq t^*(\alpha c_j(I|\alpha)|\nu, n_j, R(N_j))\} = \alpha.$$

The evaluation of this probability requires the knowledge of the correlations among all the $t_i$ for $i \in \widetilde{I}_j$ where $\widetilde{I}_j = I_1 \bigcup \cdots \bigcup I_{j-1} \bigcup N_j$. The clinical trial example in Section 6 gives an illustration of the calculation of the parametric mixing function. We conclude this section by stating a result about parametric mixing functions.

**Proposition 3** *If the component procedures $\mathcal{P}_1, \ldots, \mathcal{P}_{m-1}$ are separable and consonant then the mixture procedure $\mathcal{P}$ based on the parametric mixing function satisfies the parallel gatekeeping condition, i.e., the hypotheses in $F_{j+1}$ are testable iff at least one hypothesis is rejected in $F_j$ $(1 \leq j \leq m-1)$.*

# 5 Mixture procedures for general gatekeeping

In Dmitrienko *et al.* (2007) we introduced parallel and serial rejection sets to take into account logical restrictions among the hypotheses in tree-structured gatekeeping framework. Here we introduce a much simpler and yet a more general way of specifying logical restrictions through the so-called restriction functions. Note that logical restrictions defined here are different from those considered by Shaffer (1986).

Consider a hypothesis $H_i \in F_{s+1}$ and let $I = I_1 \bigcup \cdots \bigcup I_s$, where $I_j \subseteq N_j$ $(1 \leq j \leq s)$ and $s$ is fixed between 1 and $m-1$. The *restriction function* $L_i(I)$ is an indicator function taking a value 0 if $H_i$ is non-testable and 1 if $H_i$ is testable when the hypotheses $H_j$ for all $j \in I$ are accepted and the other hypotheses in $F_1, \ldots, F_s$ are rejected. The restriction functions satisfy the following natural conditions.

1. Monotonicity Condition: For any $H_i \in F_{s+1}$, if $L_i(I) = 0$ then $L_i(I') = 0$ for $I \subseteq I' \subseteq N_1 \bigcup \cdots \bigcup N_s$, i.e., if $H_i \in F_{s+1}$ is non-testable when the hypotheses $H_j$ for $j \in I$ are accepted then it remains non-testable if more hypotheses from $F_1, \ldots, F_s$ are accepted.

2. Parallel Gatekeeping Condition: Suppose $I = I_1 \bigcup \cdots \bigcup I_s$ and $I_j = N_j$ for some $j = 1, \ldots, s$ then $L_i(I) = 0$ for all $i \in N_k$ for $k > j$, i.e., if all hypotheses in $F_j$ are accepted then all hypotheses in $F_k$ for $k > j$ are non-testable.

**Example 3:** As an example of a logical restriction which cannot be modeled using the tree-structured gatekeeping framework but can be readily modeled using the restriction function approach, consider the FDA Guidance for Industry on rheumatoid arthritis (FDA, 1999) which states that "... trial results were considered to support a conclusion of effectiveness when statistical evidence of efficacy was shown for at least three of the four measures ...." Suppose that in addition to $n_1 = 4$ primary endpoints there are $n_2 \geq 1$ secondary endpoints, which are tested only if the efficacy requirement on the primary endpoints is met. Let $F_1 = \{H_1, H_2, H_3, H_4\}$ and $F_2 = \{H_5, \ldots, H_{4+n_2}\}$. Then the restriction functions for $H_i \in F_2$ are given by:

$$L_i(\emptyset) = L_i(\{1\}) = L_i(\{2\}) = L_i(\{3\}) = L_i(\{4\}) = 1$$

and $L_i(I_1) = 0$ for all other index sets $I_1 \subseteq N_1 = \{1, 2, 3, 4\}$. □

Once all the logical restrictions are specified through the restriction functions, the local $p$-values, $p(I)$, are modified to take into account the logical restrictions as follows (all index sets $I_j$ below are assumed to be nonempty):

- Case 1: $I = I_j$, $s = 1$. In this case $p(I) = p_j(I_j)$.
- Case 2: $I = I_1 \bigcup \cdots \bigcup I_s$ for $s \geq 2$. In this case we first recursively define the index sets of testable hypotheses as follows. Let $I_1^* = I_1$ and

$$I_j^* = \{i \in I_j : \ L_i(I_1^*, \ldots, I_{j-1}^*) = 1\} \text{ for } j = 2, \ldots, s. \tag{14}$$

Next let $p_j(I_j^*) = 1$ if $I_j^*$ is empty. Then

$$p(I) = \begin{cases} \phi_I(p_1(I_1^*), \ldots, p_s(I_s^*)) \text{ if } I_s^* \neq \emptyset \\ \phi_I(p_1(I_1^*), \ldots, p_r(I_r^*)) \text{ if } I_{r+1}^*, \ldots, I_s^* = \emptyset. \end{cases} \tag{15}$$

For the special class of mixing functions (8), we have

$$\phi_I(p_1(I_1^*), \ldots, p_s(I_s^*)) = \min\left(\frac{p_1(I_1^*)}{c_1(I|\alpha)}, \ldots, \frac{p_s(I_s^*)}{c_s(I|\alpha)}\right).$$

Note that the coefficients $c_j(I|\alpha)$ depend on the original index sets, $I_1, \ldots, I_s$ and not on the modified index sets, $I_1^*, \ldots, I_s^*$. ■

**Proposition 4** *If the component procedures $\mathcal{P}_1, \ldots, \mathcal{P}_{m-1}$ are separable and consonant then the mixture procedure $\mathcal{P}$ based on the mixing function (15) is consistent with the specified logical restrictions. In other words, for any hypothesis $H_i \in F_{j+1}$, if $L_i(A_1 \bigcup \cdots \bigcup A_j) = 0$ for any $j < m$, where $A_1, \ldots, A_j$ are the index sets of accepted hypotheses from $F_1, \ldots, F_j$, then the mixture procedure accepts $H_i$.*

To illustrate this proposition we give the following example dealing with two families. The example shows that mixture procedures in hypothesis testing problems with general gatekeeping restrictions are based on more complex rules compared to parallel gatekeeping restrictions. In particular, it is shown in Section 3 that mixture gatekeeping procedures and multistage gatekeeping procedures derived from the same component procedures are equivalent for parallel gatekeeping. In the general case, mixture gatekeeping procedures also have a multistage structure; however, they may use a different set of component procedures.

**Example 4:** Consider a two-family problem with $F_1 = \{H_1, H_2\}$ and $F_2 = \{H_3, \ldots, H_n\}$ and the corresponding index sets $N_1 = \{1, 2\}$ and $N_2 = \{3, \ldots, n\}$. Further, consider a mixture of the Bonferroni procedure ($\mathcal{P}_1$) and the Holm procedure ($\mathcal{P}_2$) using the Bonferroni mixing function. Assume the following logical restrictions:

- $H_3, \ldots, H_{n-1}$ are testable iff $H_2$ is rejected.
- $H_n$ is testable iff at least one hypothesis in $F_1$ is rejected.

The restriction functions corresponding to these logical restrictions are as follows.

- If $I_1 = \emptyset$ or $I_1 = \{1\}$, then $L_3(I_1) = L_4(I_1) = \cdots = L_n(I_1) = 1$.
- If $I_1 = \{2\}$, then $L_3(I_1) = L_4(I_1) = \cdots = L_{n-1}(I_1) = 0$ and $L_n(I_1) = 1$.
- If $I_1 = \{1, 2\}$, then $L_3(I_1) = L_4(I_1) = \cdots = L_n(I_1) = 0$.

It will be shown below that, depending on the number of hypotheses rejected in $F_1$, the mixture procedure may use the Holm procedure or a slightly different procedure in $F_2$. The $p$-values for the intersection hypotheses in $\overline{F}_1$ and $\overline{F}_2$ are given by

$$\begin{aligned} p_1(I_1) &= 2 \min_{i \in I_1} p_i, \quad I_1 \subseteq N_1, \\ p_2(I_2) &= |I_2| \min_{i \in I_2} p_i, \quad I_2 \subseteq N_2. \end{aligned}$$

Next note that the error rate function for the Bonferroni procedure is $e_1(I_1|\alpha) = |I_1|\alpha/2$. Hence $f_1(I_1) = 0$ if $I_1 = \emptyset$, $f_1(I_1) = 1/2$ if $I_1 = \{1\}$ or $\{2\}$ and $f_1(I_1) = 1$ if $I_1 = \{1, 2\}$. Finally, note that $I_2^* = \emptyset$ if $I_1 = \{1, 2\}$, $I_2^* = I_2$ if $I_1 = \{1\}$ or $I_1 = \emptyset$ and $I_2^* = \{n\}$ if $I_1 = \{2\}$. Using these facts the $p$-values for the intersection hypotheses $H(I)$ in $\overline{F}$ can be computed as follows.

- Case 1: If $I_1 = \{1, 2\}$,

$$p(I) = p_1(I_1) = 2 \min(p_1, p_2).$$

- Case 2: If $I_1 = \{1\}$,

$$p(I) = \min\left(2p_1, 2p_2(I_2^*)\right) = 2\min\left(p_1, |I_2| \min_{i \in I_2} p_i\right).$$

- Case 3: If $I_1 = \{2\}$,

$$p(I) = \min\left(2p_2, 2p_2(I_2^*)\right) = 2\min(p_2, p_n).$$

- Case 4: If $I_1 = \emptyset$,

$$p(I) = p_2(I_2^*) = |I_2| \min_{i \in I_2} p_i.$$

From these expressions the mixture procedure can be shown to consist of the following decision rules:

- Case 1 (Both $H_1$ and $H_2$ are accepted): In this case the parallel gatekeeping property ensures that all $H_i \in F_2$ are accepted without tests. To see this, consider any hypothesis $H_i \in F_2$. To reject $H_i$, it must be true that $p(I) \leq \alpha$ for all index sets $I$ with $i \in I$. However, if $I = I_1 \bigcup \{i\}$ with $I_1 = \{1, 2\}$ then $p(I) = p_1(I_1) = 2\min(p_1, p_2) > \alpha$ since both $H_1$ and $H_2$ are accepted. Therefore no $H_i \in F_2$ can be rejected and hence all are accepted without tests.

- Case 2 ($H_1$ is accepted, $H_2$ is rejected): In this case $p(I) \leq \alpha$ for all index sets $I$ containing $\{2\}$ since $H_2$ is rejected. Therefore, to derive the component procedure used to test the hypotheses in $F_2$, we only need to consider $I_1 = \{1\}$ or $I_1 = \emptyset$. If $I_1 = \{1\}$ then the local test of $H(I)$ rejects if $p(I) = 2\min\left(p_1, |I_2| \min_{i \in I_2} p_i\right) \leq \alpha$. However, $2p_1 > \alpha$ (since $H_1$ is accepted), so we must have $|I_2| \min_{i \in I_2} p_i \leq \alpha/2$. This local test of $H(I)$ is the Bonferroni test at level $\alpha/2$, which means that the hypotheses in $F_2$ are tested using the Holm procedure at level $\alpha/2$. On the other hand, if $I_1 = \emptyset$ then the local test of $H(I)$ rejects if $p(I) = |I_2| \min_{i \in I_2} p_i \leq \alpha$. This local test of $H(I)$ is the Bonferroni test at level $\alpha$ and thus the component procedure used in $F_2$ is the Holm procedure at level $\alpha$. The net result is that the mixture gatekeeping procedure uses the Holm procedure at level $\alpha/2$ in $F_2$ which was the component procedure selected for this family.

- Case 3 ($H_1$ is rejected, $H_2$ is accepted): This case is similar to Case 2. Here we only need to consider $I_1 = \{2\}$ or $I_1 = \emptyset$. If $I_1 = \{2\}$ then the local test of $H(I)$ rejects if $p(I) = 2\min(p_2, p_n) \leq \alpha$. However, $2p_2 > \alpha$ (since $H_2$ is accepted), so we must have $p_n \leq \alpha/2$. If $I_1 = \emptyset$ then, as shown above, the component procedure used in $F_2$ is the Holm procedure at level $\alpha$. The net

result is that the procedure used in $F_2$ rejects $H_n$ if $p_n \leq \alpha/2$ *and* if $H_n$ is rejected by the Holm procedure at level $\alpha$. Since there are $n-2$ hypotheses in $F_2$ and if $q_{(1)} \leq \cdots \leq q_{(n-2)}$ denote their ordered $p$-values with $p_n = q_{(k)}$ for some $k = 1, \ldots, n-2$ then the procedure used in $F_2$ rejects $H_n$ if

$$p_n \leq \alpha/2 \text{ and } q_{(i)} \leq \alpha/(n-i-1),\ i = 1, \ldots, k.$$

If $k \leq n-3$ and the above Holm procedure rejects $H_n$ then the condition $p_n \leq \alpha/2$ is automatically satisfied. If $k = n-2$, i.e., if $p_n$ is the largest $p$-value among $p_3, \ldots, p_n$ then the above Holm procedure must reject all secondary hypotheses *and* $p_n = q_{(n-2)} \leq \alpha$, which is automatically satisfied if $p_n \leq \alpha/2$. Thus the mixture gatekeeping procedure does not use a simple Holm procedure in $F_2$ even though the Holm procedure was originally specified for this family.

- Case 4 (Both $H_1$ and $H_2$ are rejected): In this case we only need to consider $I_1 = \emptyset$ since all $H(I)$, which include $H_1$ or $H_2$ or both, are rejected. As shown above, the procedure used in $F_2$ is the Holm procedure at level $\alpha$. ■

# 6 Clinical trial example

Mixture procedures with general gatekeeping restrictions introduced in Section 5 will be illustrated here using a clinical trial example with artificial data. A clinical trial in patients with pulmonary arterial hypertension (PAH) in which two doses of an experimental treatment versus placebo (the doses are labeled L and H, placebo is labeled Plac) were tested. The dose-placebo comparisons were performed with respect to three ordered endpoints, Endpoint P (six-minute walk distance), Endpoint S1 (dyspnea score) and Endpoint S2 (quality of life). The trial employed a balanced design and the sample size was $n = 110$ patients per treatment arm.

As shown in Figure 1, the six hypotheses of no treatment effect studied in this clinical trial were grouped into three families (the hypotheses were equally weighted within each family):

- Family 1 ($F_1$): L-Plac ($H_1$) and H-Plac ($H_2$) comparisons for Endpoint P.
- Family 2 ($F_2$): L-Plac ($H_3$) and H-Plac ($H_4$) comparisons for Endpoint S1.
- Family 3 ($F_3$): L-Plac ($H_5$) and H-Plac ($H_6$) comparisons for Endpoint S2.

[Insert Figure 1 here]

We assume that the joint distribution of the two-sample $t$-statistics is approximately multivariate normal and to reflect that denote them as $z$-statistics. The reason for this is two-fold: (i) the sample size per treatment arm is large, but more importantly (ii) the joint distribution of the statistics is not multivariate $t$ since the sample standard deviations of each endpoint are different and thus the statistics do not share a common denominator. The one-sided $p$-values associated with the $z$-statistics are listed in Table 1 (adjusted $p$-values shown in Table 1 are discussed below).

[Insert Table 1 here]

The hypotheses within each dose were tested subject to a fixed-sequence restriction, i.e., a hypothesis in $F_2$ or $F_3$ was tested iff higher-level hypotheses associated with the same dose were rejected. For example, $H_5$ was testable iff $H_1$ and $H_3$ were rejected. Thus the restriction functions were given by

- $L_i(I) = 0$ if $I$ contains $i-2$ and $L_i(I) = 1$ otherwise, where $I \subseteq \{1,2\}$ and $i = 3,4$.
- $L_i(I) = 0$ if $I$ contains $i-2$ or $i-4$ and $L_i(I) = 1$ otherwise, where $I \subseteq \{1,2,3,4\}$ and $i = 5,6$.

We will illustrate the process of constructing the following two mixture procedures and evaluate their performance:

- Procedure 1 is a nonparametric mixture procedure which uses the Bonferroni procedure in $F_1$ and $F_2$ and the Holm procedure in $F_3$ with the Bonferroni mixing function.
- Procedure 2 is a parametric mixture procedure which uses the single-step Dunnett procedure in $F_1$ and $F_2$ and the step-down Dunnett procedure in $F_3$ with the parametric mixing function.

Beginning with Procedure 1, the adjusted $p$-values for the six hypotheses are computed using the algorithm given in Section 5 based on the Bonferroni mixing function. This algorithm is based on a complete enumeration of all 63 non-empty intersections of the original six hypotheses. A local Bonferroni $p$-value is computed for each intersection and the adjusted $p$-values for the hypotheses are found using the closure principle. As an illustration, consider the intersection hypothesis corresponding to the index set $I = \{1,3,4,6\}$. Note that $I_1 = \{1\}$, $I_2 = \{3,4\}$ and $I_3 = \{6\}$ but $I_2$ and $I_3$ need to be modified to account for the logical restrictions. Recall that $H_3$ is non-testable if $H_1$ is accepted; similarly, $H_6$ is non-testable if $H_4$ is accepted. Thus

$I_1^* = \{1\}$, $I_2^* = \{4\}$ and $I_3^* = \emptyset$. Furthermore, the Bonferroni $p$-values for the family-specific intersection hypotheses $H(I_1^*)$ and $H(I_2^*)$ are $p_1(I_1^*) = 2p_1$ and $p_2(I_2^*) = 2p_4$, respectively. Therefore, the local $p$-value for $H(I)$ is given by

$$p(I) = \min\left(p_1(I_1^*), \frac{p_2(I_2^*)}{c_2(I)}\right) = \min(2p_1, 4p_4)$$

since $c_2(I) = 1 - f_1(I_1) = 1/2$.

The calculation of the adjusted $p$-values for Procedure 2 assumes that the joint distribution of the six test statistics is approximately multivariate normal. This calculation is based on an algorithm similar to the one above with the following two changes. First, the Bonferroni $p$-values for family-specific intersection hypotheses need to be replaced by Dunnett $p$-values and, second, the parametric mixing function needs to be applied to compute local $p$-values for all intersection hypotheses. Using the same intersection hypothesis as above, the Dunnett $p$-value for $H(I_1^*)$ is computed from the joint distributions of $z_1$ and $z_2$. Due to the balanced design, the test statistics follow a central bivariate normal distribution under $H_1 \bigcap H_2$ with correlation coefficient $\rho = 1/2$. Let $G(z|\rho)$ denote the cumulative distribution function of $\max(z_1, z_2)$. Then $p_1(I_1^*) = 1 - G(z_1|\rho)$. Similarly, $p_2(I_2^*) = 1 - G(z_4|\rho)$. The mixing function for $H(I)$ is given by

$$\phi_I(p_1(I_1^*), p_2(I_2^*)|\alpha) = \min\left(p_1(I_1^*), \frac{p_2(I_2^*)}{c_2(I|\alpha)}\right), \tag{16}$$

with the coefficient $c_2(I|\alpha)$ computed from

$$P\{z_1 \geq z^*(\alpha|\rho) \text{ or } z_4 \geq z^*(\alpha c_2(I|\alpha)|\rho)\} = \alpha,$$

where $z^*(\alpha|\rho) = G^{-1}(1 - \alpha|\rho)$. It can be shown that in this simple case the $\alpha$-consistency condition defined in Section 4.1 is satisfied and the local $p$-value for this intersection hypothesis is given by the smallest $\alpha$ for which (16) is $\leq \alpha$.

Table 1 displays the adjusted one-sided $p$-values for the six hypotheses of interest produced by the two procedures. The adjusted $p$-values for Procedure 2 were computed assuming the following correlation matrix for the three endpoints:

$$\begin{bmatrix} 1 & 0.4 & 0.2 \\ 0.4 & 1 & 0.4 \\ 0.2 & 0.4 & 1 \end{bmatrix}.$$

Using a one-sided $\alpha = 0.025$, Procedure 1 rejects three hypotheses ($H_1$, $H_2$ and $H_4$), while Procedure 2 rejects five hypotheses ($H_1$, $H_2$, $H_3$, $H_4$ and $H_6$). Furthermore, Table 1 illustrates Proposition 4 which states that both procedures are consistent

with the logical restrictions. For example, Procedure 1 cannot reject $H_5$ since it did not reject $H_3$.

The power gain of Procedure 2 is due to the fact that, unlike Procedure 1, it utilizes the knowledge of the joint distribution of the six test statistics. However, it should be noted that at least part of the gain is because the correlations between the endpoints are assumed to be known. In case of dose versus control problems, the correlations are known, being functions of the known sample sizes but in case of multiple endpoints the correlations are generally unknown.

# 7 Concluding remarks

In this paper we have provided a general theory of mixture procedures for gatekeeping. Both nonparametric and parametric mixture procedures are derived using the corresponding mixing functions. It is shown that if the component procedures for the first $m-1$ families are consonant then the mixture procedure for parallel gatekeeping is equivalent to the general stepwise procedure proposed in Dmitrienko *et al.* (2008b) based on the error rate function. Most importantly, mixture procedures can be extended to very general types of logical restrictions which were not amenable by the previous approaches, e.g., the stepwise approach proposed in Dmitrienko *et al.* (2008b) was applicable only to parallel gatekeeping restrictions while the tree-gatekeeping restrictions were dealt with using Bonferroni (nonparametric) procedures and a rather complex weight assignment algorithm embedded in the closure procedure. Thus the mixture approach offers a powerful and a generally applicable method to construct gatekeeping procedures. In fact, mixture-based gatekeeping procedures have found multiple applications in Phase III clinical trials. For example, Brechenmacher *et al.* (2011) applied the mixture method to define Hommel-based gatekeeping procedures that were successfully used in the lurasidone development program for the treatment of schizophrenia (Meltzer *et al.*, 2011).

An open problem for future research is to derive simultaneous confidence regions associated with gatekeeping procedures along the lines of Strassburger and Bretz (2008), Guilbaud (2008, 2009, 2012) and Guilbaud and Karlsson (2011).

# Acknowledgments

The authors are grateful to the editor, an associate editor and two referees for useful comments and suggestions which helped clarify a number of points in the paper.

# References

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

[2] 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.

[3] Dmitrienko, A., Kordzakhia, G., Tamhane, A.C. (2011a). Multistage and mixture parallel gatekeeping procedures for clinical trials. *Journal of Biopharmaceutical Statistics*. **21**, 726–747.

[4] Dmitrienko, A., Millen, B.A., Brechenmacher, T., Paux, G. (2011b). Development of gatekeeping strategies in confirmatory clinical trials. *Biometrical Journal*. **53**, 875-893.

[5] 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.

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

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

[8] Dmitrienko, A., Tamhane, A.C. (2011). Mixtures of multiple testing procedures for gatekeeping applications in clinical trial applications. *Statistics in Medicine*. **30**, 1473–1488.

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

[10] Dunnett, C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*. **50**, 1096–1121.

[11] Dunnett, C.W., Tamhane, A.C. (1992). A step-up multiple test procedure. *Journal of the American Statistical Association*. **87**, 162–170.

[12] FDA (Food and Drug Administration) (1999). *Guidance for Industry — Clinical Development Programs for Drugs, Devices, and Biological Products for the Treatment of Rheumatoid Arthritis (RA)*, U.S. Department of Health and Human Services.

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

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

[15] Guilbaud, O. (2009). Alternative confidence regions for Bonferroni-based closed testing procedures that are not alpha-exhaustive. *Biometrical Journal*. **51**, 721-735.

[16] Guilbaud, O. (2012). Simultaneous confidence regions for closed tests, including Holm, Hochberg and Hommel related procedures. *Biometrical Journal*. **54**, 317-342.

[17] Guilbaud, O., Karlsson, P. (2011). Confidence regions for Bonferroni-based closed tests extended to more general tests. *Journal of Biopharmaceutical Statistics*, **21**, 682-707.

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

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

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

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

[22] Lehmann, E.L., Romano, J.P. (2005). Generalizations of the familywise error rate. *Annals of Statistics*. **33**, 1138–1154.

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

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

[25] Meltzer, H.Y., Cucchiaro, J., Silva, R., Ogasa, M., Phillips, D., Xu, J., Kalali, A.H., Schweizer, E., Pikalov, A., Loebel A. (2011). Lurasidone in the treatment of schizophrenia: a randomized, double-blind, placebo- and olanzapine-controlled study. *American Journal of Psychiatry*. **168**, 957–967.

[26] Millen, B.A., Dmitrienko, A. (2011). Chain procedures: A class of flexible closed testing procedures with clinical trial applications. *Statistics in Biopharmaceutical Research*. **3**, 14–30.

[27] Naik, U.D. (1975). Some selection rules for comparing $p$ processes to a standard. *Communications in Statistics. Series A*. **4**, 519–535.

[28] Roth, A.J. (1999). Multiple comparison procedures for discrete test statistics. *Journal of Statistical Planning and Inference*. **82**, 101–117.

[29] Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association*. **81**, 826–831.

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

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

[32] 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.

[33] Wiens, B. (2003). A fixed-sequence Bonferroni procedure for testing multiple endpoints. *Pharmaceutical Statistics*. **2**, 211–215.

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

# Appendix

**Proof of Proposition 1**: If $s = 1$, i.e., $I = I_j$ then $\phi_I(p_j(I_j)) = p_j(I_j)$, so that under $H(I_j)$,

$$P\{\phi_I(p_j(I_j)) \leq \alpha\} = P\{p_j(I_j) \leq \alpha\} \leq \alpha.$$

Next suppose $s > 1$. Then under $H(I) = \bigcap_{j=1}^{s} H(I_j)$

$$\begin{aligned}
& P\{\phi_I(p_1(I_1), \ldots, p_s(I_s)) \leq \alpha\} \\
= \; & P\left\{\min\left(\frac{p_1(I_1)}{c_1(I)}, \ldots, \frac{p_s(I_s)}{c_s(I)}\right) \leq \alpha\right\} \\
= \; & P\left\{\bigcup_{j=1}^{s}[p_j(I_j) \leq \alpha c_j(I)]\right\} \\
\leq \; & \sum_{j=1}^{s} P\left\{p_j(I_j) \leq \alpha c_j(I)\right\} \\
\leq \; & \sum_{j=1}^{s-1} e_j(I_j|\alpha c_j(I)) + e_s(N_s|\alpha c_s(I)) \\
= \; & \sum_{j=1}^{s-1} \alpha c_j(I) f_j(I_j) + \alpha c_s(I) \\
= \; & \alpha\left[\sum_{j=1}^{s-1} c_j(I) f_j(I_j|\alpha) + c_s(I)\right]. \qquad (17)
\end{aligned}$$

Now from (10), $c_s(I) = c_{s-1}(I)[1 - f_{s-1}(I_{s-1})]$, and so $c_{s-1}(I)f_{s-1}(I_{s-1}) + c_s(I) = c_{s-1}(I)$. Applying this formula recursively, we see that the quantity inside the square brackets in (17) equals $c_1(I) = 1$ and thus (17) equals $\alpha$. This completes the proof of the proposition. ■

**Proof of Proposition 2**: We will show that the mixture multistage gatekeeping procedures reject exactly the same hypotheses for any set of raw $p$-values, $p_1, \ldots, p_n$. **Part 1**: First assume that the multistage procedure rejects a hypothesis $H_i \in F_j$. Then we want to show that the mixture procedure also rejects $H_i$, i.e., $p(I) \leq \alpha$ for any index set $I = I_1 \bigcup \cdots \bigcup I_s$ such that $i \in I_j \subseteq I$. Note that since $\mathcal{P}_j$ rejects $H_i \in F_j$ using the multistage procedure and it is closed, from (5) we have

$$p_j(I_j) \leq \alpha_j = \alpha \prod_{k=1}^{j-1}[1 - f_k(A_k)] \text{ for all } I_j \text{ such that } i \in I_j, \qquad (18)$$

where $p_j(I_j)$ is the local $p$-value for testing $H(I_j)$ using the closed representation of $\mathcal{P}_j$.

Consider two cases:

**Case 1** ($i \in I_1$): We have either $I = I_1$ in which case $p(I) = p_1(I_1)$ or we have $I = I_1 \bigcup \cdots \bigcup I_s$ for $s \geq 2$ in which case $p(I) \leq p_1(I_1)$. In both cases $p_1(I_1) \leq \alpha_1 = \alpha$ and so $p(I) \leq \alpha$.

**Case 2** ($i \in I_j$ for $j > 1$): Let $R_1, \ldots, R_{j-1}$ denote the index sets of the rejected hypotheses from $F_1, \ldots, F_{j-1}$. Consider the following subcases:

- **Case 2 (a)** ($I_k \bigcap R_k \neq \emptyset$ for at least one $k = 1, \ldots, j-1$): Since the component procedures $\mathcal{P}_k$ are closed and some hypotheses $H_j \in I_k \bigcap R_k$ are rejected, we have

$$p_k(I_k) \leq \alpha_k = \alpha \prod_{\ell=1}^{k-1} [1 - f_\ell(A_\ell)].$$

  Therefore

$$p(I) \leq \frac{p_k(I_k)}{c_k(I)} = \frac{p_k(I_k)}{\prod_{\ell=1}^{k-1}[1 - f_\ell(A_\ell)]} \leq \alpha.$$

- **Case 2 (b)** ($I_k \bigcap R_k = \emptyset$ for all $k = 1, \ldots, j-1$): In this case $I_k \subseteq A_k$ for all $k = 1, \ldots, j-1$. Since $e_k(I_k|\alpha_k)$ and hence $f_k(I_k)$ are monotone, we have

$$1 - f_k(A_k) \leq 1 - f_k(I_k) \text{ for all } k = 1, \ldots, j-1.$$

  Combining these inequalities with (18) and (5) we have,

$$\frac{p_j(I_j)}{\prod_{k=1}^{j-1}[1 - f_k(I_k)]} \leq \frac{p_j(I_j)}{\prod_{k=1}^{j-1}[1 - f_k(A_k)]} \leq \alpha.$$

  Therefore

$$p(I) \leq \frac{p_j(I_j)}{\prod_{k=1}^{j-1}[1 - f_k(I_k)]} \leq \alpha.$$

**Part 2**: Next we assume that the mixture procedure rejects a hypothesis $H_i \in F_j$. Then we want to show that the multistage procedure also rejects $H_i$. Thus we assume that $p(I) \leq \alpha$ for all $I$ such that $i \in I$. Again consider two cases.

**Case 1** ($i \in I_1$): Choose $I = I_1$. Then $p(I) = p_1(I_1) \leq \alpha$ for all $I_1$ containing $i$. So $\mathcal{P}_1$ and hence the multistage procedure rejects $H_i$.

**Case 2** ($i \in I_j$ for $j > 1$): Choose any $I_j$ containing $i$ and let $I = A_1 \bigcup \cdots \bigcup A_{j-1} \bigcup I_j$. Then $p(I) \leq \alpha$. We want to show that $p_j(I_j) \leq \alpha_j$ (where $\alpha_j$ is defined in (18)). It follows that since all hypotheses in $A_1, \ldots, A_{j-1}$ are accepted, $p_k(A_k) > \alpha$ for $k = 1, \ldots, j-1$. This is because if $p_k(A_k) \leq \alpha$ for some $k$ then by the consonance

property of $\mathcal{P}_k$, at least one hypothesis $H_\ell$, $\ell \in A_k$ must be rejected which is not true. Therefore

$$\frac{p_k(A_k)}{\prod_{\ell=1}^{k-1}[1-f_\ell(A_\ell)]} > p_k(A_k) > \alpha \text{ for } k = 1, \ldots, j-1.$$

On the other hand,

$$p(I) = \min\left(p_1(A_1), \ldots, \frac{p_{j-1}(A_{j-1})}{\prod_{\ell=1}^{j-2}[1-f_\ell(A_\ell)]}, \frac{p_j(I_j)}{\prod_{\ell=1}^{j-1}[1-f_\ell(A_\ell)]}\right) \leq \alpha.$$

Therefore

$$\frac{p_j(I_j)}{\prod_{\ell=1}^{j-1}[1-f_\ell(A_\ell)]} \leq \alpha \text{ or } p_j(I_j) \leq \alpha \prod_{\ell=1}^{j-1}[1-f_\ell(A_\ell)] = \alpha_j.$$

Since this is true for all $I_j$ containing $i$ and since $\mathcal{P}_j$ is closed, it follows that $\mathcal{P}_j$ and hence the multistage procedure rejects $H_i$. This completes the proof of the proposition. ■

**Proof of Proposition 3**: Since the parallel gatekeeping restriction is a special case of the general gatekeeping restrictions, Proposition 3 follows directly from Proposition 4. ■

**Proof of Proposition 4**: Consider a hypothesis $H_i \in F_{s+1}$ and assume that $L_i(A_1 \bigcup \cdots \bigcup A_s) = 0$. Let $I = A_1 \bigcup \cdots \bigcup A_s \bigcup \{i\}$, i.e., $I_j = A_j$ $(1 \leq j \leq s)$ and $I_{s+1} = \{i\}$. Further let $J = A_1 \bigcup \cdots \bigcup A_s$. Note that $I^*_{s+1}$ is empty. Therefore

$$p(I) = \phi_I(p_1(A_1^*), \ldots, p_s(A_s^*), p_{s+1}(I_{s+1}^*)) = \phi_J(p_1(A_1^*), \ldots, p_s(A_s^*)) = p(J),$$

where $A_1^* = A_1$ and $A_j^*$ for $j > 1$ are defined recursively as in (14). Now we must have $p(I) = p(J) > \alpha$ because if $p(J) \leq \alpha$ then by the consonance property of the mixture procedure, at least one hypothesis $H_k$ for $k \in J$ must be rejected; however, all hypotheses in $J$ are accepted. Since $i \in I$ and $p(I) > \alpha$, $H_i$ must be accepted. ■

Table 1. Test statistics and raw $p$-values in the pulmonary arterial hypertension clinical trial example. The asterisk identifies the adjusted $p$-values that are significant at the one-sided 0.025 level.

| Family | Null hypothesis | Test statistic | Raw $p$-value | Adjusted $p$-value | |
|---|---|---|---|---|---|
| | | | | Procedure 1 | Procedure 2 |
| $F_1$ | $H_1$ | 2.29 | 0.0115 | 0.0230* | 0.0224* |
| | $H_2$ | 2.54 | 0.0059 | 0.0118* | 0.0112* |
| $F_2$ | $H_3$ | 2.25 | 0.0127 | 0.0254 | 0.0248* |
| | $H_4$ | 2.38 | 0.0091 | 0.0230* | 0.0174* |
| $F_3$ | $H_5$ | 2.20 | 0.0144 | 0.0288 | 0.0273 |
| | $H_6$ | 2.01 | 0.0228 | 0.0457 | 0.0221* |

Figure 1. Ordering of the hypotheses in the clinical trial example. Families $F_1, F_2, F_3$ refer to the primary endpoint P, secondary endpoint S1 and secondary endpoint S2, respectively. Hypotheses $H_1, H_3, H_5$ refer to low dose versus placebo comparisons while hypotheses $H_2, H_4, H_6$ refer to high dose versus placebo comparisons.