Introduction

Multiple testing problems are common in pharmaceutical statistics and life-sciences in general. The main goal of Multiple Testing Procedures (MTP) is to (strongly) control the family wise type-I error rate. A MTP can be a single-step data-independent procedure such as Bonferroni's method or a data-dependent stepwise procedure such as Hochberg’s stepup method and Hommel’s method. It can be an $\alpha$-exhaustive or $\alpha$-conservative approach. Conceptually, stepwise procedures are usually more powerful than single-step procedures and $\alpha$-exhaustive procedures are usually more powerful than $\alpha$-conservative approach. However, consider these two aspects together, comparisons of testing procedures are not that simple, often depending on the configuration of the alternative "hypotheses" or more precisely, the truths. For example, the power of a fallback procedure is dependent on the weight and "effective sizes" in the alternative hypotheses. A fixed sequence testing procedure is a special case of the fallback procedure and its power is heavily dependent on the order of the test sequence of the hypothesis.

We develop a simple single $\alpha$-exhaustive procedure that can improve power 2-5% over Hochberg’s and Hommel’s methods in common situations when the test statistics are independent. The method can also be generalized to dependent test statistics. The idea behind our method is to construct the rejection rules using the product of marginal p-values and by controlling the upper bounds of the $k$th order terms so that $\alpha$ is controlled for any configuration of $k$ null hypotheses. Such upper bounds or critical values are determined progressively from $k = 1$ towards $k = K$, the number of null hypotheses in the problem.
mathematical derivation in the Appendix. To make the procedure ready for practical use, we have included the SAS code in the Appendix.

### Multiple Testing Procedures

Stepwise procedures are different from single-step procedures, in the sense that a stepwise procedure must follow a specific order to test each hypothesis. In general, stepwise procedures are more powerful than single-step procedures. There are three categories of stepwise procedures that are dependent on how the stepwise tests proceed: Stepup, stepdown and fallback procedure. The commonly used stepwise procedures include the Bonferroni-Holm stepdown method (Holm, 1979), the Holm stepdown method (Dmitrienko et al., 2009, p.65), Hommel’s stepup procedure (Hommel, 1988), Hochberg’s stepup method (Hochberg, 1988), the fallback procedure (Wiens, 2003) and the sequential test with fixed sequences (Westfall et al., 1999).

#### Stepdown Procedure

A stepdown procedure starts with the most significant p-value and ends with the least significant p-value. In the procedure, the p-values are arranged in an ascending order, i.e., from the smallest to the largest:

\[ p_0 \leq p_1 \leq \ldots \leq p_k \]  

(1)

with the corresponding hypotheses:

\[ H_0(1), H_0(2), \ldots, H_0(k) \]  

(2)

The test proceeds from \( H_0(1) \) to \( H_0(k) \). If \( p_k \geq C_k \alpha \) (\( k = 1, \ldots, K \)), retain all \( H_0(i) (i \geq k) \); otherwise, reject \( H_0(k) \) and continue to test \( H_0(1) \). The constants \( C_k \) are different for different procedures.

The adjusted p-values are:

\[
\{ \hat{p}_k = C_k p_k \} \\
\hat{p}_k = \min \{ \hat{p}_k, C_k p_k \}, k = 2, \ldots, n
\]  

(3)

Therefore an alternative test procedure is to compare the adjusted p-values against the unadjusted \( \alpha \). After adjusting p-values, one can test the hypotheses in any order.

#### Fallback Procedure (Wiens, 2003)

The Holm procedure is based on a data-driven order of testing, while the fixed-sequence procedure is based on a prefixed order of testing. A compromise between them is the so-called fallback procedure. The fallback procedure was introduced by Wiens (2003) and was further studied by Dmitrienko et al. (2006; Hommel and Bretz, 2008). The test procedure can be outlined as follows.

In the fallback procedure, we allocate the overall error rate \( \alpha \) among the hypotheses according to their weights \( w_k \), where \( w_k \geq 0 \) and \( \sum w_k = 1 \). For fixed sequence test, \( w_1 = 1 \) and \( w_2 = \ldots = w_K = 0 \):

- **Step 1:** Test \( H_1 \) at \( \alpha_1 = \alpha w_1 \). If \( p_1 \geq \alpha_1 \), reject this hypothesis; otherwise retain it. Go to the next step.
- **Step \( i = 2, \ldots, K-1 \):** Test \( H_i \) at \( \alpha_i = \alpha_{i-1} + \alpha w_i \) if \( H_{i-1} \) is rejected and at \( \alpha_i = \alpha_{i-1} \) if \( H_{i-1} \) is retained. If \( p_i \leq \alpha_i \), reject \( H_i \); otherwise retaintit. Go to the next step.
- **Step \( K \):** Test \( H_K \) at \( \alpha_K = \alpha_{K-1} + \alpha w_K \) if \( H_{K-1} \) is rejected and at \( \alpha_K = \alpha w_K \) if \( H_{K-1} \) is retained. If \( p_K \leq \alpha_K \), reject \( H_K \); otherwise retain it.

The formula for the adjusted p-value is complicated to be written explicitly.

#### Stepup Procedure

A stepup procedure starts with the least significant p-value and ends with the most significant p-value. The procedure proceeds from \( H_{(K)} \) to \( H_{(1)} \). If \( p_{(k)} \leq C_k \alpha \) (\( k = 1, \ldots, K \), reject all \( H_{(i)} (i \leq k) \); otherwise, retain \( H_{(k)} \) and continue to test \( H_{(k+1)} \).

The adjusted p-values are:

\[
\{ \hat{p}_k = C_k p_k \} \\
\hat{p}_k = \min \{ \hat{p}_k, C_k p_k \}, k = K-1, \ldots, 1
\]  

(4)

#### Progressive \( \alpha \)-Exhaustive Testing Procedure

An \( \alpha \)-exhaustive procedure is a closed testing procedure based on intersection hypothesis tests the size of which is exactly \( \alpha \). In other words, \( \Pr (\text{Reject } H) = \alpha \) for any intersection hypothesis \( H = \bigcap \{1, \ldots, K \} \). Put in a simple way, in an \( \alpha \)-exhaustive procedure, the supremum of the probability of false rejection in any null hypothesis configuration is equal to \( \alpha \).

Many stepwise test procedures have been developed, which are not necessarily \( \alpha \)-exhaustive. Therefore, there is a room for improvement. However, an \( \alpha \)-exhaustive procedure is not necessarily a powerful test. A fixed sequence test is an \( \alpha \)-exhaustive test, but it is often the least powerful test if the sequence of tests was inappropriately chosen.

Let’s discuss the situation of two-hypothesis testing:

\[ H_a : H_1 \cap H_2 \text{ versus } H_a : \bar{H}_1 \cup \bar{H}_2 \]  

(5)

Here \( H_a \) is the negation of \( H_k, k = 1, 2 \). In this setting, if either \( H_1 \) or \( H_2 \) is rejected, the null hypothesis \( H_a \) is
rejected. Let $p_1$ and $p_2$ be the marginal p-values for testing $H_1$ and $H_2$, respectively.

The decision rules of the progressive $\alpha$-exhaustive testing procedure are:

- If $p_1p_2\leq \alpha_1$ and $p_1\leq \alpha$, then reject $H_1$
- If $p_1p_2\leq \alpha_2$ and $p_2\leq \alpha$, then reject $H_2$

where critical value $\alpha_1>0$ and $\alpha_2>0$.

The idea behind this procedure is to borrow strength among marginal p-values. In plain language, the procedure says that we don’t have to make an $\alpha$ adjustment, as long as $p_1\leq \alpha$ and the other p-value $p_2$ is small. For example, if $p_1 = \alpha$ and $p_2 = 0.01\alpha$, we can reject $H_1$. The $\alpha_1$ and $\alpha_2$ are so determined that when both $H_1$ and $H_2$ are true, the type-I error will not exceed $\alpha$.

The procedure can control the Familywise Error Rate (FWER) strongly but at the same time exhaust all $\alpha$ under all the null hypothesis configurations: $H_1$, $H_2$ and $H_1\cap H_2$. This is done progressively as described below.

Step 1: To control the familywise error rate at $\alpha$ level, when only $H_1$ is true and $H_2$ is not true, then $p_1p_2\leq \alpha_1$ can be satisfied with probability of 1 (if, for example, the test drug is very effective, $p_2$ will be virtually always 0). Therefore, to control FWER, a necessary condition is sup $Pr(p_1p_2\leq \alpha_1|p_1p_2\leq \alpha H_1\cap \bar{H}_2) = sup Pr(p_1\leq \alpha H_1) = \alpha$. Therefore, type-I error is strongly controlled and exhausted when $H_1\cap \bar{H}_2$ is true. By the same token, we can reach the same conclusion when $H_1\cap H_2$ is true.

Step 2: Now we need to determine $\alpha_1$ and $\alpha_2$ to exhaust $\alpha$ when $H_1\cap H_2$ is true. In this step, we only consider the case when the two test statistics, $p_1$ and $p_2$, are independent.

Under the global null hypothesis $H_0$, $p_1$ and $p_2$ are iid $U(0, 1)$, which is equivalent to two standard normal test statistics: $z_1 = 1 - \Phi(p_1)$ and $z_2 = 1 - \Phi(p_2)$ under $H_0$ being independent. However, working on the p-scale, the testing procedure can be used for different endpoints (normal, binary, survival), as long as $p_1$ and $p_2$ are independent and stochastically equal to or larger than uniform $p_1$ and $p_2$.

Since $T = p_1p_2<\alpha$ implies $p_1<\frac{\alpha_1}{p_1}$, we have the conditional cdf $F_{T|H_1}(T<\alpha|p_1)$

\[
F_{T|H_1}(T<\alpha|p_1) = \begin{cases} 
1, & \frac{\alpha_1}{p_1} \geq 1 \\
\frac{\alpha_1}{p_1} & \frac{\alpha_1}{p_1} < 1 
\end{cases} \tag{6}
\]

Thus:

\[
Pr(p_1p_2\leq \alpha_1|p_1p_2\leq \alpha) = \frac{\alpha_1}{p_1} \tag{7}
\]

Consequently, the type-I error rate under $H_1\cap H_2$ is given by (for simplicity, we just use FWER ($H_1\cap H_2$)):

\[
FWER(H_1\cap H_2) = Pr(p_1p_2\leq \alpha_1|p_1p_2\leq \alpha H_1\cap H_2) = Pr(p_1p_2\leq \alpha_1) - Pr(p_1p_2\leq \min(\alpha_1, \alpha_2)) \tag{8}
\]

In (9), we have used the following result: if $\alpha^2 \leq \min(\alpha_1, \alpha_2)$, then:
However, if $\alpha^2 > \min(\alpha_1, \alpha_2)$, then the probability becomes (Fig. 2):

$$\Pr\left(p_1, p_2 \leq \alpha_1 \cap p_1 \leq \alpha \cap p_2 \leq \alpha\right)$$

$$= \Pr\left(p_1 \leq \alpha \cap p_2 \leq \alpha\right) = \alpha^2$$

The $\alpha_1$ and $\alpha_2$ are determined so that $FWER(H_1 \cap H_2) = \alpha$. We are not interested in $\alpha_1 \geq \alpha_2$ because $p_1p_2 \leq \alpha_1$ in the rejection criteria has no effect. In fact, $FWER(H_1 \cap H_2) = 2\alpha_1\alpha_2 = \alpha$ will have no solution for any $\alpha$ between 0 and 1. We are not interested in $\alpha_1 < \alpha_2$ either, because it makes the conditions $p_1 < \alpha$ and $p_2 < \alpha$ have no effect in determining the rejection boundary. In fact:

$$FWER\left(H_1 \cap H_2\right) = \alpha_1 + \alpha_2 \ln \left(\frac{1}{\alpha_1\alpha_2}\right) - \alpha_1\alpha_2$$

has no solution for $\alpha_1 < \alpha_2$ and $\alpha_2 < \alpha$. Therefore, the only scenario that we are interested in is:

$$FWER\left(H_1 \cap H_2\right) = \alpha_1 + \alpha_2 \ln \left(\frac{1}{\alpha_1\alpha_2}\right) - \alpha_1\alpha_2$$

(14)

With (14), we can determine the rejection boundaries $\alpha_1, \alpha_2$ for given $\alpha$. Here are the steps:

- Choose $\alpha_1$ so that $\alpha_2 \leq \alpha_1 < \alpha$
- Let $FWER\left(H_1 \cap H_2\right) = \alpha$ to solve for $\alpha_2$

That is, $\alpha_2$ is the solution of:

$$\alpha_1 + \alpha_2 \ln \left(\frac{1}{\alpha_1\alpha_2}\right) - \alpha_1\alpha_2 = \alpha$$

(15)

Examples of critical values from (15) are presented in Table 1 and 2.

When $\alpha_1 = \alpha_2$, (15) can be simplified as:

$$2\left[\alpha_1 + \alpha_1 \ln \left(\frac{1}{\alpha_1}\right)\right] - \alpha_1^2 = \alpha$$

(16)

The critical values for various $\alpha$ with $\alpha_1 = \alpha_2$ are presented in Table 3.

The rejection boundaries in Table 1-3 have been verified each by 10,000,000 simulations.

**Illustrative Example**

Suppose in the Statistical Analysis Plan for a cardiovascular trial with two primary endpoints (not coprimary endpoints that have to be met simultaneously), the two test statistics for the two hypotheses ($H_1$ and $H_2$) of the two endpoints are assumed to be independent and the $\alpha$-exhaustive procedure (with one-sided $\alpha_1 = \alpha_2 = 0.004855$ and $\alpha = 0.025$) was specified in the analysis for
the multiplicity adjustment to control FWER. At the end of trial, the $p$-values for the two endpoints are: Scenario (1) one-sided $p_1 = 0.024$ and $p_2 = 0.025$, scenario (2) $p_1 = 0.024$ and $p_2 = 0.2$, (3) $p_1 = 0.05$ and $p_2 = 0.02$, (4) $p_1 = 0.01$ and $p_2 = 0.26$ and (5) $p_1 = 0.012$ and $p_2 = 0.5$. Using the $\alpha$-exhaustive procedure for scenario (1), we reject both $H_1$ and $H_2$ because $p_1 \times p_2 = 0.0006 \leq \alpha$, $p_1 \times p_2 = 0.024 \leq \alpha$ to reject $H_2$ and $p_1 \times p_2 = 0.0006 \leq \alpha$, $p_1 \times p_2 = 0.025 \leq \alpha$ to reject $H_2$. For scenario (2), we reject $H_1$ but not $H_2$ because $p_1 \times p_2 = 0.0048 \leq \alpha$, $p_1 \times p_2 = 0.024 \leq \alpha$ and $p_1 \times p_2 = 0.0048 \leq \alpha$, $p_1 \times p_2 = 0.2 > \alpha$. For scenario (3), we reject $H_2$, but not $H_1$. For scenario (4), we reject $H_1$ but not $H_2$. For scenario (5), we can reject neither $H_1$ nor $H_2$.

Using the test procedures described in the previous section, we summarize the rejection status in Table 4. The $\alpha$-exhaustive procedure can reject at least one hypothesis except for scenario 5. The reason that $\alpha$-Ex method (with $\alpha_1 = \alpha_2$) cannot reject any hypothesis for scenario (5) is that the method emphasizes the consistency of the evidence against all the hypotheses and such consistency is obviously not presented in scenario (5).

The power of $\alpha$-Ex procedure ($\alpha_1 = \alpha_2$) for the two-hypothesis at one-side $\alpha$-level can be written as:

$$Power = Pr\left\{\min \left(\Phi\left(-z_1\right), \Phi\left(-z_2\right)\right) \leq \alpha \mid \bar{P}_1, \bar{P}_2\right\}$$

As a comparison, the power of Hommel’s procedure for the two-hypothesis at one-sided $\alpha$-level can be written as:

$$Power = Pr\left\{\min \left(\Phi\left(-z_1\right), \Phi\left(-z_2\right)\right) \leq \alpha \mid \bar{P}_1, \bar{P}_2\right\}$$

Table 1. Critical values for one-sided $\alpha = 0.025$

| $\alpha$ | 0.000095 | 0.000650 | 0.001000 | 0.002000 | 0.003000 | 0.004000 | 0.004855 | 0.005000 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

Table 2. Critical values for one-sided $\alpha = 0.05$

| $\alpha$ | 0.000435 | 0.002500 | 0.004000 | 0.005000 | 0.006000 | 0.007000 | 0.008000 | 0.010000 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

Table 3. Critical values for one-sided test when $\alpha_1 = \alpha_2$

| $\alpha$ | 0.005000 | 0.010000 | 0.025000 | 0.050000 | 0.075000 | 0.100000 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\alpha_1, \alpha_2$ | 0.000941 | 0.001897 | 0.004855 | 0.010000 | 0.015739 | 0.021798 |

Table 4. Rejection with different test procedures

| Method          | Scenario 1 | 2 | 3 | 4 | 5 |
|-----------------|------------|---|---|---|---|
| Fixed Sequence  | $H_1, H_2$ | $H_1$ | $H_1$ | $H_1$ | $H_1$ |
| Bonferroni      | $H_1$ | $H_1$ | $H_1$ | $H_1$ | $H_1$ |
| Fallback ($\alpha = 0.5$) | $H_1$ | $H_1$ | $H_1$ | $H_1$ | $H_1$ |
| Holm-Stepdown   | $H_1$ | $H_2$ | $H_1$ | $H_1$ | $H_1$ |
| Hommel          | $H_1$ | $H_2$ | $H_1$ | $H_1$ | $H_1$ |
| $\alpha$-exhaustive | $H_1$ | $H_2$ | $H_1$ | $H_1$ | $H_1$ |

Note: One-sided $\alpha = 0.025$, $\alpha_1 = \alpha_2 = 0.004855$ for $\alpha$-exhaustive

Table 5. Power comparisons for two-hypothesis testing ($\delta_1 = 0.3$, $\sigma = 1$)

| $\delta_1 = 0.3$ | Method Seq ($H_1, H_2$) | Power | Power | Power | Power |
|------------------|--------------------------|-------|-------|-------|-------|
|                  | 0.640 | 0.800 | 0.224 | 0.800 | 0.200 | 0.020 | 0.025 |
|                  | 0.640 | 0.800 | 0.224 | 0.800 | 0.200 | 0.800 |       |
|                  | 0.640 | 0.800 | 0.224 | 0.800 | 0.200 | 0.800 |       |
|                  | 0.640 | 0.800 | 0.224 | 0.800 | 0.200 | 0.800 |       |
|                  | 0.640 | 0.800 | 0.224 | 0.800 | 0.200 | 0.800 |       |

Note: Sample size = 90, one-sided $\alpha = 0.025$, $\alpha_1 = \alpha_2 = 0.004855$ for Progressive $\alpha$-Ex
Power Comparison of Two Hypotheses

There seems a general impression that whatever the test procedure to use the power of rejection cannot be improved significantly for two-hypothesis testing. However, this is not necessarily true. We have compared power of seven different testing methods described in section 2 and presented results in Table 5, where Power 1 is probability of simultaneously rejecting $H_1; \delta_1 \leq 0$ and $H_2; \delta_2 \leq 0$ and Power is the probability of rejecting either $H_1$ or $H_2$. For the fallback procedure the weights $w_1 = w_2 = 0.5$ are used. The fixed sequence method is equivalent to the fallback procedure with $w_1 = 1$ and $w_2 = 0$.

The progressive $\alpha$-exhaustive procedure performs overall the best, while the Holm method performs the second best. In general, Holm procedure is uniformly more powerful than the Bonferroni procedure. Hochberg’s procedure is uniformly more powerful than Holm’s procedure and Hommel’s procedure is uniformly more powerful than Hochberg’s procedure. Holm, fixed-sequence and fallback are nonparametric and control FWER for any joint distribution of test statistics. Hommel and Hochberg procedures are semi parametric and control FWER only for some joint distributions, including positively dependent test statistics such as multivariate normal test statistics. Nonparametric procedures make no assumptions about the joint distribution of test statistics which results in power loss (Dmitrienko, 2013). For two-hypothesis testing, Hochberg’s method is equivalent to Holm’s method. The power of the fallback method depends on the weights $w_1$ and $w_2$ of the hypotheses.

Comparisons of Holm’s method to use of the $\alpha$-Ex method with different $\alpha_1$ and $\alpha_2$ are presented in Table 6. For $\alpha$-Ex$^1$, $\alpha_1 = \alpha_2 = 0.004855$; $\alpha$-Ex$^2$, $\alpha_1 = 0.003798$; $\alpha_2 = 1.6\alpha_1$; $\alpha$-Ex$^3$, $\alpha_1 = 0.004332, \alpha_2 = 1.25\alpha_1$; $\alpha$-Ex$^4$, $\alpha_1 = 0.004332, \alpha_2 = 1.25\alpha_2$. From the table, we can see that all $\alpha$-Ex procedures perform very well except $\alpha$-Ex$^3$, in which the treatment effect $\delta_1$ is smaller than $\delta_2$, but alphas were set up in a wrong direction ($\alpha_1 = 1.25\alpha_2 < \alpha_2$). In general, $\alpha_1$ should be chosen larger than $\alpha_2$ if $\delta_1$ is expected larger than $\delta_2$; otherwise choose $\alpha_1 \leq \alpha_2$.

Table 6. Power comparison for two-hypothesis testing

| Method          | 0.00/0.3 | 0.03/0.3 | 0.05/0.3 | 0.1/0.3 | 0.15/0.3 | 0.2/0.3 | 0.3/0.3 |
|-----------------|----------|----------|----------|---------|---------|---------|---------|
| Hommel          | 0.732    | 0.736    | 0.741    | 0.759   | 0.792   | 0.836   | 0.933   |
| $\alpha$-Ex$^1$ | 0.712    | 0.736    | 0.752    | 0.796   | 0.843   | 0.890   | 0.962   |
| $\alpha$-Ex$^2$ | 0.745    | 0.764    | 0.777    | 0.812   | 0.852   | 0.893   | 0.962   |
| $\alpha$-Ex$^3$ | 0.735    | 0.755    | 0.770    | 0.808   | 0.850   | 0.892   | 0.962   |
| $\alpha$-Ex$^4$ | 0.723    | 0.746    | 0.761    | 0.803   | 0.846   | 0.891   | 0.962   |
| $\alpha$-Ex$^5$ | 0.700    | 0.725    | 0.742    | 0.790   | 0.839   | 0.887   | 0.962   |

The reason that $\alpha$-Ex method (with $\alpha_1 = \alpha_2$) has lower power than Hommel’s method at the extreme case when $\delta_1 = 0$ and $\delta_2 = 0.3$ is that the former emphasizes the consistency of the evidence against all the hypotheses, while $\delta_1 = 0$ and $\delta_2 = 0.3$ are clearly inconsistent. If we believe $\delta_1$ is smaller than $\delta_2$, we should use different $\alpha_1$ and $\alpha_2$ (e.g., $\alpha_2 = 1.6\alpha_1$ in $\alpha$-Ex$^3$). However, if the directional guess is wrong, it could reduce the power as seen in $\alpha$-Ex$^3$.

Formulation for Three Hypotheses

We now discuss the progressive $\alpha$-Exhaustive procedure for three-hypothesis testing:

$$H_0 : H_1 \cap H_2 \cap H_3 vs H_p : \bar{H}_1 \cup \bar{H}_2 \cup \bar{H}_3$$ (17)

Similar to two-hypothesis testing, the rejection-acceptance rules of the $\alpha$-exhaustive procedure for three-hypothesis testing are:

- If $p_1p_2p_3 \leq \alpha_1 \wedge p_1p_3 \leq \alpha_2 \wedge p_1p_2 \leq \alpha_3$ reject $H_1$; otherwise accept $H_1$.
- If $p_1p_2p_3 \leq \alpha_1 \wedge p_2p_3 \leq \alpha_2 \wedge p_1p_3 \leq \alpha_3$ reject $H_2$; otherwise accept $H_2$.
- If $p_1p_2p_3 \leq \alpha_1 \wedge p_1p_3 \leq \alpha_2 \wedge p_1p_2 \leq \alpha_3$ reject $H_3$; otherwise accept $H_3$.

Here $\alpha_1 \leq \alpha_2 \leq \alpha_3$.

Determination of Critical Values

The derivations of the critical values are placed in the Appendix. Here we described the key steps and summarized the results. The critical values $\alpha_1$, $\alpha_2$ and $\alpha_3$ are determined by all the paired null hypotheses: $H_1 \cap H_2, H_2 \cap H_3, H_1 \cap H_3$. To exhaust familywise error rate, it necessarily requires that $FWER(H_1 \cap H_2) = \alpha$, $FWER(H_2 \cap H_3) = \alpha$ and $FWER(H_1 \cap H_3) = \alpha$, which are equivalent to (Appendix), respectively:

$$\alpha + \alpha_1 \ln \left( \frac{\alpha_2}{\alpha_1} \right) + \alpha_2 + \alpha_3 \ln \left( \frac{\alpha_2}{\alpha_3} \right) - \alpha' = \alpha$$

$$\alpha + \alpha_1 \ln \left( \frac{\alpha_2}{\alpha_1} \right) + \alpha_2 + \alpha_3 \ln \left( \frac{\alpha_3}{\alpha_2} \right) - \alpha' = \alpha, \alpha' < \alpha_3 \leq \alpha_2 < \alpha (18)$$

$$\alpha + \alpha_1 \ln \left( \frac{\alpha_2}{\alpha_1} \right) + \alpha_2 + \alpha_3 \ln \left( \frac{\alpha_3}{\alpha_2} \right) - \alpha' = \alpha$$

Each equation in (18) is in the same expression as (15). Therefore, the critical values in Table 1-3 are also valid for (18).
To exhaust $\alpha$ under $H_1 \cap H_2 \cap H_3$, it requires that $FWER(H_1 \cap H_2 \cap H_3) = \alpha$, that is (assume $\alpha_1 = \alpha_2 = \alpha_3$, Appendix):

$$3\alpha_1 \left(1 + \ln \frac{\alpha_1}{\alpha_1} \right) + \frac{3}{2} \alpha_1 ^2 = \alpha$$

(19)

Now we can use (18) to determine $\alpha_1, \alpha_2$ and $\alpha_3$ and use (19) to determine $\alpha_4$ for the case when $\alpha_1 = \alpha_2 = \alpha_3$. Examples of critical values for various $\alpha$ are presented in Table 7.

The critical values can also be determined through simulations. Especially when dimension is high, simulation is a convenient way to obtain the results: For given $(\alpha_1, \alpha_2, \alpha_3)$, we can use simulation by trying different $\alpha_4$ until $FWER(H_1 \cap H_2 \cap H_3) = \alpha$. We have verified the critical values through simulations: For $\alpha = 0.025$ and $\alpha_1 = \alpha_2 = \alpha_3 = 0.004855$, $\alpha_4 = 0.002677$; the type-I error rate is 0.025003 under $H_1 \cap H_2 \cap H_3$ through 10,000,000 simulations. This progressive method to determine the critical values can be generalized to $K$-hypothesis testing.

**Power Comparison**

Let $H_i$: $\delta \leq 0$, $i = 1, 2, 3$. Using the rejection boundaries in Table 7, we can easily obtain the power of the $\alpha$-Ex method through simulations. To compare the performance of $\alpha$-Ex, we compared the best method, Hommel's method as the standard.

Since there are three hypotheses, it is meaningful to compare our approach to other more recently developed approaches. However, the gate keeping procedure is difficult to communicate with the non-statisticians and requires large set of tests when the number of individual hypotheses increases. An iterative graphical approach by Bretz et al. (2009) deals with those weakness and constructs the Bonferroni-type tests with a simple updating algorithm that fully describes a sequentially rejective test procedure. The graphic approach was then extended by dissociating the underlying weighting strategy and applied using weighted Bonferroni tests, weighted parametric tests and weighted Simes tests (Bretz et al., 2011). The existing methods controls the FWER; However, the power is addressed or compared with other methods. From Table 8, we can see that the $\alpha$-Ex procedure provides more power in all cases except the case when $\delta_1 = \delta_2 = 0$ and $\delta = 0.3$.

Again, like in the case of two-hypothesis testing, when the parameters in the alternative hypotheses (e.g., effects of the different endpoints) are very different, we should use different $\alpha_1, \alpha_2$ and $\alpha_3$ such that their trend is in opposite to the trend of parameters in the alternative hypotheses.

### K-Hypothesis Testing Procedure

We now discuss the progressive $\alpha$-exhaustive procedure for $K$-hypothesis testing. To avoid the rejection boundary being too small, causing inconvenience, we use term $\sqrt{\prod_{i=1}^{K} p_i}$ instead of $\prod_{i=1}^{K} p_i$ in the decision rules for a general $K$-hypothesis testing. It is obvious that these two test statistics are equivalent in terms of power.

For $K$-hypothesis testing, the K rejection rules are specified as: We will reject $H_i$ if and only if:

$$\sum_{\ell=i}^{K} \left( p_\ell p_i \right)^{\frac{1}{2}} \leq \alpha_i, \sum_{\ell=i}^{K} \left( p_\ell \right)^{\frac{1}{2}} \leq \alpha_i, p_i \leq \alpha.$$
one-hypothesis testing, then given \(\alpha_1\), determine \(\alpha_2\)-based on two-hypothesis testing; and given \(\alpha_1\) and \(\alpha_2\),
determine \(\alpha_3\) based on three-hypothesis testing. The process continues until \(\alpha_3\) is determined. For high
dimensional hypothesis testing problems, Partition Principle for multiple testing (Hsu, 1996) can be used to
reduce the number of null configurations to be tested. Simulation is usually more convenient than numerical
integration when the dimension is high.

Summary and Discussion

To construct a MTP, we need to consider at least three things to ensure the power: (1) \(\alpha\)-exhaustive,
(2) synergize strengths among data for local hypothesis or marginal p-values and (3) be able to use correlations
between local test statistics or local p-values. In principle, the proposed \(\alpha\)-exhaustive procedure has
considered all three aspects. To achieve \(\alpha\)-exhaustive, we use the marginal p-value product corresponding to
each null hypothesis configuration and enforce it with an upper bound in the rejection rules. Such p-value product
terms in the rejection rules also ensure the synergy between the marginal p-values. The \(K\)-hypothesis testing
algorithm can be applied to the test statistics with correlations with modifications of critical regions for
rejection (the critical values in Table 1-3 are applicable for independent test statistics), but due to its complexity
and larger applications in clinical trials (dose-finding, subgroup analysis, adaptive design), we are developing
separate manuscripts to address that.

Unlike traditional stepwise procedures, the decision rule in the progressive \(\alpha\)-exhaustive procedure explicitly
uses a set of statistics \((p_1, p_2, p_3, p_4, p_5, p_6, p_7, \cdots, p_k)\) with a set of
critical values in the decision rule for rejecting a single \(H_0(k=1, 2, \cdots, K)\). In this sense, the decision rules
in the \(\alpha\)-exhaustive procedure are expressed in the form of "adjusted p-values" and hence the order of the tests
is irrelevant. Many stepwise testing procedures also stress the importance of clinical/practical
meaningfulness since the method emphasizes the consistency among the evidences coming from different
endpoints, different doses and different populations, that is, the totality of the evidence. The test procedure is
simple and performs well in broad situations. When the true "standardized" effect size (value of the parameter) is
very different for different hypothesis, the critical values don’t have to the same for rejecting all the hypotheses.
Instead, the critical values can be different and optimized based on the prior information on effect size or
considering the importance of different endpoints.

Appendices

Derivations of Formulations of Critical Values for Three-Hypothesis Testing

\[
\text{FWER}(H_1 \cap H_2) = \sup_{\mathcal{H}_i} \left( \Pr\left( p_1 \leq \alpha \cap \cdots \cap p_2 \leq \alpha \cap \cdots \cap p_3 \leq \alpha \right) \right)
\]

where, \(\text{FWER}(H_1 \cap H_2)\) is the type-I error rate under \(\text{FWER}(H_1 \cap H_2)\) and \(\sup\) is the supreme under all
possible \(H_i\).

To control type-I error under \(H_1 \cap H_2\), \(H_3 \cap H_5\) and \(H_5 \cap H_1\), it is required that, respectively:

\[
\begin{align*}
\alpha_1 + \alpha_2 \ln \left( \frac{\alpha}{\alpha_1} \right) + \alpha_3 + \alpha_4 \ln \left( \frac{\alpha}{\alpha_2} \right) - \alpha' = 0 \\
\alpha_1 + \alpha_2 \ln \left( \frac{\alpha}{\alpha_2} \right) + \alpha_3 + \alpha_4 \ln \left( \frac{\alpha}{\alpha_2} \right) - \alpha' = 0 \\
\alpha_1 + \alpha_2 \ln \left( \frac{\alpha}{\alpha_2} \right) + \alpha_3 + \alpha_4 \ln \left( \frac{\alpha}{\alpha_2} \right) - \alpha' = 0
\end{align*}
\]

From \((A1)\), we can see that among \(\alpha_1, \alpha_2\) and \(\alpha_3\), at least two should be the same, either \(\alpha_1 = \alpha_2\) or \(\alpha_2 = \alpha_3\) (\(\alpha_1 \leq \alpha_2\)).

The type-I error rate under \(H_1 \cap H_2 \cap H_3\) can be expressed as:

\[
\text{FWER}(H_1 \cap H_2 \cap H_3) = \Pr\left( p_1, p_2, p_3, p_4, p_5, p_6, p_7, \cdots, p_k \leq \alpha \cap \cdots \cap p_2 \leq \alpha \cap \cdots \cap p_3 \leq \alpha \cap \cdots \cap p_4 \leq \alpha \right)
\]

For the purpose of critical value derivation, we rewrite \(\text{FWER}(H_1 \cap H_2 \cap H_3)\) as:
\[
\text{FWER}\left( H_1 \cap H_2 \cap H_3 \right) = \pi_1 + \pi_2 + \pi_3 - \pi_{12} - \pi_{13} - \pi_{23} + \pi_{123}
\]

where, under \( H_1 \cap H_2 \cap H_3 \):

\[
\begin{align*}
\pi_1 &= \Pr\left\{ p_1p_2p_3 \leq \alpha_1 \cap p_1p_4 \leq \alpha_1 \cap p_2p_3 \leq \alpha_1 \cap p_1 \leq \alpha_1 \cap p_3 \leq \alpha_1 \right\}, \\
\pi_2 &= \Pr\left\{ p_1p_2p_3 \leq \alpha_1 \cap p_1p_4 \leq \alpha_1 \cap p_2p_3 \leq \alpha_1 \cap p_2 \leq \alpha_1 \cap p_3 \leq \alpha_1 \right\}, \\
\pi_3 &= \Pr\left\{ p_1p_2p_3 \leq \alpha_1 \cap p_1p_4 \leq \alpha_1 \cap p_2p_3 \leq \alpha_1 \cap p_3 \leq \alpha_1 \right\}, \\
\pi_{12} &= \Pr\left\{ p_1p_2p_3 \leq \alpha_1 \cap p_1p_4 \leq \alpha_1 \cap p_2p_3 \leq \alpha_1 \cap p_1 \leq \alpha_1 \cap p_3 \leq \alpha_1 \right\}, \\
\pi_{13} &= \Pr\left\{ p_1p_2p_3 \leq \alpha_1 \cap p_1p_4 \leq \alpha_1 \cap p_2p_3 \leq \alpha_1 \cap p_2 \leq \alpha_1 \cap p_3 \leq \alpha_1 \right\}, \\
\pi_{23} &= \Pr\left\{ p_1p_2p_3 \leq \alpha_1 \cap p_1p_4 \leq \alpha_1 \cap p_2p_3 \leq \alpha_1 \cap p_3 \leq \alpha_1 \right\}, \\
\pi_{123} &= \Pr\left\{ p_1p_2p_3 \leq \alpha_1 \cap p_1p_4 \leq \alpha_1 \cap p_2p_3 \leq \alpha_1 \cap p_1 \leq \alpha_1 \cap p_2 \leq \alpha_1 \cap p_3 \leq \alpha_1 \right\}
\end{align*}
\]

We will use Fig. 3 to assist in our integration of \( \pi_1, \pi_2 \) and \( \pi_3 \). That is:

\[
\pi_1 = \int_0^1 \int_0^{\alpha_1} \int_0^{\alpha_1} d\alpha_3 d\alpha_1 d\alpha_2 + \int_0^1 \int_0^{\alpha_1} \alpha_1 d\alpha_2 d\alpha_1 + \int_0^1 \int_0^{\alpha_1} \alpha_1 d\alpha_3 d\alpha_1
\]

Similarly, we use Fig. 4 to assist in our integration of \( \pi_{12}, \pi_{23} \) and \( \pi_{31} \). That is:

\[
\pi_{12} = \int_0^1 \int_0^{\alpha_1} \int_0^{\alpha_1} d\alpha_3 d\alpha_1 d\alpha_2 + \int_0^1 \int_0^{\alpha_1} \alpha_1 d\alpha_3 d\alpha_1 + \int_0^1 \int_0^{\alpha_1} \alpha_1 d\alpha_2 d\alpha_1
\]

For \( \pi_{123} \), we just give the result for the case when \( \alpha_1 = \alpha_2 = \alpha_3 \). Assume \( \alpha_i \geq \alpha_i^* \), otherwise only \( p_1p_2p_3 \leq \alpha_4 \) has an effect in the joint probabilities, while \( p_1p_2 \leq \alpha_1 \) and other will have no effect.
In fact, the $\alpha_i$ is so small that the three curves $p_i \alpha_i = \alpha_i$ cut the cube into a smaller cube $\alpha_1 \times \alpha_2 \times \alpha_3$, as shown in Fig. 5. Therefore, we have:

$$\pi_{i3} = \int_{\alpha_1} dp_i \pi_i = \alpha_i^3 \quad (A8)$$

To summarize, we have:

$$\text{FWER}(H_1 \cap H_2 \cap H_3) = 3\alpha_1 \left( a + \ln \frac{\alpha_1}{\alpha_a} \right)^2 + \alpha_1 \left[ \frac{\frac{\alpha_1}{\alpha_a}}{a} \right] - 3(2\alpha_1 - \alpha_i) + \alpha_i$$

$$= 3\alpha_1 \left[ \frac{\left( 1 + \ln \frac{\alpha_1}{\alpha_4} \right)^2}{a} \right] - 3\alpha_1 \left( 2\alpha_1 - \alpha_i \right) + \alpha_i^3 - 3\frac{\alpha_i^2}{\alpha} \quad (A9)$$

To control, $\text{FWER}(H_1 \cap H_2 \cap H_3)$ at $\alpha_1$ level, it is required that (assume we have chosen $\alpha_1 = \alpha_2 = \alpha_3$):

$$3\alpha_1 \left[ \frac{\left( 1 + \ln \frac{\alpha_1}{\alpha_4} \right)^2}{a} \right] - 3\alpha_1 \left( 2\alpha_1 - \alpha_i \right) + \alpha_i^3 - 3\frac{\alpha_i^2}{\alpha} = \alpha \quad (A10)$$

After $\alpha_1 = \alpha_2 = \alpha_3$ is determined using (A1), we can solve $\alpha_4$ from (A10). For a general case, after $\alpha_1$ is chosen between $\alpha_2$ and $\alpha_3$, $\alpha_2$ and $\alpha_3$ can be determined using (A1) and $\alpha_4$ can be easily obtained from simulations. The power simulations are presented in Table 8.

### SAS Code for Progressive Test Procedure

/* Progressive Alpha-exhaustive Test Procedure for Two-Hypothesis */

%Macro aExTest2H(nSims, u1, u2, sigma, N, alpha1, alpha2, alpha);

* nSims = the number of simulation runs;
* u1, u2 = parameters for H1 and H2. sigma = common standard deviation;
* N = sample size;
* alpha1, alpha2, alpha = critical values on p-scale;
* Power = prob of rejecting H1 or H2;
* PowerBoth = prob of rejecting H1 and H2;

Data aEx2H;

keep u1 u2 N PowerBoth Power;
N=&N; u1=&u1; u2=&u2; sigma=&sigma; alpha1=&alpha1; alpha2=&alpha2; alpha=&alpha;

Power=0; PowerBoth=0;
Do iSim=1 To nSims;
    z1=Rand("Normal", u1, sigma*sqrt(N))/sigma*sqrt(N);
    p1=1-CDF("Normal", z1);
    z2=Rand("Normal", u2, sigma*sqrt(N))/sigma*sqrt(N);
    p2=1-CDF("Normal", z2);
    z3=Rand("Normal", u3, sigma*sqrt(N))/sigma*sqrt(N);
    p3=1-CDF("Normal", z3);
    sig1=0; sig2=0; sig3=0;
    p4=p1*p2*p3;
    If p4<=&alpha1 And p1<=&alpha1 Then sig1=1;
    If p4<=&alpha2 And p2<=&alpha2 Then sig2=1;
    If sig1=1 OR sig2=1 Then Power=Power+1/nSims;
    sig1=1 And sig2=1 Then PowerBoth=PowerBoth+1/nSims;
End;
Output;
Run;
%

Title "Checking Type-I Error under H1 and H2";
%aExTest2H(10000000, 0, 0.0, 1, 90, 0.004855, 0.004855, 0.025);
Proc print data=aEx2H;
Run;
Title "Power under H1: u1=0.3 and H2: u2= 0.3";
%aExTest2H(1000000, 0.3, 0.3, 1, 90, 0.004855, 0.004855, 0.025);
Proc print data=aEx2H;
Run;
*/

/* Progressive Alpha-exhaustive Test Procedure for Three-Hypothesis */

%Macro aExTest3H(nSims, u1, u2, u3, sigma, N, alpha1, alpha4, alpha);

* nSims = the number of simulation runs;
* u1, u2, u3 = parameters for H1, H2 and H3;
* alpha1, alpha4, alpha = critical values on p-scale;
* Power = prob of rejecting H1 or H2 or H3;
* PowerAll = prob of rejecting H1, H2 and H3 simultaneouosly;

Data aEx3H;

keep u1 u2 u3 sigma N alpha PowerAll Power;
N=&N; u1=&u1; u2=&u2; u3=&u3; sigma=&sigma; N=&N; alpha1=&alpha1; alpha4=&alpha4; Power=0; PowerAll=0;

Do iSim=1 To nSims;
    z1=Rand("Normal", u1, sigma*sqrt(N))/sigma*sqrt(N);
    p1=1-CDF("Normal", z1);
    z2=Rand("Normal", u2, sigma*sqrt(N))/sigma*sqrt(N);
    p2=1-CDF("Normal", z2);
    z3=Rand("Normal", u3, sigma*sqrt(N))/sigma*sqrt(N);
    p3=1-CDF("Normal", z3);
    sig1=0; sig2=0; sig3=0;
    p4=p1*p2*p3;
    If p4<=&alpha1 And p1<=&alpha1 And p2<=&alpha1 Then sig1=1;
    If p4<=&alpha1 And p2<=&alpha1 And p3<=&alpha1 Then sig2=1;
    If sig1=1 OR sig2=1 Then Power=Power+1/nSims;
    sig1=1 And sig2=1 Then PowerBoth=PowerBoth+1/nSims;
End;
Output;
Run;
If sig1=1 And sig2=1 And sig3=1 Then
PowerAll=PowerAll+1/&nSims;
End;
Output;
Run;
%Mend;
Title "Checking Type-I Error under H1, H2 and H3";
%aExTest3H(10000000, 0, 0, 1, 0, 0.004855, 0.002677, 0.025);
proc print data=aEx3H;
Run;
Title "Power when u1=0, u2=0.3 and u3= 0.3";
%aExTest3H(100000, 0, 0.3, 0.3, 1, 60, 0.004855, 0.002677, 0.025);
Proc print data=aEx3H;
Run;

Acknowledgment

We would like to thank the anonymous reviewers and editors for their reviews.

Author’s Contributions

Mark Chang: Involved in the methodology development, draft and approval of the manuscript.
Xuan Deng: Contributed to review, revise and approval of the manuscript.
John Balser: Discussion on idea, issues, review manuscript.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the authors have read and approved the manuscript and no ethical issues involved.

Reference

Bretz, F., W. Maurer and G. Hommel, 2011. Test and power considerations for multiple endpoint analyses using sequentially rejective graphical procedures. Stat. Med., 30: 1489-1501. DOI: 10.1002/sim.3988
Dmitrienko, A., A.C. Tamhane and F. Bretz, 2009. Multiple Testing Problems in Pharmaceutical Statistics. 1st Edn., CRC Press, ISBN-10: 1584889853, pp: 320.
Dmitrienko, A., B.L. Wiens and P.H. Westfall, 2006. Fallback tests in dose-response clinical trials. J. Biopharmaceutical Stat., 16: 745-755. DOI: 10.1080/1054340600860600
Dmitrienko, A., 2013. Multiple testing procedures in clinical trials. Proceedings of the IBS Workshop, Sept. 19-20, Berlin.
Hochberg, Y., 1988. A sharper Bonferroni procedure for multiple tests of significance. Biometrika, 75: 800-802. DOI: 10.2307/2336325
Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scandinavian J. Stat., 6: 65-70.
Hommel, G., 1988. A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika, 75: 383-386. DOI: 10.1093/biomet/75.2.383
Hommel, G. and F. Bretz, 2008. Aesthetics and power considerations in multiple testing-a contradiction? Biom. J., 50: 657-666. DOI: 10.1002/bimj.200710463
Hsu, J.C., 1996. Multiple Comparisons: Theory and Methods. 1st Edn., CRC Press, ISBN-10: 0412982811, pp: 296.
Westfall, P.H., R.D. Tobias, D. Rom, R.D. Wolfinger and Y. Hochberg, 1999. Multiple comparisons and multiple tests using SAS system. SAS Institute, SAS Campus Drive, Cary, North Carolina, USA.
Wiens, B.L., 2003. A fixed sequence Bonferroni procedure for testing multiple endpoints. Pharmaceutical Stat., 2: 211-215. DOI: 10.1002/pst.64

Bretz, F., W. Maurer and G. Hommel, 2011. Test and power considerations for multiple endpoint analyses using sequentially rejective graphical procedures. Stat. Med., 30: 1489-1501. DOI: 10.1002/sim.3988
Dmitrienko, A., A.C. Tamhane and F. Bretz, 2009. Multiple Testing Problems in Pharmaceutical Statistics. 1st Edn., CRC Press, ISBN-10: 1584889853, pp: 320.
Dmitrienko, A., B.L. Wiens and P.H. Westfall, 2006. Fallback tests in dose-response clinical trials. J. Biopharmaceutical Stat., 16: 745-755. DOI: 10.1080/1054340600860600
Dmitrienko, A., 2013. Multiple testing procedures in clinical trials. Proceedings of the IBS Workshop, Sept. 19-20, Berlin.
Hochberg, Y., 1988. A sharper Bonferroni procedure for multiple tests of significance. Biometrika, 75: 800-802. DOI: 10.2307/2336325
Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scandinavian J. Stat., 6: 65-70.
Hommel, G., 1988. A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika, 75: 383-386. DOI: 10.1093/biomet/75.2.383
Hommel, G. and F. Bretz, 2008. Aesthetics and power considerations in multiple testing—a contradiction? Biom. J., 50: 657-666. DOI: 10.1002/bimj.200710463
Hsu, J.C., 1996. Multiple Comparisons: Theory and Methods. 1st Edn., CRC Press, ISBN-10: 0412982811, pp: 296.
Westfall, P.H., R.D. Tobias, D. Rom, R.D. Wolfinger and Y. Hochberg, 1999. Multiple comparisons and multiple tests using SAS system. SAS Institute, SAS Campus Drive, Cary, North Carolina, USA.
Wiens, B.L., 2003. A fixed sequence Bonferroni procedure for testing multiple endpoints. Pharmaceutical Stat., 2: 211-215. DOI: 10.1002/pst.64

Bretz, F., W. Maurer and G. Hommel, 2011. Test and power considerations for multiple endpoint analyses using sequentially rejective graphical procedures. Stat. Med., 30: 1489-1501. DOI: 10.1002/sim.3988
Dmitrienko, A., A.C. Tamhane and F. Bretz, 2009. Multiple Testing Problems in Pharmaceutical Statistics. 1st Edn., CRC Press, ISBN-10: 1584889853, pp: 320.
Dmitrienko, A., B.L. Wiens and P.H. Westfall, 2006. Fallback tests in dose-response clinical trials. J. Biopharmaceutical Stat., 16: 745-755. DOI: 10.1080/1054340600860600
Dmitrienko, A., 2013. Multiple testing procedures in clinical trials. Proceedings of the IBS Workshop, Sept. 19-20, Berlin.
Hochberg, Y., 1988. A sharper Bonferroni procedure for multiple tests of significance. Biometrika, 75: 800-802. DOI: 10.2307/2336325
Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scandinavian J. Stat., 6: 65-70.
Hommel, G., 1988. A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika, 75: 383-386. DOI: 10.1093/biomet/75.2.383
Hommel, G. and F. Bretz, 2008. Aesthetics and power considerations in multiple testing—a contradiction? Biom. J., 50: 657-666. DOI: 10.1002/bimj.200710463
Hsu, J.C., 1996. Multiple Comparisons: Theory and Methods. 1st Edn., CRC Press, ISBN-10: 0412982811, pp: 296.
Westfall, P.H., R.D. Tobias, D. Rom, R.D. Wolfinger and Y. Hochberg, 1999. Multiple comparisons and multiple tests using SAS system. SAS Institute, SAS Campus Drive, Cary, North Carolina, USA.
Wiens, B.L., 2003. A fixed sequence Bonferroni procedure for testing multiple endpoints. Pharmaceutical Stat., 2: 211-215. DOI: 10.1002/pst.64