Current Research in Biostatistics

Original Research Paper

An Alpha-Exhaustive Multiple Testing Procedure

[1,2]Mark Chang, [1]Xuan Deng and [2]John Balser

[1]Boston University, Boston MA, USA
[2]Veristat, Southborough MA, USA

Article history
Received: 19-07-2016
Revised: 07-08-2016
Accepted: 06-11-2016

Corresponding Author:
Mark Chang
Boston University, Boston MA, USA and
Veristat, Southborough MA, USA
Email: mark.chang@veristat.com

Abstract: A multiple testing procedure can be a single-step procedure such as Bonferroni's method or a stepwise procedure such as Hochberg's stepup method and Hommel's method. It can be an α-exhaustive or α-conservative approach. We develop a single α-exhaustive procedure that can improve power 2-5% over Hochberg's and Hommel's methods in common situations when the test statistics are mutually 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 kth order terms so that α 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.

Keywords: Multiple Testing, Alpha-Exhaustive, Hypothesis Test

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 stepup method. It can be anα-exhaustive or α-conservative approach. Conceptually, stepwise procedures are usually more powerful than single-step procedures and α-exhaustive procedures are usually more powerful than α-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 α-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 kth order terms so that α is exhausted 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. Unlike common stepwise test procedures, in which every step in the decision rule will only involve one critical value for decision-making, the proposed α-exhaustive approach is a single-step method with multiple critical values involved in the decision rules.

The paper is organized as follows. In section 2, we will review several important stepwise test procedures that will be used in our power comparisons. In section 3, we elaborate our progressive α-exhaustive procedure for two-hypothesis testing, outline the idea, derive the formulations for critical values and provide examples of using this procedure in comparison with other methods. We also provide the power formulation for the α-exhaustive procedure for two-hypothesis testing. In section 4, we provide power comparisons among several different methods using simulation. In section 5, we extend the α-exhaustive procedure to three-hypothesis testing and compare with Hommel's procedure in power under broad conditions. In section 6, we further describe the α-exhaustive procedure for general K-hypothesis testing and simulation algorithms for determining the critical values. In the last section, discussion and summary are provided. We place

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:

$$p_{(1)} \le p_{(2)} \le ... \le p_{(K)} \tag{1}$$

with the corresponding hypotheses:

$$H_{(1)}, H_{(2)}, ..., H_{(K)} \tag{2}$$

The test proceeds from $H_{(1)}$ to $H_{(K)}$. If $p_{(k)} > C_k\alpha$ (k = 1,..., K), retain all $H_{(i)}$ ($i \ge k$); otherwise, reject $H_{(k)}$ and continue to test $H_{(k+1)}$. The constants C_k are different for different procedures.

The adjusted p-values are:

$$\begin{cases} \tilde{p}_1 = C_1 p_{(1)} \\ \tilde{p}_k = \max\left(\tilde{p}_{k-1}, C_k p_{(k)}\right), k = 2,...,n \end{cases} \tag{3}$$

Therefore an alternative test procedure is to compare the adjusted p-values against the unadjusted α. 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 α among the hypotheses according to their weights w_k, where $w_k \ge 0$ and $\sum_k w_k = 1$. For fixed sequence test, $w_1 = 1$ and $w_2 = ... = w_K = 0$:

- Step 1: Test H_1 at $\alpha_1 = \alpha w_1$. If $p_1 \le \alpha_1$, reject this hypothesis; otherwise retain it. Go to the next step
- Step i = 2,..., K–1: 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 $p_k \le \alpha_k$, reject H_k; otherwise retainit. 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 $p_K \le \alpha_K$, reject H_K; otherwise retainit

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, $P_{(k)} \le C_k\alpha$ (k = 1, ..., K), reject all $H_{(i)}$ ($i \le k$); otherwise, retain $H_{(k)}$ and continue to test $H_{(k-1)}$.

The adjusted p-values are:

$$\begin{cases} \tilde{p}_K = C_K p_{(K)}, \\ \tilde{p}_k = \min\left(\tilde{p}_{k+1}, C_k p_{(k)}\right), k = K-1,...,1 \end{cases} \tag{4}$$

Progressive α-Exhaustive Testing Procedure

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

Many stepwise test procedures have been developed, which are not necessarily α-exhaustive. Therefore, there is a room for improvement. However, an α-exhaustive procedure is not necessarily a powerful test. A fixed sequence testis an α-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:

$$H_o : H_1 \cap H_2 \; versus \; H_\alpha : \bar{H}_1 \cup \bar{H}_2 \tag{5}$$

Here H_k 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_o is

rejected. Let p_1 and p_2 be the marginalp-values for testing H_1 and H_2, respectively.

The decision rules of the progressive α-exhaustive testing procedure are:

- If $p_1p_2 \le \alpha_1$ and $p_1 \le \alpha$, then reject H_1
- If $p_1p_2 \le \alpha_2$ and $p_2 \le \alpha$, 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 α adjustment, as long as $p_1 \le \alpha$ 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 α_1 and α_2 are so determined that when both H_1 and H_2 are true, the type-I error will not exceed α.

The procedure can control the Familywise Error Rate (FWER) strongly but at the same time exhaust all α 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 α level, when only H_1 is true and H_2 is not true, then $p_1p_2 \le \alpha_1$ 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 $\Pr(p_1p_2 \le \alpha_1 \cap p_1 \le \alpha | H_1 \cap \bar{H}_2)$ = sup $\Pr(p_1 \le \alpha | H_1) = \alpha$. 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 $H_2 \cap \bar{H}_1$ is true.

Step 2: Now we need to determine α_1 and α_2 to exhaust α when $H_1 \cap H_2$ is true. In this study, 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: $z_1 = 1\text{-}\Phi(p_1)$ and $z_2 = 1\text{-}\Phi(p_2)$ 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 $T = p_1p_2 < \alpha_1$ implies $p_2 < \frac{\alpha_1}{p_1}$, we have the conditional cdf for T:

$$F_{T|p_1}\left(T < \alpha_1 \mid p_1\right) = \begin{cases} 1, & \frac{\alpha_1}{p_1} \ge 1 \\ \frac{\alpha_1}{p_1}, & \frac{\alpha_1}{p_1} < 1 \end{cases} \tag{6}$$

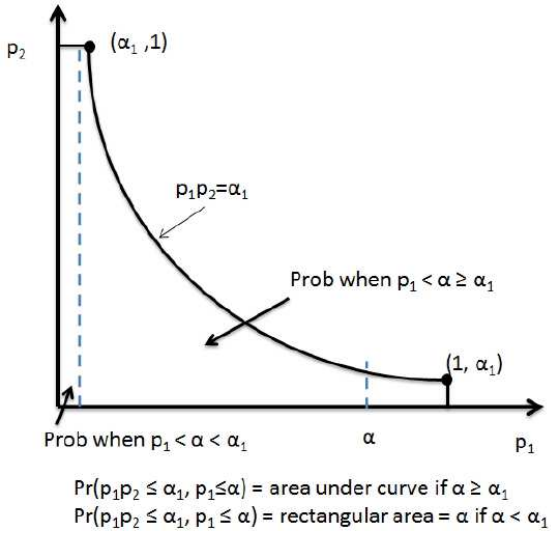

Fig. 1. Pr ($p_1p_2 \le \alpha_1 \cap p_1 \le \alpha$)

If $\alpha_1 \ge \alpha$, then Pr ($p_1p_2 \le \alpha_1 \cap p_1 \le \alpha | H_1, H_2$) = Pr ($p_1 \le \alpha | H_1, H_2$) = α. If $\alpha_1 < \alpha$, then (Fig. 1 for a geometric interpretation):

$$\begin{aligned} &\Pr\left(p_1p_2 \le \alpha_1 \cap p_1 \le \alpha\right) \\ &= \int_0^{\alpha} F_{T|p_1}\left(T < \alpha_1 \mid p_1\right) f\left(p_1\right) dp_1 = \int_0^{\alpha_1} dp_1 + \int_{\alpha_1}^{\alpha} \frac{\alpha_1}{p_1} dp_1 \\ &= \alpha_1 + \alpha_1 \ln\left(\frac{\alpha}{\alpha_1}\right) \end{aligned} \tag{7}$$

Thus:

$$\Pr\left(p_1p_2 \le \alpha_1 \cap p_1 \le \alpha\right) = \begin{cases} \alpha, & \alpha_1 \ge \alpha \\ \alpha_1 + \alpha_1 \ln\left(\frac{\alpha}{\alpha_1}\right), & \alpha_1 \ge \alpha \end{cases} \tag{8}$$

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

$$\begin{aligned} &FWER\left(H_1 \cap H_2\right) \\ &= \Pr\left(p_1p_2 \le \alpha_1 \cap p_1 \le \alpha \cup p_1p_2 \le \alpha_2 \cap p_2 \le \alpha\right) \\ &= \Pr\left(p_1p_2 \le \alpha_1 \cap p_1 \le \alpha\right) + \Pr\left(p_1p_2 \le \alpha_2 \cap p_2 \le \alpha\right) \\ &- \Pr\left(p_1p_2 \le \min\left(a_1, \alpha_2\right) \cap p_1 \le \alpha \cap p_2 \le \alpha\right) \\ &= \alpha_1 + \alpha_1 \ln\left(\frac{\alpha}{\alpha_1}\right) + \alpha_2 + \alpha_2 \ln\left(\frac{\alpha}{\alpha_2}\right) \\ &- \alpha^2, for\, \alpha^2 \le \min\left(\alpha_1, \alpha_2\right) \end{aligned} \tag{9}$$

In (9), we have used the following result: if $\alpha^2 \le \min(\alpha_1, \alpha_2)$, then:

$$\begin{aligned}&\Pr\left(p_1 p_2 \le \min\left(\alpha_1,\alpha_2\right)\cap p_1 \le \alpha \cap p_2 \le \alpha\right)\\&=\Pr\left(p_1 \le \alpha \cap p_2 \le \alpha\right)=\alpha^2\end{aligned} \tag{10}$$

However, if $\alpha^2 > \min(a_1, \alpha_2)$, then the probability becomes (Fig. 2):

$$\begin{aligned}&\Pr\left(p_1 p_2 \le \alpha_1 \cap p_1 \le \alpha \cap p_2 \le \alpha\right)\\&=\left(\int_0^{\min\left(\alpha_1,\alpha_2\right)/\alpha}\alpha dp_1+\int_{\min\left(\alpha_1,\alpha_2\right)/\alpha}^{\alpha}\frac{\alpha_1}{p_1}dp_1\right)\\&=\left(\min\left(\alpha_1,\alpha_2\right)+\min\left(\alpha_1,\alpha_2\right)\left(\ln\frac{\alpha^2}{\min\left(\alpha_1,\alpha_2\right)}\right)\right)\end{aligned} \tag{11}$$

To summarize the type-I error rates under various null configurations, we have:

$$\begin{cases}FWER\left(H_1\right)=\alpha\\FWER\left(H_2\right)=\alpha\end{cases} \tag{12}$$

$$FWER\left(H_1\cap H_2\right)=\begin{cases}2\alpha-\alpha^2, & \alpha_{\min}>\alpha\\ \alpha_1+\alpha_1\ln\left(\frac{\alpha}{\alpha_1}\right)+\alpha_2+\alpha_2\ln\left(\frac{\alpha}{\alpha_2}\right)-\alpha^2, & \alpha^2\le\alpha_{\min}\le\alpha\\ \alpha_1+\alpha_1\ln\left(\frac{\alpha}{\alpha_1}\right)+\alpha_2+\alpha_2\ln\left(\frac{\alpha}{\alpha_2}\right)-\alpha_{\min}\left(1+\ln\frac{\alpha^2}{\alpha_{\min}}\right), & \alpha_{\min}<\alpha^2\end{cases} \tag{13}$$

where, $\alpha_{\min} = \min(\alpha_1, \alpha_2)$:

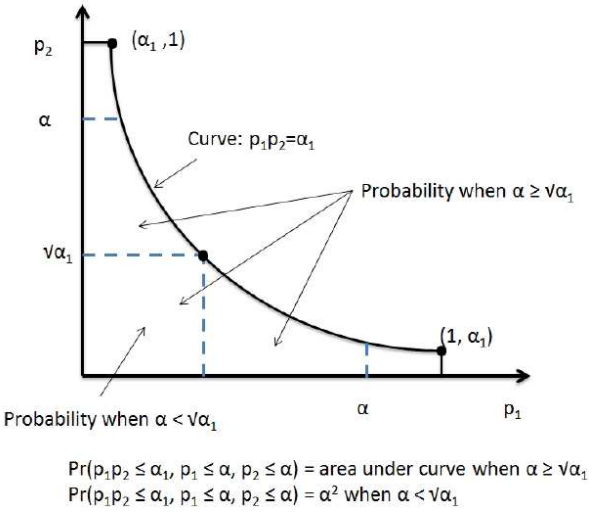

Fig. 2. Pr ($p_1p_2 \le \alpha_1 \cap p_1 \le \alpha \cap p_2 \le \alpha$)

The α_1 and α_2 are determined so that $FWER(H_1 \cap H_2) = \alpha$. We are not interested in $\alpha_1 \ge \alpha$, because $p_1p_2 \le \alpha_1$ in the rejection criteria has no effect. In fact, $FWER(H_1 \cap H_2) = 2\alpha - \alpha^2 = \alpha$will have no solution for anyα 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:

$$\begin{aligned}&FWER\left(H_1\cap H_2\right)=\alpha_1+\alpha_1\ln\left(\frac{\alpha}{\alpha_1}\right)\\&+\alpha_2+\alpha_2\ln\left(\frac{\alpha}{\alpha_2}\right)-\alpha_{\min}\left(1+\left(\ln\frac{\alpha^2}{\alpha_{\min}}\right)\right)=\alpha\end{aligned}$$

has no solution for $\alpha_1 < \alpha^2$ and $\alpha_2 < \alpha^2$. Therefore, the only scenario that we are interested in is:

$$\begin{aligned}&FWER\left(H_1\cap H_2\right)=\alpha_1+\alpha_1\ln\left(\frac{\alpha}{\alpha_1}\right)\\&+\alpha_2+\alpha_2\ln\left(\frac{\alpha}{\alpha_2}\right)-\alpha^2,\alpha^2\le\alpha_1\le\alpha\end{aligned} \tag{14}$$

With (14), we can determine the rejection boundaries α_1, α_2 for given α. Here are the steps:

- Choose α_1 so that $\alpha^2 \le \alpha_1 < \alpha$
- Let $FWER\ (H_1 \cap H_2) = \alpha$ to solve for α_2

That is, α_2 is the solution of:

$$\begin{aligned}&\alpha_1+\alpha_1\ln\left(\frac{\alpha}{\alpha_1}\right)+\alpha_2+\alpha_2\ln\left(\frac{\alpha}{\alpha_2}\right)\\&-\alpha^2=\alpha,\ \alpha^2\le\alpha_{\min}\le\alpha\end{aligned} \tag{15}$$

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

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

$$2\left[\alpha_1+\alpha\ln\left(\frac{\alpha}{\alpha_1}\right)\right]-\alpha^2=\alpha \tag{16}$$

The critical values for various α 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 co-primary 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 α-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 α-exhaustive procedure for scenario (1), we reject both H_1 and H_2 because $p_1 \times p_2 = 0.0006 \le \alpha_1 \cap p_1 = 0.024 \le \alpha$ to reject H_1 and $p_1 \times p_2 = 0.0006 \le \alpha_1 \cap p_2 = 0.025 \le \alpha$ to reject H_2. For scenario (2), we reject H_1 but not H_2 because $p_1 \times p_2 = 0.0048 \le \alpha_1 \cap p_1 = 0.024 \le \alpha$ and $p_1 \times p_2 = 0.0048 \le \alpha_2 \cap p_2 = 0.2 > \alpha$. 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 α-exhaustive procedure can reject at least one hypothesis except for scenario 5. The reason that α-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 α-Ex procedure ($\alpha_1 = \alpha_2$) for the two-hypothesis at one-side α-level can be written as:

$$Power = \Pr\left\{\begin{matrix}\Phi(-z_1)\Phi(-z_2) \le \alpha_1 \\ \cap\left(\min\left(\Phi(-z_1), \Phi(-z_2)\right) < \alpha \mid \bar{H}_1, \bar{H}_2\right)\end{matrix}\right\}$$

As a comparison, the power of Hommel's procedure for the two-hypothesis at one-sided α-level can be written as:

$$Power = \Pr\left\{\begin{matrix}\min\left(\Phi(-z_1)\Phi(-z_2)\right) \le \alpha \\ \cup \min\left(\Phi(-z_1), \Phi(-z_2)\right) < \alpha/2 \mid \bar{H}_1, \bar{H}_2\end{matrix}\right\}$$

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

α_1	0.000095	0.000650	0.001000	0.002000	0.003000	0.004000	0.004855	0.005000
α_2	0.025000	0.014884	0.012856	0.009378	0.007282	0.005814	0.004855	0.004714

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

α_1	0.000435	0.002500	0.004000	0.005000	0.006000	0.007000	0.008000	0.010097
α_2	0.050000	0.025265	0.020078	0.017610	0.015607	0.013934	0.012508	0.010097

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

α	0.005000	0.010000	0.025000	0.050000	0.075000	0.100000
α_1, α_2	0.000941	0.001897	0.004855	0.010097	0.015739	0.021798

Table 4. Rejection with different test procedures

	Scenario				
Method	1	2	3	4	5
Fixed Sequence	H_1, H_2	H_1		H_1	H_1
Bonferroni				H_1	H_1
Fallback ($w_1 = 0.5$)				H_1	H_1
Holm-Stepdown				H_1	H_1
Hochberg	H_1, H_2				H_1
Hommel	H_1, H_2			H_1	H_1
α-exhaustive	H_1, H_2	H_1	H_2	H_1	

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

Table 5. Power comparisons for two-hypothesis testing ($\delta_2 = 0.3$, $\sigma = 1$)

	$\delta_1 = 0.3$		$\delta_1 = 0.15$		$\delta_1 = 0$	
Method	Power[1]	Power	Power[1]	Power	Power[1]	Power
Fixed Seq (H_1,H_2)	0.640	0.800	0.224	0.280	0.020	0.025
Fixed Seq (H_2,H_1)	0.640	0.800	0.224	0.800	0.020	0.800
Bonferroni	0.529	0.926	0.150	0.727	0.009	0.731
Fallback (H_1,H_2)	0.590	0.926	0.168	0.784	0.010	0.730
Fallback (H_2,H_1)	0.590	0.926	0.214	0.783	0.018	0.731
Holm	0.652	0.926	0.233	0.784	0.019	0.730
Hochberg	0.660	0.933	0.241	0.791	0.020	0.732
Hommel	0.660	0.933	0.241	0.791	0.020	0.732
Progressive α-Ex	0.660	0.962	0.240	0.843	0.020	0.712

Note: Sample size = 90, one-sided $\alpha = 0.025$, $\alpha_1 = \alpha_2 = 0.004855$ for Progressive α-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 α-exhaustive procedure performs overall the best, while the Hommel 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 Hommel's method. The power of the fallback method depends on the weights w_i and the order of the hypotheses.

Comparisons of Hommel's method to use of the α-Ex method with different α_1 and α_2 are presented in Table 6. For α-Ex^1, $\alpha_1 = \alpha_2 = 0.004855$; α-Ex^2, $\alpha_1 = \alpha 0.003355$; $\alpha_2 = 2\alpha_1$; α-Ex^3, $\alpha_1 = 0.003798$, $\alpha_2 = 1.6\alpha_1$; α-Ex^4, $\alpha_1 = 0.004332$, $\alpha_2 = 1.25\alpha_1$; α-Ex^5, $\alpha_2 = 0.004332$, $\alpha_1 = 1.25\alpha_2$. From the table, we can see that all α-Ex procedures perform very well except α-Ex^5, in which the treatment effect δ_1 is smaller than δ_2, but alphas were set up in a wrong direction ($\alpha_1 = 1.25\alpha_2 > \alpha_2$). In general, α_1 should be chosen larger than α_2 if δ_1 is expected larger than δ_2; otherwise choose $\alpha_1 \leq \alpha_2$.

Table 6. Power comparison for two-hypothesis testing

	$\delta_1 = \delta_2 (\sigma = 1)$						
Method	0.0/0.3	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
α-Ex^1	0.712	0.736	0.752	0.796	0.843	0.890	0.962
α-Ex^2	0.745	0.764	0.777	0.812	0.852	0.893	0.962
α-Ex^3	0.735	0.755	0.770	0.808	0.850	0.892	0.962
α-Ex^4	0.723	0.746	0.761	0.803	0.846	0.891	0.962
α-Ex^5	0.700	0.725	0.742	0.790	0.839	0.887	0.962

The reason that α-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 δ_1 is smaller than δ_2, we should use different α_1 and α_2 (e.g., $\alpha_2 = 1.6\alpha_1$ in α-Ex^3). However, if the directional guess is wrong, it could reduce the power as seen in α-Ex^5.

Formulation for Three Hypotheses

We now discuss the progressive α-Exhaustive procedure for three-hypothesis testing:

$$H_0 : H_1 \cap H_2 \cap H_3 \; vs \; H_a : \bar{H}_1 \cup \bar{H}_2 \cup \bar{H}_3 \tag{17}$$

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

- If $p_1p_2p_3 \leq \alpha_4 \cap p_1p_2 \leq \alpha_1 \cap p_1p_3 \leq \alpha_1 \cap p_1 \leq \alpha$, reject H_1; otherwise accept H_1
- If $p_1p_2p_3 \leq \alpha_4 \cap p_2p_1 \leq \alpha_2 \cap p_2p_3 \leq \alpha_2 \cap p_2 \leq \alpha$, reject H_2, otherwise accept H_2
- If $p_1p_2p_3 < \alpha_4 \cap p_3p_1 \leq \alpha_3 \cap p_3p_2 \leq \alpha_3 \cap p_3 \leq \alpha$, 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 α_1, α_2 and α_3 are determined by all the paired null hypotheses: $H_1 \cap H_2$, $H_2 \cap H_3$ and $H_1 \cap H_3$. To exhaust familywise error rate, it necessarily requires that $FWER(H_1 \cap H_2) = \alpha$, $FWER(H_2 \cap H_3) = \alpha$ and $FWER(H_3 \cap H_1) = \alpha$, which are equivalent to (Appendix), respectively:

$$\begin{cases} \alpha_1 + \alpha_1 \ln\left(\dfrac{\alpha}{\alpha_1}\right) + \alpha_2 + \alpha_2 \ln\left(\dfrac{\alpha}{\alpha_2}\right) - \alpha^2 = \alpha \\ \alpha_2 + \alpha_2 \ln\left(\dfrac{\alpha}{\alpha_2}\right) + \alpha_3 + \alpha_3 \ln\left(\dfrac{\alpha}{\alpha_3}\right) - \alpha^2 = \alpha, \; \alpha^2 < \alpha_1 \leq \alpha_2 \leq \alpha_3 < \alpha \\ \alpha_3 + \alpha_3 \ln\left(\dfrac{\alpha}{\alpha_3}\right) + \alpha_1 + \alpha_1 \ln\left(\dfrac{\alpha}{\alpha_1}\right) - \alpha^2 = \alpha \end{cases} \tag{18}$$

Each equation in (18) is in the same expression as (15). Therefore, the critical values in Table 1-3 are also valid for (18).

Table 7. Critical values for one-sided test ($\alpha_1 = \alpha_2 = \alpha_3$)

α	0.010000	0.025000	0.050000	0.075000	0.100000
α_1, α_2,α_3	0.001897	0.004855	0.010097	0.015739	0.021798
α_4	0.001105	0.002677	0.005157	0.007566	0.009966

Table 8. Power comparison

	$\delta_1 = \delta_2$ ($\delta_3 = 0.3$, $\sigma = 1$)						
Method	0.0/0.0	0.0/0.3	0.03/0.3	0.2/0.3	0.1/0.2	0.1/0.1	0.3/0.3
Hommel	0.482	0.735	0.737	0.794	0.612	0.533	0.869
Dunnet Step-up	0.478	0.725	0.728	0.781	0.602	0.528	0.862
Graphic Approach	0.478	0.721	0.724	0.773	0.508	0.525	0.854
α-Ex	0.470	0.756	0.775	0.885	0.698	0.599	0.941

Note: $\alpha = 0.025$, $\alpha_1 = \alpha_2 = \alpha_3 = 0.004855$, $\alpha_4 = 0.002677$, sample size = 60

To exhaust α 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):

$$3\alpha_4\left[\left(1+\ln\frac{\alpha_1}{\alpha_4}\right)^2+1\right]-3\alpha_1\left(2\alpha-\alpha_1\right)+\alpha^3-3\frac{\alpha_1^2}{\alpha}=\alpha \qquad (19)$$

Now we can use (18) to determine α_1, α_2 and α_3 and use (19) to determine α_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 (α_1, α_2,α_3), we can use simulation by trying different α_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: $\delta_i \leq 0$, $i = 1, 2, 3$. Using the rejection boundaries in Table 7, we can easily obtain the power of the α-Ex method through simulations. To compare the performance of α-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 α-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 α_1, α_2 and α_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 α-exhaustive procedure for *K*-hypothesis testing. To avoid the rejection boundary being too small, causing inconvenience, we use term $\sqrt[k]{\prod_{i=1}^{k} p_i}$ 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:

$$\sum_{k>l,k\neq i,l\neq i}\left(p_i\,p_k\,p_l\right)^{1/3}\leq\alpha_{kl},\ \ \sum_{k\neq i}\left(p_i\,p_k\right)^{1/2}\leq\alpha_k,\ \ p_i\leq\alpha.$$

We didn't include higher order term of p-product because through simulations, we find that even we set $p_i p_j p_k p_m \leq 1$, the FWER is controlled. This means that for a multiple testing problem with more than four hypothese, the proposed procedure may not be an alpha-exhaustive one.

The rejection boundaries α_1, α_2,...α_K are progressively determined: Determine $\alpha_1 = \alpha$ based on

one-hypothesis testing, then given α_1, determine $\alpha_2$based on two-hypothesis testing; and given α_1 and α_2, determine α_3 based on three-hypothesis testing. The process continues until α_K 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) α-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 α-exhaustive procedure has considered all three aspects. To achieve α-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 α-exhaustive procedure explicitly uses a set of statistics (p_1,p_1p_2,p_1p_3,$p_1p_2p_3$,etc.) with a set of critical values in the decision rule for rejecting a single H_k(k = 1, 2,..., K). In this sense, the decision rules in the α-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 have the feature of borrowing strengths among data for local hypotheses, but such dependencies are realized through a discrete function. Theα-exhaustive procedure uses a continuous dependency function of marginal p-values, i.e., product of p-values, which is more effective. We have also tried other functions such as average p-values or linear combination of normal-inverse p-values, the results are similar.

In summary, the proposed progressive α-exhaustive procedure is not only statistically powerful, but it also stresses 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

$$
\begin{aligned}
&FWER\left(H_1 \cap H_2\right)\\
&= \sup_{H_3} \Pr\begin{pmatrix}\left(p_1p_2p_3 \le \alpha_4 \cap p_1p_2 \le \alpha_2 \cap p_1p_3 \le \alpha_2 \cap p_1 \le \alpha\right)\\ \cup\left(p_1p_2p_3 \le \alpha_4 \cap p_2p_1 \le \alpha_2 \cap p_2p_3 \le \alpha_2 \cap p_2 \le \alpha\right)\end{pmatrix}\\
&= \sup_{H_3} \Pr\left(\left(p_1p_2 \le \alpha_2 \cap p_1 \le \alpha\right) \cup \left(p_2p_1 \le \alpha_2 \cap p_2 \le \alpha\right)\right)\\
&= \Pr\left(p_1p_2 \le \alpha_2 \cap p_1 \le \alpha\right) + \Pr\left(p_2p_1 \le \alpha_2 \cap p_2 \le \alpha\right)\\
&- \Pr\left(p_1p_2 \le \alpha_2 \cap p_1 \le \alpha \cap p_2 \le \alpha\right)\\
&= \alpha_1 + \alpha_1 \ln\left(\frac{\alpha}{\alpha_1}\right) + \alpha_2 + \alpha_2 \ln\left(\frac{\alpha}{\alpha_2}\right) - \alpha^2, \alpha^2 \le \alpha_1, \alpha_2 < \alpha
\end{aligned}
$$

where, $FWER(H_1 \cap H_2)$ is the type-I error rate under $FWER(H_1 \cap H_2)$ and $\sup_{H_3}$ is the supreme under all possible H_3.

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

$$
\begin{cases}
\alpha_1 + \alpha_1 \ln\left(\frac{\alpha}{\alpha_1}\right) + \alpha_2 + \alpha_2 \ln\left(\frac{\alpha}{\alpha_2}\right) - \alpha^2 = \alpha\\
\alpha_2 + \alpha_2 \ln\left(\frac{\alpha}{\alpha_2}\right) + \alpha_3 + \alpha_3 \ln\left(\frac{\alpha}{\alpha_3}\right) - \alpha^2 = \alpha\\
\alpha_3 + \alpha_3 \ln\left(\frac{\alpha}{\alpha_3}\right) + \alpha_1 + \alpha_1 \ln\left(\frac{\alpha}{\alpha_1}\right) - \alpha^2 = \alpha
\end{cases} \qquad (A1)
$$

From (A_1), we can see that among α_1,α_2 and α_3, at least two should be the same, either$\alpha_1 = \alpha_2$ or$\alpha_2 = \alpha_3$ ($\alpha_1 \le \alpha_2$).

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

$$
\begin{aligned}
&FWER\left(H_1 \cap H_2 \cap H_3\right)\\
&= \Pr\begin{pmatrix}\left(p_1p_2p_3 \le \alpha_4 \cap p_1p_2 \le \alpha_1 \cap p_1p_3 \le \alpha_1 \cap p_1 \le \alpha\right)\\ \cup\left(p_1p_2p_3 \le \alpha_4 \cap p_2p_1 \le \alpha_2 \cap p_2p_3 \le \alpha_2 \cap p_2 \le \alpha\right)\\ \cup\left(p_1p_2p_3 \le \alpha_4 \cap p_3p_1 \le \alpha_3 \cap p_3p_2 \le \alpha_3 \cap p_3 \le \alpha\right)\end{pmatrix}
\end{aligned}
$$

For the purpose of critical value derivation, we rewrite $FWER(H_1 \cap H_2 \cap H_3)$as:

$$FWER\left(H_1 \cap H_2 \cap H_3\right) = \pi_1 + \pi_2 + \pi_3 - \pi_{13} - \pi_{23} - \pi_{33} + \pi_{123} \quad \text{(A2)}$$

where, under $H_1 \cap H_2 \cap H_3$:

$$\begin{cases} \pi_1 = \Pr\left(p_1 p_2 p_3 \le \alpha_4 \cap p_1 p_2 \le \alpha_1 \cap p_1 p_3 \le \alpha_1 \cap p_1 \le \alpha\right), \\ \pi_2 = \Pr\left(p_1 p_2 p_3 \le \alpha_4 \cap p_2 p_1 \le \alpha_2 \cap p_2 p_3 \le \alpha_2 \cap p_2 \le \alpha\right), \\ \pi_3 = \Pr\left(p_1 p_2 p_3 \le \alpha_4 \cap p_3 p_1 \le \alpha_3 \cap p_3 p_2 \le \alpha_3 \cap p_3 \le \alpha\right) \end{cases} \quad \text{(A3)}$$

$$\begin{cases} \pi_{12} = \Pr\left(\begin{array}{l} p_1 p_2 p_3 \le \alpha_4 \cap p_1 p_2 \le \alpha_1 \\ \cap p_1 p_3 \le \alpha_1 \cap p_2 p_3 \le \alpha_2 \cap p_1 \le \alpha \cap p_2 \le \alpha \end{array}\right), \\ \pi_{23} = \Pr\left(\begin{array}{l} p_1 p_2 p_3 \le \alpha_4 \cap p_2 p_1 \le \alpha_2 \\ \cap p_2 p_3 \le \alpha_2 \cap p_3 p_1 \le \alpha_3 \cap p_2 \le \alpha \cap p_3 \le \alpha \end{array}\right), \\ \pi_{31} = \Pr\left(\begin{array}{l} p_1 p_2 p_3 \le \alpha_4 \cap p_3 p_2 \le \alpha_3 \\ \cap p_1 p_3 \le \alpha_1 \cap p_1 p_2 \le \alpha_1 \cap p_3 \le \alpha \cap p_1 \le \alpha \end{array}\right), \end{cases} \quad \text{(A4)}$$

and:

$$\pi_{123} = \Pr\left(\begin{array}{l} p_1 p_2 p_3 \le \alpha_4 \cap p_1 p_2 < \alpha_1 \cap p_2 p_3 < \alpha_2 \\ \cap p_3 p_1 < \alpha_1 \cap p_1 < \alpha \cap p_2 < \alpha \cap p_3 < \alpha \end{array}\right) \quad \text{(A5)}$$

We will use Fig. 3 to assist in our integration of π_1, π_2 and π_3. That is:

$$\begin{aligned} \pi_1 &= \int_{\Omega_1} a dp_2 dp_1 + \int_{\Omega_2} \frac{\alpha_1}{p_1} dp_2 dp_1 + \int_{\Omega_3} \frac{\alpha_1}{p_2} dp_1 dp_2 \\ &+ \int_{\Omega_4} \frac{\alpha_4}{p_1 p_2} dp_2 dp_1 \\ &= \alpha \left(\frac{\alpha_1}{\alpha}\right)^2 + 2\left(\int_{\frac{\alpha_1}{\alpha}}^{\frac{\alpha_4}{\alpha_1}} \int_0^{p_1} \frac{\alpha_1}{p_1} dp_2 dp_1 + \int_{\frac{\alpha_4}{\alpha_1}}^{1} \frac{\alpha_1}{p_1} dp_2 dp_1\right) \\ &+ \int_{\frac{\alpha_4}{\alpha_1}}^{1} \int_{\frac{\alpha_4}{\alpha_1}}^{1} \frac{\alpha_4}{p_1 p_2} dp_2 dp_1 \\ &= \alpha_4 \left(1 + \ln \frac{\alpha_1}{\alpha_4}\right)^2 + \alpha_4 - \frac{\alpha_1^2}{\alpha} \end{aligned} \quad \text{(A6)}$$

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

$$\begin{aligned} \pi_{12} &= \int_{\Omega_1} dp_2 dp_1 + \int_{\Omega_2} \frac{\alpha_1}{p_1} dp_2 dp_1 + \int_{\Omega_3} \frac{\alpha_1}{p_2} dp_1 dp_2 \\ &= \alpha_1^2 + \left(since\ \alpha < \sqrt{\alpha_1}\right)^2 \int_{\alpha_1}^{\alpha} \int_0^{p_1} \frac{\alpha_1}{p_2} dp_2 dp_1 \\ &= a_1^2 + 2a_1\left(\alpha - \alpha_1\right) = 2\alpha\alpha_1 - \alpha_1^2 \end{aligned} \quad \text{(A7)}$$

For π_{123}, we just give the result for the case when $\alpha_1 = \alpha_2 = \alpha_3$. Assume $\alpha_4 \ge \alpha_1^3$, otherwise only $p_1p_2p_3 \le \alpha_4$ has an effect in the joint probabilities, while $p_1p_2 \le \alpha_1$ and other will have no effect.

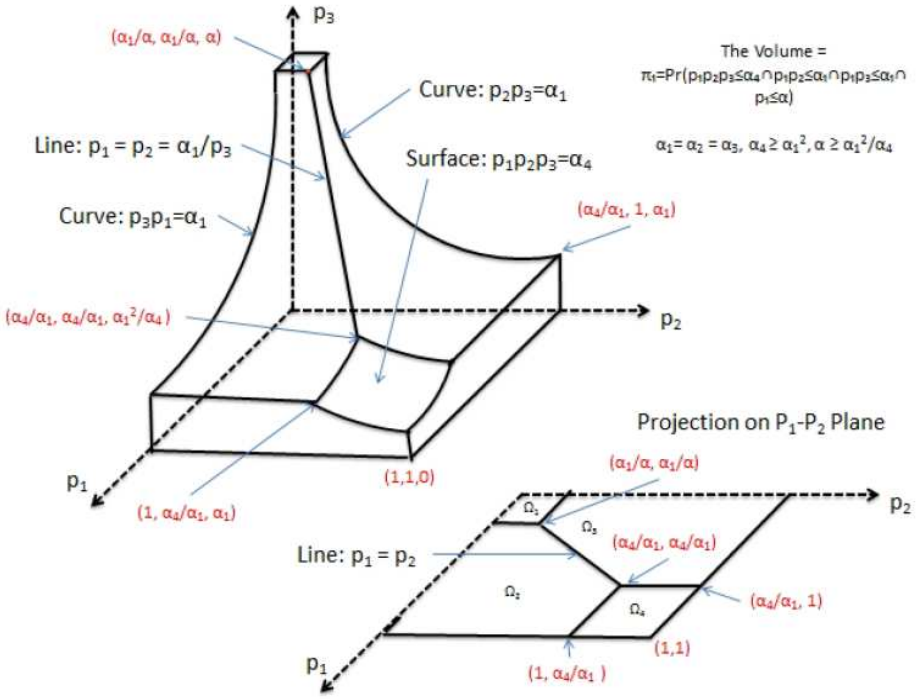

Fig. 3. $\pi_1 = \pi_2 = \pi_3$

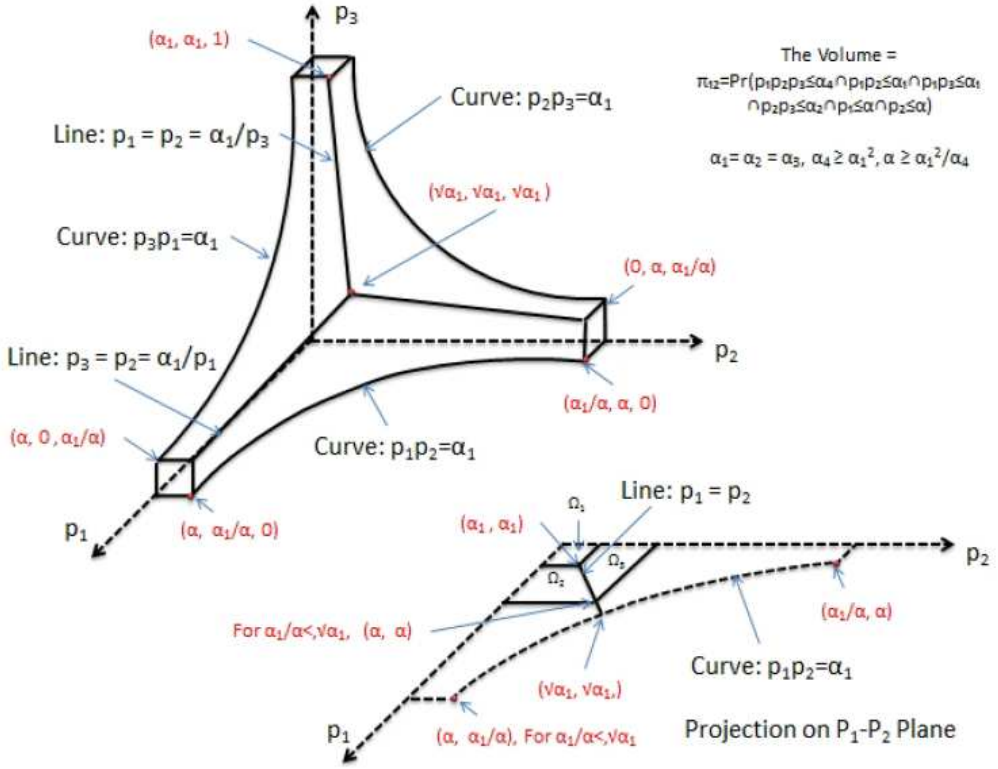

Fig. 4. π_{12}

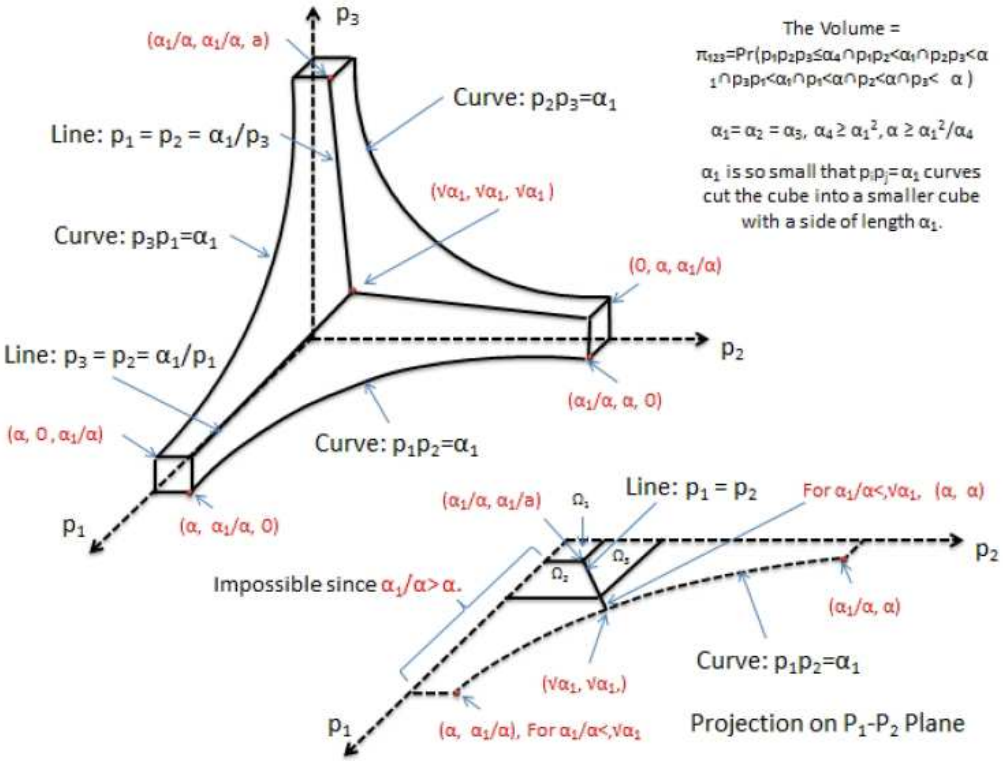

Fig. 5. π_{123}

In fact, the α_1 is so small that the three curves $p_i p_j=\alpha_1$ cut the cube into a smaller cube $\alpha_1\times\alpha_1\times\alpha_1$, as shown in Fig. 5. Therefore, we have:

$$\pi_{123}=\int_{\Omega_1} dp_1\,dp_1=\alpha_1^3 \tag{A8}$$

To Summarize, we have:

$$\begin{aligned} &FWER\left(H_1\cap H_2\cap H_3\right)\\ &=3\left[\alpha_4\left(a+\ln\frac{\alpha_1}{\alpha_4}\right)^2+\alpha_4-\frac{\alpha_1^2}{\alpha}\right]-3\left(2\alpha\alpha_1-\alpha_1^2\right)+\alpha^3 \\ &=3\alpha_4\left[\left(1+\ln\frac{\alpha_1}{\alpha_4}\right)^2+1\right]-3\alpha_1\left(2\alpha-\alpha_1\right)+\alpha^3-3\frac{\alpha_1^2}{\alpha}\end{aligned} \tag{A9}$$

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

$$3a_4\left[\left(1+\ln\frac{\alpha_1}{\alpha_4}\right)^2\right]-3\alpha_1\left(2\alpha-\alpha 1\right)+\alpha^3-3\frac{\alpha_1^2}{\alpha}=\alpha \tag{A10}$$

After $\alpha_1=\alpha_2=\alpha_3$ is determined using (A1), we can solve α_4 from (A10). For a general case, after α_1 is chosen between α_2 and α,α_2 and α_3 can be determined using (A1) and α_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; 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);
sig1=0; sig2=0;
If p1*p2<=&alpha1 And p1<=alpha Then sig1=1;
If p1*p2<=&alpha2 And p2<=alpha Then sig2=1;
If sig1=1 OR sig2=1 Then Power=Power+1/&nSims;
If sig1=1 And sig2=1 Then PowerBoth=PowerBoth+1/&nSims;
End;
Output;
Run;
%Mend;
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;
* N = sample size;
* u1, u2, u3 = parmeters for H1, H2 and H3;
* alpha1, alpha2, alpha = cretical values on p-scale;
* Power = prob of rejecting H1 or H2 or H3;
* PowerAll = prob of rejecting H1, H2 and H3 simutanneuosly;
Data aEx3H;
keep u1 u2 u3 sigma N alpha PowerAll Power;
u1=&u1; u2=&u2; u3=&u3; sigma=&sigma; N=&N;
alpha=&alpha; 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<=alpha4 & p1*p2<=alpha1 & p1*p3<=alpha1 & p1<=alpha Then sig1=1;
If p4<=alpha4 & p2*p1<=alpha1 & p2*p3<=alpha1 & p2<=alpha Then sig2=1;
If p4<=alpha4 & p3*p1<=alpha1 & p3*p2<=alpha1 & p3<=alpha Then sig3=1;
If sig1=1 OR sig2=1 Or sig3=1 Then Power=Power+1/&nSims;
```

```
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, 0, 1, 60, 0.004855, 0.002677, 0.025);
proc print data=aEx3H;
Run;
Title "Power when u1=0, u2=0.3 and u3= 0.3";
%aExTest3H(1000000, 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 reveiw, 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, W. Brannath and M. Posch, 2009. A graphical approach to sequentially rejective multiple test procedures. Stat. Med., 28: 586-604. DOI: 10.1002/sim.3495

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: 10.1080/10543400600860600

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 Drivew, 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