SMGroup

SM Journal of Biometrics & Biostatistics

Research Article

Progressive Alpha-Exhaustive Multiple Testing Procedure with Independent Test Statistics

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

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

Article Information

Received date: Nov 17, 2016
Accepted date: Dec 26, 2016
Published date: Dec 30, 2016

***Corresponding author**

Mark Chang, Boston University, Boston MA, USA, Email: Mark.Chang@veristat.com

Keywords Multiple Testing; Alpha-Exhaustive; Hypothesis Test

Abstract

A multiple testing procedure can be a single-step data-independent procedure, such as Bonferroni's method, or a data-dependent stepwise procedure such as Hochberg's step-up method and Hommel's method. It can be an *α*-exhaustive, where the maximum type-I error rate under all configurations of null hypotheses equals *α*, or *α*-conservative, where the type-I error rate falls below the nominal level. We develop a simple one-step a-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 correlated test statistics. In our method we construct the stopping rules using the product of marginal p-values and control the upper bounds of the k^{th} order terms so that *α* is exhausted for any configuration of *k* null hypotheses. Such upper bounds are determined progressively from *k* = 1 towards *k* = K, the number of null hypotheses in the problem. The method can be used in different multiple testing problems, including adaptive clinical trial designs.

Introduction

Multiple testing problems are common in pharmaceutical statistics and in lifesciences research in general. It is well known that testing multiple hypotheses (multiple testing) can inflate the type-I error dramatically if performed without proper adjustments of the p-values or significance level cutoffs. This is referred to as a "multiplicity issue". In clinical trials, multiplicity can come from different sources, including (1) multiple treatment comparisons, (2) tests performed at different times during the study, such as in adaptive trials, (3) tests for several endpoints, and (4) tests conducted for multiple populations using the same treatment(s) within a single experiment [1]. In their review of multiple testing in clinical trials [2], provided motivating examples where multiplicity issues may exist. A few of these examples include: (A) A clinical trial that compares two doses of a new treatment to a single control with respect to the primary efficacy endpoint may have a multiplicity issue. (B) In a clinical trial, there may be two endpoints. Multiplicity may exist if at least one is required to be statistically significant to claim efficacy versus if both need to be statistically significant. (C) In a trial with three specified primary endpoints E1, E2 and E3, it may be required that either E1 needs to be statistically significant or both E2 and E3 need to be statistically significant to claim efficacy. (D) In a trial with two endpoints, it may be required that one of the two specified endpoints must be statistically significant with superiority and the other one shown to be noninferior.

The main purpose of multiple testing procedures is to (strongly) control the Familywise type-I Error Rate (FWER), i.e., to control the type-I error rate at or below the nominal level under any combination of hypotheses. A multiple testing procedure can be a single-step data-independent procedure, such as Bonferroni's method, or a data-dependent stepwise procedure such as Hochberg's step-up and Hommel's stepwise procedures. The methods may be described as *α*-exhaustive or *α*-conservative. An *α*-exhaustive procedure is a closed testing procedure based on intersection hypothesis tests the size of which is exactly *α* [3,4]; in other words, the maximum type-I error rate under any combination of hypotheses is equal to *α*. In an *α*-conservative approach, the type-I error rate is controlled at or below the nominal level, often by dividing *α* between the hypothesis tests.

Conceptually, stepwise procedures are typically more powerful than single-step procedures; while *α*-exhaustive procedures tend to be more powerful than *α*-conservative approaches.

In this paper we proposed a simple one-step *α*-exhaustive procedure that can improve power 2%-5% over Hochberg's and Hommel's methods in common study designs where the test statistics are independent. The method can also be generalized to be applied to dependent test statistics. The idea behind our method is to construct the stopping rules using the product of marginal *p*-values and to control the upper bounds of the k^{th} 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, where K is the total number of null hypotheses in the problem. Unlike more common

OPEN ACCESS

ISSN: 2573-5470

stepwise test procedures where 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 at the same time.

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. We 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. In the last section, discussion and summary are provided. 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 tend to be more powerful than single-step procedures because the stepwise methods test hypotheses sequentially, allowing a data-dependent α-sharing between the tests, resulting in larger significance cutoffs than the single-step procedures where all hypotheses are tested simultaneously.

There are three categories of stepwise procedures that are dependent on how the stepwise tests proceed: step-up, stepdown, and fallback procedures. The commonly used stepwise procedures include the Bonferroni-Holm stepdown method [5], the Holm stepdown method [4], Hommel's step-up procedure [6], Hochberg's step-up method [3], the fallback procedure [7], and the sequential test with fixed sequences [8].

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:

$$p_{(1)} \le p_{(2)} \le \ldots \le p_{(K)} \qquad (1)$$

with the corresponding hypotheses

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

The test proceeds from $H_{(1)}$ to $\mathrm{H}_{(K)}$. If $p_{(k)} > C_k \alpha \left(k = 1, \ldots, K\right)$, retain all $H_{(i)} \left(i \ge k\right)$; otherwise, reject $H_{(k)}$ and continue to test $H_{(k+1)}$. C_k values are differ for the different multiple testing procedures.

The adjusted p-values are

$$\begin{cases} \overline{p}_1 = C_1 p_{(1)} \\ \overline{p}_k = \max\left(\overline{p}_{k-1}, C_k p_{(k)}\right), \; k = 2, \ldots., K. \end{cases} \qquad (2)$$

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.

Holm Stepdown procedure

Suppose there are K hypothesis tests $H_i \left(i = 1, \ldots, K\right)$. The Holm stepdown procedure [4,5] can be outlined as follows:

Step 1. If $p_{(1)} \le \alpha / K$, reject $H_{(1)}$ and go to the next step. Otherwise retain all hypotheses and stop.

Step i (i = 2,….,K-1). If $\Pr\left(p_1 p_2 \le \alpha_1 \cap p_1 \le \alpha\right)$, reject $H_{(i)}$ and go to the next step. Otherwise retain $H_{(i)}, \ldots., H_{(K)}$ and stop.

Step K. If $p_{(K)} \le \alpha$, reject $H_{(K)}$. Otherwise retain $H_{(K)}$.

The adjusted p-values are given by

$$\overline{p}_k = \begin{cases} p_{(K)} & \text{if } k = K \\ \min\left(\overline{p}_{k+1}, \left(K - k + 1\right) p_{k+1}\right) & \text{if } k = K - 1, \ldots, 1. \end{cases} \qquad (3)$$

The adjusted p-values can be used for hypothesis testing as was described for stepdown procedures in general.

Fallback procedure [7]

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 [7] and was further studied by Dmitrienko, Wiens, and Waterfall [9] and Hommel & Bretz [10]. 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 $w_2 = \ldots = w_k = 0$.

Step 1: Test H_1 at α_1 = αw_1. If $p_1 \le \alpha_1$, reject this hypothesis; otherwise retain it. Go to the next step.

Step $i = 2, \ldots., K - 1$: Test H_k at $\alpha_k = \alpha_{k-1} + \alpha w_k$ if $H_{k\text{-}1}$ is rejected and at α_k = αw_k if $H_{k\text{-}1}$ is retained. If $p_k \le \alpha_k$, reject H_k; otherwise retain it. Go to the next step.

Step K: Test H_K at $\alpha_K = \alpha_{K-1} + \alpha w_K$ if $H_{K\text{-}1}$ is rejected and at $\alpha_K = \alpha w_K$ if $H_{K\text{-}1}$ is retained. If $p_K \le \alpha_K$, reject H_K; otherwise retain it.

The formula for the adjusted p-value is complicated to be written explicitly, therefore, only the significance level is being adjusted, not the p-value.

Step-up procedure

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

The adjusted p-values are

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

Hochberg Step-up procedure

Step 1: If $p_{(K)} > \alpha$, accept $H_{(K)}$ and go to Step 2; otherwise reject all null hypotheses and stop.

Steps $k = 2,....,K-1$: If $p_{(K-k+1)} > \alpha/k$, accept $H_{(K\text{-}k+1)}$ and go to Step k+1, otherwise reject all remaining null hypotheses and stop.

Step K: If $p_{(1)} > \alpha/K$, accept $H_{(1)}$; otherwise reject $H_{(1)}$.

The constants C_k for the Hochberg step-up procedure are C_k = K- k +1 (k = 1,…,K). The Hochberg step-up method does not control FWER for all correlations, and is conservative when p-values are independent (Westfall, et al., 1999, p.33) [8].

As a reference for determination of C_k for various methods, one can refer to the book by Dmitrienko, Tamhane, and Bretz, F. (2010) [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 (Grechanosky and Hochberg, 1999; Demitrienko, et al. 2010)[3,4]. In other words, in an α-exhaustive procedure, the supremum of the probability of false rejection in any null hypothesis configuration is equal to α; equivalently, Pr (Reject H_I) = α for any intersection hypothesis H_I, $I \subseteq \{1,...,K\}$.

Many of the currently available stepwise test procedures are not α-exhaustive which provides a natural area for improvement; however it is worth noting that an α-exhaustive procedure is not necessarily a powerful test. For example, a fixed sequence test is an α-exhaustive test but if the sequence of tests is chosen inappropriately, it may also be the least powerful test.

First we will discuss the situation of testing two-hypotheses:

$$H_0 : H_1 \cap H_2 \textit{ Versus } H_a : \bar{H}_1 \cup \bar{H}_2. \tag{5}$$

Here $\bar{H}_k$ 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 marginal p-values for testing H_1 and H_2, respectively. For the development of a progressive α-exhaustive testing procedure, we will borrow strength from marginal p-values to aid in the decision for rejection or fail to rejection of the null hypothesis.

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

If $p_1 p_2 \le \alpha_1$ and $p_1 \le \alpha$, then reject H_1,

If $p_1 p_2 \le \alpha_2$ and $p_2 \le \alpha$, then reject H_2,

where critical value α_1 > 0 and α_2 > 0 and are determined such that when both H_1 and H_2 are true, the type-I error will not exceed α.

The idea behind this procedure is to borrow strength among marginal p-values. In plain language, the procedure does not require an α adjustment, as long as $p_1 \le \alpha$ and the other p-value p_2 is small. For example, if p_1=α and p_2 = 0.01α (see the next section), we can reject H_1.

With appropriate selection of α_1 and α_2, the procedure will control the FWER strongly while simultaneously exhausting 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: Consider when only H_1 is true and H_2 is not true. If, for example, the test drug is very effective with respect to H_2, p_2 will be very close to 0.

The resulting probability that $p_1 p_2 \le \alpha_1$ will be equal to 1. Therefore, to control FWER, a necessary condition is $\sup \Pr\left(p_1 p_2 \le \alpha_1 \cap p_1 \le \alpha \middle| H_1 \cap \bar{H}_2\right) = \sup \Pr\left(p_1 \le \alpha \middle| H_1\right) = \alpha$. Type-I error is strongly controlled and exhausted when $H_1 \cap \bar{H}_2$ is true. 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 paper, 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 independent and identically distributed as U (0, 1) random variables, which can be equivalently expressed as two independent standard normal test statistics: z_{1-p_1} and z_{1-p_2} under H_0. However, working on the p-scale, the testing procedure can be used for different endpoints (normal, binary, survival).

Since $T = p_1 p_2 < \alpha_1$ implies $p_2 < \frac{\alpha_1}{p_1}$, we have the conditional cdf for T:

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

If $\alpha_1 \ge \alpha$, then $\Pr\left(p_1 p_2 \le \alpha_1 \cap p_1 \le \alpha \middle| H_1, H_2\right) = \Pr\left(p_1 \le \alpha \middle| H_1, H_2\right) = \alpha$.

If $\alpha_1 < \alpha$, then (see Figure 1 for a geometric interpretation)

$$\begin{aligned} &\Pr\left(p_1 p_2 \le \alpha_1 \cap p_1 \le \alpha\right) \\ &= \int_0^{\alpha} F_{T|p_1}\left(T < \alpha_1 \middle| p_1\right) f\left(p_1\right) dp_1 = \int_0^{\alpha_1} dp_1 + \int_{\alpha_1}^{\alpha} \frac{\alpha_1}{p_1} dp_1 \\ &= \alpha_1 + \alpha_1 \ln\left(\frac{\alpha}{\alpha_1}\right). \end{aligned} \tag{7}$$

Thus,

$$\Pr\left(p_1 p_2 \le \alpha_1 \cap p_1 \le \alpha\right) = \begin{cases} \alpha, & \alpha_1 \ge \alpha \\ \alpha_1 + \alpha_1 \ln\left(\frac{\alpha}{\alpha_1}\right), & \alpha_1 < \alpha \end{cases} \tag{8}$$

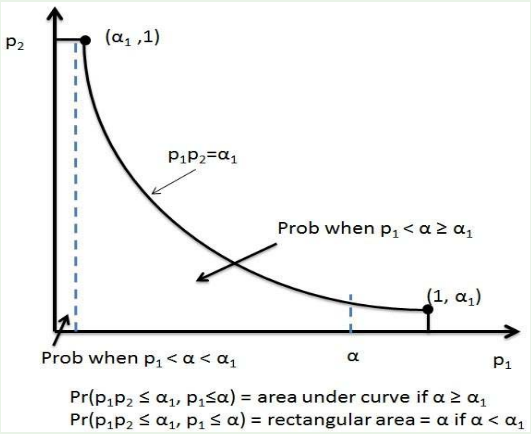
p₂
(α₁ ,1)
p₁p₂=α₁
Prob when p₁ < α ≥ α₁
(1, α₁)
Prob when p₁ < α < α₁
α
p₁
Pr(p₁p₂ ≤ α₁, p₁≤α) = area under curve if α ≥ α₁
Pr(p₁p₂ ≤ α₁, p₁ ≤ α) = rectangular area = α if α < α₁

Figure 1: Area under Curve represents the probability, $\Pr\left(p_1 p_2 \le \alpha_1 \cap p_1 \le \alpha\right)$.

Citation: Chang M, Deng X, Balser J and Bliss R. Progressive Alpha-Exhaustive Multiple Testing Procedure with Independent Test Statistics. SM J Biometrics Biostat. 2016; 1(1): 1003.

Consequently, the type-I error rate under $H_1 \cap H_2$, denoted FWER $(H_1 \cap H_2)$ is given by

$$\begin{aligned} &FWER(H_1 \cap H_2) \\ &= \Pr(p_1p_2 \le \alpha_1 \cap p_1 \le \alpha \cup p_1p_2 \le \alpha_2 \cap p_2 \le \alpha) \\ &= \Pr(p_1p_2 \le \alpha_1 \cap p_1 \le \alpha) + \Pr(p_1p_2 \le \alpha_2 \cap p_2 \le \alpha) \quad (9) \\ &\quad - \Pr(p_1p_2 \le \min(\alpha_1, \alpha_2) \cap p_1 \le \alpha \cap p_2 \le \alpha). \end{aligned}$$

Assuming that $\alpha^2 \le \min(\alpha_1, \alpha_2)$,

$$\Pr(p_1p_2 \le \min(\alpha_1, \alpha_2) \cap p_1 \le \alpha \cap p_2 \le \alpha) = \Pr(p_1 \le \alpha \cap p_2 \le \alpha) = \alpha^2.$$

As a resut,

$$FWER(H_1 \cap H_2) = \alpha_1 + \alpha_1 \ln\left(\frac{\alpha}{\alpha_1}\right) + \alpha_2 + \alpha_2 \ln\left(\frac{\alpha}{\alpha_2}\right) - \alpha^2,\ for\ \alpha^2 \le \min(\alpha_1, \alpha_2). \quad (10)$$

However, if $\alpha^2 > \min(\alpha_1, \alpha_2)$ (see Figure 2), then the probability becomes

$$\begin{aligned} &\Pr(p_1p_2 \le \alpha_1 \cap p_1 \le \alpha \cap p_2 \le \alpha) \\ &= \left(\int_0^{\min(\alpha_1,\alpha_2)/\alpha} \alpha dp_1 + \int_{\min(\alpha_1,\alpha_2)/\alpha}^{\alpha} \frac{\alpha_1}{p_1} dp_1 \right) \quad (11) \\ &= \left(\min(\alpha_1,\alpha_2) + \min(\alpha_1,\alpha_2) \left(\ln \frac{\alpha^2}{\min(\alpha_1,\alpha_2)} \right) \right) \end{aligned}$$

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

$$\begin{cases} FWER(H_1) = \alpha \\ FWER(H_2) = \alpha \end{cases} \quad (12)$$

$$FWER(H_1 \cap H_2)$$

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

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

The next step is to determine α_1 and α_2 such that $FWER(H_1 \cap H_2) = \alpha$.

Note that we do not consider $\alpha_1 \ge \alpha$, 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 any α between 0 and 1. Similarly, we don't consider $\alpha_1 < \alpha^2$ either, because it makes the conditions, $p_1 < \alpha$ and $p_2 < \alpha$, have no effect in determining the rejection boundary. The equation,

$$FWER(H_1 \cap H_2) = \alpha_1 + \alpha_1 \ln\left(\frac{\alpha}{\alpha_1}\right) + \alpha_2 + \alpha_2 \ln\left(\frac{\alpha}{\alpha_2}\right) - \alpha_1\left(1 + \left(\ln \frac{\alpha^2}{\alpha_1}\right)\right) = \alpha$$

has no solution for $\alpha_1 < \alpha^2$ *and* $\alpha_2 < \alpha^2$. Therefore, the only relevant scenario is where

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

From (14), we can determine the rejection boundaries α_1, α_2 for given α using the following procedure:

(1) choose α_1 such that $\alpha^2 \le \alpha_1 < \alpha$. (2) let $FWER(H_1 \cap H_2) = \alpha$ to solve for α_2, where α_2 is the solution of

$$\alpha_1 + \alpha_1 \ln\left(\frac{\alpha}{\alpha_1}\right) + \alpha_2 + \alpha_2 \ln\left(\frac{\alpha}{\alpha_2}\right) - \alpha^2 = \alpha, \alpha^2 \le \min(\alpha_1, \alpha_2) \le \alpha \quad (15)$$

Examples of stopping boundaries from (15) are presented in Tables 1 and 2.

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

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

Table 1: Stopping Boundaries for One-Sided α = 0.025.

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

Table 2: Stopping Boundaries for One-Sided α = 0.05.

α_1	0.000435	0.002500	0.004000	0.005000	0.006000	0.007000	0.008000	0.010097
α_2	0.50000	0.025265	0.020078	0.017610	0.015607	0.013934	0.012508	0.010097

The stopping boundaries for various α with $\alpha_1 = \alpha_2$ are presented in Table 3.

The rejection boundaries in Tables 1, 2 and 3 have been verified each by 10,000,000 simulations.

When $\alpha_1 = \alpha_2$, the power of the α-exhaustive procedure for the two-hypothesis at one-side α-level can be written as

$$Power = \Pr\left\{\Phi(-z_1)\Phi(-z_2) \le \alpha_1 \cap \left(\min\left(\Phi(-z_1), \Phi(-z_2)\right)\right) < \alpha \middle| \bar{H}_1, \bar{H}_2\right\}.$$

Table 3: Stopping Boundary for One-Sided Test when $\alpha_1 = \alpha_2$.

α	0.005000	0.010000	0.025000	0.050000	0.075000	0.100000
$\alpha_{1,}\ \alpha_2$	0.000941	0.001897	0.004855	0.010097	0.015739	0.021798

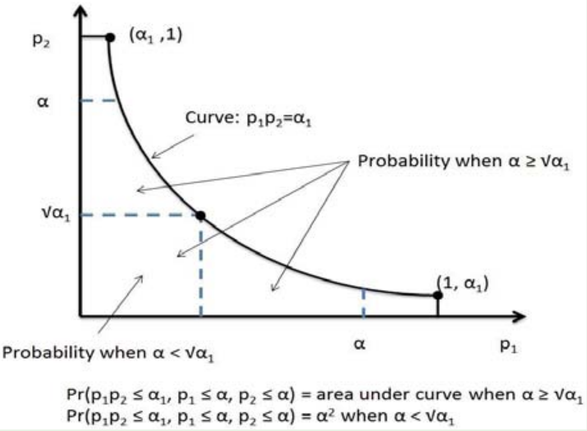
p₂
(α₁ ,1)
α
Curve: p₁p₂=α₁
Probability when α ≥ √α₁
√α₁
(1, α₁)
Probability when α < √α₁
α
p₁
Pr(p₁p₂ ≤ α₁, p₁ ≤ α, p₂ ≤ α) = area under curve when α ≥ √α₁
Pr(p₁p₂ ≤ α₁, p₁ ≤ α, p₂ ≤ α) = α² when α < √α₁

Figure 2: Area under Curve represents the probability, $\Pr(p_1p_2 \le \alpha_1 \cap p_1 \le \alpha)$.

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

$$Power = \Pr\left\{\max\left(\Phi\left(-z_1\right),\Phi\left(-z_2\right)\right)\le\alpha\cup\min\left(\Phi\left(-z_1\right),\Phi\left(-z_2\right)\right)<\alpha/2\middle|\bar{H}_1,\bar{H}_2\right\}$$

Illustrative example

Suppose in the Statistical Analysis Plan for a cardiovascular trial with two primary endpoints (note these are 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 endpoints are assumed to be independent, and the α-exhaustive procedure (with a one-sided $\alpha_1 = \alpha_2 = 0.004855$ and $\alpha = 0.025$) was specified 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 $p_1 p_2 = 0.0006 \le \alpha_1 \cap p_1 = 0.024 \le \alpha$ to reject H_1 and $p_1 p_2 = 0.0006 \le \alpha_1 \cap p_2 = 0.025 \le \alpha$ to reject H_2. For scenario (2), we reject H_1 but not H_2 because $p_1 p_2 = 0.0048 \le \alpha_1 \cap p_1 = 0.024 \le \alpha$ and $p_1 p_2 = 0.0048 \le \alpha_2 \cap p_2 = 0.2 > \alpha$. 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 Section 2, we summarize the rejection status in Table 4. The α-exhaustive procedure can reject at least one hypothesis except for scenario (5), where the method fails to reject a hypothesis because it emphasizes the consistency of the evidence against all the hypotheses, and such consistency is not presented in this scenario.

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
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 α= 0.025, α_1= α_2= 0.004855 for α-exhaustive.

Trial example

In a recent Phase 3, Randomized, Double-Blind Trial of a PARP1/2 Inhibitor Drug versus Placebo in Patients with Ovarian Cancer, the objective was to evaluate the primary endpoint of Progression Free Survival (PFS) in a cohort of patients with Germline Breast Cancer Susceptibility Gene (BRCA) mutation tumors (gBRCAmut cohort) and in a separate, independent cohort of patients with high grade serous or high grade predominantly serous histology who were not gBRCAmut carriers (non-gBRCAmut cohort). Over 100 international study sites were planned to enroll a total of 540 patients with 180 in the gBRCAmut cohort and 360 to non-gBRCAmut cohort. Patients were randomized to the treatment and placebo with randomization ratio 2:1. The drug or placebo was administered once daily continuously during a 28-day cycle. The log rank test will be used for the analysis of PFS. The overall FWER on the primary efficacy endpoint of PFS was controlled at a one sided 0.025 significance level. The hazard ratio between the two treatment groups was 0.3 for gBRCAmut Cohort with p-value p_1 = 0.001 and 0.5 for the non-gBRCAmut with p-value p_2 = 0.002. Using $\alpha_1 = \alpha_2 = 0.004855$ for the rejection criteria, because $p_1 p_2 < \alpha_1$ and $p_1 < \alpha$, H_1 is rejected. The treatment effect is significance in gBRCAmut cohort. Similarly, because $p_1 p_2 < \alpha_1$ and $p_2 < \alpha$, H_2 is rejected as well. The treatment effect is also significance in non-gBRCAmut cohort. (Due to confidentiality concerns, we adopt changes to mask the data and to fit the purpose of the methodological illustrations).

Power Comparisons of Two Hypotheses

There may be a general impression that regardless of the multiple testing procedure applied, the power of rejection will be approximately the same and cannot be improved in a two-hypothesis testing scenario. 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 the probability of simultaneously rejecting $H_1 : \delta_1 \le 0$ and $H_2 : \delta_2 \le 0$, 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 the best overall, 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. The Holm, fixed-sequence, and fallback procedures are nonparametric and control FWER for any joint distribution of test statistics. The Hommel and Hochberg procedures are semiparametric and control FWER for only 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 a loss of power [11]. 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.

Table 5: Power Comparisons for Two-Hypothesis Testing (δ_2 = 0.3, σ= 1).

Method	δ_1 = 0.3		δ_1 = 0.15		δ_1 = 0	
	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 α = 0.025, α_1 = α_2 = 0.004855 for Progressive α-Exhaustive.

A comparison of Hommel's method to use of the α-exhaustive method with different α_1 and α_2 is presented in Table 6. For α-Ex[1], $\alpha_1 = \alpha_2 = 0.004855$; α-Ex[2], $\alpha_1 = 0.003355$, $\alpha_2 = 2\alpha_1$; 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-exhaustive procedures perform well except α-Ex[5], in which the treatment effect δ_1 is smaller than δ_2, and alphas were set up in the 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 $\alpha_1 \le \alpha_2$.

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-exhaustive 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. When $\delta_1 = 0$ and $\delta_2 = 0.3$, the results are 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]). Even if we are wrong about the direction of δ_1 versus δ_2, the power of the α-exhaustive method will be still higher than Hommel's method as seen in α-Ex[5]. For instance, we use $\alpha_1 = 1.25\ \alpha_2$ as in α-Ex[5], when in fact $\delta_1/\delta_2 = 0.05/0.3$, the proposed method has higher power. If, however, we are completely wrong about the direction of δ_1 versus δ_2, e.g., when $\delta_1/\delta_2 = 0/0.3$, the proposed method will have lower power (70%) than the Hommel's power (73%). In general, if we do not know which delta is larger, we should use equal α, i.e., $\alpha_1 = \alpha_2$.

Formulation for Three Hypotheses

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

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

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

- If $p_1 p_2 p_3 \le \alpha_4 \cap p_1 p_2 \le \alpha_1 \cap p_1 p_3 \le \alpha_1 \cap p_1 \le \alpha$, reject H_1; otherwise accept H_1.
- If $p_1 p_2 p_3 \le \alpha_4 \cap p_2 p_1 \le \alpha_2 \cap p_2 p_3 \le \alpha_2 \cap p_2 \le \alpha$, reject H_2, otherwise accept H_2.
- If $p_1 p_2 p_3 \le \alpha_4 \cap p_3 p_1 \le \alpha_3 \cap p_3 p_2 \le \alpha_3 \cap p_3 \le \alpha$, reject H_3, otherwise accept H_3.

Here $\alpha_1 \le \alpha_2 \le \alpha_3$.

Determination of critical values

The derivations of the critical values are placed in the Appendix. In this section we describe 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 FWER, 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,

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

The equations in (18) are of the same expression as we saw in equation (15). Therefore, the critical values in Tables 1, 2 and 3 can also be used as solutions to (18).

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 such stopping boundaries for various α are presented in Table 7.

The critical values can also be determined through simulations which can provide a convenient solution when dimension is high: 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 stopping boundaries can be generalized to K-hypothesis testing using a similar procedure.

The critical values in Table 7 are typical values for the multiple testing, from which further optimization can be by select critical values that maximize the power. If we prior distribution, g (δ_1, δ_2, δ_3), we can simulations by trying different values for α_1, α_2, and α_3 to maximize the expected (predictive) power

$$\int Power\left(\delta_1, \delta_2, \delta_3 \middle| \alpha_1, \alpha_2, \alpha_3\right) g\left(\delta_1, \delta_2, \delta_3\right) d\delta_1 d\delta_2 d\delta_3,$$

where the integration can be approximated using summation over various δ_1, δ_2, and δ_2 in practice.

Power comparison

Let $H_i : \delta_i \le 0$, $i = 1, 2, 3$. Using the rejection boundaries in Table 7, we can obtain the power of the α-exhaustive method through simulations and compare it to the performance of Hommel's method as a standard.

The power of the two methods is compared in Table 8. We can see that the α-exhaustive procedure provides more power in all cases except the case when $\delta_1 = \delta_2 = 0$ and $\delta = 0.3$.

Again, as was observed in the case of two-hypothesis testing, when the parameters in the alternative hypotheses (e.g., effects of the different endpoints) are very different, for the α-exhaustive method we should use different α_1, α_2, and α_3 such that their trend is consistent with the trend of parameters in the alternative hypotheses.

Citation: Chang M, Deng X, Balser J and Bliss R. Progressive Alpha-Exhaustive Multiple Testing Procedure with Independent Test Statistics. SM J Biometrics Biostat. 2016; 1(1): 1003.

Table 7: Stopping Boundary 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 between Hommel's and α-exhaustive Procedures.

	$\delta_1/\ \delta_2(\delta_3=0.3,\ \sigma=1)$							
Method	0.0/0.0	0.0/0.3	.03/0.3	0.1/0.3	0.2/0.3	0.1/0.2	0.1/0.1	0.3/0.3
Hommel	0.482	0.735	0.737	0.750	0.794	0.612	0.533	0.869
α-Ex	0.470	0.756	0.775	0.821	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.

K-Hypothesis Testing Procedure

We will 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 rejection rules for H_1 can be specified as

$$p_1 \le \alpha \cap \left(p_1 p_2\right)^{1/2} \le \alpha_{12} \cap \left(p_1 p_3\right)^{1/2} \le \alpha_{13} \cap \ldots \cap \left(p_1 p_2 \ldots p_K\right)^{1/K} \le \alpha_{12\ldots K}\ ,(20)$$

reject H_1; otherwise accept H_1.

The rejection rules for H_2, H_3,…. H_K can be constructed in similar way.

When K increases, the number of terms increases quickly, but any of the terms that is a product of more than 5 p-values will be automatically satisfied based simulations. The main problem is not the computational time, but the loss of efficacy for a large K. Therefore, the advantage of the α-exhaustive methods is not clear when $K > 6$. In a clinical trial, however, often the most important (clinically meaningful and commercial viable) hypotheses are usually limited between 1 and 6.

Summary and Discussion

To construct a multiple testing procedure, we need to consider at least three things to ensure sufficient power: (1) α-exhaustive, (2) synergize strengths among data for local hypothesis or marginal p-values, and (3) 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-exhaustion, we use the marginal p-value product corresponding to each null hypothesis configuration and enforce it with an upper bound in the rejection rules. Using such p-value product terms in the rejection rules also ensures the synergy between the marginal p-values. The K-hypothesis testing algorithm can be applied to the test statistics with correlations, however, due to its complexity and larger scale applications in clinical trials (dose-finding, subgroup analysis, adaptive design), the details of such an expansion will be considered in future research.

The proposed progressive a-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, it uses the totality of the evidence to make a conclusion. The test procedure is simple and performs well in broad situations. When the true "standardized" effect size, the true value of the parameter, is very different for different hypothesis, the choice of the set of critical values should be consistent with the trend of alternative $\bar{H}_k\left(k = 1, 2, \ldots, K\right)$ to boost the power as shown in Table 6.

Appendices

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 simultaneously;
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;
```

References

1. Chang M. Modern Issues and Methods in Biostatistics. Springer, New Yourk, New York. 2011.
2. Wright SP. Adjusted *P*-values for simultaneous inference. Biometrics. 1992; 48: 1005-1013.
3. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika. 1988; 75: 800-802.
4. Dmitrienko A, Tamhane AC, Bretz F. Multiple Testing Problems in Pharmaceutical Statistics. Chapman and Hall/CRC. FL. 2010.
5. Holm S. A simple sequentially rejective multiple test procedure. Scand J Statist. 1979; 6: 65-70.
6. Hommel G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika. 1988; 75: 383-386.
7. Wiens BL. A fixed sequence Bonferroni procedure for testing multiple endpoints. Pharmaceutical Statistics 2003; 2: 211-215.
8. Westfall PH, Tobias RD, Rom D, Wolfinger RD, Hochberg Y. Multiple comparisons and multiple tests using SAS system. SAS Institute. SAS Campus Drive, Cary, North Carolina, USA. 1999.
9. Dmitrienko A, Wiens B, Westfall P. Fallback tests in dose-response clinical trials. J Biopharm Stat. 2006; 16: 745-755.
10. Hommel G, Bretz F. Aesthetics and power considerations in multiple testing-a contradiction? Biom J. 2008; 50: 657-666.
11. Dmitrienko A. Multiple Testing Procedures in Clinical Trials. IBS workshop, Berlin. 2013.