

The Commonly Used Statistical Methods to Control the Probability of an Overall Type I Error and the Application in SAS in Clinical Trial

Lixiang Yao, ICON, Beijing, China

Abstract

In a clinical study, we often set up multiple hypotheses with regard to the cost of getting study result. However, the multiplicity problem arises immediately when they are performed in a univariate manner. Some methods to control the rate of the overall type I error are applied widely, and they are discussed in this paper. Except the methodology, we will introduce its application in one study case and provide the SAS code using PROC MULTITEST.

Background

During statistical analysis, we conclude some statistical significant results when p-Value is less than 0.05. As a matter of fact, the conclusions are not accurate if they are based on un-adjusted p-Value because here it can increase the type I error probability. Table 1 are the unadjusted Alpha values when the compared time is from 1 to 10, among these numbers, we know if the compared time is 10, the type I error is increased to 0.4, eight times of pre-defined Alpha value.

Table 1: unadjusted Alpha value when the compared time is from 1 to 10 (original Alpha=0.05)

Compared times(m)	Formula	Unadjusted Alpha value
1	$1-0.95^{**1}$	0.05
2	$1-0.95^{**2}$	0.0975
3	$1-0.95^{**3}$	0.1426
4	$1-0.95^{**4}$	0.1855
5	$1-0.95^{**5}$	0.2262
6	$1-0.95^{**6}$	0.2649
7	$1-0.95^{**7}$	0.3017
8	$1-0.95^{**8}$	0.3366
9	$1-0.95^{**9}$	0.3698
10	$1-0.95^{**10}$	0.4013

Moreover, regulatory agencies mandate a strict control of the overall type I error rate in clinical trials, but give some clear guidelines that are not required to control overall type I error rate for the following cases:

1: there is only one primary efficacy endpoint and other endpoints serve as supportive purpose, and there is no interim analysis in this study.

2: there are more than 1 primary endpoints, however to prove clinically relevant treatment benefits, all endpoints should show statistical significance, take a study to treat hypertension for example, the study can be proved successfully if the data for change from baseline in blood pressure and LDL-C shows statistical significance, in this case, Alpha level stay the same.

3: two or more primary variables are ranked according to clinical relevance. However, no confirmatory claims can be based on variables that have a rank lower than or equal to that variable whose null hypothesis was the first that could not be rejected. And the hierarchical order for testing null hypotheses, however, has to be pre-specified in the study protocol.

4: As is pointed out in E9 guideline, the set of subjects whose data are to be included in the main analyses should be defined in the protocol, from these sets of subjects one(usually FAS) is selected for the primary analysis.

5: For a disease, where a commonly acknowledged reference drug therapy exists, it is often recommended to demonstrate the efficacy and safety of a new substance in a three arm study with three treatments: the reference drug, placebo and investigational drug. Usually the aims of such a study are manifold:

- 1) To demonstrate superiority of the investigational drug over placebo
- 2) To demonstrate superiority of the reference drug over placebo
- 3) To demonstrate non-inferiority of the investigational drug over placebo.

All three comparisons must show statistical significance at the required level, and no formal adjustment is necessary.

Except the above five cases, any multiple treatment comparisons, multiple primary endpoints, multiple secondary (subgroup) endpoints and interim analysis, if any of them have been involved in the study, multiple issue should be taken into consideration, pre-defined effective method should be documented in the protocol to control the overall type I error rate within 0.05 to meet agency's requirement. In the following sessions, we will introduce some commonly used methods in clinical to control the overall type I error.

Statistical methodology and case study

1: **Bonferroni method:** Bonferroni method can strictly control type I error rate within 0.05. the Alpha adjusted by Bonferroni method is calculated by $0.05/m$ (m is the compared time), it rejects null hypothesis when the p-value is less than the adjusted Alpha.

Table 2: the Alpha value adjusted by Bonferroni method (original Alpha=0.05)

The compared time (m)	formula	Alpha value adjusted by Bonferroni method
1	0.05/1	0.05
2	0.05/2	0.025
3	0.05/3	0.0167
4	0.05/4	0.0125
5	0.05/5	0.01
6	0.05/6	0.0083
7	0.05/7	0.0071
8	0.05/8	0.0063
9	0.05/9	0.0056

10	0.05/10	0.005
----	---------	-------

Case study:

In a phase III study compared A drug to B drug on hypertension patients, according to study design, three hypotheses are selected in the study, and Bonferroni method will be used for controlling the overall type I error rate:

1): according to the statistical method in protocol, we got the following p values for each hypothesis:

Hypothesis	p-Value
H(1): L = P	0.0421
H(2): L = M	0.0152
H(3): M = P	0.0243

2): the adjusted Alpha by Bonferroni method is 0.0167, and it will be compared with p values to determine the results:

Hypothesis	p-Value	Algorithm
H(1): L = P	0.0421	P value 0.0421>0.0167, retain H(1)
H(2): L = M	0.0152	P value 0.0152<0.0167, reject H(2)
H(3): M = P	0.0243	P value 0.0243>0.0167, retain H(3)

3): conclusion:

Using Bonferroni method to control overall type I error rate, it rejects H(1), H(3) and retain H(2)。

2: **Šidák method:** the methodology is similar to Bonferroni method. the Alpha adjusted by Šidák method is calculated by $1-(1-\alpha)^{1/m}$ (m is the compared time), it rejects null hypothesis when the p-value is less than the adjusted Alpha.

Table 3: the Alpha value adjusted by Šidák method (original Alpha=0.05)

The compared time (m)	formula	Alpha value adjusted by Šidák method
1	$1-(1-0.05)^{1/1}$	0.05
2	$1-(1-0.05)^{1/2}$	0.0253
3	$1-(1-0.05)^{1/3}$	0.0170
4	$1-(1-0.05)^{1/4}$	0.0127
5	$1-(1-0.05)^{1/5}$	0.0102

6	$1-(1-0.05)^{**1/6}$	0.0085
7	$1-(1-0.05)^{**1/7}$	0.0073
8	$1-(1-0.05)^{**1/8}$	0.0064
9	$1-(1-0.05)^{**1/9}$	0.0057
10	$1-(1-0.05)^{**1/10}$	0.0051

Case study:

Using the case we mentioned in the previous section, Šidák method will be used for controlling the overall type I error rate:

1): according to the statistical method in protocol, we got the following p values for each hypothesis:

Hypothesis	p-Value
H(1): L = P	0.0421
H(2): L = M	0.0152
H(3): M = P	0.0243

2): the adjusted Alpha by Šidák method is 0.0170, and it will be compared with p values to determine the results:

Hypothesis	p-Value	Algorithm
H(1): L = P	0.0421	P value $0.0421 > 0.0170$, retain H(1)
H(2): L = M	0.0152	P value $0.0152 < 0.0170$, reject H(2)
H(3): M = P	0.0243	P value $0.0243 > 0.0170$, retain H(3)

3): conclusion:

Using Šidák method to control overall type I error rate, it rejects H(1), H(3) and retain H(2).

3: **Holm stepwise test:** the Holm multiple test is based on a sequentially rejective algorithm that tests the ordered hypotheses $H(1), \dots, H(m)$ corresponding to the ordered p-values $p(1), \dots, p(m)$. The multiple testing procedure begins with the hypothesis associated with the most significant p-value. This hypothesis is rejected if $p(1) < \alpha/m$, further, $H(i)$ is rejected at the i th step if $p(i) < \alpha/(m-i+1)$ for all $i=1, \dots, m$, otherwise, $H(1), \dots, H(m)$ are retained and the algorithm terminates, the details of the Holm test are shown table 4:

Table 4: Holm stepwise test algorithm

Steps	condition	Met condition	Not met condition
-------	-----------	---------------	-------------------

1	$p(1) < 0.05/m$	Reject $H(1)$, go to step 2	Retain $H(1)$, stop
2	$p(2) < 0.05/(m-1)$	Reject $H(2)$, go to step 3	Retain $H(2)$, stop
.....
i	$p(i) < 0.05/(m-i+1)$	Reject $H(i)$, go to step i+1	Retain $H(i)$, stop
.....
m	$p(m) < 0.05$	Reject $H(m)$	Retain $H(m)$

Case study:

Using the case we mentioned in the previous section, Holm stepwise test will be used for controlling the overall type I error rate:

1): according to the statistical method in protocol, we got the following p values for each hypothesis:

Hypothesis	p-Value
$H(1): L = P$	0.0421
$H(2): L = M$	0.0152
$H(3): M = P$	0.0243

2): rank the p values from minimum to maximum:

Hypothesis	p-Value
$H(2): L = M$	0.0152
$H(3): M = P$	0.0243
$H(1): L = P$	0.0421

3): according to the rule of Holm stepwise test, the following decisions were made:

Hypothesis	p-Value	Algorithm
$H(2): L = M$	0.0152	P value $0.0152 < 0.05/3$, reject $H(2)$, go to the next step
$H(3): M = P$	0.0243	P value $0.0243 < 0.05/2$, reject $H(3)$, go to the next step
$H(1): L = P$	0.0421	P value $0.0421 < 0.05/1$, reject $H(1)$.

4): conclusion:

Using Holm stepwise test to control overall type I error rate, it rejects $H(1)$, $H(2)$ and $H(3)$.

4: **Hochberg stepwise test:** Contrast with Holm test, this test examines the ordered p -values $p(1), \dots, p(m)$ starting with the largest one and thus falls into the class of *step-up tests*. the details of the Hochberg test are shown table 5:

Table 5: Hochberg stepwise test algorithm

Steps	condition	Met condition	Not met condition
1	$p(1) < 0.05$	Reject $H(1)$, go to step 2	Retain $H(1)$, stop
2	$p(2) < 0.05/2$	Reject $H(2)$, go to step 3	Retain $H(2)$, stop
.....
i	$p(i) < 0.05/(m-i+1)$	Reject $H(i)$, go to step $i+1$	Retain $H(i)$, stop
.....
m	$p(m) < 0.05/m$	Reject $H(m)$	Retain $H(m)$

Case study:

Using the case we mentioned in the previous section, Hochberg stepwise test will be used for controlling the overall type I error rate:

1): according to the statistical method in protocol, we got the following p values for each hypothesis:

Hypothesis	p-Value
$H(1): L = P$	0.0421
$H(2): L = M$	0.0152
$H(3): M = P$	0.0243

2): rank the p values from maximum to minimum:

Hypothesis	p-Value
$H(1): L = P$	0.0421
$H(3): M = P$	0.0243
$H(2): L = M$	0.0152

3): according to the rule of Hochberg stepwise test, the following decisions were made:

Hypothesis	p-Value	Algorithm
$H(1): L = P$	0.0421	P value $0.0421 < 0.05/1$, reject $H(1)$. go to the next step
$H(3): M = P$	0.0243	P value $0.0243 < 0.05/2$, reject $H(3)$,

		go to the next step
H(2): L = M	0.0152	P value $0.0152 < 0.05/3$, reject H(2),

4): conclusion:

Using Hochberg stepwise test to control overall type I error rate, it rejects H(1), H(2) and H(3)。

SAS codes:

Be noted that PROC MULTITEST only adjust p-value not Alpha level, and the examples provided above adjust Alpha level, the methodology is still the same whatever adjust p-value or alpha.

```
data antihyp2;
  input test $ raw_p @@;
  cards;
  l 0.0421
  m 0.0243
  p 0.0152
  ;
run;

proc multtest pdata=antihyp2 bonferroni sidak holm hochberg out=adjp;
run;
```

Summary:

Bonferroni method, Šidák method, Holm stepwise test, Hochberg stepwise test are frequently chosen in clinical trial to deal with multiple issue. Bonferroni method is very conservative, and it can strictly control type I error within 0.05, but it is not recommended to choose if the compared time is greater than 3, and Hochberg has more power than the other three methods.

Reference

- [1] Alex Dmitrienko, Geert Molenberghs, Christy Chuang-Stein, Wakter Ofen, et al. Analysis of Clinical Trials Using SAS, 2002, 67–129.

Contact Information

Lixiang Yao
 ICON Clinical Research
 Beijing China
 +86 10 5825 6060
 A123s123f123@tom.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.