December 2020

I. Overview

The problem of multiplicity is common in clinical trials. It refers to a complete

In the research, it is necessary to go through more than one statistical inference (multiple test) to determine the

On issues related to decision making. For example, multiple endpoints (such as the primary endpoint and off

Key secondary endpoints), comparisons between multiple groups, multi-stage overall decision-making (such as

Interim analysis for the purpose of strategy), multiple time point analysis of longitudinal data, sub-groups

Analysis, analysis of different parameter combinations or different data sets of the same model, sensitivity anal

Analyze and so on. For confirmatory clinical trials, control the total type I error rate (FWER)

At a reasonable level is the basic criterion of statistics. Some of the above multiplicity problems can be

Cause FWER to expand, some will not. For the former, appropriate

Strategies and methods to control FWER at a reasonable level, this process is called multiplicity

Adjustment; for the latter, there is no need for multiplicity adjustment. Therefore, in the formulation of clinical

When testing plans and statistical analysis plans, adopt appropriate strategies and methods to control

FWER is very important.

This guideline focuses on common multiplicity issues and corresponding decisions

Strategies, introducing commonly used multiplicity adjustment methods and multiplicity analysis methods, aim

To provide guidance on how to control FWER in clinical trials of confirmatory drugs, so

The general principles discussed are also applicable to other types of clinical research.

2. Type I error, total type I error rate and type II error in multiple inspection

(1) Type I error and total type I error rate

1

Page 4

Type I error means that the null hypothesis (or null hypothesis) is correct but the test result is rejected

Eliminating the error of the original hypothesis is equivalent to statistically inferring drugs that are actually ine

The error of drawing a valid conclusion. Its probability needs to be controlled at a certain level, which is called

Is the test level, or significance level, expressed by α ; for a certain

The test level of a hypothesis test is called the nominal test level, also called partial test

Level, expressed by α_i .

The total type I error rate refers to multiple hypothesis tests that are concerned in the same clinical trial.

In the test, the probability that at least one true null hypothesis is rejected. No matter how many hypothesis che

In the test, which null hypothesis or hypotheses are true, the FWER can be controlled at the α level,

Is called the strong control FWER; under the condition that all the null hypotheses are true, the FWER

Control at the αλπηα level is called weak control FWER. Weak control FWER can only get

The overall conclusion does not support the conclusion of a single hypothesis test, so it is confirming

The application in sexual clinical trials is of little significance. The “control

"FWER" refers to strong control FWER.

(2) Type II error

Type II error means that the null hypothesis is incorrect, but the test result fails to reject the original false

The mistake of assumption is equivalent to statistically inferring the actually effective drug to be invalid

The probability of the error of the conclusion is expressed by β , and correspondingly $1-\beta$ *is* called the test po

For confirmatory clinical trials, provided that Type I errors are effectively controlled,

The risk of Type II errors also requires attention. For multiple tests that need to be adjusted, by

Controlling FWER reduces the α_i of a single hypothesis test in multiple tests , correspondingly

It also reduces the inspection efficiency. Therefore, when it comes to multiplicity adjustments, develop researcl

2

Page 5

The plan should consider controlling the impact of FWER on the effectiveness of the inspection, for example b

Increase the sample size to ensure sufficient test efficiency.

Three, common multiplicity problems

Common multiplicity problems in clinical trials are generally reflected in multiple endpoints, multiple

Comparison between groups, subgroup analysis, interim analysis, longitudinal data at different time points

Analysis and other aspects.

(1) Multiple endpoints

1. Primary endpoint

The primary endpoint refers to the main concern of the clinical trial (main purpose)

Directly relevant and capable of providing the most clinically meaningful and convincing evidence

Points, often used in main analysis, sample size estimation and evaluation of whether the test has reached the n

Purpose. In confirmatory clinical trials, a single primary endpoint is more common, but some
In this case, multiple primary endpoints will be involved. For the study of multiple primary endpoints, through

There are often two types of research hypotheses, namely, multiple primary endpoints are required to be signifi

At least one of the endpoints is significant.

(1) Multiple primary endpoints are required to be significant. That is, all primary endpoints are required t

The study drug is considered effective when it is significant (this situation is often referred to as the common p

point). For example, in a confirmatory clinical trial for the treatment of chronic obstructive pulmonary disease

Set two separate primary efficacy endpoints, forced expiratory volume in the first second and patient report

Report the symptom score, and the decision-making stipulates that the two primary endpoints are significant be

Material is effective. In this case, FWER will not inflate, because this strategy

There is no opportunity to choose one or several primary endpoints that are most beneficial to the study drug,

3

There is only one possibility to conclude that the drug is effective (that is, both null hypotheses are rejected

Absolutely). However, this will increase Type II errors and reduce inspection efficiency. Test efficiency

The degree of reduction is correlated with the number of primary endpoints and the primary endpoint

The greater the number and the weaker the correlation, the greater the reduction in test efficiency.

(2) At least one of the multiple primary endpoints is required to be significant. I.e. at least one

The study drug is considered effective when the primary endpoint is significant. For example, a certain corrobc

The clinical trial aims to verify a drug for treating burn wounds, setting two separate

The primary endpoints: wound closure rate and scar formation, the clinical trial protocol stipulates only

If one of the endpoints is significant, or both endpoints are significant, the drug can be considered

Overall clinically effective. In this case, the FWER will swell, because the medicine

The conclusion that the material is effective includes the following three possible combinations: ①The wound

The scar formation is not significant; ②The wound closure rate is not significant but the scar formation is sign

③The wound closure rate and scar formation are both significant. Due to at least

There is a significant end-point combination that is not the same, whether it will lead to FWER expansion

Depends on the specific research hypothesis.

2. Secondary endpoint

There are usually multiple secondary endpoints in clinical trials, and in most cases they provide Support for the primary endpoint. But in some cases, some secondary endpoints can be Can be used to support the claimed benefits of the drug insert, generally referred to as the critical secondary en point. At this time, the key secondary endpoints and the primary endpoint should be included in the FWER con system. Only after the hypothesis test of the primary endpoint is considered significant as a whole, the key is c Hypothesis testing of secondary endpoints.

4

3. Composite Endpoint

The composite endpoint refers to the combination of multiple clinically relevant outcomes into a single cl Such as the composite endpoint of cardiovascular events, as long as myocardial infarction, cardiac Any of these events, such as congestive failure, sudden coronary death, etc., will be regarded as the end-point e Health; or combine the scores of several symptoms and signs into one through a certain method A single variable, such as the ACR20 scale for evaluating rheumatoid arthritis. If a certain As a single primary endpoint, a composite endpoint will not involve multiple issues. but, If a certain component of the composite endpoint (such as an event or component quantity One dimension of the table) is used to support the claimed benefits of the drug insert, which should be defined Located at the main or key secondary end point, and then based on the above positioning to the main involved Or the multiplicity of secondary endpoints should be considered.

4. Exploratory endpoint

The exploratory end point can be pre-set or non-pre-set (e.g. (Such as data-driven) end points, generally including the expected frequency of occurrence is very low and dif Clinically important events that show the effect of treatment, or are considered improbable for other reasons The end point that can show effect but is included in the exploratory hypothesis, the results may help Design future new clinical trials. Such endpoints do not involve multiplicity issues.

5. Safety Endpoint

If the safety endpoint (event) is part of the confirmatory strategy, use To support the claimed benefits of the package insert, the multiplicity should be determined in advance and cor

problem. It should be noted that in the practice of clinical trials, due to safety incidents
There is a lot of uncertainty, and sometimes it is difficult to specify the main safety assumptions in advance. Th

Confirmatory strategies for multiple safety endpoints (usually serious adverse reactions)
The strategy may adjust the strategy based on the multiplicity after the fact. At this time, it should be fully expl
Be rational and reach consensus with regulatory agencies.

(2) Comparison among multiple groups

Comparisons between multiple groups in clinical studies are quite common, such as three-arm design, dos
-Reaction relationship research, evaluation of combination drugs and compound drugs, etc.

1. Three-arm design

The three-arm design is mostly used for non-inferiority trials, and the three groups arranged are trials
Group, positive control group and placebo group. At this time, the research hypothesis should consider three ty
Situation: ①The superiority of the test group compared with the placebo group; ②The positive control group a
Superiority compared with placebo group; ③Non-inferiority between test group and positive control group
Effectiveness. For the above multiplicity problem, only if the three hypothesis tests are significant
Believe that the experimental drug is effective, or based on a weak research hypothesis that only
To satisfy ①, the test drug can be considered as effective (need to be approved by the regulatory agency before
Can be implemented), or use a fixed sequence method, such as the hypothesis test sequence is ① → ② →
③ At this time, it will not cause FWER to expand. If the other three-arm design does not follow
Follow the above multiple testing strategy and do not satisfy that all hypothesis tests are significant,
It is necessary to consider whether it will cause FWER to expand according to the situation.

2. Dose-response relationship

Dose-response relationship studies are important for finding safe and effective therapeutic doses or agents
The amount range is critical. The method and purpose of dose exploration are exploratory testing and confirma
It is different in the syndrome test.

In exploratory trials, when a dose-response relationship is used for dose exploratory research,
It is up to the sponsor to decide whether to control FWER. Confirmatory clinical trial
In order to select and confirm that the test drug is recommended for use in a specific patient population
One or more dose levels must control FWER.

3. Combination drugs and compound drugs

Combination medication refers to the simultaneous use of two or more drugs in the treatment medication,
Prescription refers to a combination of two or more drugs. Combination medication
Or the purpose of clinical trials of compound drugs is mainly to verify the benefits-risks of combination drugs
Is it better than the single drug, or the benefit of the compound drug-whether the risk is better than its compone
Medicine.

Taking the combination of two single drugs as an example, the trial design will set at least three
Group, namely the combination group, single-drug A group and single-drug group B, the latter two groups are
Photo group. If another placebo group is added, it is a 2 · 2 factorial design.
Whether it's a three-group design or a four-group factorial design, its hypothesis test is used to infer
Whether the combination drug group is better than the other groups, it will not cause FWER expansion
Inflation, because the joint treatment can only be proved if all hypothesis tests are significant.
The efficacy of treatment.

(3) Analysis of longitudinal data at different time points

Longitudinal data, that is, repeated measurement data based on time points, are clinical trials
Common data types. There are two types of analysis related to this type of data and time point.
One is to compare between groups at different time points; the other is to compare within the treatment group
Effects at different time points.

Take a study design with only one primary endpoint and only two treatment groups
For example, if the primary endpoint evaluation is defined as one of multiple time points

Time point (such as the last visit point) for comparison between treatment groups, other time
Inter-group comparisons between points are regarded as secondary endpoint evaluations, which do not involve
If the primary endpoint evaluation is defined as the treatment group at more than one time point
The comparison between the groups, if the comparison between the groups at all relevant time points is signific
If it is effective, it will not cause FWER to expand, otherwise it will cause expansion.

For the case of comparing effects at different time points within the treatment group, if the purpose is
Through the comparison between time points to confirm the effect of the best time point, that is, when the time
Should be part of the confirmatory strategy, multiplicity issues need to be considered, whether
No need to consider.

For more than one primary endpoint or more than two treatment groups and involving longitudinal
The multiplicity of the research design that analyzes the data at different time points is more complicated.
Need to consider comprehensively.

If you want to avoid the multiplicity of longitudinal data, a possible solution
The solution is to convert the effects at different time points into the area under the broken line, such as treatme
The pain VAS score at different time points can be converted into the area under the broken line to replace
Table of the total pain score after treatment, that is, convert multiple variables into one variable, but
Correspondingly, after this conversion, the comparison between groups at each time point cannot be realistic.
Shit. Another possible solution is to use a single model for repeated measurement data
Analysis, such as repeated measures analysis of variance or mixed effects models.

(4) Subgroup analysis

Subgroup analysis is usually used to explain that the test drug is in a target subgroup of people
The curative effect in each subgroup, or the consistency of the curative effect among the subgroups. If the targe
The analysis used to support the claimed benefits of the drug insert requires a comprehensive consideration of t
The multiplicity of populations and subgroups, and at the same time, we must also pay attention to ensuring tha
The quantity has sufficient inspection power. Conversely, if subgroup analysis is not used to support drugs
For the benefits claimed in the instructions, there is no need to consider the issue of multiplicity.

(5) Interim analysis

Interim analysis for monitoring effectiveness, because the research process requires
Many decisions need to be made, and the multiplicity issues are complex and diverse, so control the FWER dis
It is especially important. When formulating a clinical trial protocol, it should be carefully considered and set u
Determine appropriate control strategies and methods for FWER.

(6) Complex design

For basket design, umbrella design, platform design for the purpose of confirmation
And other complex designs covering multiple disease fields, multiple drugs, and cross-research, due to the sam
Many sub-topic studies are carried out at the time, which may involve multiple issues. But because of this
Some subtopic studies are mostly independent studies and answer specific clinical questions, if applicable
Diseases, target populations, etc., generally do not cause FWER inflation.

When there is a large overlap in the target population of a complex design sub-topic study, or
Will it cause FWER when using the same control group for multiple sub-topic studies?
Expansion should depend on specific circumstances. At this time, it is recommended that the sponsor and the r
Communicate adequately.

4. Common strategies and methods of multiplicity adjustment

9

Page 12

In response to the multiple issues that may cause FWER swelling in clinical trials,
The strategies and methods used for multiplicity adjustment depend on the purpose, design, and
Research hypotheses and their testing methods. The sponsor needs to select
Make the necessary assessments of the strategies and methods of major adjustments, and in the clinical trial pr
And detailed in the statistical analysis plan.

The strategies and methods of multiplicity adjustment can be divided into decision-making strategies, adj
The analysis method is considered at three levels.

(1) Decision-making strategies for multiple problems

The research conclusions of clinical trials are mainly based on the analysis of all trial data
The result of inference is a process from local decision-making to overall decision-making. many
Decision-making strategies for serious problems can be divided into parallel strategies and sequential strategies

In addition to the process from departmental decision-making to overall decision-making, there is also a phase
The research purpose and test plan sort out possible multiplicity problems, and a certain

A strategy or a combination of multiple strategies, and then determine according to the selected strategy or stra

Each test the hypothesis corresponding to a statistical analysis and nominal significance level *[alpha]* $_I$ of

Allocation strategy (if required).

1. Parallel strategy

Parallel strategy means that the various hypothesis tests included are independent of each other and progr

Line, has nothing to do with the order of testing, just like a parallel relationship, each hypothesis test

The inference result does not depend on the inference result of other hypothesis tests.

2. Sequential Strategy

Sequential strategy refers to testing the null hypothesis in a certain order until it is satisfied

Stop testing related conditions, like a series relationship, according to the set conditions,

The result of the previous hypothesis test will determine whether to perform the subsequent hypothesis test. se

The order of hypothesis testing in the strategy and the corresponding multiplicity adjustment method are differ

The same effect on the overall conclusion is also different, which should be paid special attention to in the desi

3. Phased overall decision-making strategy

The phased overall decision-making strategy means that the overall decision-making is determined in adv

The sequence is carried out in stages, and its typical representative is the mid-term analysis for the purpose of e

Analysis. Make an overall decision at each stage to determine whether the test is valid or invalid

To terminate early or continue. The overall decision-making at each stage can use multiple questions

Parallel strategy or sequential strategy in question decision strategy. Multi-stage decision-making requires mult

Adjustability, that is, a certain amount of α is consumed in each stage, and the nominal test water of each stage

The quasi α_i can be the same or different, depending on the adopted α consumption strategy.

(2) Multiplicity adjustment method

The multiplicity adjustment method is essentially to adjust each of the overall decision-making

The nominal test level of independent hypothesis testing is α_i in order to control FWER at α water.

Flat purpose. The method of determining the nominal inspection level α_i can be based on the multiplicity prob

The choice of decision-making strategies.

1. Multiplicity adjustment method of parallel strategy

(1) Bonferroni method. The basic idea of the Bonferroni method is that each individual

The sum of the nominal test level α_i of hypothesis testing is equal to α , namely

$$\alpha_1 + \alpha_2 + \ldots + \alpha_i \ldots + \alpha_m = \alpha$$

The nominal inspection level α_i can be the same ($\alpha_i = \alpha / m$) or different, the latter

It is often used when the importance of each hypothesis test is different. For example, a clinical trial

The test has 3 primary endpoints, and 3 hypothesis tests are required, and $\alpha = 0.05$.

If the importance of the 3 primary endpoints is the same, then the $\alpha_{i\,for}$ each hypothesis test is the same,

Both are 0.0167 (=0.05/3), then the *P* value of each hypothesis test is less than 0.0167

Is considered significant; if the importance of the three primary endpoints is different, such as

Set α_1 , α_2 and α_3 to 0.030, 0.015 and 0.005 respectively, then each hypothesis check

The *P* value of the experiment is less than the corresponding $\alpha_{i\,to}$ be considered significant.

(2) Forward-looking *alpha* distribution method. Prospective *Alpha* Allocation Method (PAAS) and

The Bonferroni method has similar ideas and can be understood as the nominal test of each hypothesis test

The product of the reciprocal of the level α_i is equal to the reciprocal of α , namely

$$(1-\alpha_1)(1-\alpha_2)\ldots(1-\alpha_i)\ldots(1-\alpha_m) = (1-\alpha)$$

Each α_i can be the same or different, if the same, it can be obtained according to the Šidák method

$$\alpha_i = 1-(1-\alpha)^{1/m}$$

For example, a clinical trial with 3 endpoints, two of which are assigned points

Equipped with the value of α_i , $\alpha_1 = 0.02$, $\alpha_2 = 0.025$, if α is set to 0.05, according to the above formula

There is $0.98 \times 0.975 \times (1-\alpha_3) = 0.95$, the α_3 of the third endpoint is found to be 0.0057.

If the $\alpha_{i\,of\,the}$ 3 null hypotheses are distributed with equal weights, then the α_i is obtained based on the Šidák me

0.01695. It should be noted that the PAAS method is independent or positive in meeting multiple tests.

FWER can be controlled only when it is off.

2. Multiplicity adjustment method of sequential strategy

(1) Holm method. Holm method is a test based on Bonferroni method

The multiple adjustment method in which the statistics gradually decrease (the *P* value gradually increases). Th

12

Page 15

After calculating the *P* value of each hypothesis test, sort the *P* values from small to large,

Denoted as $P_1 < P_2 < \ldots < P_m$, the corresponding null hypothesis is H_{01} , H_{02} ,… H_{0m} ,

Then compare with the corresponding α_i according to the order of *P* value from small to large , according to

The second test H_{0i} , $1 \le i \le m$. The first step starts with the smallest *P* value and tests the null hypothesis

H_{01} , if $P_1 > \alpha_1 (= \alpha / m)$, do not reject the null hypothesis H_{01} , and stop testing

All remaining hypotheses; if $P_1 \le \alpha_1$, reject H_{01} , H_{A1} holds, enter

The next step is hypothesis testing. The second hypothesis test $\alpha_2 = \alpha /(m -1)$, the hypothesis

Compare the tested *P* value with α_2 , if $P_2 > \alpha_2$, stop testing the remaining hypotheses;

Otherwise, H_{A2} is established, and proceed to the next hypothesis test. More generally, when testing

For the *i*- th null hypothesis H_{0i} , if $P_i > \alpha_i (= \alpha /(m - i +1))$, stop

Check and accept H_{0i} ,..., H_{0m} ; otherwise, reject H_{0i} (accept H_{Ai}), and

Proceed to the next hypothesis test; and so on.

(2) Hochberg method. The Hochberg method is based on the Simes method

The multiple adjustment method in which the test statistics gradually increase (the *P* value gradually decreases

The method first calculates the *P* value of each hypothesis test , and ranks the *P* values from large to small

Order, denoted as $P_1 > P_2 > \ldots > P_m$, and then follow the order of *P* value from large to small with

Compare the corresponding α_i . The first step starts with the largest *P* value and checks the original false

Suppose H_{01} , if $P_1 \le \alpha_1 (= \alpha)$, reject all null hypotheses and stop testing,

All alternative hypotheses H_{Ai} are established; otherwise, H_{01} is not rejected , and the next hypothesis is entered

test. The second hypothesis test of *[alpha]* $_2$ = *[alpha]* / 2, the hypothesis test *P* value *[alpha]* $_2$

Comparison, if $P_2 \le \alpha /2$, stop testing the remaining hypotheses, except for H_{A1} , the rest

All of the alternative hypotheses are valid; otherwise, H_{02} is not rejected , and the next hypothesis check is ente

13

Test. More generally, when testing the *i*- th null hypothesis H_{0i} , if $P_i \leq \alpha_i$ (= α /i),

The remaining test is stopped, reject H_{01} , ..., H_{0m} ; if $P_I >$ [αλπηα] $_I$, not

Reject H_{0i} and proceed to the next hypothesis test; and so on. requires attention,

The Hochberg method can only achieve control when the multiple tests are independent or positively correlatec

System FWER.

(3) Fixed sequence method. The fixed sequence method refers to the pre-defined sequence

Perform hypothesis testing, the nominal test level of each hypothesis test α_i is the same as α , only

When the previous hypothesis test rejects the null hypothesis, proceed to the next hypothesis test.

Until a certain hypothesis test does not reject the null hypothesis, and the final conclusion

All previous significance conclusions are accepted for this hypothesis test. For example, in order

The three null hypotheses are H_{01} , H_{02} and H_{03} respectively . If the first and second hypotheses are checked

All tests rejected the null hypothesis at the α level, but the third hypothesis test failed to reject the original hypc

Assuming H_{03} , the alternative hypotheses H_{A1} and H_{A2} are both valid, but H_{A3} is not valid.

(4) Back-off method. The fallback method needs to check each hypothesis in advance according to the fix

And determine the nominal test level α_i for each hypothesis test , and then follow

Perform hypothesis testing in order. The method first checks H_{01} at the level of α_1 , if it is not rejected

H_{01} , then test H_{02} at the level of α_2 ; if reject H_{01} , then at the level of $\alpha_1 + \alpha_2$

Check H_{02} , and so on. For example, one item has 2 primary endpoints (O_1 and O_2)

Of clinical trials, using the fallback method, corresponding to the nominal inspection level of O_1 and O_2

Instead of α_1 =0.04 and α_2 =0.01, if the *P* value of the hypothesis test is P_1 =0.062,

P_2 =0.005, the final decision-making conclusion is that the experimental drug has a significant benefit on O_2

(P_1 =0.062> α_1 , P_2 =0.005< α_2); if the *P* values of the hypothesis test are respectively

If P_1 =0.032, P_2 =0.015, the final decision-making conclusion is that the test drug is in O_1

And O_2 on both benefit significantly ($P_{.1} = 0.032 <$ *[alpha]* $_{.1}$, $P_2 = 0.015 <$ *[alpha]* $_{.1} +$ *[alpha]* $_2$).

3. Interim analysis of common *alpha* segmentation methods

Interim analysis The more classic α segmentation methods include Pocock method, O'Brien-Fleming method and Haybittle-Peto method. One of the three segmentation methods

The premise is that the calendar time or cumulative data ratio of each interim analysis is the same, only

It is that the allocation of α_i for each hypothesis test has a different focus. More flexible *alpha* split

The law is the α consumption function, such as Lan-DeMets α consumption function, the method is

The extension of the classic method described above is more flexible in setting the time point of interim analys

For example, a confirmatory clinical trial evaluating an anti-tumor drug with immune target inhibitors,

The main evaluation index is all-cause death. An interim analysis is planned, which can be based on

The effectiveness of the early termination of the trial. Considering that the onset time of immune target inhibitc

Is delayed, so it is planned to observe 75% at a relatively late point in the study

In the event of death, an interim analysis is carried out. Use approximate O'Brien Fleming edges

The Lan-DeMets α consumption function of the world, and the two-sided FWER is required to be controlled at

The two-sided nominal inspection levels for the interim analysis and final analysis are 0.019 and 0.044.

When the multiplicity of clinical trials is more complicated, multiple combinations can be used

The multiplicity of strategy adjustment methods. It should be noted that the multiple multiplicity adjustments

A simple combination of methods may not be able to control FWER. Therefore, in complex situations

When using multiple multiplicity adjustment methods in combination, in order to ensure that FWER can be cor

Consider adopting the gatekeeping method or the graphic method.

15

Page 18

(3) Multiplicity analysis method

For the multiplicity problems that need to be solved, most of them are based on specific statistical analysi

The analysis method is combined with the multiplicity adjustment method to achieve. For example, for differer

Types of multiple endpoints (such as quantitative, qualitative, survival time), comparisons between groups will

To different statistical analysis methods (such as analysis of covariance, Mantel-Haenszel χ^2

Test, Kaplan-Meier test), at the same time, it also depends on multiple endpoints

Multiplicity adjustment method (such as Bonferroni method, etc.) to determine each hypothesis test
The inspection level α_i, and then the decision-making conclusion can be made.

For a single endpoint variable and multiple group comparisons in the same research stage, some statistics

The calculation analysis method is to solve the problem of multiple comparisons on the basis of overall hypoth

The basic idea is that the standard error involved in the pairwise comparison is the overall hypothesis test

The standard error. For example, the pairwise comparison of quantitative outcome variables based on analysis

LSD method, SNK method, etc., Dunnett method, etc. are compared between multiple groups and the reference

Multiple comparisons of qualitative outcome variables can be transformed by variable transformation (such as

Become a quantitative variable, and then use the analysis method of the above quantitative variable; survival ti

Inter-outcome variables are based on the log rank test of Kaplan-Meier method (Mantel-

Cox method), Breslow method (extended Wilcoxon method), etc. It should be noted that

Some methods may not be able to control FWER. For the basis of the overall hypothesis test

The statistical analysis method that cannot achieve multiple comparisons on the above requires the use of local

Test (pairwise comparison) combined with the method of αλπηα allocation (such as Bonferroni method, etc.).

Multivariate parameter methods (such as multivariate analysis of variance) are to solve the problem of mu

One of the methods of the problem, especially for the case of multiple endpoints, but this type of method is one

16

Page 19

It is required to meet the multivariate normal distribution, and the second is that the interpretation of the analys

Limits its application.

Repeated sampling (such as bootstrap method and permutation method) is also a solution

One of the means of multiplicity problem, the advantage of this kind of method is to control the FWER

At the same time, it can guarantee a high inspection efficiency; its disadvantage is that it is based on

The empirical distribution is difficult to verify, which leads to insufficient accuracy of the estimation. In additio

Rely on large samples. Therefore, this type of method is rarely practiced in clinical trials, and caution is require

For re-use, it is recommended to fully communicate with the regulatory agency in advance.

As there are many statistical analysis methods to solve the multiplicity problem, each method has

It has its advantages and disadvantages. The sponsor needs to plan for clinical trial or statistical analysis.

The plan stipulates in advance the statistical analysis methods used for multiplicity problems.

Five, other considerations

(1) Conditions that do not require multiplicity adjustment

Circumstances that do not require multiplicity adjustment include but are not limited to the following situ
Including interim analysis of validity):

1. Multi-group comparisons for a single primary endpoint (such as the target of a non-inferiority trial
Quasi-three-arm design), when all hypothesis tests are significant, it is considered valid;

2. For a single primary endpoint, the research hypothesis is that the efficacy of the test drug is at least
Not inferior to the positive control drug, when hypothesis testing is performed in a fixed order, that is, the first
The first step is to verify the hypothesis that the efficacy of the test drug is not inferior to the positive control d
Assuming that H_0 is rejected, the second step is to verify that the efficacy of the test drug is better than that of
Hypothesis of taking medicine;

3. For multiple primary endpoints, if and only if the hypothesis tests for all endpoints are equal
When it is significant, it is considered effective;

4. For multiple secondary endpoints, none of them will be used in the drug label
When said to benefit;

5. For the complexity of cross-research such as basket design, umbrella design, platform design, etc.
Design, if the sub-topic study is an independent study and answers the respective clinical questions,
If applicable diseases, target population, etc.;

6. In the process of statistical analysis, for the same primary endpoint, it may be
Different analysis data sets are used for analysis, as long as which analysis data is defined in advance
Set as the main conclusion basis;

7. Use different statistical models or the same model with different parameter settings
As long as the main analysis model is defined in advance;

8. Carry out sensitivity analysis based on different assumptions, such as using different shortcomings
The analysis after the missing data estimation method is filled, the outliers are treated with different
Analysis etc.

(2) The parameter estimation problem of multiple testing

The corresponding confidence interval should be estimated according to the multiplicity adjustment meth

There are many major adjustment methods, some of which are relatively simple but relatively conservative and

Line interval estimation, for example, the Bonferroni method is used to adjust the confidence interval; some

The method is more complicated, and it may be difficult to make a corresponding interval estimation.

The multiplicity adjustment may also bring about the selection bias of the point estimate. E.g,

In confirmatory clinical trials with multiple dose groups, if the problem of multiplicity is

The decision-making strategy selected the most different from placebo in the drug label

The effect size of the dose group may overestimate the efficacy of the drug. Similar selectivity

Bias can also arise from the choice of subgroups. Therefore, it is necessary to assess the

The selection bias that may be brought about by the adjustment.

(3) Communication with regulatory agencies

Multiplicity should be clarified in advance in clinical trial protocol and statistical analysis plan

Strategies and methods for problem and multiplicity adjustment. For complex and multiplicity issues,

Do you need multiplicity adjustment and how to adjust, the existing strategies and methods may

Faced with challenges, sponsors are therefore encouraged to be proactive in the design phase of confirmatory c

Communicate with regulatory agencies. During the test, if you change the multiplicity of adjustments

Strategies and methods to make major adjustments to the clinical trial program, should be discussed with regul

Communicate in a timely manner.

6. References

[1] Qian Jun, Chen Pingyan. Multiple comparisons of multiple sample rates. China Health Statistics, 2008; 25 (2): 206-212.

[2] Alosh M, Bretz F, Huque M. Advanced multiplicity adjustment methods in clinical trials. Statistics in Medicine, 2014; 33 (4): 693-713.

[3] Bretz F, Tamhane AC, Pinheiro J, et al. Multiple Testing in Dose-Response Problem, Chapter 3 of Multiplicity Testing Problem in Pharmaceutical Statistics. CRC Press, 2010.

[4] Bretz F, Maurer W, Brannath W, et.al. A graphical approach to sequentially rejective multiple test procedures. Statistics in Medicine, 2009; 28 (4): 586-604.

[5] Chen J, Luo JF, Liu K, et al. On power and sample size computation for multiple testing procedures. Computational Statistics and Data Analysis, 2011; 55 (1): 110-122.

[6] Collignon O, Gartner C, Haidich AB, et al. Current statistical considerations and regulatory perspectives on the planning of confirmatory basket umbrella and platform trial. Clinical Pharmacology & Therapeutics, 2020; 107 (5): 1059-1067.

[7] Dmitrienko A, Tamhane AC, Bretz F, et al. Multiple Testing Methodology, Chapter 2 of Multiplicity Testing Problem in Pharmaceutical Statistics. CRC Press, 2010.

[8] Dmitrienko A, Tamhane AC, Bretz F, et al. Gatekeeping

Procedures in Clinical Trials, Chapter 5 of Multiplicity Testing Problem in Pharmaceutical Statistics. CRC Press, 2010.

[9] Dunnett CW. A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association, 1955; 50 (272): 1096-1121.

[10] European Medicines Agency. Guidance on Multiplicity Issues in Clinical Trials.

[11] Freidlin B, Korn EL, Gray R, et.al. Multi-arm clinical trials of new agents: some design considerations. Clinical Cancer Research, 2008; 14 (14): 4368-4371.

[12] Hochberg Y, Tamhane A. Multiplicity Comparison Procedure. New York: Wiley, 1987.

[13] Howard DR, Brown JM, Todd S, et.al. Recommendations on multiple testing adjustment in multi-arm trials with a shared control group. Statistical Methods in Medical Research, 2018; 27 (5): 1513-1530.

[14] Huque MF, Rohmel J. Multiplicity Problem in Clinical Trials, Chapter 1 of Multiplicity Testing Problem in Pharmaceutical Statistics. CRC Press, 2010.

[15] International Conference on Harmonization (ICH). E9 guideline “Statistical Principles for Clinical Trials”.

[16] International Conference on Harmonization (ICH). E8 guideline “General Considerations for Clinical Trials”.

[17] International Conference on Harmonization (ICH). E17 guideline “General Principles for Planning And Design Of

Multi-Regional Clinical Trials".

[18] Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika, 1983; 70 (3):659-663.

[19] O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics, 1979; 35 (3): 549-556.

[20] Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observations of each patient. I. Introduction and design. British Journal of cancer, 1976; 34 (6): 585-612.

[21] Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika, 1977; 64 (2): 191-199.

[22] Sen PK. Some remark on Simes-type multiple tests of significance. Journal of statistical Planning and Inference, 1999; 82 (1-2): 139-145.

[23] US Food and Drug Administration. Multiple Endpoints in Clinical Trials – Guidance for the Industry.

[24] Wang DL, Li YH, Wang X, et al. Overview of multiple testing methodology and recent development in clinical trials. Contemporary Clinical Trials, 2015; 45 (Pt A): 13-20.

Appendix 1: Glossary

I like Error (**Type I Error**): refers to the null hypothesis (also known as null hypothesis) correct

But the test result rejected the error of the null hypothesis, which is equivalent to removing the actually ineffec

An error in which a valid conclusion is reached by statistical inference. The probability needs to be controlled i

Level, this level is called inspection level, or significance level, expressed by α .

II class error (**Type II Error**): refers to the null hypothesis is not correct, but the test results did not

Being able to reject the error of the null hypothesis is equivalent to statistically inferring the actually effective c
The mistake of drawing an invalid conclusion.

***α* Spending Function** (***α* Spending Function**): When a clinical study is divided into several

During the overall decision-making phase (such as an interim analysis based on effectiveness), each

All phases will consume a certain amount of α . As the research progresses, the proportion of the research com

Such as 1/3, 1/2, 3/5, etc.) show a certain functional relationship with the cumulative type I error rate,

As shown below.

Multiplicity Issues (**Multiplicity Issues**): Refers to a complete clinical research

In the research, it is necessary to go through more than one statistical inference (multiple test) to the research c
Issues related to decision making.

Multiple adjustment (**Multiplicity Adjustment**): the use of appropriate strategies and

Methods The process of controlling the total type I error rate at a reasonable level.

The key secondary endpoint (**Key Secondary Endpoint**) **:** secondary endpoints in

Indicators used to support the claimed benefits of the drug insert.

Nominal Level (**Nominal Level**): For a certain hypothesis in multiple testing

The inspection level of the inspection is called the nominal inspection level, also called the local inspection lev

Expressed by α_i .

Total **I** Type I error rate (**Familywise Rate Error** , **FWER**): refers to the same

At least one true null hypothesis among multiple hypothesis tests focused on a clinical trial

The probability of being rejected. It should be controlled at a reasonable level.

Primary Endpoint (**Primary Endpoint**): refers to the main

The main problem (main purpose) is directly related, can provide the most clinical significance and

The endpoint of convincing evidence is often used in primary analysis, sample size estimation, and evaluation

Whether the price test achieves the main purpose.

Appendix 2: Chinese-English comparison table

Chinese	English
α distribution	α Allocation
Alpha consumption	α Spending
α consumption function	α Spending Function
Type I error	Type I Error
Type II error	Type II Error
Multiplicity	Multiplicity
Multiplicity adjustment	Multiplicity Adjustment
Multiplicity problem	Multiplicity Issue
Multiple endpoints	Multiple Endpoints
Sub-topic research	Substudies
Key secondary endpoint	Key Secondary Endpoint
Fallback method	Fallback Method
Dose-response relationship	Dose-response Relationship

Nominal inspection level	Nominal Level
Prospective α distribution method	Prospective Alpha Allocation Scheme, PAAS
Gatekeeping	Gatekeeping Procedure
Graphical method	Graphical Approach
Significance level	Significance Level
Total type I error rate	Familywise Error Rate, FWER