Research on the Clinical Trials for Group Sequential Design of *α* Consumption Function Based on Condition Test for Efficiency

Kong Yujia, An Hongqing, Bian Weiwei and Shi Fuyan*

School of public health, Weifang Medical College, Shandong, China

*Corresponding author: Shi Fuyan

Keywords: α Consumption Function, Clinical trials, Condition Test, Group Sequential Design

Abstract. It takes the group sequential design as a method during the period of having interim analysis, in the clinical trial period, it can have multiple statistical analysis, which also can make an early decision to stop testing, so as to save manpower, material and financial resources and so on. In this paper, it takes the definition of group sequential design as the starting point, with the help of the the interpretation of the effectiveness of group sequential testing theory, making analysis on the application of α consumption function as the method in clinical trials.

Introduction

Clinical trials refers to the prospective research that takes the human body as the object of study, studying the intervention effect as well as its value. It has become a standard method to appraise the effectiveness and safety of new drugs, biological products, medical devices and other medical operation measures such as: the prevention, diagnosis and treatment of diseases. However, in any clinical trials, the reasonable design method is the key to ensure the scientific test.

The Interpretation of Group Sequential Design

Group sequential design (GSD) means after the sample completing a certain proportion in every certain time period or after the cumulative completed trials finishing its interim analysis on all subjects, it can determine whether the experiment get valid or invalid conclusion, so as to finish the test in advance. Group sequential design method inherits the advantages of traditional sequential design, its advantages are mainly embodied in the following aspects: (1) When the tested drug is obvious better than that of the contrasted drug, group sequential design can be got the valid conclusion before the interim analysis, which can shorten the test cycles and save funds at the same time; (2) On the other hand, it can be more in line with the requirements of ethics. By shortening the test cycles, group sequential design can make the effective drugs go on public as soon as possible, which also can meet the needs of patients at the same time.

The Theoretical Basis for Group Sequential Test Based on the Effectiveness of Test Conditions

When an ongoing test goes on with the repeated interim analysis, it will meet a theoretical problem, namely, how to determine the probability of making the type of I errors during the period of the interim analysis (i.e. the level of significance). Armitage and other people have already pointed out that, if every time we use the same value α as the significant criterion, when the null hypothesis is true, the probability of obtaining at least one significant result in inspection at one time is higher than that of the given fixed value α. In other words, the repeated significant inspection can increase the probability of "overall significant level" or the probability of making type I error. For example, every time $\alpha=0.05$, while it goes on with 10 times consecutive significant test for the accumulation of the experimental data, at this time, the overall significant level is no longer *0.05*, instead, it is increased to *0.19*, which can be shown in table 1.

Table 1 When H0 is the amount of the repeated inspection time under the true condition, which is also the probability of refusing H_0

The times of the repeated test (m)	1	2	3	4	5	10	20	50	100	∞
The probability of refusing H_0	0.05	0.08	0.11	0.13	0.14	0.19	0.25	0.32	0.37	1.00

The Application of Method of α Consumption Function in Clinical Trials

In 1969, Armitage as well as other people proposed a method, namely, when the given fixed threshold value of the interim analysis is C_1,C_2, ,...,C_k, and the assumed statistics of the interim analysis is Z_1, Z_2,..., Z_k approximately obeyed the normal distribution, it can calculate the total size of type I error by using iteration method of numerical integral for the group sequential test. This kind of method can become the theoretical basis for α consumption function to calculate the threshold value of the interim analysis and the nominal inspection level.

Stallard and Facey extended the α consumption function method to the method with situation that the interim analysis can be considered as valid or invalid ahead of the end of the test. They suggest that group sequential test should set up two increasing functions respectively $\alpha_U(t)$ and $\alpha_L(t)$, which can respectively meet the conditions of $\begin{cases}\alpha_U(0)=0\\ \alpha_U(1)=\alpha\end{cases}$ and $\begin{cases}\alpha_L(0)=0\\ \alpha_L(1)=1-\alpha\end{cases}$, while the valid field *Uk* and the invalid field *1k* in the interim analysis of No.*k* can meet the followings:

$P(\text{refuse } H_0,\ t \leqslant t_k \mid H_0)=\alpha_U(t_k)$

$P(\text{accept } H_0,\ t \leqslant t_k \mid H_0)=\alpha_L(t_k)$

Having estimation.

The Proposed Method

As for the group sequential test during the period of *k* stage, according to the pre- specified α consumption function form $\alpha(t)$, the calculated threshold value of the first interim analysis is $c_1 = \Phi^{-1}[1-\alpha(t_1)]$, while the nominal inspection level at this time point t_1 is:

$\alpha_1 = \alpha(t_1)$

In other word,the nominal inspection level of the trial during the first midterm analysis is equal to the size of the error of I class, which is consumed at the time point t_1. Thus, we can make estimation for the size of type I error of the test at any time point *t* with the inspection of the first time interim analysis.

In order to make the overall significant level be maintained constant as α, we must adjust the significant level of each interim analysis. This kind of method is used for the analysis of the adjustment during the period of the interim analysis can be called the nominal significance level, which can be represented by α_m.

Table 2 Under the situation that the overall significant level is *0.05*, the comparison of the threshold value of three kinds of group sequential design

Interim analysis			Three kinds of group sequential design		
The number of times		Threshold value	Pocock method	O'Brien-Fleming method	Peto method
2	1	Z	2.177	2.796	3.289
		P	0.028	0.005	0.001
	2	Z	2.177	1.976	1.961
		P	0.028	0.047	0.050
3	1	Z	2.288	3.470	3.289
		P	0.021	0.0005	0.001
	2	Z	2.288	2.453	3.289
		P	0.021	0.013	0.001
	3	Z	2.288	2.003	1.963
		P	0.022	0.044	0.050

Assuming the efficacy of the drug reaction level for the observed patient is an independent random variable within normal distribution, we can go on with the continuous observation on the multiple statistics to compare these two kinds of drugs, so as to get the statistics $U_m(m=1,\ldots k$; m can represent the No.m inspection, k is the maximum time of the inspection), thus the corresponding significant name level is α_m, the corresponding value to α_m is the standard normal difference Z_m (which is also called threshold value), then the nominal significant level α_m can meet the following formula:

$P_r(|U_1| <Z_1)\ldots(|U_k| <Z_k)|H_0=1-\alpha$

There are two kinds of methods to determine the threshold value Z_m which can be corresponding to the value of α_m, depending on the difference of the bilateral inspection and unilateral inspection. As for the data of normal distribution, using the standard normal deviation Z_m as boundary. As for the data of obeying the two distribution (i.e., success and failure), it can use X^2_m as boundary.

Estimation on the Size of Sample

There are some relevant parameters that can determine the sample content, which can be including: the probability of type I error α, effectiveness of inspection $1-\beta$, the theoretical value difference between group $\delta=(\mu_A-\mu_B)$ or $\delta=(P_A-P_B)$ and the standard deviation σ. Here, μ_A、μ_B is the average value of the A medicine group and B medicine group for the medicine curative effect (P_A, P_B is the rate of being positive). During the period of making group sequential design, it should consider the expected analysis to do how times tests to discuss the possibility of monitoring so as to early stop the test in the whole test process. Usually the sooner the test is stopped, the more amount of the samples can be saved, the better the benefits are. There are few advantages both in the aspect of statistical superiority, ethics or practical aspects, when the interim analysis is more than five times. Both group A and group B can increase n observed patients during the period of two times consecutive interim analysis. 2_n patients are increased for the two groups, which means that the test did not meet the requirements in the last analysis, therefore, it needs to increase the number of the patients until the number of the treated patients increased to 2_n, it can go on with the next interim analysis. If the maximum number of the whole test can have k times analysis, then the total number of the patients that are included in the test is $(2_n)\ k$.

The followings will list out the maximum number of inspection k under the fixed nominal significant level and the changeable nominal significant level, as well as the amount of 2_n samples between the required two consecutive analysis (both A group and B group has n cases).

Table 3 Group sequential design, with the required increased number of the observed cases between two consecutive analysis, namely, 2_n* (two-sided test, $\alpha=0.05$)

The maximum number of analysis	Effectiveness of inspection $1-\beta$					
k	Fixed name	Level of significance		Changeable name	Level of significance	
	0.50	0.80	0.90	0.50	0.80	0.90
2	8.72	17.42	23.11	7.74	15.81	21.15
3	6.17	12.21	16.10	5.22	10.64	14.22
4	4.79	9.43	12.44	3.94	8.03	10.73
5	3.94	7.22	10.13	3.16	6.44	8.62

Conclusion

In clinical trials, the estimation of sample is an important and sensitive problem, which is also an unavoidable problem. If the amount of sample is too small, the inspection efficiency is too low, thus it can not get the correct conclusion, which may lead to false negative; if the amount of sample is too large, it will not only cause great waste of manpower, material and financial resources, but also can make result have false positive result. Therefore, how to determine an appropriate amount of

sample is one of the most important factors that can be related to the success or failure of the clinical experiment.

Acknowledgments

This paper was supported by the Fund (Funding information: K11QC1011)

Reference

[1] Bretz F, Schmidli H, König F, et al. 2006. Confirmatory seamless phase II/III clinical trials with hypothesis selection at interim: general concepts. Biometrical Journal, vol.48, pp623-634.

[2] Stallard N, Yodd S. 2011. Seamless phase II/III designs. Statistical Methods in Medical Research, vol.20, pp623-634.

[3] Maxwell SE, Kelley K, Rausch R. 2008. Sample size planning for statistical power and accuracy in parameter estimation. The Annual Review of Psychology. vol.59, pp537-563.

[4] Miettinen OS, Nurminen M. 1985. Comparative analysis of two rates. Statistics in Medicine, vol.4, pp213-226.

[5] Anderson KM, Clark JB. 2010. Fitting spending functions. Statistics in Medicine vol.29, pp321-327.