Estimation of sample size and testing power (Part 2)

Liang-ping Hu, Xiao-lei Bao, Li-xin Tao, Shi-guo Zhou, Yue Guan
Consulting Center of Biomedical Statistics, Academy of Military Medical Sciences, Beijing 100850, China

Abstract: This article introduces definitions of three special tests, namely, non-inferiority test (to verify that the efficacy of the experimental drug is clinically not inferior to that of the positive control drug), equivalence test (to verify that the efficacy of the experimental drug is equivalent to that of the control drug), and superiority test (to verify that the efficacy of the experimental drug is superior to that of the control drug), and methods of sample size estimation under the three different conditions. By specific examples, the article introduces formulas of sample size estimation for the three special tests, and their SAS realization in detail.

Keywords: statistics, medical; research design; sample size; testing power; parametric estimation

For the design of one factor with two levels, the corresponding $t$ test of quantitative data or Wilcoxon two-sample test is frequently used for hypothesis testing. In fact, the hypothesis testing mentioned above belongs to common difference test. However, in clinical trials of new drugs or medical applications, there are three special tests, namely, non-inferiority test, equivalence test and superiority test. This article introduces methods of sample size estimation of the three special tests.

1 Sample size estimation of equivalence test for univariate quantitative data with the design of one factor with two levels

Equivalence test refers to the test of which the objective is to test whether two drugs, doses or applications have clinically equivalent curative effect. The equivalence is reflected through two critical values ($\delta_L$ and $\delta_U$) which should be stipulated during the research design phase.

1.1 Formula For the equivalence test, suppose that the two-sided testing power is set to be $\alpha$ (the equivalence test includes two single-sided tests, and the testing power of each single-sided test is $\alpha/2$), and the largest allowed probability of making type II error is $\beta$, the relationship between the sample size of each group and the testing power of quantitative data is as follows:

$$
power = \Phi \left[ \frac{\delta_U - (\mu_T - \mu_B)}{\sqrt{(\sigma_T^2 + \sigma_B^2)/n_T} + u_{1-\alpha/2}} \right] - \Phi \left[ \frac{\delta_L - (\mu_T - \mu_B)}{\sqrt{(\sigma_T^2 + \sigma_B^2)/n_T} + u_{1-\alpha/2}} \right]
$$

DOI: 10.3736/jcim20111105
http://www.jcimjournal.com
Received September 12, 2011; accepted September 15, 2011; published online November 15, 2011.
Correspondence: Prof. Liang-ping Hu; Tel: 010-66931130; E-mail: lphu812@sina.com

Journal of Chinese Integrative Medicine (JCIM) or Zhong Xi Yi Jie He Xue Bao is an international, peer-reviewed, open access journal for the study of complementary and alternative medicine or integrative medicine from all regions of the world. JCIM is indexed in PubMed and Directory of Open Access Journals (DOAJ). JCIM is a member journal of CrossRef. Articles published in JCIM have maximum exposure to the international scholarly community.
Submit your manuscript here:
http://mc03.manuscriptcentral.com/jcim-en (for manuscripts written in English)
http://mc03.manuscriptcentral.com/jcim-en (for manuscripts written in Chinese)
• No submission and page charges for manuscripts written in English
• Quick decision and rapid publication

Send your postal address by e-mail to jcim@163.com, we will send you a complimentary print issue upon receipt.
ISSN 1672-1977. Published by JCIM Press, Shanghai, China.
Formula 1 is a transcendental equation, which can be resolved by iteration in order to get \( n_T \), the sample size of the experimental group. \( \mu_T \) and \( \mu_R \) in formula 1 refer to the population mean of the experimental group and the population mean of the control group respectively, which can be replaced by their estimated values when unknown; \( \mu_T - \mu_R \) is denoted as \( \theta \) occasionally; \( \delta_L \) and \( \delta_U \) are the lower limit and the upper limit of the equivalence critical value, and \( \delta_L < 0 \), \( \delta_U > 0 \); \( u_{1-\alpha/2} \) is the deviation value of a standard normal distribution; \( n_T \) refers to the sample size of the experimental group; \( k \) refers to the sample size ratio of the control group and the experimental group, namely, \( n_R/n_T = k \) or \( n_R = k n_T \); \( \sigma^2_T \) and \( \sigma^2_R \) are the population variances of the experimental group and the control group respectively, which can be replaced by their estimated values if unknown.

1.2 Example

In order to observe the influence of losartan and ibraesartan on serum uric acid level in treating patients suffering high blood pressure with hyperuricemia, and to evaluate whether the two drugs have the same efficacy of lowering blood pressure, researchers expected to adopt a multicenter, double-blind, parallel-control and randomized design. The research subjects were equally divided into the losartan group and the ibraesartan group. After six weeks of treatment, the lowered values of blood pressure in the two groups were measured. Stipulate the equivalence critical value \( \delta_L = -5 \) mmHg and \( \delta_U = 5 \) mmHg based on clinical experience. The result of pilot test showed that the sample standard deviations of the two groups were both 6 mmHg and the difference of the lowered value between the two groups was 1.5 mmHg. Set \( \alpha = 0.05 \), \( \beta = 0.20 \), how many subjects were needed for the research?

Analysis: Based on the research objective, the research is known as an equivalence test of quantitative data with the design of one factor with two levels. The researchers wanted to estimate the sample size. The known information includes the sample size ratio of the experimental group and the control group, the sample standard deviations of the two groups, the equivalence critical value and the allowed probabilities of making type I error and type II error when performing hypothesis testing. The SAS program based on formula 1 to estimate the sample size of each group is as follows:

```
data a;
  alpha=0.05; beta=0.20; s_t=6; s_r=6; delta_L=-5; delta_U=5;
  theta=1.5; k=1;
  n_t=2;
p=probnorm((delta_U-theta)/sqrt((s_t^2+s_r^2)/n_t))-probit(1-alpha/2))-
  probnorm((delta_L-theta)/sqrt((s_t^2+s_r^2)/n_t))+probit(1-alpha/2));
do while(p<1-beta);
  n_t=n_t+1;
p=probnorm((delta_U-theta)/sqrt((s_t^2+s_r^2)/n_t))-probit(1-alpha/2))-
  probnorm((delta_L-theta)/sqrt((s_t^2+s_r^2)/n_t))+probit(1-alpha/2));
end;
  n_r=n_t*k;
file print;
put 2 the sample size of the experimental group is: n_t=;
put 3 the sample size of the control group is: n_r=;
run;
```

Output: the sample size of the experimental group is 47 and the sample size of the control group is 47.

Program explanation: The first eight statements in the beginning of the above program specify the main information, including the allowed probabilities of making type I error and type II error, the standard deviations of the experimental group and the control group, the lower and upper limits of the equivalence critical value, the estimated difference of the population means of the experimental group and the control group, and the sample size ratio of the two groups. The ninth statement “n_t = 2” means to assign an initial value to the sample size of the experimental group. The above program computes the sample size.
sizes of the experimental group and the control group based on formula 1. In similar situations of sample size estimation, readers only need to change the values of the first 8 options to get the result.

The above computation can be realized by the POWER procedure in SAS.

```
PROC POWER:
   twosamplemeans test=equiv_diff
   lower = -5 upper = 5 meandiff = 1.5
   stdev = 6 alpha=0.025
   groupweights = (1 1) */
   ntotal = . */
   npergroup = .
   power = 0.8;
run;
```

Output: the sample size needed for each group is 48.

Program explanation: The above program invokes the POWER procedure to realize the sample size estimation. The option “test=equiv_diff” means to perform the equivalence test to the two population means; the options “lower” and “upper” specify the lower limit and the upper limit of the equivalence critical value respectively; the option “meandiff” specifies the estimated difference of the two population means; “stdev” specifies the estimated standard deviation of the population; the option “alpha” specifies the equivalence testing level (“alpha=0.025” refers to half of the equivalence testing level, meaning that the testing level of each single-sided test is 0.025); the option “power” refers to the testing power $1-\beta$, which should not be lower than 80%; “npergroup” refers to the sample size of each group. Besides, the option “groupweights =” and “ntotal =” can not be used together in the same procedure. In similar situations, readers only need to modify the values of the above options to get the result.

2 Sample size estimation of superiority test for univariate quantitative data with the design of one factor with two levels

Superiority test refers to the test of which the research objective is to demonstrate that the efficacy of an experimental drug is better than the efficacy of a control drug (placebo or positive control drug). During the research design phase, researchers are required to stipulate the critical value $\delta_v$ based on specific situations and clinical experience. If the efficacy of the experimental drug is higher than the efficacy of the control drug and the difference is larger than the superiority critical value $\delta_v$, the clinical efficacy of the experimental drug is considered better than the control drug\endnote{1}.

2.1 Formula Superiority test belongs to one-sided test. Set the testing level $\alpha$ and the largest allowed probability of making type I error $\beta$ beforehand, the formula of sample size estimation for each group is as follows\endnote{2-6}:

$$n=\frac{(\alpha-\alpha)+(1-\alpha)^{\frac{1}{2}}\sigma}{(\delta_v-\delta)^2}$$ (2)

In formula 2, $n$ is the sample size of each group; $u_{1-\alpha}$ and $u_{1-\beta}$ are both one-sided deviation values of a standard normal distribution; $\sigma$ is the combined standard deviation of the two populations, which can be replaced by its estimated value if unknown; $\delta_v$ is the superiority critical value, and $\delta_v>0$; $\theta$ is the population mean deviation of the experimental group and the control group, which can also be replaced by its estimated value if unknown.

2.2 Example In order to evaluate the efficacy of Xiaooyujiangzhi Capsules in treatment of hyperlipidemia, researchers planned to conduct an experiment comparing the efficacy of Xiaooyujiangzhi Capsules and placebo by measuring the lowered value of cholesterol. If the lowered value of cholesterol by using Xiaooyujiangzhi Capsules was 0.28 mmol/L higher than that by using placebo, the efficacy of Xiaooyujiangzhi Capsules was considered better than the efficacy of placebo. According to the general requirement of efficacy and statistics, $\alpha$ was stipulated as 0.05, and $\beta$ was set to be 0.20. The standard deviation of each group was 1 mmol/L, and the efficacy difference of the two groups was 0.6 mmol/L. Suppose the sample size ratio of the experimental group and the control group is 1:1, how many patients were required in this research?

Analysis: According to the research objective, the research involves the sample size estimation of clinical superiority test for univariate quantitative data with the design of one factor with two levels. The known information includes the sample size ratio, the estimated lowered cholesterol value of the experimental group and the control group, the same standard deviation of the two groups, the superiority critical value and the allowed probabilities of making type I error and type II error. The research expected to estimate the minimum sample size of each group. Therefore, the corresponding SAS program is as follows:

```
data a:
   alpha=0.05; beta=0.20; s=1; delta_U=0.28; theta=0.6;
   n=CEIL(2*(probit(1-alpha)+probit(1-beta)) + 2
   *s**2/(delta_U-theta) + s**2); file print;
put # 2 @5 the sample size needed for each group is n.; run;
```

Output: the sample size needed for each group is 121.

Program explanation: The first five statements are used to specify the allowed probability of making type I error ($\alpha$), the allowed probability of making type II error ($\beta$), the combined standard deviation of the two groups, the superiority critical value and the estimated difference of the two population means. The program computes the estimated value of sample size based on formula
2. Readers only need to change the values of the first five options in similar situations for sample size estimation.

3 Sample size estimation of non-inferiority test for univariate quantitative data with the design of one factor with two levels

Non-inferiority test refers to the research of which the objective is to show that the clinical efficacy of an experimental drug is not inferior to that of a positive control drug. During the research design phase, it is required to stipulate the critical value $\delta_L$. If the efficacy of the experimental drug is worse than that of the control drug, however, the difference is less than the non-inferiority critical value $\delta_L$, the clinical efficacy of the experimental drug is considered non-inferior to the control drug\(^{[1]}\).

3.1 Formula Non-inferiority test also belongs to one-sided test. Stipulating that the testing significant level is $\alpha$, and the allowed probability of making type I error is $\beta$, the sample size estimation for each group of non-inferiority test is as follows\(^{[4\text{-}6]}\):

$$n = \frac{2(\mu_{1-\alpha} + \mu_{1-\beta})^2}{\sigma^2} / (\delta_L - \theta)^2$$  \hspace{1cm} (3)

In formula 3, $n$ refers to the sample size of each group; $\mu_{1-\alpha}$ and $\mu_{1-\beta}$ are both one-sided deviation values of a standard normal distribution; $\sigma$ is the combined standard deviation of the two populations, which can be replaced by its estimated value if unknown; $\delta_L$ is the non-inferiority critical value, and $\delta_L < 0$. $\theta$ is the population mean difference of the two groups, which can also be replaced by its estimated value if unknown.

3.2 Example In order to examine the efficacy of Jintong Capsule in treatment of tic disorders in children, researchers planned to conduct a randomized, double-blind experiment using tiapride as a positive control drug. The research subjects were divided equally into the Jintong Capsule group and the control group. The former took Jintong Capsule for 6 weeks, while the latter took tiapride for the same time period. If the lowered value of motor tic integral in the Jintong Capsule group was 1.77 point lower than that in the control group or more, it was considered that the Jintong Capsule was not worth promotion. In other words, the research aimed to examine whether the efficacy of Jintong Capsule in treatment of tic disorders in children is inferior to the efficacy of tiapride. According to general experience, the standard deviation of each group was 3.8, and the difference of the lowered value of tic disorders in children in the Jintong Capsule group and in the control group was $-1$. Set $\alpha$ as 0.05, and $\beta$ as 0.20, how many patients were needed in the experimental group and the control group respectively?

Analysis: According to the research objective, the research involves the issue of sample size estimation of clinical non-inferiority test for univariate quantitative data with the design of one factor with two levels. Since the known information includes the sample size ratio of the experimental group and the control group, the estimated difference of the lowered value of the motor tic integral in the two groups, the same standard deviation, the non-inferiority critical value and the allowed probabilities of making type I error and type II error, the SAS program needed to estimate the minimum sample size for each group is as follows:

```
data a;
alpha = 0.05; beta = 0.20; s = 3.8; delta_L = -1.77; theta = -1;
n = CEIL(2 * (probit(1 - alpha) + probit(1 - beta)) * 2 * s * s * 2 / (delta_L - theta) * 2);
file print;
put #2 @15 ‘the sample size needed for each group is’ n.;
runtime;
```

Output: the sample size needed for each group is 302.

Program explanation: The first five statements in the beginning of the program are used to respectively specify the values of the allowed probability of making type I error, the allowed probability of making type II error, the combined standard deviation of the two groups, the non-inferiority critical value and the estimated difference of the two population means. The program is written based on formula 3 for sample size estimation. Readers only need to modify the values of the first five options in the beginning of the program in similar situations for sample size estimation.

REFERENCES


样本量估计与检验效能分析（二）

胡善平，鲍晓蕾，陶丽新，周诗国，吴雪
军事医学科学院生物医学统计学咨询中心，北京 100850

摘要：本文介绍了3种特殊检验的基本概念和所需样本量估计方法，即非劣效性检验（主要目的为检验试验药的疗效在临床上不比阳性对照药差）、等效性检验（主要目的为检验两种药物的疗效在临床上是否等效）和优效性检验（主要目的为检验试验药的疗效在临床上优于对照药）。本文还通过具体的实例，详细介绍了进行前述3种特殊检验所需设计的试验研究之前所需要的样本量的计算公式和用SAS实现计算的方法。

关键词：统计学；医学；研究设计；样本大小；检验效能；参数估计

国产原研药消渴丸首次进入国家用药指南

为2010版《中国2型糖尿病防治指南》唯一中西结合药物

2011年10月16日，中华医学会在北京举办的“2010版《中国2型糖尿病防治指南》”新闻发布会上宣布，包括广药集团中一药业消渴丸在内的一批治疗糖尿病药物，经过专家审核甄选，入选国家用药指南。

据了解，这是国内原研民族药物首次进入糖尿病国家用药指南，打破了防治指南药物被化学生物制药一统天下的局面，也是出现在本指南中的唯一一个非化学名称的药物。其他药物都是以化学名称出现。在国内，生产糖尿病化学药物的厂家就有上千家。

2010年版《中国2型糖尿病防治指南》与2007版相比，有一个显著差异，就是在“2型糖尿病治疗程序图”中，将磺脲类药等胰岛素促泌剂作为一线备选药物及二线首选药物。这样的修订，既与国际医学界的主流趋势相符，也与欧美等国的糖尿病防治指南更具有“兼容性”。

新版指南中首次列入了国产原研的糖尿病治疗药物消渴丸。该药迄今已上市销售30年，久经临床验证，并通过了纳入国家863计划的循证医学研究，其显著疗效和良好安全性得到严格证实，并经过严格的临床试验和循证医学证实，是目前销量最大的具有自主知识产权的民族糖尿病药物。

据悉，今年6月，广药集团中一药业的消渴丸还作为中国唯一的民族药物在美国糖尿病学会年会（ADA会议）上进行专题学术报告，引起与会专家的强烈反响；今年7月，消渴丸与片仔癀、云南白药等一同被列为国家首批九大“传统名优中成药”。