Biomedical Statistics and Informatics
2017; 2(3): 103-106
http://www.sciencepublishinggroup.com/j/bsi
doi: 10.11648/j.bsi.20170203.14

Optimal Sample Size Determination for Medium or Large Clinical Study

Thomas Jyh-Ming Jiang

Department of Mathematical Sciences, National Chengchi University, Wen-Shan, Taipei, Taiwan

Email address:
jiangt@nccu.edu.tw

To cite this article:
Thomas Jyh-Ming Jiang. Optimal Sample Size Determination for Medium or Large Clinical Study. *Biomedical Statistics and Informatics.* Vol. 2, No. 3, 2017, pp. 103-106. doi: 10.11648/j.bsi.20170203.14

Received: February 26, 2017; **Accepted**: March 13, 2017; **Published**: March 29, 2017

Abstract: Clinical trials are often costly, and time consuming. The ability to get new products into the market early is critical to the success of pharmaceutical and medical device companies. Most practitioners use Fisher's exact tests to determine the required sample size for testing efficacy rates. We shall argue that when the sample size is not too small, normal approximation tests should be used instead of Fisher's exact tests. Several different sets of hypotheses and their corresponding formulas to compute sample size for clinical trial based upon normal approximation test are given.

Keywords: Fisher's Exact Test, Normal Approximation Test, Clinical Trial, Clinical Significance, Efficacy Rate

1. Introduction

The objective of most clinical studies is to evaluate and to compare the safety and effectiveness of an experimental treatment with either another treatment or no treatment. In principle, the sponsor must scientifically show that an experimental treatment is effective in order for the (US) FDA to approve its release. In other words, to gain approval of a new product, one major step is that a statistical test must show that it is unlikely (the chance is less than a constant α, e.g. 5%) to have the observed clinical data if the effectiveness of the new product is not clinically significant. The constant α is the probability of the FDA approving the new treatment when in fact it should not be approved. Although safety is also important, in practice the statistical significance of safety rarely needs to be shown in order to gain the approval of the product. Therefore, the sample size required for the clinical trial is usually based upon the need to demonstrate effectiveness.

Many, if not most, statisticians (practitioners) use "cook-books" to determine the required sample size. As a result, most clinical trials use Fisher's exact test (e.g. [1, 2]) to determine the sample size for testing the efficacy rates. The purpose of this paper is to show that, from the sponsor's point of view, the normal approximation test (e.g. [3, 4]) is a better test as long as the sample size is not too small. In the next section, we shall state the steps needed to show the effectiveness of new products. Some of these steps (testing hypotheses, test statistics, and sample size) are discussed with detail in sections 3, 4, and 5. In section 6, we make a concluding remark.

2. Steps to Show the Effectiveness of the New Products

In order to show whether the effectiveness of a new product is clinically significant, the first step is to set up the testing hypothesis based upon clinical knowledge. Two competing hypotheses, a null hypothesis and an alternative hypothesis, are required for each test. Usually, the alternative hypothesis is the statistical expression for the claim that the sponsor is making. For example, the alternative hypothesis can be that the new product is more effective than an existing one in the market. The null hypothesis is the complement of the alternative hypothesis. Continuing the previous example, the null hypothesis is that the new product is no more effective than the existing one in the market. Once the hypotheses are made, sample clinical data (i.e. a small part of the population clinical data) are used to make a decision to accept either the null hypothesis or the alternative hypothesis.

Since the decision is based upon only partial information (i.e. sample), it could be an incorrect decision (a risk). Typically, there are two types of risk. One is called Type I error with probability α. The second one is called Type II

error with probability β. Type I error is the error that the alternative hypothesis is accepted given that the null hypothesis is true. Type II error is the error that the null hypothesis is accepted given that the alternative hypothesis is true. Again continuing our previous example, α is the probability that the sample data will show significance (i.e. the new product is effective) given that the new product is not effective, and β is the probability that the sample data will not show significance (i.e. the new product is not effective) given that the new product is truly effective.

The second step is to decide which test statistic to be used. A test statistic can be thought of as a formula used by the statisticians or practitioners to see whether the null hypothesis should be rejected (i.e. statistical significance). Depending on the model and the testing hypotheses, there may be several test statistics available. If there are multiple test statistics available, one must be chosen. Typically, the test statistic with the most power (note that the power is 1-β) should be used, given a fixed α and sample size. Exceptions may exist for certain reasons, such as computational difficulty. Alternatively, if α and β are fixed, the test statistic with the smallest sample size should be used.

The third step is then to determine the sample size based upon the formula to be used for the statistical test as decided in the second step. The last few steps are to conduct the clinical trials, to perform statistical tests, and then to make statistical decisions and interpretations.

We shall not discuss the last few steps here. Instead, we will focus only on the first three steps.

3. Testing Hypotheses

In this section, we shall set up the testing hypotheses. For the clinical trials, it is very common to have the following set up:

$$\begin{aligned} H_0&: P_N = P_S \\ H_a&: P_N \neq P_S, \end{aligned} \tag{1}$$

where P_N and P_S are the efficacy rates for the new and standard treatment groups respectively. Of course, we may have the following one-side set ups:

$$\begin{aligned} H_0&: P_N \leq P_S \\ H_a&: P_N > P_S \end{aligned} \tag{2}$$

Or

$$\begin{aligned} H_0&: P_N \geq P_S \\ H_a&: P_N < P_S. \end{aligned} \tag{3}$$

However, it is also possible and reasonable to have the following testing hypotheses:

$$\begin{aligned} H_0&: P_N \leq P_S - d \\ H_a&: P_N > P_S - d, \end{aligned} \tag{4}$$

where $0 < d < 1$. For example, if a new treatment has the additional benefits of reducing the risk of complications and/or pains and the hospitalization time, then it is very likely that the physicians and/or patients would prefer the new treatment even if the new treatment has a lower efficacy rate. For the rest of the paper, we shall focus on the hypotheses (4), although we shall also discuss the hypotheses (1), (2), or (3).

4. Test Statistic

For testing the efficacy of two groups, Fisher's exact test has been the most popular test for clinical trials with small and intermediate size samples. It is likely that the word "exact" made this test popular. As a matter of fact, Fisher's exact test is not the "unconditional" exact test. It is just a conditional test, which is conditioned on the sample sizes of each of two groups and numbers of each of successes and failures of two groups. These conditions made the test extremely conservative (i.e. a test that rejects the null hypothesis too rarely (relative to α) given the null hypothesis is true) and inappropriate [5, 6, 7, 8, 9, 10]. If the sponsor is willing to have a few more patients with about the same power and the hypothesis set-up is (1), it would be fine to use Fisher's exact test. However, this is not the case for some studies. A more reasonable test is the Pearson's chi-square test, which is equivalent to the normal approximation test. The difference between Pearson's chi-square test (e.g. [11, 12, 13]) and the normal approximation test is that the former can only be used for two-sided tests (e.g. (1)) and the latter can be used for either one-sided (e.g. (4)) or two-sided tests.

Computing the Fisher's exact test under our hypotheses set-up (4) is quite complicated, and there is a lack of commercial solutions. On the other hand, the computation for normal approximation under either hypotheses set-up (1) or (4) is much easier.

5. Sample Size

Based upon the discussion above, it is better for sponsors to use the normal approximation test to compute the sample size for clinical trials that are not too small. It can be shown that the number of patients (one treatment group) required is given in equation (5) based upon the normal approximation test and test hypotheses (4). We have

$$n=\frac{1}{d^2}\left[z_\alpha \sqrt{(P_S-d)[1-(P_S-d)]+P_S(1-P_S)}+z_\beta \sqrt{2P_S(1-P_S)}\right]^2, \tag{5}$$

where P_S is the efficacy rate of the standard treatment group and *d* is the clinical significant value for the efficacy rate in the sense that physicians would be willing to use the new treatment even if its efficacy rate is reduced by *d* due to the other benefits of the new treatment. On the other hand, if we use testing hypotheses (1), the sample size required (for one treatment group) based upon the normal approximation test is

$$n=\frac{1}{d^2}[z_{\frac{\alpha}{2}}\sqrt{2P_S(1-P_S)}+\ z_\beta\sqrt{(P_S-d)[1-(P_S-d)]+P_S(1-P_S)}]^2. \quad (6)$$

Here d is again the clinical significance value for the difference of efficacy rates ($P_S - P_N$) and could be positive or negative. If the hypotheses test was a one-sided test (2) or (3), then the sample size required (for one treatment group) based upon the normal approximation test is

$$n=\frac{1}{d^2}[z_\alpha\sqrt{2P_S(1-P_S)}+\ z_\beta\sqrt{(P_S-d)[1-(P_S-d)]+P_S(1-P_S)}]^2. \quad (7)$$

Therefore, if the hypotheses set-up (4) or (1) is given, then the sample size required (for one treatment group) based upon the normal approximation test is formula (5) or (6), respectively. However, if either the hypotheses set-up (2) or (3) is given, then the sample size required (for one treatment group) based upon the normal approximation test is formula (7).

For the examples below, we consider $P_S = 0.8$, $\alpha = 0.05$, $\beta = 0.2$ and the clinical significance value $d = 0.2$.

Example 1. If we use the hypotheses set-up (2), then $z_\alpha = z_{.05} = 1.645$ and $z_\beta = z_{0.2} = 0.84$. Using formula (7), the sample size n can be shown to be 54 for one treatment group. Therefore, total sample size required for two treatment groups is 108. Note that the total sample size required (for two treatment groups) based on Fisher's exact test is 148. Hence, the required sample size is reduced drastically by using the normal approximation test.

Example 2. If we use the hypotheses set-up (1), then $z_{\frac{\alpha}{2}} = z_{\frac{.05}{2}} = 1.96$ and $z_\beta = z_{0.2} = 0.84$. Using formula (6), n can be shown to be 68 for one treatment group. Therefore, total sample size required for two treatment groups is 136.

Example 3. If we use the hypotheses set-up (4), then $z_\alpha = z_{.05} = 1.645$ and $z_\beta = z_{0.2} = 0.84$. Using formula (5), n can be shown to be 58 for one treatment group. Therefore, total sample size required for two treatment groups is 116.

Although it is preferable to have the same sample size for each of two treatment groups, it is possible to have different sample sizes. We'll generalize formula (6) for hypotheses set-up (1) when different sample sizes may be required for two different treatment groups.

Let n and m denote the required sample sizes for the new treatment group and the standard treatment group, respectively. In addition, let $m = c\,n$, where c is any given positive constant, then the sample size required for the new treatment group based upon the normal approximation test is

$$n = \frac{1}{d^2}\left[z_{\frac{\alpha}{2}}\sqrt{\left(1+\frac{1}{c}\right)P_S(1-P_S)}\ + z_\beta \sqrt{(P_S-d)[1-(P_S-d)]+\frac{1}{c}P_S(1-P_S)}\right]^2. \quad (8)$$

As can be seen and expected from formula (8), if we assume $c = 1$, then formula (8) is the same as formula (6). Hence, formula (6) is just a special case of formula (8).

Example 4. In Example 2, we consider the hypotheses set-up (1) and assume that m = n. Here, we assume that $c = 1/2$, that is m = n/2. Again, $z_{\frac{\alpha}{2}} = z_{\frac{.05}{2}} = 1.96$ and $z_\beta = z_{0.2} = 0.84$. Using formula (8), n can be shown to be 98 for the new treatment group. In addition, m can be shown to be 49. Therefore, total sample size required for two treatment groups is 147.

6. Conclusion

In this paper, we give four different hypotheses set-ups for possible clinical trials. Each of these set-ups is given a formula to compute the sample size required for clinical trial based upon normal approximation test.

The Fisher's exact test, which is very popular in the pharmaceutical industry for statistical hypotheses set-up (1), is extremely conservative (therefore, less powerful) and inappropriate. From the discussion above and examples in Section 5, we can see that the required sample size can be reduced drastically without changing the assumptions if we use the normal approximation test. This means that the sponsor may be able to save millions of dollars on the cost of clinical trials. More importantly, the sponsor may also be able to reduce the time to get the new product into the market.

Acknowledgements

This research was supported in part by the Ministry of Science and Technology, Taiwan. The author thanks a referee for his/her comments.

References

[1] Casella, G. and Berger, R. L., "Statistical Inference" 2nd ed. Pacific Grove, CA: Duxbury. 2002.

[2] Pocock, S. J., "Clinical Trials: A Practical Approach" New York, NY: John Wiley & Sons. 1983.

[3] Rohatgi, V. K. and Saleh, A. K. M. E., "An Introduction to Probability and Statistics" 3rd ed. New York, NY: John Wiley & Sons. 2015.

[4] Hogg, R. V. and Tanis, E. A., "Probability and Statistical Inference" Upper Saddle River, NJ: Pearson. 2010.

[5] D'Agostino, R. B., Chases, W., and Belanger, A., "The Appropriateness of Some Common Procedures for Testing the Equality of Two Independent Binomial Populations," The American Statistician, Vol. 42, No. 3, 198-202. 1988.

[6] Upton, U. J. G., "A Comparison of Alternative Tests for the 2x2 Comparative Trial, "Journal of the Royal Statistical Society, Ser. A, 145, 86-105. 1982.

[7] Berkson, J., "In Dispraise of the Exact Test," Journal of Statistical Planning and Inference, 2, 27-42. 1978.

[8] Conover, W. J., "Some Reasons for Not Using the Yates Continuing Correction on 2 x 2 Contingency Tables, " Journal of the American Statistical Association, 69, 374-376. 1974.

[9] Grizzle, J. E., "Continuity Correction in the X^2 Test for 2 x 2 Tables," The American Statistician, 21, 28-32. 1967.

[10] Kempthorne, O., "In Dispraise of the Exact Test: Reactions," Journal of Statistical Planning and Inference, 3, 199-213. 1979.

[11] DeGroot, M. H. and Schervish, M. J. "Probability and Statistics" 3rd ed. Boston, MA: Addison-Wesley, 2002.

[12] Fienberg, S. E., "The Analysis of Cross-Classified Categorical Data" 2nd ed. Cambridge, MA: MIT Press, 1980.

[13] Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. "Discrete Multivariate Analysis: Theory and Practice" Cambridge, MA: MIT Press, 1975.