Special Proceeding of the 21st Annual Conference of SSCA held at SV Agricultural College (ANGRAU), Tirupati, during January 29-31, 2019; pp 93-99

Sample Size Determination in Clinical Research – Perceptions and Practice

K.V.S. Sarma
Formerly Professor of Statistics
S.V. University, Tirupati – 517 502
Biostatistician & Research Coordinator
Sri Venkateswara Institute of Medical Sciences (SVIMS), Tirupati

Received: 24 June 2019; Revised: 18 July 2019; Accepted: 22 July 2019

Abstract

One of the basic requirements of any scientific study is to establish the truth or false of a belief with strong evidence. Both surveys and laboratory experiments aim at estimating the effect of a treatment or intervention basing on a sample that represents the target group. The sample size or the number of replications of the experiment shall be chosen carefully in order to balance risks of committing false positives or false negatives. This article is an appraisal of the need for statistically designed sampling plans, some specific procedures, software support and a remark on the practices adopted by users.

Key words: Sample size, Power analysis.

1 Approach for Scientific Research

Every scientific research starts with objectives like (a) inventing principles/products that are unknown earlier,(b) discovering facts that were masked by history,(c) verifying the truth of a belief (claim of the researcher),(d) developing alternative theories, principles and methods, and (e) predicting the occurrence of specific events like weather conditions.

Sometimes *exploratory research* is carried out as a passion to find novel things. In these studies we do not have a predefined hypothesis to be verified. In fact we go on observing a set of phenomena over a large number of subjects or few subjects over a long time. This is mainly to understand the *happenings* and latter generate *hypotheses* which will be verified by subsequent researchers.

Research in life sciences and medicine needs more accurate results / findings when compared to social science projects. The broad categories can be laboratory studies, plant and animal studies, clinical studies on human subjects, assessment of new procedures/interventions, community based epidemiological and public health studies etc.

Corresponding Author: K.V.S. Sarma
Email: smskvs@gmail.com

2 Some Pertinent Issues

In every study we need clarity on (a) the outcomes of interest, and (b) the factors affecting the outcomes. These two aspects indirectly speak about the purpose of the study and the approach. While writing the protocol the following aspects need clarity.

1. Is the study descriptive or comparative?

In a descriptive study we aim at estimating the characteristics of a population/cohort by observing a sample. In a comparative study, we estimate the parameters in two or more groups and also test the truth of a hypothesis on the group wise summaries (lie means or proportions)

2. What is the Target group?

The *target group* or populations is the collection of all subjects of interest on whom we wish to draw conclusions. Sample is called *study group*. Observations made on the sample will be applied for the target group.

3. Can we cover all the subjects of the target group?

Like the daily census of in-patients of a hospital. We need to report the mortality status of every in-patient and there is no sampling.

4. How long the study goes?

When the sample subjects are followed up for a long period say 5 years, there will be issues of environmental variables on the outcomes. It is called a longitudinal study. A cross sectional study on the other hand, covers all members of the sample once, and collects information.

5. Primary objective/outcome

There can be several objectives in a study but usually it is possible to identify few of them as primary and others as secondary. The statistical design may be focused on the primary objective. Similarly the primary outcome may a single one but other outcomes can also be studied with the same sample. The outcomes could be observations (like presence or absence of a skin rash) or measurements (like the mean systolic blood pressure).

Given the above dimensions the next issue is on finding out the sample size needed for the study.

3 Sample Size Determination

Finding the optimal sample size for the study is the key step in writing the protocol. A small sample may save resources but may not pick up a real effect when it exists. A sample that is fairly large is bounded by constraints on time, budget and other resources.

There is a belief among practitioners that a sample of size 30 or slightly above is enough in many clinical studies.

Similarly sample of 10% from the population is also used as a thumb rule. Sometimes all consecutive subjects that visit a hospital for 6 months will be taken as a sample. The reliability of the findings however, depend on the sample size, method of sampling, randomization /blinding, data recording and statistical analysis.

The statistical basis for determining the minimum required sample size emanates from the principles of statistical inference focused on achieving either a) the desired margin of error or ii) balancing the risks of committing type-1 and type-2 risks.

4 Sample Size for Descriptive Studies

In the case of descriptive studies, we wish to estimate a parameter μ (like population mean) which is unknown. However, with some prior information or by using historical data we may hypothesize that $\mu = \mu_0$. We draw an unbiased sample of n observations from the population and compute sample mean $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ of the characteristic x, which we claim to be the value of μ_0. Due to sample variations this $\bar{x}$ may not be exactly equal to μ_0, and the absolute difference $\delta = |\bar{x} - \mu_0|$, which is called the *margin of error* (or bias), is attributable to sampling.

For large samples under the influence of the central limit theorem, it is true that $Z = \frac{\delta\sqrt{n}}{s}$ follows standard normal distribution and this gives the relationship $n = \frac{Z^2 s^2}{\delta^2}$, where s denotes the population standard deviation (hypothetical) for the characteristic under question.

If we wish to limit the type-1 risk to like $\alpha = 0.05$, then we can get $Z_{1-\alpha/2} = 1.96$, which ensures a 95% confidence level for two sided comparison (positive or negative error). Suppose we wish to fix δ as some value like (5 grams of weight). Then the minimum required sample size is given by $n = \frac{Z_{1-\alpha/2}^2 s^2}{\delta^2}$. We can round up this to avoid fractional values. For instance, with $\alpha = 0.05$, $s = 20$ and $\delta = 5$, we get $n = 62$. In practice, 5 to 10% attrition is added to the n to make up for loss of samples/data.

It is important to note that the researcher has to provide information on s and δ failing which the formula cannot be applied. Usually a trial and error is required when the information on s is obtained from previous studies.

Similarly, in the case of estimating a proportion p (like prevalence of a disease) the formula for minimum sample size becomes $n = \frac{Z_{1-\alpha/2}^2 p(1-p)}{\delta^2}$. Here p and δ are the mandatory inputs.

Remark: If there are multiple parameters to be estimated we have to determine n for each parameter and pick up the largest of all.

5 Sample Size for Comparative Studies

The logic here is based on the comparison of the means or proportions of two independent or related target groups. In addition to the confidence level (α) we have to account for the power of comparison denoted by ($1-\beta$) and the corresponding Z values are used to determine n. In the case of two sample comparison of means, knowledge on the standard deviations is mandatory. We need to input the means and standard deviations in the respective groups. Again, when two groups are being compared for proportions, we need to input the two proportions p_1 and p_2 for the two groups.

Here are the steps of the procedure.

(a) Proportions in two independent groups

1) Study has two arms or two groups
2) The response is a percentage (incidence of pressure sores) p_1 and p_2 in the two groups
3) We need a two samples of size n_1 and n_2
4) Aim is to test the significance of (p_1-p_2). This is called the 'effect'.
5) With prior knowledge on p_1 and p_2, with $\alpha = 0.05$ we get the sample size n_1.
6) $n_1 = \frac{(Z_\alpha+Z_\beta)^2 p(1-p)}{(p_1-p_2)^2}$ where $p = \frac{(p_1+p_2)}{2}$
7) $n_2 = k*n_1$, where k is a constant. For equal samples take $k = 1$

(b) Means of Two independent groups

1) We need two samples of size n_1 and n_2.
2) We need to have hypothetical values of μ_1,μ_2, s_1 and s_2 (we may get them from previous studies).
3) We to find combined (pooled) standard deviation $s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2}}$ of the study variable from the two groups.
4) The sample size for one group is given by $n_1 = \frac{2(Z_\alpha+Z_\beta)^2}{\left(\frac{\delta}{s}\right)^2}$, $where\ \delta = (\mu_1 - \mu_2)$.
5) $n_2 = k*n_1$, where k is a constant is the size of the second sample. Unless otherwise required we can take equal sized samples by taking $k = 1$.

6 Cohen's Approach

Several issues are linked to the choice of the power of the sample which indirectly measures the ability to pick up the effect when it exists. Jacob Cohen (1977, 1992) observed that in reality several studies lead to about 50% of the power even though it is desirable to have around 95%. As a compromise Cohen suggested to use 80% as a convention. For this reason many software keep power as 0.80 as default!

In many practical situations it is difficult to obtain consistent values of mean or standard deviation from previous studies.

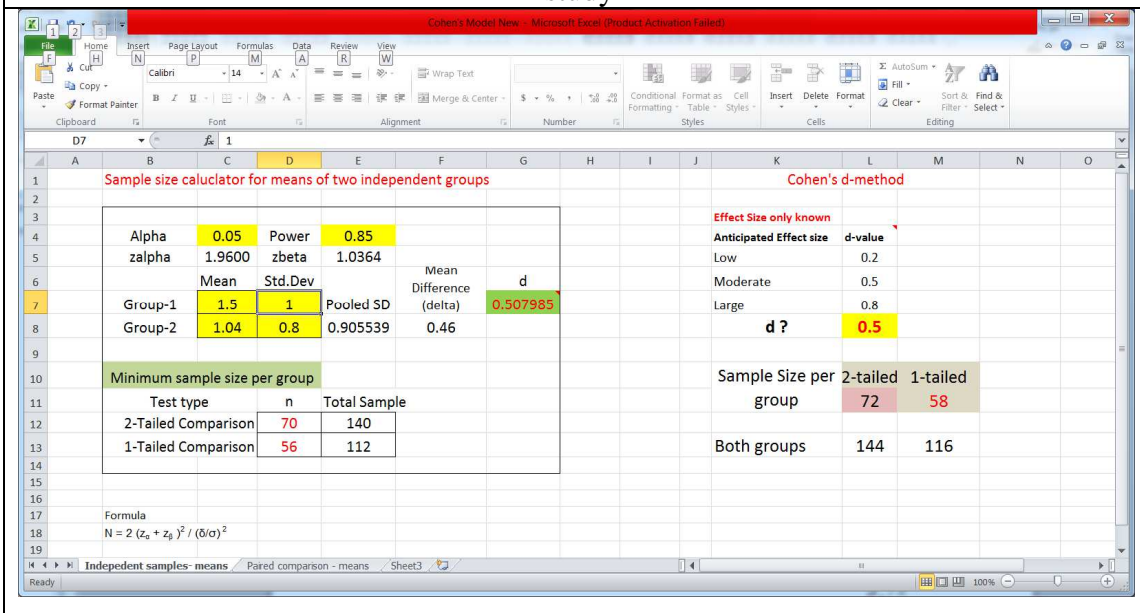
Jonathan Berkowit (http://www.columbia.edu/~mvp19/RMC/6_SampleSize.htm), summarized various issues related to sample size determination and referred to Cohen's method of *effect size*. It is the ratio of the difference between two means to the pooled standard deviation (s). Further, $d=\delta/s$ is the standardized effect known as Cohen's 'd'.

For practical implementation Cohen classified the effects as *low, moderate* and *large*. He suggested d = 0.2 for a low effect, d = 0.3 for a moderate and d = 0.5 for a large effect. Instead of specifying mean and standard deviation for the study variable in the two groups, researchers can express the effect low/moderate/large which helps as good approximation. Further, in order to get these mean, one has to refer to published articles and the issues of unmatched backgrounds will be a problem using the published findings.

If the researcher believes that the difference can be either positive or negative, we use a two-tailed comparison. If it is required to test the inferiority of one mean over the other we have to use one-sided comparison. In either case the sample size differs because the Z-values are different.

The value of Cohen's-d need not be limited to the above three values and admits leverage while other values like α, β and the number of tails can be fixed. Figure-1 shows a simple Excel template to work out the sample size by the two methods. We can change the values shown in the cells shown in Yellow colour.

Figure-1: Excel template to find the minimum sample size for means in a two-arm study



Web sites like www.danielsoper.com and www.sample-size.netprovide online calculators to determine the sample. Some software like MedCalc or EZR have modules for sample

size determination.EZR is a convenient package of medical researchers, based on the contemporary statistical programming using R-software. Apart from several tools for data analysis it has modules for sample size determination and power calculation. It has more than 15 different tools to determine the sample size or to find the power of the sample. However, the input windows for some tools need to be more user-friendly than what they are. For instance to find the sample size for comparing two independent means, we need to input the hypothesized values of a) two means and b) two standard deviations. We may however input the difference in the means. The window shall provide two input boxes for the standard deviations and there is only one box for entry. This is useful only when the two groups have (by hypothesis) same standard deviation. All this is difficult to the perceptions of a medical researcher! Similarly the same input window appears for paired comparison of means. Perhaps the right way of using EZR is under the directions of a statistician.

5 Sample Size – Three or More Means

When we have to compare the mean values among 3 or more groups, the principle works with ANOVA setup and the formulas are based on multiple comparison tests using 'family wise error rate'. Let τ be the number of comparisons to be made. For instance with 5 treatments in the study we get ${}^5C_2 = 10$ pairs for comparison but we may be interested in comparing 4 pairs and τ becomes 4. The formula for sample size for each pair of comparison (like A and B) is given by

$$n = 2\left(\sigma\left\{\frac{Z_{1-(\alpha/2\tau)} + Z_{1-\beta}}{\mu_A - \mu_B}\right\}\right)^2$$

where the parameters have their usual meaning. In the case of 4 comparisons, we have to find out the sample size for each of the 4 pairs and pick up the largest. One can work out at this formula at www.powerandsamplesize.com.Karadaand Altunay (2012) contains some interesting issues of sample size determination in the context of ANOVA.

Daniel (2014) is one useful reference that contains both theoretical reasoning and applications of sample size determination.

6 Other Calculators

For every context where estimation or comparison is made based on sample data, the issue of sample size arises. We have specific web based calculators to handle the various issues like the following.

1) Estimation/comparison of correlation coefficient(s)
2) Comparison of paired means/paired proportions
3) Comparison of proportions in community clinical trials. The formula makes use of Intra Class Correlation coefficient (ICC). One can we use online calculator at http://statulator.com/SampleSize/ss1P.html.
4) Sample size with desired sensitivity/specificity (epidemiological studies)

5) Sample size for multiple regression/logistic regression, etc.

Conclusion

Sample size determination is pre-requirement in the design of a study protocol. If the inputs are meaningfully given, we can work out the minimum sample size required for the study. When previous data is not available, it is suggested to conduct a pilot survey and arrive at the values of the parameters. At the end of the study it is necessary whether the sample had achieved the anticipated/desired power. Power analysis and effect size estimation is a part of *Meta-Analysis.* One should be careful to avoid unusual values for α,β margin of error etc., in order to justify a pre-conceived sample size without a statistical basis. For instance changing the margin of error from 5% to 10% drastically decreases the sample size but the reliability of the study is at question. A trained statistician plays a vital role at this stage.

References

Cohen, Jacob (1988). *Statistical Power Analysis for the Behavioral Sciences. Second Edition,*Lawrence Erlbaum Associates.

Cohen, Jacob (1992). A power primer. *Psychological Bulletin,* **112**(**1**), 155-159.

Jonathan Berkowitz. *Sample Size Estimation*. PERC Reviewer: Timothy Lynch, (http://www.columbia.edu/~mvp19/RMC/6_SampleSize.htm)

Daniel W. Wayne and Cross L. Chad (2014).*Biostatistics: Basic Concepts and Methodology for the Health Sciences*.10th Edition International Student Version, Wiley Series in Probability and Statistics.

Özge, Karada and Serpil, Aktas Altunay (2012). Optimal sample size determination for the ANOVA designs. *International Journal of Applied Mathematics and Statistics*,**25**(**1**), 127-134.