SPECIAL COMMUNICATION
SAMPLE SIZE DETERMINATION IN HEALTH RESEARCH

Waqas Sami, Mohammed Othman Alrukban*, Tayyaba Waqas**, Mohammad Rehan Asad***, Kamran Afzal***

Department of Public Health & Community Medicine, College of Medicine, Majmaah University-Saudi Arabia, *Department of Family & Community Medicine, College of Medicine, King Saud University-Saudi Arabia. ** Biostatistics and Research Consultancy Lahore-Pakistan. *** Department of Basic Medical Science, College of Medicine, Majmaah University-Saudi Arabia

One of frequently asked question by medical and dental students / researchers is how to determine the sample size. Sample size calculations is necessary for approval of research projects, clearance from ethical committees, approval of grant from funding bodies, publication requirement for journals and most important of all justify the authenticity of study results. Determining the sample size for a study is a crucial component. The goal is to include sufficient numbers of subjects so that statistically significant results can be detected. Using too few subjects' will result in wasted time, effort, money; animal lives etc. and may yield statistically inconclusive results. There are numerous situations in which sample size is determined that varies from study to study. This article will focus on the sample size determination for hypothesis testing that involves means, one sample t test, two independent sample t test, paired sample and one-way analysis of variance.

Keywords: Sample size; Hypothesis testing; Ane sample *t*-test; Two independent sample tests; Paired sample *t* test; One-way analysis of variance

INTRODUCTION

The ever-increasing demand for research has created a need for an efficient method of determining the sample size needed to be representative of a given population.[1] Sample size calculations is necessary for approval of research projects, clearance from ethical committees, approval of grant from funding bodies, publication requirement for journals and most important of all justify the authenticity of study results.[2] Sample size determination is a mathematical process of deciding, before a study begins, how many subjects should be studied'. Whatever type of research design is used for a study, the student/researcher will face the problem of sample size determination.[3] Healthcare researchers, however, give little thought to sample size and choose the most convenient numbers (30, 50, 100, etc.) or time period (one month,6 months, one year, etc.) for their studies. They, and those who approve such studies, should realize that there are important statistical and ethical implications in the choice of sample size for a study.[4]

A study with an overlarge sample may be deemed unethical through the unnecessary involvement of extra subjects and the correspondingly increased costs.[5] On the other hand, a study with a sample that is too small will be unable to detect clinically important effects. Such a study may thus be scientifically useless, and hence unethical in its use of subjects and other resources.[6] Using the appropriate number of subjects optimizes the probability that a study will yield interpretable results and minimizes research waste.[7]

From a statistical perspective, studies with the optimal number of subjects have sufficient (neither too much nor too little) statistical "power" to detect findings.[8] There are numerous situations in which sample size is determined that varies from study to study. This article will focus on the sample size determination for hypothesis testing that involves means, one sample t test, two independent sample t test, paired sample and one-way analysis of variance with examples.

Parameters for sample size determination:
An appropriate sample size is usually dependent upon six parameters; objectives of the study, variables of interest (categorical or quantitative), desired significance level, desired power, effect size/clinical important difference, measurement variability and one-tailed or two-tailed tests.[9,10]

Objectives of the study:
This information is a key in calculating the sample size because every situation has a separate formula for calculating the sample size, for example if you are studying correlations, you cannot apply the two-independent sample t-test sample size formula. Moreover, if you are comparing means of two groups you cannot calculate the sample size through one sample proportion formula. Hence, the objectives of the study are the gateway for sample size calculation.[11]

Variable(s) of interest:
The nature of variables whether categorical (gender, socioeconomic status, occupation etc.) or quantitative (height, BMI, age etc.) also plays a key role in determining the sample size. For determination of

sample size for comparing two proportions the variable(s) of interest should be categorical, and for determining the sample size for one sample t test the variable(s) of interest should be quantitative.

Desired level of significance:

Level of significance is the probability at which null hypothesis is rejected. It is denoted by α. For example, if someone argues that "there's only one chance in a thousand this could have happened by coincidence", a 0.001 level of statistical significance is being implied. The most frequently used values of **α** are 0.05, and 0.01, i.e., 5% and 1%. To reiterate, by 5% we mean that there are about 5 chances in 100 of incorrectly rejecting the null hypothesis or we can say that we are 95% confident in making the correct decision.[12] It is also called type I error. The values against corresponding α can also be taken form published tables.

Table-1: Normal deviates α values

Alpha	Z_{α} (one sided)	$Z_{\alpha/2}$ (two sided)
0.20 (0.80)	0.84	1.28
0.10 (0.90)	1.28	1.64
0.05 (1.64)	1.64	1.96
0.01 (0.99)	2.33	2.81

Desired Power:

The ability of a study to demonstrate an association or causal relationship between two variables given that an association exists. For example, 80% power in a clinical trial means that the study has an 80% chance of ending up with a p value of less than 5% in a statistical test (i.e. a statistically significant treatment effect) if there really was an important difference (e.g. 10% versus 5% mortality) between treatments.[13,14] If the statistical power of a study is low, the study results will be questionable (the study might have been too small to detect any differences). By convention, 80% is an acceptable level of power. The values against corresponding Z values can also be taken form published tables.

Table-2: Normal deviates 1-β values

Power	$Z_{1-\beta}$
0.80	0.84
0.85	1.03
0.90	1.28
0.95	2.64
0.99	2.33

Effect size/clinical important difference:

This parameter is the measured difference between comparisons groups that the investigator would like to detect. Suppose a study is designed to compare a standard diagnostic procedure of 80% accuracy with a new procedure of unknown but potentially higher accuracy.[15] Suppose that the investigator believes that it would be a clinically important improvement if the new procedure were 90% accurate. Therefore, the investigator would choose a minimum expected difference of 10% (0.10). Generally, the setting of this parameter is subjective and is based on clinical judgment and experience with the problem being investigated. The results of literature review or a pilot study can also guide the selection of a minimum difference.[16]

Measurement Variability:

This parameter is represented by the expected standard deviation (SD) in the measurements made within each comparison group. A review of the literature can provide estimates of this parameter. For example, suppose we want to know the average TSH level in hypothyroidism patients. We'd like to be 99% confident about our result. From a previous study, we know that the standard deviation for such population is ±2.9. If preliminary data are not available, this parameter may have to be estimated on the basis of subjective experience, or a range of values may be assumed.[17,18]

One tailed or two tailed tests:

One important concept in significance testing is whether you use a one-tailed or two-tailed test of significance. The answer is that it depends on your hypothesis. When your research hypothesis states the direction of the difference or relationship, then you use a one-tailed probability. For example, a one-tailed test would be used to test these null hypotheses: Females BMI level will be significantly higher than males. CAD patients will have significantly higher cholesterol level than controls. In each case, the null hypothesis (indirectly) predicts the direction of the difference. On the other hand, a two-tailed test would be used to test these null hypotheses: The BMI of males and females does not vary significantly. The cholesterol level of CAD patients and healthy controls does not vary significantly.[19] The recommendation is that if there is any doubt, a two-tailed test should be done as it yields a larger sample size.

Various effects in determining the sample size:

- As the desired power increases ⇑ the sample size also increases ⇑
- As the effect size decreases ⇓ the sample size increases ⇑
- As the treatment effect increases ⇑ the sample size decreases ⇓
- As measurement variability decreases ⇓ the sample size also decreases ⇓
- As α (level of Significance) decreases ⇓ the sample Size increases ⇑
- 2-tailed tests have higher sample size as compared to 1-tailed tests

Sample Size Determination Formulas:

One Sample t test:

Example-1

We want to know whether birth weights of full-term infants who ultimately died of Sudden Infant Death

Syndrome (SIDS) is significantly different from that of other full-term births. A Sample of n=10 SIDS cases demonstrated that the average birth weight of 10 cases was 3000 grams with a Standard Deviation =720.0 gram. The average birth weight of other full-term births was 3300 grams. How large a sample is needed to test the SIDS data with 90% power at α =0.05 (two sided). We want to detect the mean difference in birth weight of 300 grams. The standard deviation is 720.0.

Formula for determination of sample size for one sample t test is;

$$n = \frac{\sigma_2 (Z_{1-\alpha/2} + Z_{1-\beta})_2}{(\mu_1 - \mu_2)_2}$$

Where;
Desired Power ($Z_{1-\beta}$) = 90% (1.28)
Desired Level of Significance ($Z_{1-\alpha/2}$) = 0.05 (1.96)
Mean Difference ($\mu_1 - \mu_2$) = 300
Standard Deviation (σ) = 720.0
Putting the values in the formula the required sample size is 61 infants.

Two independent sample *t* test:
There are two different situations for determining the sample size when comparing the means of two groups. First is when common Variance or Standard Deviation is given for both groups. The second is when separate Variance or Standard Deviation is given for both groups.[20]

Example-2. Situation I (Common Variance or Standard Deviation):
A study tested the cholesterol-lowering potential of dietary linoleic acid in mildly hypercholesterolemia subjects and healthy subjects. Values are plasma cholesterol levels (mmol/L) in two independent groups. Suppose that you are looking for a Mean difference ($\mu_o - \mu_1$) of 1 mmol/L, α = 0.05, $1 - \beta$ = 90%, σ^2 = 2.55. How many individuals must be studied to achieve these conditions?

Formula for determination of sample size for situation I is;

$$n_1 = \frac{2\sigma^2 (Z_{1-\beta} + Z_{1-\alpha/2})^2}{difference^2}$$

Where;
Desired Power ($Z_{1-\beta}$) = 90% (1.28)
Desired Level of Significance ($Z_{1-\alpha/2}$) = 0.05 (1.96)
Mean Difference ($\mu_1 - \mu_2$) = 1mmol/L
Standard Deviation (σ^2) = 2.55
Putting the values in the formula the required sample size is 54 subjects and 54 controls.

Example-3: Situation II (Separate Variance or Standard Deviation):
A clinical dietician wants to compare two different diets, A and B, for diabetic patients. She hypothesizes that diet A (Group 1) will be better than diet B (Group 2), in terms of lowering blood glucose. She plans to get a random sample of diabetic patients and randomly assign them to one of the two diets. At the end of the experiment, which lasts 6 weeks, a fasting blood glucose test was conducted on each patient. She also expects that the average difference in blood glucose measure between the two groups will be about 10 mg/dl. Furthermore, she also assumes the standard deviation of blood glucose distribution for diet A to be 15 and the standard deviation for diet B to be 17. The dietician wants to know the number of subjects needed in each group.

Formula for determination of sample size for situation II is;

$$n_1 = \frac{(Z_{1-\beta} + Z_{1-\alpha/2})^2 (\sigma_1 2 + \sigma_2 2)}{(\mu_1 - \mu_2)^2}$$

Where;
Desired Power ($Z_{1-\beta}$) = 90% (1.28)
Desired Level of Significance ($Z_{1-\alpha/2}$) = 0.05 (1.96)
Mean Difference ($\mu_1 - \mu_2$) = 10 mg/dl
Standard Deviation (σ_1^2) for group A = 15
Standard Deviation (σ_2^2) for group B = 17
Putting the values in the formula the required sample size is 54 subjects for diet A and 54 subjects for diet B.

Paired sample *t* test:
Example-4:
A researcher sought to learn whether oat bran cereal lowered low density lipoprotein (LDL) cholesterol in hypercholesterolemic men. How large a sample is required if he wants to detect a mean change of 0.08 mmol /L with a Standard Deviation of 0.4 mmo/L using level of significance = 0.05 and power = 90%.

Formula for determination of sample size for paired sample t test is;

$$n = \frac{\sigma_d^2 (Z_\beta + Z_{\alpha/2})^2}{difference^2}$$

Where;
Desired Power (Z_β) = 90% (1.28)
Desired Level of Significance ($Z_{\alpha/2}$) = 0.05 (1.96)
Mean Difference ($\mu_{before} - \mu_{after}$) = 0.08 mmol/L
Standard Deviation (σ_d) = 0.4 mmo/L
Putting the values in the formula the required sample size is 262 subjects.

One-way analysis of variance:
The formula for two independent sample t test has limitation to determine the sample size for 2 groups only, whereas, with the help of the formula for one-way analysis of variance one can determine the sample size for 3 or more groups.

Example-5:
A four-arm (k=4) parallel group double blind randomized clinical trial is to be conducted to compare 04 treatments. The comparison is made at a significance level of α = 0.05. Assume that the standard deviation within each group is σ = 3.5 and the true mean

responses for the four treatment groups are: $\mu_1 = 8.25$, $\mu_2 = 11.75$, $\mu_3 = 12.00$ and $\mu_4 = 13.00$. Calculate the sample size at 80% power of study.

Formula for determination of sample size for one-way analysis of variance is:

$$n = \frac{\lambda}{\Delta}$$

Where $\quad \Delta = \frac{1}{\sigma^2} \sum_{i=1}^{k} (\mu_i - \bar{\mu})^2, \qquad \bar{\mu} = \frac{1}{k} \sum_{j=1}^{k} \mu_j.$

The values of λ can be taken from the following table correspondingly to number of groups significance level and power of study.

Table-3: λ values with k, 1-β and α

k	$1-\beta = 0.80$		$1-\beta = 0.90$	
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$
2	11.68	7.85	14.88	10.51
3	13.89	9.64	17.43	12.66
4	15.46	10.91	19.25	14.18
5	16.75	11.94	20.74	15.41
6	17.87	12.83	22.03	16.47
7	18.88	13.63	23.19	17.42
8	19.79	14.36	24.24	18.29
9	20.64	15.03	25.22	19.09
10	21.43	15.65	26.13	19.83
11	22.18	16.25	26.99	20.54
12	22.89	16.81	27.80	21.20
13	23.57	17.34	28.58	21.84
14	24.22	17.85	29.32	22.44
15	24.84	18.34	30.04	23.03
16	25.44	18.82	30.73	23.59
17	26.02	19.27	31.39	24.13
18	26.58	19.71	32.04	24.65
19	27.12	20.14	32.66	25.16
20	27.65	20.56	33.27	25.66

Putting the values in the formula and extracting the values from the above table the required sample size is 11 per group.

CONCLUSION

Sample size calculation is an extremely important step in designing and conducting the research. If not, enough subjects are available then the study should not be carried out or some additional source of subjects should be found. The accuracy of sample size calculations obviously depends on the accuracy of the estimates of the parameters used in the calculations. It is of prime importance to carefully consider the parameters to be used in determining the sample size. Sample size is best considered early in the planning of a study, when modifications in study design can still be made. Attention to sample size will hopefully result in a more meaningful study whose results will eventually receive a high priority for publication.

REFERENCES

1. Daniel WW. Biostatistics: a foundation for analysis in the health sciences 7th ed. New York NY: Wiley, 1999; p.180–5, 268–7.
2. Pagano M, Gauvreau K. Principles of biostatistics. Pacific Grove, Calif: Duxbury, 2000; p.246–9, 330–1.
3. Altman DG. Practical statistics for medical research. London, England: Chapman & Hall; 1991.
4. Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. N Engl J Med 1978;299(13):690–4.
5. Rosner B. Fundamentals of biostatistics 5th ed. Pacific Grove, Calif: Duxbury, 2000; p.308.
6. Feinstein AR. Principles of medical statistics. Boca Raton, Fla: CRC, 2002; p.503.
7. Snedecor GW, Cochran WG. Statistical methods 8th ed. Ames, Iowa: Iowa State University Press, 1989; p.52, 439.
8. SPSS. Sample Power. [internet]. [cited 2012 March 3]. Available from: www.spss.com/SPSSBI/SamplePower/
9. Frison L, Pocock S. Repeated measurements in clinical trials: analysis using mean summary statistics and its implications for design. Stat Med 1992;11(13):1685–1704.
10. Fleiss JL. Statistical methods for rates and proportions 2nd ed. New York, NY: Wiley, 1981; p.45.
11. Hoenig JM, Heisey DM. The abuse of power: the pervasive fallacy of power calculations for data analysis. Am Stat 2001;55:19–24.
12. Kirkpatric LA, Feeney BC. A Simple Guide to SPSS for Windows: Versions 8, 9, and 10. 4th ed. Wadsworth; 2000.
13. Roberts I, Kwan I, Evans P, Haig S. Does animal experimentation inform human healthcare? Observations from a systematic review of international animal experiments on fluid resuscitation. Br Med J 2002;324(7335):474–6.
14. Thornley B, Adams C. Content and Quality of 2000 Controlled Trials in Schizophrenia over 50 Years. BMJ 1998;317(7167):1181–4.
15. Wheeler RE. Portable Power. Technometrics 2001;16:193–201.
16. Wright T. A simple algorithm for tighter exact upper confidence bounds with rare attributes in finite universes. Stat Probab Lett 1999;36(1):59–67.
17. Lipsey MW. Design Sensitivity: Statistical Power for Experimental Research. Newbury Park, Calif: Sage Publications, 1990; p.207.
18. Mace AE. Sample-Size Determination, New York: Reinhold; 1964.
19. Muller KE, Benignus VA. Increasing Scientific Power with Statistical Power. Neurotoxicol Teratol 1992;14(3):211–9.
20. Gerstman BB. Basic Biostatistics: Statistics for public health practice. Boston: Jones & Bartlett Publishers; 2008.

Address for Correspondence:
Waqas Sami, Biostatistician, College of Medicine, Majmaah University, 11952-Kingdom of Saudi Arabia.
Cell: +966 534971951
Email: w.mahmood@mu.edu.sa