HOW MANY SUBJECTS SHOULD BE STUDIED: SAMPLE SIZE DETERMINATION THROUGH HYPOTHESIS TESTING AND CONFIDENCE INTERVAL

*Waqas Sami[1], Mohammed O. Al-Rukban[2], Mohammed Almansour[3], Tayyaba Waqas[4], Kamran Afzal[5], Rehan Asad[6]

ABSTRACT

One of frequently asked question by health researchers is how many individuals will I need to study. Sample size determination is one of the central canons of health research. A study is always better when planned scientifically and determining the sample size for a study is a prime component as it will help to determine optimum number of subjects so that statistically significant results can be detected. If the sample size is larger than what is needed, the study will become cumbersome and ethically exorbitant. On the converse, using too few subjects' will eventually result in wasted time, effort and money etc. Literature is full of examples in which sample size is incorrectly determined for health studies thus resulting in bias conclusions. To ensure the reliability of the results, the significance level and power of study must be fixed before the sample size determination. Sample size determination is very important and always a difficult process to handle. It requires the collaboration of a specialist who has good scientific knowledge in the art and practice of medical statistics. 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 and confidence interval situations commonly used in health studies.

KEY WORDS: Hypothesis testing, confidence interval, significance level, power of study, effect size, sample size

ملخص:

واحد من أهم الأسئلة المتكررة من الباحثين الصحيين هو"كم مشاركا نحتاج للقيام بهذه الدراسة؟".ولذلك فإن حجم العيينة يعتبر من الركائز الأساسية للقيام ببحث علمي رصين,فدائما ماتكون الدراسة أفضل وأكثر مصداقية في نتائجها الإحصائية إذا بني إختيار حجم العينة على أساس علمي. إن الأدبيات والمصادر العلمية مليئة بنماذج لدراسات كان حجم العينة الغيرمناسب زيادة أو نقصا سببا في عدم جدوى الدراسة,ولذلك فإن إختيار حجم العينة المناسب مهم وليس بالأمر السهل فهو يحتاج تعاون بين الخبير في الفن الذي تنطوي تحته الدراسة والمختص في الإحصاء الحيوي.

هدف الدراسة: التعرف على الطريقة المناسبة لتحديد حجم العينة لإختبار فرضية البحث وتحديد وضع فاصل الثقة في الأبحاث الصحية .

Received: 2 April, 2013; Accepted: 12 June, 2013

**Correspondence*: biostatistician1@gmail.com

[1]Lecturer of Biostatistics, Department of Public Health & Community Medicine, College of Medicine, Majmaah University, Saudi Arabia, [2]Professor of Family Medicine, King Saud University, Saudi Arabia. [3]Assistant Professor of Family Medicine, College of Medicine, Majmaah University, Saudi Arabia. [4]Biostatistician, Punjab Institute of Cardiology Lahore, [5]Lecturer of Physiology, Department of Basic Medical Sciences, College of Medicine, Majmaah University, Saudi Arabia. [6]Assistant Professor of Anatomy, Department of Basic Medical Sciences College of Medicine, Majmaah University, Saudi Arabia

INTRODUCTION

Progress in almost every field of science depends upon the contributions made by systematic research, thus research is often viewed as the cornerstone of scientific progress.[1] A scientific research is a one which is a carefully planned investigation, addressing all the important aspects; from generation of research ideas till the publication.[2] The ever increasing demand for research has created a need for an efficient method of determining the sample size to be representative of a given population[3]. Sample size determination is a process of deciding before a study begins, how many subjects should be studied. Whatever type of research design is used for a study, the health researcher will face the problem of sample size determination.[4] The sample size is now easily calculated with the help of computer statistical programs, but the principles underlying the calculation and the limitations must be clearly understood by the investigators.[5] It is not necessarily true that the bigger the sample, the better the study. Beyond a certain point, an increase in sample size will not improve the study in fact, it may do the opposite.[6] Calculation of an appropriate sample size is usually dependent upon six parameters; objectives of the study, variables of interest (qualitative or quantitative), desired significance level, desired power, effect size / clinical important difference, measurement variability and one tailed or two tailed tests.[7] Although a biostatistician may do the necessary exercise to determine the sample size, he/she can only do it with guidance from the investigator on the level of uncertainty that is considered acceptable.[8] In addition, the aforementioned six parameters have to be taken into consideration while calculating the sample size. Since the data are not available before the study begins, the investigators will have to make some assumptions about the data, and provide these assumptions to the biostatistician to be able to calculate the desired sample size. The procedure for estimating sample size is not as precise as investigators may be led to think. One such assumption is about the prevalence, incidence or frequency of the condition or event. If the rate of the event is large, statistical power will be high with a smaller number of cases. If the event is rare, a larger sample size will be needed. Some information may be available from previous studies to guide the estimates. If not, it is up to the investigators to come up with a tentative estimate which the biostatistician can use.[9-11] The ability of a study to demonstrate an association or causal relationship between two variables given that an association exists is known as power of study, 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.[12] The effect size in a study refers to the actual size of the differences observed between groups or the strength of relationships between variables. The likelihood that a study will be able to detect an association between a predictor and an outcome variable depends on the magnitude of the association we decide to look for. The choice of effect size is difficult and arbitrary, but it must be set beforehand and must make a meaningful difference.[13,14] 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 and confidence interval situations commonly used in health studies.

Various effects while determining the sample size[15]:

The information below is a guideline about how sample size is affected by variations that incur due to change in different parameters.

Values for power of study and significance level[9]:

Power of study and significance level's are the core components for determining sample size. Earlier, it was cumbersome to calculate the power of study and significance level due to complex formulas and tables respectively, but now due to recent advancements in the field, experts have prepared tables of normal deviates for corresponding exact values of power of study and significance level.

Sample Size Determination through Hypothesis Testing approach:

One Group Proportion:

One group proportion is use to compare a hypothesized value of population with the sample value, for example, Drug A is administered to 100 patients to treat a particular disease, after administration the health of 50 patients improved. Test whether this drug is better than drug B, which is known to produce improvement in 45% of patients.

Power	$Z_{1-\beta}$
0.80	0.84
0.85	1.03
0.90	1.20
0.95	1.61
0.99	2.33

Table 1: Normal Deviates 1-β 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 (0.95)	1.64	1.96
0.01 (0.99)	2.33	2.81

Table 2: Normal Deviates α values

Example for sample size determination:

The five-year cure rate for a particular cancer is reported in the literature to be 50%. An investigator wishes to test the hypothesis that this cure rate can be applied in a certain local health district. What is the required sample size if the investigator is interested in detecting a true rate of 40%. The level of significance is set at 5% with 90% power of study.

Formula for determination of sample size for one group proportion is[4]:

$$n = \frac{\left(Z_{1-\beta}\left[\sqrt{p_1(1-p_1)}\right] + Z_{1-\alpha/2}\left[\sqrt{p_o(1-p_o)}\right]\right)2}{(p_1 - p_o)2}$$

Where;

P_1 is the anticipated proportion = 40%

P_o is the Null Hypothesis value or actual proportion (test cure rate) = 50%

$p_1 - p_o$ is the difference between proportions = 0.10

$Z_{1-\beta}$ is the desired power of study = 90%

$Z_{1-\alpha/2}$ is the desired level of significance = 5%

Alternate hypothesis is two sided

Putting the values in the formula the required sample size is 259.

Two Groups Proportion:

Compares responses in two groups but the responses in both groups are independent, for example, two preparations of same drug in tablet form were tested for their efficacy in alleviating headache. Preparation A was given to 25 patients of which 17 claiming its effectiveness, while drug B was given to 20 patients of which 16 claiming its effectiveness. Compare the effectiveness of preparation A and B.

Example for sample size determination:

It is believed that the proportion of patients who developed complications after undergoing one type of surgery is 5% while the proportion of patients who developed complications after second type of surgery is 15%. How large should be the sample size in each of the two groups of patients if the investigator wishes to detect with a power of 90% at 5% level of significance.

Formula for determination of sample size for two group proportion is[19]:

$$n = \frac{\left(Z_{1-\alpha/2}\sqrt{2\bar{p}(1-\bar{p})} + Z_{1-\beta}\sqrt{p_1(1-p_1)p_2(1-p_2)}\right)^2}{(p_1 - p_2)^2}$$

$$Where\ \bar{p} = \left(\frac{p_1 + p_2}{2}\right)$$

Where;
P_1 and P_2 are the anticipated proportions = 5% & 15%
$p_1 - p_2$ is the difference between proportions = 10%
$Z_{1-\beta}$ is the desired power of study = 90% (1.28)
$Z_{1-\alpha/2}$ is the desired level of significance = 5% (1.96)
$\overline{P}$ is the average of proportions = 0.10
Alternate hypothesis is two sided

Putting the values in the formula the required sample size is 93 patients in each group.

Paired Groups Proportion:
Two proportions are paired if they share a common feature that affects the outcome. For example, when comparing two laboratory methods (culture media) to detect bacteria in samples of blood, if blood from the same sample is put into both methods, this is the "pairing". Pairs of results from multiple samples can then be compared as a pair of proportions:

Example for sample size determination:
Two drugs A and B were used to treat patients of depression and were compared in-terms of possible side-effects, nausea. The drugs were given to patients on two different occasions. It is believed that incidence of nausea from drug A is 20% less than drug B. Calculate the number of patients required in each group assuming 90% power of study and 5% level of significance.

Formula for determination of sample size for paired group proportion is[20]:

$$n = \frac{\bar{p}(1-\bar{p})(Z_{\beta} + Z_{\alpha/2})^2}{(p_1 - p_2)^2}$$

Where;
P_1 and P_2 are the anticipated proportions = 20% & 50%
$p_1 - p_2$ is the difference between proportions = 30%
$Z_{1-\beta}$ is the desired power of study = 90% (1.28)
$Z_{1-\alpha/2}$ is the desired level of significance = 5% (1.96)
$\overline{P}$ is the average of proportions = 0.35
Alternate hypothesis is one sided

Putting the values in the formula the required sample size is 47 participants.

Correlation:
Correlation is the degree to which two or more quantities are linearly associated. This association may be positive or negative. A positive correlation exists where the high values of one variable are associated with the high values of the other variable(s). A 'negative correlation' means association of high values of one with the low values of the other(s). For example, age and blood pressure are positively correlated, whereas, volume and pressure and negatively correlated.

Example for sample size determination:
According to the literature, the correlation between salt intake and systolic blood pressure is around 0.3. A study is conducted to test the correlation in a population, with the significance level of 1% and power of 90%. How many participants should be there in the study?

Formula for determination of sample size for correlation is[10]:

$$N = \frac{\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2}{\frac{1}{4}\left[\log_e\left(\frac{1+r}{1-r}\right)\right]^2} + 3$$

Where:
$Z_{1-\beta}$ is the desired power of study = 90% (1.28)
$Z_{1-\alpha/2}$ is the desired level of significance = 1% (2.81)
r is the correlation coefficient = 0.30

Putting the values in the formula the required sample size is 158 participants.

Sample Size Determination through (Confidence Interval):

There is only one method of determining sample size that allows the researcher to pre-determine the accuracy of the sample results that is "The Confidence Interval Method of Determining the Sample Size". This method allows the investigator to draw conclusions about the true value of the outcome measure on the basis of information in the sample.[16,17] The sample size is calculated in such a way that there is a 95% probability that it includes the true value. To calculate the sample size from confidence interval the following information is necessary; *Anticipated Population Proportion* which is usually reported in the form of prevalence, incidence or rate etc. *Confidence Level,* tells you an estimated range of values that is likely to include the population parameter being calculated. For example, how confident are we that our confidence interval of 23 – 27 years of age contains the mean age of our population? If this range of ages was calculated with a 95% confidence level, we could say that we are 95% confident that the mean age of our population is between 23 and 27 years [18-19]. *Absolute or Relative Precision* which is a measure of consistency i.e. if a survey is repeated several numbers of times the same result is yielded every time; it is usually reported in percentage.[20]

Example for sample size determination (one group):

A local health department wishes to estimate the prevalence rate of tuberculosis infection among children less than five years of age in its locality. How many children should be included in the sample so that the rate may be estimated to within 5 percentage points of the true value with 95% confidence if it is known that the true rate is unlikely to exceed 20%.

Formula for determination of sample size for one group is[18]:

$$n = \frac{z^2(pq)}{e^2}$$

Where:

Anticipated population proportion = 20%
Confidence Level = 95%
Precision = 5%
(15% - 25%)

Putting the values in the formula the required sample size is 244 children.

Example for sample size determination (two groups):

Suppose that in a pilot study of 50 agricultural workers in an irrigation project. It was observed that 40% had active schistosomiasis. A similar pilot study of 50 agricultural workers outside the project demonstrated that 32% had active schistosomiasis infection. If we would like to carry out a larger study to estimate the true schistosomiasis risk difference to within 5 percentage points of the true value of the confidence interval, how many people must be studied in each of the two groups.

Formula for determination of sample size for two groups is:

$$n = \frac{Z^2 1-\alpha/2[P_1(1-P_1)+P_2(1-P_2)]}{d^2}$$

Where:

Anticipated population proportion= 40% & 32%
Confidence Level = 95%
Precision = 5%age points

Putting the values in the formula the required sample size is 354 people in each group.

Tips in calculation of sample size for Confidence Interval approach:

For situations in which no anticipation is possible a figure of 0.5 should be used. If the anticipation is given in range the closest to 0.5 should be used. The sample size required will be largest when P is equal to 0.5 and in two groups calculation the larger proportion should be kept first.[16]

CONCLUSION

Sample size is best considered early in the planning of a study. The accuracy of sample size calculations depends on the accuracy of the estimates of the parameters used in the calculations. Determining sample size is necessary so that optimum numbers of subjects are available to achieve statistically significant results. On the converse, If the sample size is larger than what is needed, the study will become cumbersome and ethically exorbitant. For successful resolution of sample size issue close and honest collaboration of biostatisticians and subject-matter experts is required.

REFERENCES

1. Waqas Sami, M. O. AlRukban, Tayyaba Waqas, Rehan Asad, Kamran Afzal. Sample size determination in health research. J Ayub Med Coll Abbottabad 2012; 24:4.
2. Feinstein AR. Principles of medical statistics. Boca Raton 2003, Fla: CRC; 503.
3. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. JAMA 1994; 272:122–124.
4. Naing, NN. Determination of sample size. Malaysian Journal of Medical Science 2003; 10(2): 84-86.
5. Uitenbroek DG. Sample size: SISA—simple interactive statistical analysis. Available at: http://home.clara.net/sisa/samsize.html Accessed March 3, 2003.
6. Armitage P, Berry G, Matthews JNS. 2002, Statistical Methods in Medical Research, 4th ed. Blackwell, Oxford.
7. Naing L, Winn T and Rusli BN. Sample Size Calculator for Prevalence Studies. 2006. Available at: http://www.kck.usm.my/ppsg/stats_resources.htm
8. Bland M. An Introduction to Medical Statistics, 2000, 3rd. ed. Oxford University Press, Oxford.
9. Eng J. Sample size estimation: how many individuals should be studied? Radiology 2003; 227: 309–13.
10. Gerstman BB. Basic Biostatistics: statistics for public health practice: 2008. Jones and Bartlett Publishers, Sudbury, Massachusetts.
11. Van Belle G. Statistical Rules of Thumb. Wiley: 2002, New York, NY.
12. Moher D, Dulberg CS, Wells GA. Statistical Power, Sample Size, and Their Reporting in Randomized Controlled Trials. JAMA 1994; 272:122-124
13. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials. In Bayesian Biostatistics. 1996, New York.
14. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York, NY: Wiley, 1981; 45.
15. Casagrande JT, Pike MC, Smith PG. An improved approximate formula for calculating sample sizes for comparing two binomial distributions. Biometrics 1978;34: 483–6.
16. Gardner, MJ., Altman D.G. Statistics with confidence-confidence intervals and statistical guidelines. British Medical Journal 1986; 5:1-13.
17. Daniel WW. Biostatistics: A Foundation for Analysis in the Health Sciences. 1999, 7th edition. New York: John Wiley & Sons.
18. Lwanga, SK., and Lemeshow, S. Sample size determination in health studies: A practical manual. 1991, World Health Organization, Geneva.
19. Chan YH. Randomized Control Trials: The magic number. Singapore Med J 2003; vol 44(4): 172 – 174.
20. Schulz KF, Grimes DA. Sample size calculations in randomized trials: mandatory and mystical. Lancet 2005; 365: 1348–53.