IJBCS
ISSN: 2147-1428

Review Article
International Journal of Basic and Clinical Studies (IJBCS)
2014;3(1): 1-8 Celik MY

Power Analysis for Highlight Clinical Research : How Many Responses Do You Really Need?

Celik M.Yusuf[1]

[1]PhD, Prof. Dr. Biruni University, Medical Faculty, Department of Biostatistics, ycelik@biruni.edu.tr

Abstract

The power of the test is useful when making research plans such as criteria for determining sample size or defining level of statistical significance. Power analysis and sample size calculations are increasingly recognised as essential parts of study design and as important concepts to allow correct interpretation of the scientific and medical literature. In general, power and sample size are positively related; the bigger the sample size, the greater the power.

Key concepts that affect the power of a study are presented as follows; Effect size, Significance level or p value, Nature of the outcome to be measured, Variability of the data, 1-tailed or 2-tailed hypothesis, Sample size.

Three components of power are sample size, effect size and p value. It has been shown that larger effect sizes (population proportion differences) will have more power. If the effect size of two sample proportions are between 93% vs 95% (i.e. very small effect size), in this case the sample size per group was calculated as 1353 individuals. If the effect size of two sample proportions are between 75% vs 95% (i.e. large effect size), in this case the sample size per group was calculated as 47 individuals. The effect size effects the sample size strongly. Larger effect size needs smaller samle size.

Researchers should be aware of the assumptions in power calculations made by different statistical software packages. There are many softwares to calculate sample size in the Internet. The main point to note in using a software is to understand the proper instructions for sample size. The researchers when they calculate the sample size and the power of their study have to consider the main highlight subjects reported in this review.

Key words: Power analysis, Effect size, Sample size, Type I error, Type II error

Introduction

Cohen was the first statistician who gave attention and emphasis to the power of tests in the researche studies. Cohen indicated that test power statistics show that a null hypothesis will be rejected when the null hypothesis is false. Moreover, it is also the most important part of the information on test statistics that can be applied to research data. The power of the test refers to the probability that a null hypothesis will be rejected, which can be written as (1 - β). The power of the test has a relationship with the Type I error or level of statistical significance, represented by (Į), which is the probability that a null hypothesis will be rejected when the null hypothesis is true. The Type II error, represented by (β), is the probability that a null hypothesis will be accepted when the null hypothesis is false. The power of the test is useful when making research plans such as criteria for determining sample size or defining level of statistical significance (1,2).

Power analysis and sample size calculations are increasingly recognised as

IJBCS
ISSN: 2147-1428

Review Article
International Journal of Basic and Clinical Studies (IJBCS)
2014;3(1): 1-8 Celik MY

essential parts of study design and as important concepts to allow correct interpretation of the scientific and medical literature (3). Power analysis, which helps to determine required sample size as well as to interpret research results. One important aspect of good planning is that the study must be of adequate size, relative to the goals of the study. The sample should be large enough to be sensitive to the differences that may exist between treatments. However, it should not be so large that the analysis produces results that are statistically significant yet practically trivial. In an undersized experiment the subjects are exposed to potentially harmful treatments without advancing knowledge, while in an oversized experiment an unnecessary number of subjects are exposed to a potentially harmful treatment, or are denied to a potentially beneficial result. Frequently asked question in these days is how many subjects are really needed for a thermal comfort or productivity study (4).

In general, power and sample size are positively related; the bigger the sample size, the greater the power. However, there are 2 other important factors: the baseline rate and size of the difference between the 2 groups that the study seeks to detect. In general, a big difference is easier to find than a small one: an analogy would be that it is easier to find a lost pair of glasses than a lost contact lens, and you would probably have to have many more tries at finding the contact lens because of its small size. In epidemiology, one can either calculate the needed sample size in advance, or, calculate how much power there is to find a statistical difference with a given sample size. (5)

Researchers should be aware of the assumptions in power calculations made by different statistical software packages. Assumptions should be explicitly stated in grant proposals and manuscripts and should match proposed analyses. Misuse of statistical methodology often fueled by the ease of generating inferential statistics using modern software (6,7).

The purpose of this review is to make available for understanding the power analysis for research and to discusse the component of power analysis like as effect size, sample size and p value.

Key Concepts of Power

Estimating a larger effect size will allow a smaller sample size to be calculated for the
same power of 80%. Key concepts that affect the power of a study are presented as follows (8);

1. Effect size: The size of the difference that the study is designed to detect. The minimum clinically important difference (MCID) is the smallest difference between 2 or more groups that would be clinically worth detection.

2. Significance level or p value: The probability of committing type I error. A level of 0.05 is most commonly used as a criterion standard. A lower level than the criterion standard results in a smaller probability of making a type I error and is usually denoted statistically significant.

3. Nature of the outcome to be measured: The power of a study is calculated using a different formula when the outcome is a mean rather than a proportion.

4. Population variability: Power decreases with an increase in variability of the population to be sampled. For normally distributed data, the variability is expressed as standard deviation.

5. 1-tailed or 2-tailed testing: A decision to specify a 1-tailed or 2-tailed test will

affect power. A 2-tailed test is used rather than 1-tailed test in a power calculation when there is doubt as to the direction of the effect.

6. Sample size: Once other parameters are kept constant, power varies directly in proportion to the number of participants. The larger the sample size, the greater the power, and more
information about the true difference is obtained.

7. Compliance: Compliance refers to the proportion of participants who remain in the study and receive treatment as specified in the protocol. Drop-outs and drop-ins tend to affect the
treatment effect.

8. Allocation ratio: Allocation ratio is the ratio of participants to be recruited to each group of the study (8).

According to Srivastava et al (9); the calculation of optimum sample size is a necessary prerequisite for any research work. There are several methods given in standard textbooks but often the researchers find it very difficult to understand such calculations. The developing countries often tend to look at the research done in other parts of the world. The lack of financial resources/infrastructure to carry out large population based studies in India often prompt the researchers to review the database already existing in the Medline / Google websites. The Sample size calculations detect the minimum number of participants required to be able to detect clinically relevant effect.

Type I (alpha) is the chance when the reasearcher states the that two groups differ when in reality they do not. Most commonly, alpha is taken as 0.05, which means that there is less than 5% chance of drawing a false- positive conclusion.

Type II error (beta) is the chance of a false negative conclusion . Beta is conventionally fixed at 0.20, that means there is less than 20% chance of a false negative conclusion.

Power: The power is the component of beta. i.e. 1-beta. It represents the chance of avoiding a false negative conclusion, or in other words, the chance of detecting a specified effect if it really exists.

Variability: In sample size calculations, variability is based on the population variance of
the outcome variable. The more the variability of the outcome variable, the larger is the sample size required to assess whether the effect is a true effect. In case of continuous outcome variable, the variability is estimated by means of the standard deviation (SD).The variance is generally unknown, and therefore the researchers use an estimate from a plot study or previously done research (9).

Sathian B at al (10) presented that if the study is well designed with a desired sample size then the standard error will be less and the power and precision will be good. All statistical procedures become valid in this context. Every researcher must strive for the proper sample size and the protocol should contain its details. The suggestions are as follows;

1. If the effect of a clinical treatment is not marked when compared to a placebo, or power of the study is low, or a lower significance level (lower ‘p’ value) is expected, the sample size should be increased.
2. If the measurements are highly varying, use the average of repeated measurements.
3. Determine the scientifically acceptable power and level of significance.
4. Estimate the event rate form similar population.

IJBCS
ISSN: 2147-1428

Review Article
International Journal of Basic and Clinical Studies (IJBCS)
2014;3(1): 1-8 Celik MY

5. In research protocols, statistically determined sample size, power of the study, significance level, event rate, duration of the study, and compliance should be mentioned.
6. The sample size should be increased to adequate level for each sub-group when dealing with multiple sub-groups in a population.
7. Always aim for a cost effective sample size.
8. In small negative trials, meta analysis can be tried.
9. When a study requires very large sample size net working with other researchers engaged in similar projects and Multi-centre trials will be beneficial.
10. A study which needs a large sample size to prove any significant difference in two treatments must ensure the required sample size. Otherwise such studies may not provide much information by any method and are better terminated so that the money and time are at least saved.

Lenth RV (11) claimed that there is simple intuition behind results like these: If my car made it to the top of the hill, then it is powerful enough to climb that hill; if it didn't, then it obviously isn't powerful enough. Retrospective power is an obvious answer to a rather uninteresting question. A more meaningful car is powerful enough to climb a particular hill never climbed before; or whether a different car can climb that new hill. Such questions are prospective not retrospective. harmless in its own right. However, in typical practice, it is used to exaggerate the validity of a significant result ("not only is it significant, but the test is really powerful!"), or to make excuses for a nonsignificant one ("well, Power is .38, but that's only because the test isn't very powerful"). The latter case is like blaming the messenger.

Generally speaking, as your sample size increases, so does the power of your test. This should intuitively make sense as a larger sample means that you have collected more information -- which makes it easier to correctly reject the null hypothesis when you should. To ensure that your sample size is big enough, you will need to conduct a power analysis calculation. Unfortunately, these calculations are not easy to do by hand, so unless you are a statistics whiz, you will want the help of a software program. Several software programs are available for free on the Internet and are described below. For any power calculation, you will need to know:
1) What type of test you plan to use (e.g., independent t-test, paired t-test, ANOVA, regression, etc. See Step 6 if you are not familiar with these tests.),
2)The alpha value or significance level you are using (usually 0.01 or 0.05. See the next section of this page for more information.),
3)The expected effect size (See the last section of this page for more information.),
4)The sample size you are planning to use (12).

For understanding the principles of sample size calculation and power analysis, one should have an understanding what is the meaning of effect size.

Effect Size

The effect size refers to the magnitude of the effect under the alternative hypothesis. It should represent the smallest difference that would be of clinical or biological significance. It varies from study to study (13).

An effect size is a measure that describes the magnitude of the difference between two groups. Effect sizes are particularly valuable in best practices

IJBCS
ISSN: 2147-1428

Review Article
International Journal of Basic and Clinical Studies (IJBCS)
2014;3(1): 1-8 Celik MY

research because they represent a standard measure by which all outcomes can be assessed. For example, we can compare effect sizes of dropout, graduation, and academic outcomes on the same scale. An effect size is typically calculated by taking the difference in means between two groups and dividing that number by their combined (pooled) standard deviation. Intuitively, this tells us how many standard deviations' difference there is between the means of the intervention (treatment) and comparison conditions; for example, an effect size of .25 indicates that the treatment group outperformed the comparison group by a quarter of a standard deviation (14).

The effect size could be calculated by the formula as follows;

$$\text{Effectsize} = \frac{\bar{x}_{\text{Experiment}} - \bar{x}_{\text{Control}}}{SD}$$

As its seen that the formula is depends the mean of experiment and control groups and also to the value of standard deviation.

If available, it may be useful to use the effect size found from prior studies. Where no previous study exists, the effect size is determined from literature review, logical assertion, and conjecture. The variance or standard deviation for sample size calculation is obtained either from previous studies or from pilot study. Larger the standard deviation, larger is the sample size required in a study (15). Maybe realized a pilot study could be given an opinion to use the the formula of effect size. In the formula of effect size which the mean is calculated, the measurements are assumed to have normal distributions.

Sample Size

The ideal power for any study is considered to be 80%. Table 1. repesent the sample size for various effect sizes at a fixed power (80%) and level of significance (two tailed, α=0.05) by using the "Sample Size Calculator: Comparing Two Proportions" (16)

Table 1. Sample size for various effect sizes at a fixed power (80%) and level of significance (two tailed hypothesis and α=0.05).

Effect size (Two sample proportions)	Sample size per group
93% vs 95% (i.e. very small effect size)	1353
90% vs 95% (i.e. small effect size)	432
85% vs 95% (i.e. medium effect size)	138
75% vs 95% (i.e. large effect size)	47

According to Table 1. with larger effect sizes (population proportion differences) will have more power.

A Sample size calculation is an essential step in research protocols and is a must to justify the size of clinical studies in papers, reports etc. Nevertheless, one of the most common error in papers reporting clinical trials is a lack of justification of the sample size, and it is a major concern that

important therapeutic effects are being missed because of inadequately sized studies (17).

The three components of power are sample size, effect size and p value. There is more power for larger effect size and less power for smaller effect size. Figure 1 shows three components of power are sample size, effect size and p value.

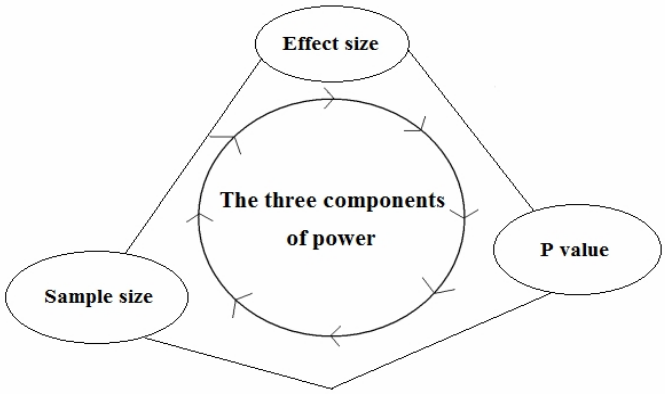


Figure 1. Three components of power are sample size, effect size and p value.

Statistical power is positively correlated with the sample size, which means that given the level of the other factors viz. alpha and minimum detectable difference, a larger sample size gives greater power. However, researchers should be clear to find a difference between statistical difference and scientific difference. Although a larger sample size enables researchers to find smaller difference statistically significant, the difference found may not be scientifically meaningful. Therefore, it is recommended that researchers must have prior idea of what they would expect to be a scientifically meaningful difference before doing a power analysis and determine the actual sample size needed (17).

It has been reported about the topics of "Alpha's influence on power", "Effect size's influence on power" and "Power" as follows (18):

Alpha's influence on power: A small alpha (α) makes the critical value more extreme so that less of the alternative distribution is allocated to the rejection region. Hence, we have less power with smaller alphas. A larger α makes the critical value less extreme so that more of the alternative distribution is allocated to the rejection region. Hence, we have more power with larger alphas.

Effect size's influence on power: A small effect size makes the critical value more extreme on the alternative distribution so that less of that distribution's area is allocated to the rejection region. Hence, we have less power with smaller effect sizes. A larger effect size makes the critical value less extreme on the alternative distribution so that more of that distribution's area is allocated to the rejection region. Hence, we have more power with larger effect sizes.

Power: A small sample size makes the critical value more extreme on the alternative distribution so that less of that distribution's area is allocated to the rejection region. Hence, we have less

power with smaller sample sizes. A larger sample size makes the critical value less extreme on the alternative distribution so that more of that distribution's area is allocated to the rejection region. Hence, we have more power with larger sample sizes (18).

Conclusions

Advice on the power analysis, effect-size, sample size and p value can be summarised as follows:

- Power analysis and sample size calculations are increasingly recognised as essential parts of study design
- The ideal power for any study is considered to be 80%.
- There are four key factors that influence the power of a statistical test: The alpha (α) that a researcher chooses, the effect size,the sample size and the statistical test used.
- The researcher will have less power with smaller alphas and they have more power with larger alphas.
- The researcher will have less power with smaller effect sizes and they have more power with larger effect sizes.
- The researcher will have less power with smaller sample sizes and they have more power with larger sample sizes

There are many softwares to calculate sample size in the Internet. The main point to note in using a software is to understand the proper instructions for sample size. The researchers when they calculate the sample size and the power of their study have to consider the main highlight subjects reported in this review.

References

1) Cohen, J. Statistical power analysis of the behavioral science. 2 nd ed.New Jersey : Larence Erlbaum. 1988

2) Natcha Mahapoonyanonta N, Mahapoonyanontb T, Pengkaewa N, Kamhangkita R. Power of the test of One-Way Anova after transforming with large sample size data. Procedia Social and Behavioral Sciences. 2010 (9): 933–937

3)Columb MO, Stevens A: Power analysis and sample size calculations. Current Anaesthesia & Critical Care. 2008 (19): 12–14

4) Lan L , Lian Z. Application of statistical power analysis – How to determine the right sample size in human health, comfort and productivity research. Building and Environment. 2010 (45): 1202–1213

5)Janet P. Haas DN. Sample size and power. American Journal of Infection Control. 2012 (40): 766-7

6) Bella ML, Teixeira-Pintoc A, McKenzied JE, Oliviere J. A myriad of methods: Calculated sample size for two proportions was dependent on the choice of sample size formula and software. J. Clin. Epidemiol. 2014 (67): 601-605

7) Marino MJ. The use and misuse of statistical methodologies in pharmacology Research. Biochemical Pharmacology.2014 (87): 78–92

8) Cadeddu, M, Farrokhyar F, Thoma A, Haines T, Garnett A, Goldsmith CH, for the Evidence-Based Surgery Working Group. Users' guide to the surgical literature: how to

assess power and sample size. J can chir. 2008 51 (6)

9) Shruti Srivastava1 S, Manjeet S. Bhatia MS, Rajoura OP, Kumari A. Sinha VK. A Pilot study to calculate the sample size for the prevalence of aggression in psychiatric outpatient setting. Delhi Psychiatry Journal. 2012; 15:(2)

10)Sathian B, Sreedharan J, Mittal A. Importance of sample size calculation in the original
medical research articles from developing countries . Webmed Central Epidemiology. 2012;3(5)

11) Lenth RV Two Sample-Size Practices that don't recommend. http:// homepage.math.uiowa.edu/~rlenth/Power/ 2badHabits.pdf

12) MEERA; My Environmental Education Evaluation Resource Assistant. http://meera.snre.umich.edu/plan-an-evaluation/related-topics/power-analysis-statistical-significance-effect-size

13) Zodpey SP. Sample size and power analysis in medical research. Indian J Dermatol Venereol Leprol 2004;70:123-8

14) TEA; Texas Education Agency: How to Interpret Effect Sizes. http:// www.tea.state.tx.us/Best_Practice_Standards/How_To_Interpret_Effect_Sizes.aspx

15)Suresh KP, Chandrashekara S. Sample size estimation and power analysis for clinical research studies. J Hum Reprod Sci (serial online]) 2012 (cited 2014 Jun 28);5:7-13.

16) Sample size calculator: Comparing two proportions. http://www.select-statistics.co.uk/sample-size-calculator-two-proportions

17) KP Suresh KP and S Chandrashekara S. Sample size estimation and power analysis for clinical research studies. J Hum Reprod Sci. 2012 5(1): 7–13

18) Effect Sizes: homepages.luc.edu/~ rwill5 /Lecture_8.ppt