Review Article

International Journal of Basic and Clinical Studies (IJBCS)

2015; 4(1): 1-5, Celik MY

Power of a Clinical Study

M.Yusuf Celik[1],

Editor-in-Chief

[1] Prof.Dr. Biruni University, Medical Faculty, Dept of Biostatistics, Topkapi, Istanbul.

Abstract

The probability of *not* committing a Type II error is called the power of a hypothesis test. Making the wrong decisions can have serious results and risks. Sample size is closely tied to statistical power, which is the ability of a study to enable detection of a statistically significant difference when there truly is one. A trade-off exists between a feasible sample size and adequate statistical power. Sample size is important primarily because of its effect on statistical *power*. To increase power of the study depends to the main factors as follows; Increase alpha, conduct a one-tailed test, increase the effect size, decrease random error, increase sample size. Underestimation of sample size may result in drug turning out to be statistically non-significant even though clinical significance exists.

At the result, underpowered study may not yield useful results and consequently unnecessarily put respondents at risk. Overall, researchers can and should attend to power and calculate it at the beginning of the study. More importantly, supplementary computer programs are developed to calculate it.

Key worlds: Power, Sample Size, Type I Error, Type II Error

Researchers are often disregard the power of his research. However, the results of researchers with little power is not reliable. Statistical methods and techniques are a tool. Making the wrong decisions can have serious results and risks. It is essential that the results of the analysis are prepared in a consistent and standard way so that the comparison of the information is meaningful and leads to the right decision.

Underestimation of sample size may result in drug turning out to be statistically non-significant even though clinical significance exists.

There are two main reasons why a study may not show a significant difference between groups being studied (e.g. in a randomized trial of a new drug, or a case-control study testing the effect of an exposure on a disease).

1)There really was no significant difference (hence a true negative result), 2)There was a difference but the study failed to detect it (false negative result). This may arise because the study was poorly designed (e.g. used imprecise measurements) or because the study was too small (in statistical jargon, it "lacked power").

Statistical power is affected by 3 factors:

1)The difference in outcome rates between the two groups. A smaller difference requires exponentially more power.

2)The level of significant difference you are hoping to show (e.g. $p < 0.05$ or <0.001). Chasing after a small p value takes more study power,

3)The frequency of the outcome in the two groups. Imagine an exposure that increases incidence by a third: it is easier to show a difference between 30 and 45 percent than between 10 and 15 per cent. Maximum power is reached when roughly half of the people studied have the outcome of interest (1).

Sample size is closely tied to statistical power, which is the ability of a study to enable detection of a statistically significant difference when there truly is one. A trade-off exists

between a feasible sample size and adequate statistical power. Sample size is important primarily because of its effect on statistical *power*. Statistical power is the probability that a statistical test will indicate a significant difference when there truly is one. Statistical power is analogous to the sensitivity of a diagnostic test and one could mentally substitute the word “sensitivity” for the word “power” during statistical discussions (2).

Power is the probability of rejecting the null hypothesis when the alternative hypothesis is true. It measures the ability of a test to reject the null hypothesis when it should be rejected. At a given significance level, the power of the test is increased by having a larger sample size. The minimum accepted level is considered to be 80%, which means there is an eight in ten chance of detecting a difference of the specified effect size (3).

The power could be defined as (4):

Power

1) Probability of finding true significance

2) True positive

3) 1 – beta, where beta is :

- Probability of not finding significance when it is there
- False negative
- Probability of a Type II error

4) Usually set to .80

Power should not be less that 80%. Type II error is directly proportional to sample size.

The power could be define as follows (5);

The probability f *not* committing a Type II error is called he power of a hypothesis test.

Increasing beta, the probability of a Type II error will reduce the power of a hypothesis test (Figure 1). Increasing sample size makes the hypothesis test more sensitive - more likely to reject the null hypothesis when it is, in fact, false. Increasing the significance level reduces the region of acceptance, which makes the hypothesis test more likely to reject the null hypothesis, thus increasing the power of the test. Since, by definition, power is equal to one minus beta, the power of a test will get smaller as beta gets bigger.

Increasing sample size makes the hypothesis test more sensitive - more likely to reject the null hypothesis when it is, in fact, false. Thus, it increases the power of the test. The effect size is not affected by sample size. And the probability of making a Type II error gets smaller, not bigger, as sample size increases.

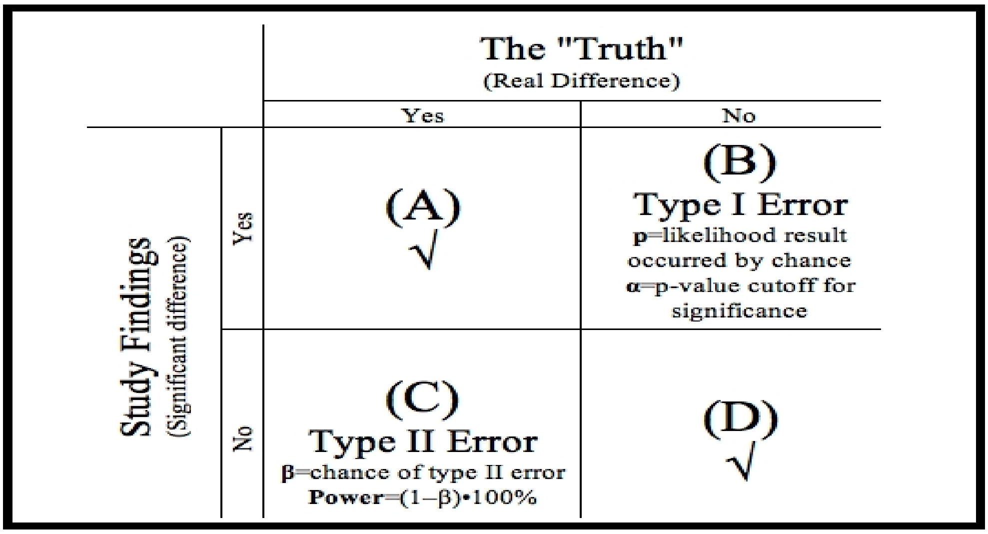

Figure 1. Type 1 and Type 2 errors

β depends on additional factors from your sample, like the standard error and the effect size. Effect size is the raw distance from the null hypothesis you're considering when you set the type II error rate. In the coin flipping scenarios, the farther we got from 5 heads out of 10, the larger the effect size. β decreases when effect size increases.

If a new medicine causes someone's heart to explode (large effect size), that's easier to notice than a small increase in blood pressure (small effect size). 1 – β is called the power of the test. If we're considering a large effect, our test is going to have a lot of power to detect the effect and reject the null based on that. β increases when standard error increases. We can't usually get a small type 1 error rate and a small type 2 error rate at the same time. When we make α small, we're less likely to reject the null. That means we're more likely to fail to reject the null. Failing to reject the null when we should is a type 2 error.

α is the type 1 error rate. αis the rate that type 1 errors will occur. That means if we're testing many hypotheses at once, then α of rejected nulls will be falsely rejected. In courts, some people are wrongly convinced. Type 2 error: The act of failing to reject the null when it is false. β is the type 2 error rate. β is the chance of failing to reject the null hypothesis when we should reject it (6,7).

To increase power of the study depends to the main factors as follows (8):

1. Increase alpha
2. Conduct a one-tailed test
3. Increase the effect size
4. Decrease random error
5. Increase sample size

It should also be remembered that using non-parametric tests have less power and categorizing continuous variables should be avoided as it reduces statistical power. Parametric tests will have more power than a nonparametric counterpart if the assumptions are met. However, the distributional assumptions are often strict or undesirable for the parametric tests and deviations can lead to misleading results.

The ability of statistical tests to demonstrate that a difference or association of a particular magnitude would be unlikely if the null hypothesis is true, under circumstances in which the null hypothesis is actually false, is referred to as statistical power. The more statistical power that you have in your research study; the more likely you are to get the evidence you are looking for if your research hypothesis is actually true (9).

It has to remembered that multivariate techniques have emerged as a powerful tool to analyse data represented in terms of many variables. The main reason being that a series of univariate analysis carried out separately for each variable may, at times, lead to incorrect interpretation of the result. This is so because univariate analysis does not consider the correlation or inter-dependence among the variables. As a result, during the last fifty years, a number of statisticians have contributed to the development of several multivariate techniques. Today, these techniques are being applied in many fields such as economics, sociology, psychology, agriculture, anthropology, biology and medicine. These techniques are used in analyzing social, psychological, medical and economic data, specially when the variables concerning research studies of these fields are supposed to be correlated with each other and when rigorous probabilistic models cannot be appropriately used. Applications of multivariate techniques in practice have been accelerated in modern times because of the advent of high speed electronic computers (10).

At the result, underpowered study may not yield useful results and consequently unnecessarily put respondents at risk. Overall, researchers can and should attend to power and calculate it at the beginning of the study. More importantly, supplementary computer programs are developed to calculate it.

References

1) Statistical power of a study, http://www.med.uottawa.ca/sim/dat a /Study_Design_ Power_e.htm, June 2015
2) Eng J. Sample Size Estimation: How Many Individuals Should Be Studied? Radiology 2003; 227:309–313
3) McCrum-Gardner E. Sample size and power calculations made simple. International Journal of Therapy and Rehabilitation, January 2010, Vol 17, No 1
4) A Researcher's Guide to Power Analysis, Anne Hunt, OMDS, Utah State University. https://rgs.usu.edu/irb/files/uploads/A_Researchers_Guide_to_Power_Analysis_USU.pdf
5) Power of a Hypothesis Test, Ipsa Software, Power System Analysis Software. Fast, Flexible and Powerful. http://stattrek.com/hypothesis-test/power-of-test.aspx?tutorial=ap
6)Taylor C. What Is the Difference Between Type I and Type II Errors? http://statistics.about.com/od/Inferential-Statistics/a/Type-I-And-Type-II-Errors.htm
7) Review P-values Type I and Type II Errors, http://www.sfu.ca /~jackd/ Stat203/ Wk07_3_Full.pdf
8) Grace-Martin K. 5 Ways to Increase Power in a Study. The Analysis Factor, http://www.theanalysisfactor.com/5-ways-to-increase-power-in-a-study

9) Dewberry C. Statistical Methods for Organizational Research. Taylor & Francis e-Library, 2005.

10) Kothari CR. Research Methodology Copyright © 2004, New Age International (P) L