

Is the Sample Size Big Enough? 4 Things You Need to Know!

Global Spine Journal
2022, Vol. 12(5) 1027–1028
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/21925682211055711
journals.sagepub.com/home/gsj



Joseph R. Dettori, PhD¹ , Daniel C. Norvell, PhD¹, and Jens R. Chapman, MD²

Keywords

sample size, power, type II error

One of the first questions asked when embarking on a new research project is, “How many patients do I need to recruit for my study?” In a descriptive study, this is less important. However, when making inferences (e.g., comparing 2 treatments or evaluating the effect of specific risk factors on outcomes), this is an important question because if the sample is too small, a study will not be able to find true results, and if too big, it will find inconsequential ones. Studies too small or too large waste energy, time, and money, either by using resources without finding results or using more resources than necessary. Both expose participants unnecessarily to experimental risks.

The key, then, is to size a study optimally so that it is just large enough to detect an important clinical effect. In order to do so, 4 key components are needed.

1. Level of significance. Also known as the alpha, this corresponds to the P -value, usually set by convention at 5% (i.e., $P = .05$). It represents a false positive risk of finding a difference between 2 treatments when in reality, no difference exists. This is also referred to as a Type I error.
2. Power. Failure to detect a difference between treatments when one actually exists is referred to as a Type II error. It results in a false negative conclusion, also referred to as beta (β). The power of a study ($1 - \beta$) equals the probability to detect a difference between 2 treatments when one does in fact exist. Typically, power threshold is set at 0.80 or 0.90, that is, there is an 80% or 90% probability that we have a true positive result.
3. Expected effect size. The effect size or the magnitude of effect is simply the minimum difference in outcome means, rates, or proportions between treatment A and treatment B that the investigators wish to

detect. This difference should be clinically relevant from a patient’s and clinician’s perspective. And the effect size should be realistic, that is, it should be biologically plausible. How does one choose an appropriate effect size before a study begins? This is done based on prior evidence (a pilot study or previous published clinical studies) or from clinical experience.

4. Underlying event risk in the population. The underlying event risk is the risk in a patient population that goes untreated or receives the standard of care for a specific condition. This is estimated from previous studies. For example, the underlying risk of nonunion 1 year following a single level cervical fusion has been reported to be approximately 8.6%.¹ Future studies evaluating nonunion in a similar population might reasonably expect a similar underlying event risk. The smaller the underlying risk, the more patients will need to be included in the study.

The table demonstrates how changes in the different parameters described above affect the sample size in Example #1. An increase in the power of the study requires a larger sample size (Example #2). However, increasing the effect size (Example #3) or increasing the underlying risk (Example #4) reduces the sample size needed.

¹Spectrum Research, Inc., Steilacoom, WA, USA

²Swedish Neuroscience Institute, Seattle, WA, USA

Corresponding Author:

Joseph R. Dettori, PhD, Spectrum Research, Inc., Box 88998, Steilacoom 98388, WA, USA.

Email: joe@specri.com



Creative Commons Non Commercial No Derivs CC BY-NC-ND: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits non-commercial use, reproduction and distribution of the work as published without adaptation or alteration, without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Example	Significance Level (α)	Power (1- β)	Expected Effect Size	Underlying Event Risk	Sample Size
#1	.05	80%	5%	15%	1452
#2 (changing power)	.05	90%	5%	15%	1914
#3 (changing effect size)	.05	80%	10%	15%	948
#4 (changing underlying risk)	.05	80%	5%	25%	118

Other factors investigators consider in the final sample size include the expected dropout rate, an unequal allocation ratio between a treatment and control group, and the objective and design of the study.

Calculating the sample size is a process that investigators should conduct prior to initiating a study. However, once a study has been reported, it may be important for you to determine whether the results of the study are over, or more likely, under powered depending on the sample size used.

Using the principles described above should provide you with the basics to help in your interpretation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Joseph R. Dettori  <https://orcid.org/0000-0002-0216-8363>

References

1. Luszczuk M, Smith JS, Fischgrund JS, Ludwig SC, Sasso RC, Shaffrey CI, et al. Does smoking have an impact on fusion rate in single-level anterior cervical discectomy and fusion with allograft and rigid plate fixation? *J Neurosurg Spine* 2013;19(5):527-531.