

REVIEW ARTICLE—EDUCATIOAL TRACK

Sample Size Considerations: Basics for Preparing Clinical or Basic Research

David N. Williams, PhD, MA¹⁾ and Kathryn A. Williams, MS²⁾

Received: April 30, 2020/Revised manuscript received: July 29, 2020/Accepted: July 31, 2020

© The Japanese Society of Nuclear Cardiology 2020

Abstract

Background: Sample size estimation is an important and integral part of a research protocol. While “how large a sample?” is a simple question, the answer is only meaningful within the context of the research question.

Methods: Clear definitions of the variable of interest and target population parameters are key to estimating sample size. In turn the sample must be sized such that it can accurately detect the ‘effect’ of interest, adequately represent the target population and maintain maximum design efficiency. Four basic pieces of information are utilized in most sample estimation across all clinical research: significance level, power, magnitude of effect and variability of the variable of interest.

Results/Discussion: Preliminary determination of these will greatly facilitate work with a biostatistician or a computer application to create a sample size estimation. While applications can support relatively simple sample size calculations consultation with a biostatistician is recommended.

Keywords: Magnitude of effect, Power, Sample size, Sampling

Ann Nucl Cardiol 2020; 6 (1): 81–85

In research design, an important (and certainly very frequently asked) question is “how large a sample is needed”. Perhaps this is because it seems like such a simple question or because it can seem shrouded in mystery or both. But any discussion of sample size must begin with the research question and then with the sample design; both of which are derived from and a critical part of a full study protocol. This is critical because sampling design and, in turn, sample size are based on and only meaningful within the context of the research question and a concern with effectively and accurately answering that question. The goal of this paper is to introduce the basic concepts needed to understand sample size estimation. Our objective is, in part, that the basics introduced here will apply and enlighten even in the wide range of designs required for medical research. Although the examples used are drawn from clinical research, the approach is the same in all basic research.

Once the research question and variables (i.e. effects) of interest are clearly defined, the target population, i.e. a group

of people who share a common character or condition (the variables of interest) (1), can be described. The sample to be taken should be, as much as possible, representative of this target population designed (for purposes of this paper we will assume the researcher will work with a sample rather than the entire target population). The researcher is entirely responsible for assuring this representation through the sample design adopted. The sample design has a minimum of four key factors: an effective sample methodology, identification of the sample frame, sampling methods and sample size estimation (2)¹.

Methods

So how big should the sample be? The answer has two parts; *first*: “large enough to detect the key variable, effect or association of interest.” This depends on the magnitude of that variable, effect or association in the target population (3). If the magnitude is large, e.g. in an imaging study where we might wish to show that brain natriuretic peptide (BNP) (at

¹ This paper deals with a limited range of research design and needed sample size estimations to introduce and address basic sampling concepts. While there are a wide variety of alternative study designs and many different approaches than those discussed here, the basics that are introduced still apply. Specifically addressing the wide variety of approaches, however, is beyond the scope of this paper.

baseline) in the granulocyte-colony-stimulating-factor (G-CSF) group measurement of 149.5 units is larger than the control group measure of 184.7 (a difference of 35.2 percentage points) (4), it requires a smaller sample. If the magnitude is small, e.g. a difference of 3.1 BNP (at 6 month) units between the control (106.6) and G-CSF groups (103.5) (4), it requires a larger sample. Hence, if the sample is too small the study can (easily) fail to detect a clinically important effect or association (5). Too large a sample and the study can be too costly (i.e. waste of live animals in animal research), difficult, and possibly detect an effect that is not clinically important.

Second: “large enough to effectively represent the target population”; this depends in part on the kind of study being designed and homogeneity of the target population. Generally, there are two types of clinical research studies: clinical trials (which include the introduction of an intervention) and observational. While random assignment is the ideal, in observational studies (arguably the most common type of clinical research study) the investigator does not assign treatment to subjects but instead examines pre-existing treatments, exposures or policies and their effects (6). Because treatment is not randomly assigned observational studies normally experience greater heterogeneity in the target population; this heterogeneity must be ‘coped with’ in the sampling plan because representativeness is key (1). Heterogeneity in a target population generally requires a larger sample size to accurately capture the variable of interest. And then two different samples are required, one for the control (or sometimes alternative treatment) group, which often represents the population that originated the cases, and another for the case/treatment group, based on diagnosis of the disease or condition of interest (2). The inclusion and exclusion criteria in the study plan should make it very clear how heterogeneous or homogeneous the target population is.

In contrast, clinical trials aim to obtain a homogeneous sample, often those who will benefit from (and will consent to) an intervention (2) and may not always represent a general target population. Multicenter studies are conducted, in part, to off-set for these limitations.

Four basic pieces of needed information

So how big should the sample be? In general, there are four basic pieces of information needed to estimate sample size:

1. The level of significance that will be acceptable for rejecting the null hypothesis. Commonly called ‘p-value’, it reflects the probability of detecting a difference when none actually exists. This is known as Type 1 Error; Type 1 Error decreases as sample size increases (7). The symbol for the probability of a Type 1 Error is α . 0.05 is commonly used for this level but some areas of research

use 0.01; check research publications in your area for standards.

2. Power: the probability of not committing a Type 2 Error, i.e. incorrectly accepting a null hypothesis. Power analysis estimates the number of patients required to avoid a Type 2 Error. The symbol for the probability of a Type 2 Error is β ; power is equal to $1-\beta$ (8). 0.80 (i.e. 80%) is a common power level for research but don’t assume this should be used; review publications in your area to verify this. Some fields of research utilize or situations will require a more stringent standard of 0.90 (i.e. 90%).
3. Magnitude of the effect or association of interest. For continuous variables, this will often be a numerical difference, e.g. 52.7 (184.7 vs. 237.4), but the difference also may be presented as a percentage of the control measure, e.g. a 5% or 10% difference between the treatment and the control group. For binomial variables magnitude of difference will be presented as a percent, for example 10% in the control group, 25% in the treatment group, a 15-point difference between the two groups. In some cases, where a single sample may be used to capture estimates of prevalence or incidence, magnitude of difference may be expressed as a confidence interval width. As discussed above, smaller magnitudes will require a larger sample sizes.

Estimating a meaningful magnitude of difference may be a challenge. While typically estimated as the difference between treatment and control outcomes, observational, epidemiological, and other study designs require alternative methods. In single sample, one-arm, designs magnitude of difference may be estimated based on difference between the treatment arm and a known population proportion, findings from a previous study or a historical control. Difference may have to be based not on previous data but on estimates of what would be a ‘minimal clinically important difference (MCID)’, that is “something significant enough to change patient management (9).” Also there may be several ‘variables of interest;’ (in this situation sample size could be based on the one variable that requires that largest sample) which can be confusing.

4. Variability of the variable or condition of interest: greater variability requires a larger sample. In a homogenous population variability will be lower than in a heterogeneous population. Estimates of variation are most often based on previous research and are most often represented by the standard deviation. Note that if standard deviation is not presented in a study it may be included as (and can be estimated from) the standard error, percentiles, or coefficient of variation. If no comparable previous study is available an option that will give a rough estimate is based

on an expected population range: divide the expected range (high-low values) by 6. For example: a range of 60/6=10, an estimated standard deviation (10).

Any one of these four pieces of information can influence the estimated size of the needed sample; selection of a lower level of power; a higher level of significance, a larger magnitude of difference can decrease the needed sample (11). While critical peer reviews (for a journal article) may 'catch' poorly designed and sized samples, it is the responsibility of the researcher to develop meaningful estimates and, in turn, correct sample size estimates.

Two basic sample size calculations

These four pieces of information, or some portion of them, are generally utilized in most sample size formulations across all fields (human, animal and others of clinical research (12)). They may take on different configurations, but researching them and developing estimates and/or determining standards (e.g. for significance and power levels) before you go see the biostatistician (or before you access a sample size application) will result in a far more meaningful and correct estimate of the needed sample. Remember, if you don't calculate the sample size correctly, you may not be able to prove your hypothesis.

We present two examples, first an example of estimating sample size for comparing two groups with a continuous variable (a two independent samples test) assuming equal variance and samples (per group), illustrates how these four pieces of information are utilized. The second deals with estimating sample size for two proportions.

Example 1. Two continuous measures

For the first example, a basic formula for calculating sample size (n) is (13, 14):

$$n=2[(a+b)^2 \sigma^2]/(\mu_1 - \mu_2)^2$$

For this example, we will use the statistics from the results section of Toyama, et al, paper in ANC Vol. 5, p. 24, (4) as the basis to calculate the sample size for a new study. if BNP represents our variable of interest and the 6-month control group represents our target population, we would set:

1. $a=(1-\alpha/2)$ 100 percentile of a standard normal distribution (where α =level of significance=0.05)=1.96
2. $b=(\beta)$ 100 percentile of a standard normal distribution (where β =power=0.8)=0.842
3. $(\mu_1 - \mu_2)$ =minimum magnitude of difference. μ_1 =the mean of the control group (i. e. 184.7), μ_2 =the minimum hypothesized mean of the treatment group (e.g. 237.4). The difference between the two is the magnitude 52.7.
4. σ =variability of the variable of interest (i. e. standard

deviation estimated to be 91.7).

Applying these parameters to our example formula:

$$n=2[(1.96+0.842)^2 91.7^2]/(184.7-237.4)^2$$

delivers an estimated sample of 48 per group, 96 total, required to detect a difference of 52.7 with the desired power of 80%. (11) Note that if the magnitude were smaller, 20 instead of 52.7, a sample size of approximately 331 per group would be required.

Example 2: Two proportions

For the second example, a basic formula for estimating sample size (n) for two proportions is (15):

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

For this example, we will use statistics from the results section of the Muramatsu, et. al. paper, Table 1, in ANC Vol. 5, p 35 (16); if "B blocker (%)" represents our treatment of interest and Aortic cal +/- our variable of interest, which provides two proportions for the B blocker and the control groups for comparison, we would set

1. $Z_{\alpha/2}$ =1.96, the critical value of a standard normal distribution with a 95% confidence level
2. where α =0.05, the level of significance at which the null hypothesis will be rejected Z_{β} =0.842, for a power=0.8
3. p_1 and p_2 are the expected proportions of the two groups, i.e. 24.7% and 32.4% respectively.

Note that the only different information needed, compared with example 1, are the expected proportions of the two groups. These proportions are comparable to the minimum magnitude of difference as we discussed.

Applying these values to our example formula:

$$n=((1.96+.842)^2 * (.247(1-.247)+.324(1-.324)))/(.247-.324)^2$$

delivers an estimated sample size of 460 for groups 1 and 2, 920 total, is needed to achieve 80% power to detect a difference between the two groups of 7.7%. The sample sizes are large due to the small difference we wish to detect.

While these examples are drawn from clinical research, it is not difficult to substitute population and variable of interest and the treatment vs control groups from a basic science research study into the formulas.

Support: Biostatisticians and applications

The optimal support you can obtain is to consult with a biostatistician early in your research design. A biostatistician

Figure 1 Four basic information pieces commonly needed in sample size estimates.

Information	Purpose	Common names and/or values
1. The level of significance	Level of probability of the outcome at which the null hypothesis is rejected, reflects the probability of detecting a difference when none actually exists. Type 1 Error decreases as sample size increases.	p-value, 0.05 is most common, 0.01 is also used in some fields.
2. Power	The probability of not committing a Type 2 Error, i. e. incorrectly accepting a null hypothesis.	Power, 80% is the most common value, 90% may be applied in some fields.
3. Magnitude of the effect	An estimation of the size of effect, difference, association that is expected from the treatment.	Typically, the difference between the measured outcomes of the treatment and control groups but may be based on previous research, levels of minimal clinically important difference (MCID).
4. Variability of the variable or condition of interest	A measure of heterogeneity of the variable of interest in the target population. Greater variability requires a larger sample.	Commonly represented by the standard deviation but may be represented by the standard error, percentiles or coefficient of variation.

can help you to:

- define your research question in statistical terms and determine the best study design
- identify and estimate the parameters needed for the sample size calculation,
- assist with or complete calculation of the sample size
- support working with sample size applications (for simple sample size calculations) or post-hoc power calculations

A number of applications are available on-line, free of charge or to purchase, which can facilitate sample size estimations as well as estimations of power if sample size is fixed. They work well for simple estimations such as the above example. But as the study design and statistical methods become more complex these applications can quickly meet their limits. Given the importance of sample size as an integral part of the research plan, we strongly advise that an experienced biostatistician be consulted.

Discussion/Conclusions

Sample size is an essential part of a meaningful study protocol. Derived from the research question and the “who” of the target population, how the sample is selected and how large the sample needs to be are equally important components. The four basic pieces of information needed for sample size calculations need to be understood for their importance as well. While basic formulas and applications exist, estimation of sample sizes can become complex with increasing complexity of study design and analysis tools used. There a variety of sample size estimation applications that can adequately facilitate estimations in less complex study designs. It is recommended that, if possible, a biostatistician be consulted throughout the study and especially for more complex designs.

Acknowledgments

None.

Sources of funding

None.

Conflicts of interest

None.

Reprint requests and correspondence:

David N. Williams, PhD, MA

Boston Children’s Hospital, 300 Longwood Avenue,
Boston, MA, 02115 USA

E-mail: David.williams3@childrens.harvard.edu

References

1. Elfi M, Negida A. Sampling methods in clinical research; an educational review. *Emerg (Tehran)* 2017; 5: e52.
2. Martinez-Mesa J, González-Chica DA, Duquia RP, Bonamigo RR, Bastos JL. Sampling: how to select participants in my research study? *An Bras Dermatol* 2016; 91: 326–30.
3. Hulley SB, Cummings SR, Warren SB, Grady DG, Newman TB. *Designing clinical research*. 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2001.
4. Toyama T, Hoshizaki H, Kasama S, Miyaishi Y, Kan H, Yamashita E, et al. Usefulness of hyper early granulocyte-colony-stimulating factor therapy for patients with acute myocardial infarction. *Ann Nucl Cardiol* 2019; 5: 21–7.
5. Suresh K, Thomas SV, Suresh G. Design, data analysis and sampling techniques for clinical research. *Ann Indian Acad Neurol* 2011; 14: 287–90.
6. Dinan AM. In: Curtis LH, Thomas LE, editors. *Understanding Clinical Research*. second ed: McGraw Hill Education, 2013: 14.
7. Gupta KK, Attri JP, Singh A, Kaur H, Kaur G. Basic concepts for sample size calculation: critical step for any clinical trials! *Saudi J Anaesth* 2016; 10: 328–31.
8. Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. *Emerg Med J* 2003; 20: 453–8.
9. Cook CE. Clinimetrics corner: the Minimal Clinically Important Change Score (MCID): a necessary pretense. *J Man*

- Manip Ther 2008; 16: E82–3.
10. Cochran W. Sampling Techniques. 3rd ed. New York: John Wiley & Sons; 1977.
 11. Noordzij M, Dekker F, Zoccali C, Jager KJ. Sample size calculations. *Nephron Clin Pract* 2011; 118: c319–23.
 12. Boston University Research Support. Sample Size Calculations (IACUC). Research with Animals. Boston: Boston University, 2019.
 13. Florey CD. Sample size for beginners. *BMJ* 1993; 306: 1181–4.
 14. Julious SA. Sample sizes for clinical trials with Normal data. *Stat Med* 2004; 23: 1921–86.
 15. Wang H, Chow S. Sample Size Calculation for Comparing Proportions. *Wiley Encyclopedia of Clinical Trials*. New Jersey: John Wiley & Sons, 2007.
 16. Muramatsu T, Nishimura S, Takeishi Y, Nishimura. Prognostic risk stratification of cardiac events evaluated by aortic calcification in elderly patients with chronic kidney disease: sub-analysis of a J-ACCESS 3 study. *Ann Nucl Cardiol* 2019; 5: 35–43.