"Small" sample size

Shengping Yang PhD, Gilbert Berdine MD

I am planning a pre-clinical study to compare the effect of two active drug compounds on a metabolic biomarker. We will be using 8 mice in each group. Is this sample size considered to be small?

Sample size is one of the critical considerations in various types of biomedical research. Whether it is a pre-clinical study, a randomized controlled clinical trial, or an epidemiological investigation, the size of the sample has profound implications for the reliability, validity, and generalizability of study findings. In fact, sample size is not only a key consideration that influences the study's statistical power and precision, but also determines the study's capacity to draw meaningful conclusions and to extend the findings to a broader context. Although the absolute size of a study may differ, the interpretation of a "small" sample size is contingent upon the unique characteristics of distinct study type, outcome measurement, and study design. This article primarily explores what factors potentially contribute to whether a study is considered as "small" in the context of conducting effective biomedical research, rather than delving deep into statistical methodologies for analyzing data with a limited number of observations.

1. Determinates of a "small" sample size

A reasonable sample size for a biomedical study is often determined after considering several factors. While a large sample size is often favorable, depending on the nature of a study, at certain times, a "small" sample size might also be a reasonable option when compromises must be made.

Corresponding author: **Shengping Yang**
Contact Information: Shengping.Yang@pbrc.edu
DOI: 10.12746/swrccc.v11i49.1251

1.1 Pre-clinical vs. clinical studies

Pre-clinical studies are often conducted before clinical trials and involve laboratory-based research using cells, tissues, or animal models. The primary goal of pre-clinical research is to gather preliminary data and understand the safety, efficacy, and mechanisms of action of potential interventions, such as new drugs, therapies, or medical devices. It is noteworthy that while results from pre-clinical studies may not always transition to subsequent clinical trials, they remain an efficient and cost-effective means of obtaining valuable information. Pre-clinical studies are often a prerequisite for clinical investigations.

Compared to randomized controlled clinical trials, pre-clinical trials often feature a "small" sample size, typically preset at 6 to 20 animals per group, and frequently without a formal power calculation. However, this seemingly modest sample size is justifiable for several reasons:

- *No Previous Work*: For instance, pre-clinical studies are designed to explore novel treatments without pre-existing work, and there is often no reference available for power calculation, making a formal power calculation challenging.
- *Genetic Homogeneity and Controlled Environments*: Experimental animals in pre-clinical studies are usually not only genetically homogeneous but are also exposed to similar environments. This controlled setting reduces variability and thus enhances the ability to detect treatment effects.
- *Focus on Large Effect Sizes*: While pre-clinical studies provide preliminary data on the effect, there is often a preference for detecting treatments with large effect sizes to improve the chance of real effectiveness in subsequent

clinical trials. This emphasis on treatments with a substantial impact justifies the use of a "small" sample size in the pre-clinical phase. There is also a consideration on biological vs. technical replicates,[1] however, this is beyond the scope of this article.

Therefore, in the proposed study, 8 mice per group is often not considered as a "small" sample size. Conversely, subjects in a clinical study are often more heterogeneous, depending on the inclusion/exclusion criteria, and more subjects are often expected. Besides, epidemiological studies are designed to investigate health-related phenomena at the population level and are typically characterized by their large sample sizes and observational nature.

1.2 Outcome measurement and statistical power

While statistical power requirement has a crucial role in determining the sample size of a study, the interpretation of a "small" sample size varies across different outcome measurements. For instance:

- *Continuous Outcome*: In the case of a continuous outcome, given the same effect size, the statistical power is determined by the number of subjects included in a study. If this total number is small, then given the same effect size, the study is deemed "small" indicating lower statistical power.
- *Time-to-Event Outcome*: Conversely, for a time-to-event (survival) outcome, statistical power is influenced by the number of events rather than by the total number of subjects. In other words, the ability to detect differences or effects in survival analysis depends primarily on having a sufficient number of events.[2,3] If the event rate is low, even a study with a large total number of subjects may lack the power to detect differences in survival times. For example, if no subjects have an event, it becomes impossible to identify factors associated with that event in a study. Therefore, the number of events might be more pertinent to determine if a study is "small." Additionally, in the case of a binary outcome, the statistical power, and thus whether a study is considered "small," is also influenced by the number of events, assuming the same effect size.

1.3 Different types of study

The determination of sample size holds distinctive considerations for studies of different natures, such as pilot studies compared to confirmative studies.

- *Pilot Studies*: The emphasis is often on feasibility, exploration, and obtaining initial insights into the research question. As such, sample sizes are typically "smaller", and the primary goal is not statistical significance but rather refining study procedures and assessing the practicality of the research design.
- *Confirmative Studies*: Aiming for robust statistical evidence, demand larger sample sizes calculated based on power analysis to detect meaningful effects. These studies prioritize precision, reliability, and the ability to draw definitive conclusions.

Striking the right balance in sample size is essential, ensuring that pilot studies provide valuable groundwork for confirmative studies while the latter possess the statistical power needed for meaningful and generalizable results. Therefore, depending on the goal of a study, the same sample size can be considered as "small" for a confirmatory trial, or acceptable for a pilot study.

2. Study design

At times, what might be regarded as a "small" study within a specific design framework could be deemed acceptable when transitioning to a different design with the same number of subjects. For example, a crossover design is often considered more efficient than a parallel design due to several factors.[4]

- *Within-Subject Comparison*: In a crossover design, each participant serves as his/her own control, receiving all available treatments in a randomized order. This within-subject

comparison reduces variability attributed to individual differences, leading to increased precision and statistical power. In addition, crossover designs inherently control variability between subjects, as each subject undergoes all treatments. This can be particularly advantageous in situations in which there is significant variability in responses among individuals.

- *Balanced Treatment Allocation*: For the same reason, the treatment assignments are seldom imbalanced in crossover design, which is also associated with improved efficiency of a study.

Therefore, a crossover design with the same number of subjects might not be considered as "small" compared to a corresponding parallel design. On the other hand, it is important to note that crossover designs have limitations, such as the potential carryover effects and complexity in implementations.

As another comparison, there are differences in sample size considerations between cohort and case-control study designs:[5]

- *Cohort Study*: Efficient for studying rare exposures but may require a large sample size for rare outcomes.
- *Case-Control Study*: Efficient for studying rare outcomes but can be less efficient for rare exposures.

Therefore, in studies related to a rare event/disease, a cohort study with a large number of participants might be deemed "small" due to the scarcity of cases. Conversely, a case-control study with the same total number of subjects could be considered a respectable sample size. It is crucial to note that there are other distinctions between a cohort study and a case-control study, such as duration, cost, selection biases, and confounding, which, though pertinent, are beyond the scope of this article.

Now, although our focus is not comparing statistical methodologies, it is important to understand that statistical analysis can be more challenging for studies with a "small" sample size.

3. Statistical analysis for "small" sample sizes

3.1 Continuous outcome

A small sample size makes it difficult to examine the assumptions for statistical modeling, including, data distribution, equal variance, etc. Should a non-parametric method be used, it might result in further reduced statistical power and in limited choices of available modeling options.

3.2 Categorical outcome

Statistical power is often lower, with limited data analysis options, if the outcome is categorical. While a Chi-squared test is commonly used, a Fisher's exact test[6] is often more appropriate for a "small" sample size. Nevertheless, interpreting the results from some other models can be complex, and there may be instances in which assumptions of the models are violated.

As an alternative, a categorical outcome can be converted into a binary outcome if not already binary and analyzed using a logistic regression. Additionally, for situations with small sample sizes, options such as the *firth* option in SAS *Proc Logistic* can be considered.[7]

3.3 Bayesian analysis

Bayesian analysis facilitates the integration of prior knowledge or information concerning the parameters under estimation. This capability is particularly valuable when working with small sample sizes, as it enables the utilization of existing knowledge to enhance parameter estimation. By incorporating prior information, Bayesian methods offer a means to supplement limited data and provide more informed and robust inferences. However, there might be a learning curve in applying Bayesian analysis for data with a small sample size.

Small sample sizes can introduce several limitations, including limited generalizability, the risk of random variability, lack of precision and reliability, and limited exploration of heterogeneity. It is important to be aware of these limitations when interpreting and generalizing study findings. On the other hand,

whether a study is considered as "small" depends on many factors other than its absolute number of subjects, including the nature, type and outcome of a study, the study design, etc. For data with a "small" sample size, there might be limited options for data analysis, and caution needs to be taken to ensure the validity of the analytical methods used.

Article citation: Yang S, Berdine G. Small sample size. The Southwest Respiratory and Critical Care Chronicles 2023;11(49):52–55
From: Department of Biostatistics (SY), Pennington Biomedical Research Center, Baton Rouge, LA; Department of Internal Medicine (GB), Texas Tech University Health Sciences Center, Lubbock, Texas
Submitted: 10/3/2023
Accepted: 10/8/2023
Conflicts of interest: none

Reference

1. NIH Rigor and Reproducibility Training Module 3: Biological and Technical Replicates (last access: October 1, 2023).
2. Hsieh FY, Lavori PW. Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. Control Clin Trials 2000;21(6):552–60. doi: 10.1016/s0197-2456(00)00104-5.
3. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. Biometrics 1983;39:499–503.
4. Lim CY, In J. Considerations for crossover design in clinical study. Korean J Anesthesiol. 2021 Aug;74(4):293–299. doi: 10.4097/kja.21165.
5. Song JW, Chung KC. Observational studies: cohort and case-control studies. Plast Reconstr Surg. 2010 Dec;126(6):2234–2242. doi: 10.1097/PRS.0b013e3181f44abc.
6. Fisher RA. On the interpretation of X^2 from contingency tables, and the calculation of P. Journal of the Royal Statistical Society 1922;85(1):87–94. doi: 10.2307/2340521.
7. Firth D. Bias reduction of maximum likelihood estimates. Biometrika 1993;80:27–38.