

Sampling Methods: A guide for researchers

Ann Eshenaur Spolarich, RDH, PhD, FSCDH

ABSTRACT

Sampling is a critical element of research design. Different methods can be used for sample selection to ensure that members of the study population reflect both the source and target populations, including probability and non-probability sampling. Power and sample size are used to determine the number of subjects needed to answer the research question. Characteristics of individuals included in the sample population should be clearly defined to determine eligibility for study participation and improve power. Sample selection methods differ based on study design. The purpose of this short report is to review common sampling considerations and related errors.

Keywords research design, sampling, sample size
NDHRA priority area, **Professional development: Education** (evaluation).

Submitted for publication: 7/1/2023; accepted 7/27/2023

BACKGROUND

Sampling is a critical element of research design. Non-experimental designs (e.g., observational studies) typically involve one group of subjects. Quasi-experimental designs include two groups, one of which is a comparison group, but subjects are not randomly assigned. Without randomization, it is possible that subjects in both groups may differ in measured and/or unmeasured variables that can influence the results of the study. Experimental designs use randomization to assign subjects to either the treatment (“intervention”) or comparison (“control”) group. The underlying assumption is that the groups are equal on unmeasured variables or characteristics, allowing the investigator to attribute observed effects to the study intervention.¹

An investigator should carefully think about who their study participants should be, including how to source those individuals very early in the study design process. Participants should be representative of the **target population**, or the group of people to which the study results will apply. To further refine subject selection, the **source population**, or the subset of representative people, must be well-described. An investigator should

consider the characteristics of the population they wish to study, using specific criteria to identify eligibility to participate. Eligibility is based on these **inclusion criteria** as well as **exclusion criteria** that eliminate potential study candidates who do not possess the desired characteristics. From the source population, the investigator selects the **sample population**, which includes the people who are asked to participate in the study. The actual **study population** consists of eligible members from the sample population who consent to participate in the study.²

When the number of individuals in the source population is small, an investigator may choose to invite all individuals to participate. In this case, the source population and the sample population are the same. For example, when an investigator wants to document a craniofacial malformation in patients with a rare genetic syndrome, any individuals who present with the syndrome should be considered for eligibility, as the number of potential study candidates is limited. For most studies, the size of the source population is much larger than what is needed for the sample population, so only some members of the source population will be needed as participants.²

Sampling Methods

There are numerous ways to select members from the source population to comprise the sample population. Sampling methods are described as either *probability sampling* or *non-probability sampling*. Probability sampling methods are based on chance for selection and include:^{2,3}

- **Random sampling**, where chance for selection is equal for all members of the source population;
- **Systematic sampling**, where the investigator selects a random starting point after which, every “nth” person is chosen from a list of members of the source population;
- **Stratified sampling**, where the source population is divided into different sections or “strata” and then, people are randomly selected from each strata; and
- **Cluster sampling**, where the source population is divided into geographic sections and different sections are randomly selected for inclusion.^{2,3}

The sample population is most likely to be representative of the source and target populations if it is randomly selected.³

Non-probability sampling methods are used when investigators choose specific populations based on availability, ease of access or specific characteristics.⁴ Often these methods are more cost-effective and time-efficient to maximize limited resources. These sampling methods may be used for both quantitative and qualitative research studies. A limitation is that not every potential study candidate will have an equal chance of being selected, posing risk for *non-random sampling bias*, a form of selection bias.^{2,5} There is also a risk for *ascertainment bias*, meaning that because there is a higher likelihood that members of the sample population will not be representative of the source and/or target population, results will not be generalizable.⁶ Lack of generalizability negatively impacts the external validity of the study.⁵ Types of non-probability sampling methods include:⁷

- **Convenience sampling**, where subjects are chosen based on availability. There are two subtypes of convenience sampling:
 - **Consecutive sampling**, where subjects are enrolled based on availability. Each subject completes the research, one after the other, until a conclusion is reached. Results are analyzed after each subject before moving on to the next.
 - **Self-selection**, also known as **Volunteer sampling**, where volunteers are recruited until the desired sample size has been reached.
- **Quota sampling**, where people are divided into strata with defined characteristics and then are selected until a predetermined number of subjects representing each strata has been filled. Quotas may be:
 - **Proportional**, where selected subjects represent a broad population (so characteristics of the source and/or target population must be known), or
 - **Nonproportional**, where only a minimum sample size is selected from each strata.
- **Snowball sampling**, where the investigator selects a small sample of participants, who then go on to recruit other participants until the sample size is reached. Snowball sampling is often used with recruitment via social media when investigators are looking for specific people with a defined problem, experience or characteristic who may be hard to locate and/or when the purpose of the study is very sensitive (e.g., a survey to understand involvement in illegal activities or a study that involves individuals who might be put at risk for harm if their identities or their participation are known);
- **Purposive sampling**, where the investigator seeks individuals who meet a specific set of criteria to determine who should comprise the sample based on the purpose of the study;

subjects are then selected based on these criteria. The subtypes of purposive sampling include:

- **Heterogeneity sampling**, where the investigator recruits subjects with a wide range of knowledge, attitudes, opinions or beliefs as related to the research topic;
- **Homogeneity sampling**, where subjects who possess shared knowledge or viewpoints are selected;
- **Deviant or Extreme sampling**, where subjects are selected who possess unique or unusual characteristics. The investigator must often work hard to identify these individuals; and
- **Expert sampling**, where subjects are selected who possess specialized knowledge or experience related to the research topic.⁷

The sampling method used is based on the research question.⁴ Regardless of the sampling method chosen, the investigator's goal remains the same: to diligently work to create a sample that is representative of the source population, and if possible, the target population as well.²

To determine the number of people needed for the sample, known as the *sample size*, a *power analysis* is performed. A power analysis calculates the minimum sample size needed for:

- a desired power level (i.e., the likelihood that a statistical test will detect an effect of a certain size if there is one),
- significance level (i.e., the alpha level, which is Type I error probability), and
- an expected effect size (i.e., the magnitude of the difference of the expected study result between groups or the relationship between the variables).

Power and sample size are used to determine the number of subjects needed to answer the research question (null hypothesis)⁸ For many studies, the

significance level is set to 5% and the desired power level to 80%. Variability among the characteristics of the sample population can affect study power. Large variability will reduce power, while having a well-defined population will improve power.⁹ Determining the appropriate sample size is necessary to detect clinically relevant differences.¹⁰

Sample Selection and Study Design

Sample selection differs by study design.^{1,2} Cross-sectional designs gather a snapshot of data from one timepoint in a reference population. The sample population must be representative of the source population, which in turn, must be representative of the target population. This design is used for studies that describe a specific target population.² The National Health and Nutrition Examination Survey (NHANES) is a longitudinal, cross-sectional study that assesses health and nutritional status of adults and children in the United States. Probability sampling is used to gather a nationally representative sample of approximately 5,000 people from 15 counties. A new sample is selected from 15 different counties each year.¹¹

Case-control designs include subjects who are specifically chosen because they have a certain disease or health condition that is the focus of the study. Investigators need to establish a *case definition* that clearly defines the characteristics that individuals must possess for eligibility as a "case" for the study. Case definitions include both inclusion and exclusion criteria to ensure that selected individuals are representative of the disease or health condition under study. Cases are then compared to controls who must be free of the disease or health condition. Investigators must use care to ensure that cases have similar stages of disease (e.g., mild versus severe active periodontitis). Both cases and controls should be drawn from populations with similar demographics, such as age, sex, and geographic location. Case-control designs are either cross-sectional or retrospective.²

Cohort studies involve groups of people who share common characteristics and are selected based on exposure status. For example, a prospective cohort

study includes a group of subjects who are selected because of exposure to a risk factor or disease, and a control group of subjects who are not exposed. A variable of interest is studied across time and compared between the groups. Certain characteristics of subjects may be “matched” in both groups, such as age or disease status, or unmatched. Processes involved in identifying potential study candidates who were exposed and not exposed mimic the selection process for cases and controls in case-control designs. A *retrospective cohort study* looks at existing data among a particular group or groups of individuals who share defined characteristics, such as risk factor exposure or a disease state, at a specified timepoint. The investigator then examines what happened to the subjects from that point forward up until the present time.^{1,2}

Experimental designs require sampling from a source population that is representative of the target population.^{1,2} Typically, subjects are randomly assigned to a treatment (intervention) group or a control group. Investigators use well-defined inclusion and exclusion criteria to determine eligibility for participation. Other important considerations with this design include the risk for harm associated with the intervention, requiring careful screening to ensure that subjects are healthy enough to engage in or tolerate the intervention, unknown adverse events associated with the intervention, and the ethics of withholding the intervention from subjects in the control group. Randomized controlled clinical trials (RCTs) are considered the gold standard design for determining causal relationships.

Arguably, the tight design parameters of RCTs can make it challenging to replicate the findings in clinical practice, which may make practitioners question the applicability and generalizability of the results to the broad population. There are often differences in characteristics between study subjects and patients who are treated in the clinical setting, as subjects are selected based on a well-defined set of criteria. The setting where the RCT was conducted may also differ from a clinical practice setting to which results would be translated. It is not always easy to apply findings from RCTs into the clinical

practice setting. One review reported that clinicians should have confidence in generalizing the results from well-conducted clinical trials to their patients, especially from RCTs that have large sample sizes, test a simple intervention, and demonstrate good evidence of the benefit of the intervention.¹²

CONCLUSIONS

Proper sample selection is an essential element of study design. Investigators should develop well-defined inclusion and exclusion criteria to ensure proper subject selection for eligibility for study participation. Power analysis should be performed to determine the appropriate sample size needed. Smaller sample sizes increase the risk that the result obtained is due to chance, and negatively affect internal and external validity. Sample sizes that are too large may result in statistical significance which may overinflate the clinical relevance of the findings. Investigators should use their best efforts to ensure that the selected sample population is representative of the source and target populations to ensure generalizability of study findings.

Ann Eshenaur Spolarich, RDH, PhD, FSCDH

Arizona School of Dentistry & Oral Health
A.T. Still University
Mesa, AZ, USA

Corresponding author:

Ann Eshenaur Spolarich, RDH, PhD, FSCDH;
aspolarich@atsu.edu

REFERENCES

1. Brunette DM. Research strategies and threats to validity. In: Brunette DM. Critical thinking: Understanding and evaluating dental research. 2nd ed. Chicago (IL): Quintessence Publishing Co., Inc.: 2007. p.185-96.
2. Jacobsen KH. Primary studies: selecting a sample population. In: Jacobsen KH. Introduction to health research methods: a practice guide. Sudbury (MA): Jones & Bartlett Learning:2012. p. 105-15
3. Andrade C. The inconvenient truth about convenience and purposive samples. Indian J Psychol Med. 2021 Jan;43(1):86-88.

4. Setia MS. Methodology series module 5: sampling strategies. *Indian J Dermatol*. 2016 Sep-Oct;61(5):505-9.
5. Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiol Community Health*. 2004 Aug;58(8):635-41.
6. Guillot G, Foll M. Correcting for ascertainment bias in the inference of population structure. *Bioinformatics*. Feb 2009;25(4):552-54.
7. Qualtrics. What is non-probability sampling? Everything you need to know [Internet]. Provo: QualtricsSM; 2023 [cited 2023 Jul 25]. Available from: <https://www.qualtrics.com/experience-management/research/non-probability-sampling/>
8. Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. *Emerg Med J*. 2003 Sep;20(5):453-8.
9. Bhandari P. Statistical power and why it matters. A simple introduction [Internet]. Amsterdam (NL): Scribbr; 2023 [cited 2023 Jul 25]. Available from: <https://www.scribbr.com/statistics/statistical-power/>
10. Faber J, Fonseca LM. How sample size influences research outcomes. *Dental Press J Orthod*. 2014 Jul-Aug;19(4):27-9.
11. CDC. National Center for Health Statistics. About the National Health and Nutrition Examination Survey [Internet]. Atlanta: Centers for Disease Control and Prevention; 2023 [cited 2023 Jul 25] Available from: https://www.cdc.gov/nchs/nhanes/about_nhanes.htm
12. Flather M, Delahunty N, Collinson J. Generalizing results of randomized trials to clinical practice: reliability and cautions. *Clin Trials*. 2006;3(6):508-512.