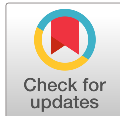

Headache
© 2019 American Headache Society

ISSN 0017-8748
doi: 10.1111/head.13707
Published by Wiley Periodicals, Inc.

Guest Editorials

Sampling Methods in Research Design

Sampling is the selection of a subset of the population of interest in a research study. In the vast majority of research endeavors, the participation of an entire population of interest is not possible, so a smaller group is relied upon for data collection. Sampling from the population is often more practical and allows data to be collected faster and at a lower cost than attempting to reach every member of the population. However, because the sample will be used to make inferences about the population, understanding the means by which the data arrived in the database is an important aspect of analyzing and drawing conclusions from that data.

This installment continues the ongoing series of methods and statistics editorials in the Journal.[1-5] This editorial is focused on common ways that data are observed. Several common sampling designs, uses of sampling designs in headache research, and common problems encountered with sampling are described. While the topic of sampling is vast, this serves as a simple introduction to the concepts.

DEFINING THE ISSUE

In many areas of research, statistical estimates are obtained from a sample and used to make inferences, called population parameters, about a population of interest. When researchers wish to design a study, they must define the target population about which they would like to make these inferences (eg, individuals who are diagnosed with migraine headache). Often, precisely defining a target population is a challenging task, but it is of the utmost importance for determining how the research findings will eventually be used.[6] When designing sampling plans, the units of a population are referred to as *elements* and lists of all population elements are called *sampling frames*.[7,8] An example of a sampling frame could be all individuals who seek care for a migraine attack in an emergency department in the United States in the current year. In this case, the study population is the distinctly defined group from which elements (ie, individuals) will be selected and studied. Although other classification systems exist, sampling strategies are often categorized as probability sampling and non-probability sampling (Table 1).

Probability Sampling.—In a probability sample, each element in a sampling frame has a known and nonzero chance of selection (ie, a probability of being sampled), and random selection is used to choose elements.[6,7] Several methods of probability sampling exist. Some of the most commonly used are described here.

Simple Random Sampling.—When simple random sampling is used, all elements have an equal probability of being selected. Because this sampling method gives equal probability to all elements, it is useful when researchers are interested in associations that would apply to the whole population.[6] However, a weakness of this design is that the sample may not reflect the population, especially the population composition of specific groups. For example, when randomly selecting 10% of individuals who present for treatment at a tertiary headache clinic, each individual has an equal chance of being included. Though, due to random sampling error, this sample of individuals may not be representative of older individuals in the general population who by chance were not selected in accordance

Address all correspondence to D.P. Turner, Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA, email: dpturner@mgh.harvard.edu

Accepted for publication October 29, 2019.

Conflict of Interest: None

Table 1.—Examples of Sampling in Headache Research

Sampling Method	Definition	Common Types	Examples in *Headache*
Probability sampling	Each element has a known and nonzero chance of selection Random selection is used based on laws of probability	Simple random Systematic Stratified Cluster	"Stratified sampling methods were designed to yield a final sample of respondents who would be demographically representative of the U.S. population by gender, age, household income, race, marital status, and U.S. Census region."[9] "Analyses were done using SAS procedures … to account for the stratified, multistage probability cluster sampling design of NHANES."[10]
Non-Probability sampling	Elements of the population do not have a known or equal probability of selection Random selection is not used	Convenience Purposive Quota Snowball	"A convenience sample of 395 participants was recruited for the study."[11] "Quota sampling was used from September 2012 to November 2013 to generate a sample of 489,537 panel members, representative of the U.S. population."[12]

with their levels of population membership (ie, this sample is younger than the population as a whole).

Systematic Sampling.—In systematic random sampling, elements are chosen based on a predefined interval. In its simplest form, all elements are listed, a random number is chosen as the starting point, and every kth element after that number is selected.[7] This sampling interval, k, is calculated by dividing the population size by the desired sample size. For example, if the start point is randomly chosen to be element 8 and the sampling interval is calculated as 12, elements included in the sample would be 8, 20, 32, 44, 56, 68, etc. This method ensures that sampling is spread more evenly across the population. However, a disadvantage occurs when the list of elements contains some form of temporal periodicity.[6] If similar elements occur in cycles, these may be selected more often, which would lead to a sample that contains a greater proportion of these elements than actually present in the population. For example, if patients are referred in cycles from a traumatic brain injury clinic, and these referrals coincidentally correspond to the sampling interval, then the sample will contain more of these individuals than if simple random sampling were employed.

Stratified Sampling.—Stratified sampling allows for the inclusion of population subgroups of interest and is useful when there are differences in a variable across different groups.[6] The population is divided into groups, or strata, based on a stratification variable.[7] A simple random sample is then taken from each stratum with sampling fractions used to determine the number chosen from each in relation to population numbers. For example, when studying migraine, an investigator could stratify the population based on sex and then randomly sample within each group. While an advantage of stratified sampling is that it is often useful when variability within strata is small and variability between strata is large, it can be complicated and easily misused. For instance, trouble can arise when elements of the population do not fall into only one distinct subgroup, which can lead to a sample that does not accurately represent the population.

Cluster Sampling.—In cluster sampling, the population is divided into groups, or clusters, and these groups are sampled for inclusion in the study.[7] This method is useful because many times, researchers will be able to identify clusters but will not easily be able to identify individual elements within those clusters.[6] For example, all hospitals in a city can be identified, but all patients within all hospitals are much more difficult to identify. When using this method, clusters are randomly selected from all possible clusters. Within each cluster, the elements are expected to be heterogeneous and representative of the population. In single-stage cluster sampling, all elements in the cluster are included in the research.[8] In two-stage cluster sampling, random

sampling is used to select some elements within each cluster. While cluster sampling has logistic and financial advantages, it can be susceptible to sampling error and biased samples.

Non-Probability Sampling.—In non-probability sampling, elements of the population do not have a known or equal probability of selection. Some elements have no chance of being selected, and some have a probability of selection that is impossible to know. This type of sampling, though limited in generalizability, is still useful for collecting information, especially for exploratory purposes and in qualitative investigation.

Convenience Sampling.—Convenience sampling involves sampling from those elements of the population that are easiest, or most convenient, to access. Such studies may include those who are the first to respond to advertisements or any eligible patients presenting to a clinic. Convenience sampling is common due to the ease in recruitment of participants. However, studies relying on this type of sampling can produce biased estimates of population parameters because of the nature of the individuals included in the sample (ie, lack of generalizability to the population). For example, those most likely to respond to recruitment efforts may not be representative of others who do not respond.[6]

Purposive Sampling.—Purposive sampling is used when researchers wish to target certain individuals with characteristics of interest in the study. While the sample is not likely to be representative of the population, those included in the study can provide a great deal of information on the topic of the specific research question.[6] One use of this method is in the study of rare disorders (eg, a homogeneous sample consisting only of individuals with hemicrania continua). However, researchers should be aware that studies using this technique are prone to bias based on the judgments researchers must make when recruiting participants.

Quota Sampling.—Quota sampling can be described as a non-probability version of stratified sampling. In this method, the population is divided into mutually exclusive subgroups (eg, individuals with migraine or severe headache versus individuals without migraine or severe headache). Then, individuals are selected from each subgroup in proportions that reflect the target population. Based on a 3-month migraine or severe headache prevalence of 15.3%,[13] the quota-driven sample will consist of 15.3% of individuals with a migraine diagnosis and 84.7% of individuals without a diagnosis of migraine or severe headache. The difference from stratified sampling is that this selection process is not random. Often, each subgroup is sampled until enough members are chosen.[6] While this method is easily conducted and inexpensive, selection bias can be a problem because those individuals the researcher selects to include may be the most accessible but may not be representative of the population.

Snowball Sampling.—When snowball sampling is employed, one member of the population introduces the researcher to other members of that population. This technique is useful for rare or hidden populations where members are difficult to identify and contact.[6] As more people are introduced, more data are collected. Although very useful for hard to reach populations (eg, illegal drug users), this method is subject to bias due to the lack of control over recruitment and the fact that those with more connections are more likely to be included.

SAMPLING IN HEADACHE RESEARCH

Headache researchers use a variety of sampling methods in their studies. Much of headache research is observational, which is reflected in the sampling strategies that are most often employed. Based on the availability of treatment-seeking individuals, convenience samples are commonly used, with patients recruited from outpatient clinics, tertiary care centers, and hospitals. Of notable exception, several definitive large survey efforts designed to elicit population-based estimates have used random sampling techniques. Examples of some of these sampling strategies are included in Table 1.

Another consideration when sampling individuals with headache is the episodic nature of most headache attacks. This creates a unique problem for sampling because individuals can be observed either ictally or inter-ictally, and this must be considered when designing the sampling strategy. Headache researchers must often decide both *whom* to sample and *when* to observe or assess these individuals and base their sampling strategy on this unique aspect of the headache experience.

COMMON PROBLEMS IN SAMPLING

While the details of the various sampling methods are not difficult to comprehend, their implementation

may not be as straightforward. Details of these methods and their consequences for interpretation of results are often easy to overlook. Below are common pitfalls and strategies for avoiding them.

1. Failure to Report Sampling Strategy.—All too often, authors report limited information about their sampling methods. When the sampling strategy is not fully reported, readers will not be aware of the precise sampling frame or who was included and how they were chosen for observation. Without this important information, it is not possible to fully understand the study and how the methods and results fit into the existing body of literature. Further, this prevents assessment of the generalizability of the study and of the conclusions. To assist readers in understanding important findings, authors are encouraged to report their sampling methods with a high level of detail.

2. Failure to Appreciate Consequences of Sampling Strategy.—As stated above, different sampling strategies have different strengths and weaknesses. To fully interpret study findings, these attributes must be made explicit and fully considered. Sampling strategies are associated with various types of potential bias, and this must be incorporated into the conclusions to appropriately understand the study findings. If the sampling design systematically prevents the inclusion of some sampling elements (eg, sampling from an adult outpatient neurology clinic may not include pediatric or geriatric populations), this can introduce sampling bias in particular estimates.[7] If those not included have different characteristics than those included, the resulting estimates can be inaccurate, and selection bias will occur when part of the population is not included in the sample or not included as intended.[6] For example, it is extremely common to misinterpret findings from convenience samples (eg, treatment-seeking individuals) as though they necessarily generalize to the overall population. Authors are encouraged to incorporate the strengths and weaknesses of their sampling methods into the conclusions of their studies.

3. Conducting Subgroup Analyses Not Accounted for in Sampling Strategy.—Subgroup analyses are complicated, and interpreting them is often difficult. While some sampling methods are designed to accommodate subgroups (eg, stratified), others are not (eg, simple random, convenience). In some cases, subgroups are developed based on ad hoc criteria after data collection. For example, if a subgroup of individuals with positive imaging findings is extracted from a larger sample obtained through convenience sampling, any observed prevalence rates (eg, allodynia rates) in the subgroup cannot be thought to be representative of the population of that subgroup. Stated differently, analysis of subgroups derived from samples that do not reflect subgroups within the population may produce misleading results. When analyzing data, especially in secondary analyses, acknowledging the capabilities of the sampling strategy through which the data were collected will help prevent biased conclusions.

4. Failure to Consider Generalizability.—As described in previous sections, the nature of the sampling strategy is associated with the generalizability of the results obtained from a sample. Probability samples are more generalizable than non-probability samples, which means the results can be readily applied to the greater population. This must be taken into consideration when interpreting data and forming conclusions. Shrewd researchers will be aware of the generalizability of their findings and report their work within that context.

5. Ignoring or Misusing Survey Weights.—In many survey studies and publicly available databases, survey or sampling weights have been derived from the sampling plan used in data collection. These survey weights, when applied properly, can be used to create estimates that represent the population. Weights are intended to reflect the probability that an element was included in the sample, and these weights are applied to calculate the point estimates and their variances.[6] When such weighting is ignored or misused, the results do not accurately reflect the population, and conclusions also may be inaccurate. For example, the National Health Interview Survey (NHIS) contains weights that can be used to properly estimate the prevalence of migraine or severe headache in the past 3 months for individuals in the United States.[13] However, the yearly sample contains only thousands of individuals whose responses must be weighted to produce the population estimates that reflect the millions of individuals in the United States. When working with such databases, investigators are encouraged to consult with a statistician to ensure that survey weights are properly applied.

CONCLUSIONS AND RECOMMENDATIONS

An understanding of sampling design and its consequences is essential for conducting and evaluating most types of research. While the sampling methods described in this editorial are not applicable to some types of research designs (eg, controlled laboratory experiments, meta-analyses), they are widely applicable to any design that requires a researcher to purposely select a sample from a larger population. When designing such studies, researchers are encouraged to base their sampling methods on these considerations and interpret the results of their completed studies in line with the sampling methods used. Likewise, readers and reviewers of studies should be aware of these concepts when evaluating new research. This editorial provides a basic introduction to the concepts of sampling. Those interested in implementing any of these sampling strategies in their more complex forms are encouraged to consult additional resources.[6-8]

Dana P. Turner, MSPH, PhD
Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

REFERENCES

1. Turner DP. P-hacking in headache research. *Headache*. 2018;58:196-198.
2. Turner DP, Houle TT. The importance of statistical power calculations. *Headache*. 2018;58:1187-1191.
3. Turner DP, Houle TT. Conducting and reporting descriptive statistics. *Headache*. 2019;59:300-305.
4. Turner DP, Houle TT. Assessing and interpreting reliability. *Headache*. 2019;59:653-658.
5. Turner DP, Houle TT. Observational study designs. *Headache*. 2019;59:981-987.
6. Lohr SL. *Sampling: Design and Analysis*, 2nd edn. Boston, MA: Brooks/Cole Cengage Learning; 2009.
7. Groves RM, Fowler FJ Jr, Couper MP, Lepkowski JM, Singer E, Tourangeau R. *Survey Methodology*, 2nd edn. Hoboken, NJ: John Wiley & Sons; 2011.
8. Thompson SK. *Sampling*. Hoboken, NJ: John Wiley & Sons; 2012.
9. Lipton RB, Munjal S, Alam A, et al. Migraine in America Symptoms and Treatment (MAST) study: Baseline study methods, treatment patterns, and gender differences. *Headache*. 2018;58:1408-1426.
10. Pogoda JM, Gross NB, Arakaki X, Fonteh AN, Cowan RP, Harrington MG. Severe headache or migraine history is inversely correlated with dietary sodium intake: NHANES 1999–2004. *Headache*. 2016;56:688-698.
11. Kubik SU, Martin PR. The headache triggers sensitivity and avoidance questionnaire: Establishing the psychometric properties of the questionnaire. *Headache*. 2017;57:236-254.
12. Buse DC, Rains JC, Pavlovic JM, et al. Sleep disorders among people with migraine: Results from the chronic migraine epidemiology and outcomes (CaMEO) study. *Headache*. 2019;59:32-45.
13. Burch R, Rizzoli P, Loder E. The prevalence and impact of migraine and severe headache in the United States: Figures and trends from government health studies. *Headache*. 2018;58:496-505.