JOURNAL OF CLINICAL RESEARCH BEST PRACTICES

Vol. 10, No. 9, September 2014 "Happy Trials to You"

Making Sense of Biostatistics: Types of Probability Sampling

By Kathleen Mathieson

Sampling methods have important implications for the quality of research results.[1] Ideally, a clinical research study could draw on a database of all potentially eligible patients. However, clinical studies are usually limited to *interested* patients drawn from site populations and advertising programs, a small fraction of the total population that may, further, be biased in a variety of ways. Nevertheless, this column will discuss three variations on the ideal case as a baseline for the next column, which will deal with the real world of clinical research.

Probability sampling is considered the best method for reducing bias and increasing generalizability because it involves a process of random selection, in which each patient in the population has an equal probability of being selected.[2,3] Therefore, probability sampling techniques help ensure that important characteristics of the target population are accurately represented in the study sample.

The main types of probability sampling are simple random sampling, systematic sampling, and stratified random sampling.

Simple random sampling is the most basic method of probability sampling. It involves random selection from a comprehensive list of eligible subjects.[2] For example, investigators might compile a list of patients who meet study eligibility criteria from a breast cancer registry. A random number table or other method of generating random numbers (e.g., http://www.random.org) can then be used to select the desired number of patients for study recruitment from the compiled list.

The most common barrier to conducting a simple random sample is lack of availability of a comprehensive list of eligible patients that is arranged in a pre-numbered fashion. When such a list is unavailable, *systematic sampling*, a minor variation, may be a useful alternative. This sampling method involves selecting every Nth patient from the list. Unlike a simple random sample, this method does not require that the list be pre-numbered. For example, if investigators wish to select a sample of 100 and there are 1,000 total patients on the compiled list, every 10th patient would be selected. Systematic sampling is considered equivalent to simple random sampling "as long as no recurring pattern or particular order exists in the listing."[3]

Although simple random sampling and systematic sampling should yield unbiased samples, they may, just by chance, result in samples that over- or under-represent population characteristics. When it is important to have representation of specific characteristics in a study, *stratified random sampling* is an effective alternative. In this method, the population is divided into strata (groups) according to the characteristic(s) of interest. For example, if investigators wish to ensure representation of both younger and older patients in a trial, they may divide the population into younger and older strata based on specific age criteria. Then, a random sample is drawn from each of the strata.

In stratified random sampling, whether the strata are proportional to the distribution of characteristics in the population depends on the study purpose. For example, the registry of breast cancer patients noted above may be comprised of 200 younger patients and 800 older patients. To achieve a proportional sample of n=100, the investigator would divide the list into older and younger patients and randomly select 20 from the younger and 80 from the older group. However, if the investigator wishes to ensure statistical power to compare the age groups, she could use disproportional sampling to select an equal number (n=50) from each group. Because this method would result in a disproportional number of younger

women compared to the population, the data would need to be weighted appropriately during analysis.[3]

References

1. Field, L, Pruchno, RA, Bewley, JA, Lemay Jr., EP, Levinsky, NG. Using Probability Versus Nonprobability Sampling to Identify Hard-to-Access Participants for Health-Related Research: Costs and Contrasts. Journal of Aging Health. 2006; 18(4). DOI: 10.1177/0898264306291420
2. Mathieson, K. Making Sense of Biostatistics Probability Versus Nonprobability Sampling. Journal of Clinical Research Best Practices. 2014; 10(8). Available from https://firstclinical.com/journal/
3. Portney, LG, Watkins MP. Foundations of Clinical Research: Applications to Practice (Third Edition). Upper Saddle River, NJ: Pearson Prentice Hall; 2009.

Author

Kathleen Mathieson, PhD, CIP, is an Associate Professor in the Doctor of Health Sciences Program at A.T. Still University (ATSU) in Mesa, Arizona. She is also a Vice-Chair of the ATSU-Arizona Institutional Review Board and a Still Research Institute Scientist. Contact her at 1.602.573.6547 or kmathieson@atsu.edu.