Edward A. Panacek, MD, MPH, and Cheryl Bagley Thompson, PhD, RN

Sampling Methods: Selecting Your Subjects

Editors' note: This article is the fifth in a multipart series designed to improve the knowledge base of readers, particularly novices, in the area of clinical research. A better understanding of these principles should help in reading and understanding the application of published studies. It should also help those involved in beginning their own research projects.

After a basic research design has been selected for a proposed study, a number of details must be determined before the study is initiated. This involves developing the entire study protocol (i.e., fleshing out the protocol). This article in the Basics of Research series discusses decisions related to sampling and enrolling and randomizing study subjects. Because this area can be complex, definitions of some of the new terms used within the paper are given in Table 1.

Defining the Study Population

The first step in developing a full study protocol is having a clear understanding of the research question to be answered. The next step is to have explicit definitions of the independent and dependent variables of interest in the study. Then the target population can be defined (i.e., who will qualify for the study?). This process involves generation of a list of inclusion and exclusion criteria for potential study subjects (Table 2).

Inclusion criteria define the type of subjects that fulfill the needs of the researcher for the study. Common inclusion criteria include demographic parameters, clinical characteristics, geographic considerations, and the temporal setting. Demographic parameters help to ensure a degree of homogeneity in the sample. For example, when studying the effect of surfactants on neonatal respiratory distress, an upper age limit will be necessary as part of the definition of a neonate. Clinical characteristics help to narrow the sample to subjects appropriate to the study. For example, subjects with mild asthma may not be good candidates for a study on the effect of a new drug on asthma hospitalization rates. Geographic considerations may help to limit subjects to an area accessible to the researchers or to ensure geographic diversity. Temporal setting may be important in a number of ways. For example, sleep-research subjects may need to be available in the evening, or a surgical intensive care unit (SICU) study could specify that patients be at least 24 hours after surgery. Subjects can also can be stratified in the enrollment phase (or the analysis phase) based on temporal factors. For example, patients whose asthma symptoms lasted less than 24 hours could be in one group, and individuals whose symptoms lasted more than 24 hours could be in another group. Finally, temporal (time frame) requirements could be part of the randomization plan. For example, only patients seen during the morning hours of the month might be included in a study that requires several hours of monitoring while study personnel are present.

A final consideration for inclusion criteria is informed consent. Ethics will be discussed in further detail in a future segment of the series. However, a common inclusion criterion is that subjects must be able to provide verbal or written consent to be eligible for the study.

Exclusion criteria are as important as inclusion criteria because they help to predict and/or to eliminate potential study problems. Potential confounding variables commonly are used as exclusion criteria. For example, if patients taking digoxin are known to react differently to the new medication being studied, all patients on digoxin could be excluded from the study. Exclusion criteria can also can help to facilitate the research process. Subjects who may provide poor quality data or who are difficult to recruit into or keep in the study can create problems. Exclusion criteria often are written to keep these individuals out of the sample. Two common examples are the ability to speak English and the ability to read. Such individuals may not be able to comply with a research protocol and might be excluded from the study. An example of subjects at risk of "lost to follow-up" might be patients transported by air to a facility other than the base hospital. Such patients might be excluded as potential subjects.

Finally, ethical constraints may dictate specific exclusion criteria. Prisoners often are viewed as individuals at particular risk for violation of their personal rights. Because of the risk that prisoners may not feel free to refuse to participate in a study, they are commonly excluded to eliminate potential ethical violations.

Table 1. Definition of Terms

Population—all subjects of interest to the researcher for the study
Sample—the small portion of the population selected for participation in the study
Sampling—the process used for selecting a sample from the population
Simple random sampling—a process in which a sample is selected randomly from the population with each subject having a known and calculable probability of being chosen
Stratified random sampling—a process in which a population is divided into subgroups and a predetermined portion of the sample is randomly drawn from each subgroup
Systematic random sample—a process in which a sample is drawn by systematically selecting every *n*th subject from a list of all subjects in the population. The starting point in the population must be selected randomly.
Cluster sampling—a process in which the sample is selected by randomly choosing smaller and smaller subgroups from the main population
Convenience sampling—a process in which a sample is drawn from conveniently available subjects
Snowball sampling—a process in which the first subjects are drawn by convenience and these subjects then recruit people they know to participate, and they recruit people they know, etc.
Quota sampling—a process in which subjects are selected by convenience until the specified number of subjects for a specific subgroup is reached. At this point, subjects are no longer selected for that subgroup, but recruitment continues for subgroups that have not yet reached their quota of subjects.
Purposive sampling—a process in which subjects are selected by the investigator to meet a specific purpose
Judgmental sampling—another name for purposive sample

Table 2. Defining the Target Population

Inclusion Criteria Goals
Specific and focused on a target population
Demographic parameters
Clinical characteristics
Geographic considerations
Temporal setting
Informed consent ability
Exclusion Criteria Goals
Attempt to predict and eliminate analysis problems
Probable confounding variables
High risk of "lost to follow-up"
Inability to provide good data
Ethical constraints

Inclusion and exclusion criteria should be considered carefully before initiating a study. There are two general approaches. One is to be "inclusive" and enroll a large, heterogeneous population that replicates real life. This is sometimes referred to as an "effectiveness" study, in which the results are broadly generalizable. However, this can put the study at greater risk for confounding variables and difficulty in obtaining an accurate data set.

The second approach is to be "exclusive" and only enroll a tightly controlled, homogenous population. This is often called an "efficacy" study. Such an approach is often used when studying a new drug or therapy, i.e. under "ideal" conditions. However, such strict criteria make enrolling sufficient numbers of subjects more difficult, and the final results are not as applicable to real clinical practice.

Sampling Techniques

After the potential or target study subjects have been defined, the next step is to decide how to enroll subjects in the study. This starts with determining the target subject population. The term *population* refers to all potential subjects for the study. For example, if a researcher is interested in stress levels of health care providers who transport patients by air, all nurses, paramedics, emergency medical technicians (EMTs), physicians, and technicians employed by air transport programs would be included in the population. However, the actual *study population* of interest is usually much narrower. The researcher may wish to investigate only United States air medical personnel or alternately just nurses and paramedics employed by US transport programs. However, in the actual study, the *sample* used for the research project contains only the individual *subjects* who actually will participate in the study.[1] In other words, of the overall population, the sample contains only a small portion of the target population, which is selected for analysis. How this sample is selected from the entire population of subjects can impact the quality of the study. A poorly selected sample may yield biased results that cannot be applied to individuals outside of the sample (i.e., the results do not apply to the entire target population).This is what is referred to as the *external validity* of the study.

There are two main categories of sampling techniques: random and non-random.

Random Sampling Methods

Of all the methods that can be used to select a sample (Table 3), the most powerful sample is one that is selected *randomly* from the population. Random selection means that

Table 3. Sampling Methods

Probability (true random)
Simple random sample
Stratified random sample
Systematic random sample
Cluster sampling
Nonprobability (non-random)
Convenience sample
Snowball sampling
Quota sampling
Purposive sample
Judgmental sampling

every potential subject in the target population has a known and equal probability of being selected for participation. That probability is quantifiable (i.e., it can be calculated). For this reason, these are often called "true-probability" samples.

If there is concern about enrolling sufficient subjects within select subgroups, a population can be divided into major subgroups before random selection is applied. This is called *stratified random sampling*. This ensures that any groups of interest are adequately included in the sample. For example, an investigator may be interested in high school students. The investigator wants to be sure that some of the students are from the special education class. However, if only 2% of the students are in special education in the high school of interest, a simple random sample may not enroll any special education students. Consequently, a stratified random sample may be drawn in which 98% of the subjects are selected randomly from the general student body and 2% of the subjects are selected randomly from the group of special education students. This approach ensures that both groups are included proportionally within the sample.

The researcher also can elect to alter the proportions in the sample from the proportions present in the population. In the example above, the researcher may instead select 90% from the general student body and 10% from the special education students. This approach would provide more information in a subgroup that constitutes a small portion of the population.

Stratification could occur by sex, age, racial group, etc. In this example, although the chance of being selected is not equal for all students, the probability of being selected is equal within each stratum and known for each individual. Thus, the study sample is considered to be a true random sample.

Random selection can be accomplished in a variety of ways. One common way is to draw names from a hat. If the researcher is interested in members of the National EMS Pilots Association, the name of each member is placed on a piece of paper and put into the hat. One slip of paper is drawn for every subject required for the study. Another method is to use a table of random numbers to select individuals from the list of the population. Computer randomization programs are increasingly available on the Internet to help generate random lists.

Another method uses the list of all possible subjects but divides the total number of potential subjects by the number of subjects needed. The answer is used as the interval from which to pick names on the list. For example, if there were 1,000 names on the list and 50 subjects were needed, 1,000 divided by 50 is 20. Consequently, every 20th name from the list would be selected. If the starting point for the selection is determined randomly (e.g., drawing one of the numbers 1 to 20 out of a hat), and the list does not have a pre-established non-random order (e.g., if males and females were listed alternately), this method of sample selection is considered to produce a *systematic random sample*. If the researcher starts at a self-selected name in the first 20 and then picks every 20th person, bias may be present because of the way the first subject is chosen.

Another method of obtaining a random sample is *cluster sampling*. Most commonly used in survey research, this approach may be used in cases in which a list of all subjects in the population is not available or it would be too difficult to randomly select from the total population. Instead of randomly selecting subjects, smaller subgroups of subjects are selected. For example, to select a random sample of nurses employed in emergency departments (ED) of major cities, a list of all states could be created and the desired number of states randomly selected. Next, a list of representative cities in the selected states would be created and a set of cities selected randomly. A list of all hospitals in the selected cities would be created and a sample of hospitals selected randomly. Finally, from that list of hospitals a complete list of ED nurses would be constructed and the final sample randomly drawn. The advantage of this method is that random selection is preserved, at each step, without having to obtain a list of every US nurse employed in an ED. Consequently, study sample selection is not only easier but less expensive. The final study sample may also be more representative of the overall population, because of expected low response rates/non-participation rates common in very large studies.

Non-random Sampling Methods

Unfortunately, true random samples are often difficult to obtain. Often the investigator does not have access to all subjects of interest. Getting accurate lists of all candidate study subjects may be difficult or impossible. Obtaining access to all of the individuals even if their identity is known also may be difficult. Consequently, although non-random sampling techniques are less scientifically valid, they are the type most commonly used for health care research.

Convenience samples are the most common type of non-random samples. As the name suggests, they are subjects who are convenient to the researcher, for one reason or another. In the case of adult trauma patients, a convenience sample could consist of patients transported by the teams participating in the study. A variant of convenience sampling is *snowball sam-*

pling. In this case, the initial subjects identify other individuals who qualify and also may be interested in participating. For example, a sample could be obtained by recruiting air medical personnel at the annual conference who then talk with friends and encourage them to participate.

Quota sampling is similar to stratified random sampling, in that a specific number of subjects from different subgroups are recruited. The difference is that subjects are recruited by convenience rather than randomly. Once the quota for a subgroup is met, subjects are no longer recruited for that subgroup. So if 40 gunshot wounds, 40 abdominal blunt trauma, and 20 head injuries are required for the study, patients with gunshot wounds no longer will be recruited once 40 subjects meeting gunshot criteria have been enrolled, even if the overall study is still active. The advantage of quota sampling is that the researcher can be more specific about the type of subjects desired for the study and can be assured that specific subgroups are represented adequately in the final study sample. As with convenience sampling, bias in the method of selection of subjects for the subgroups still may exist. An additional disadvantage is that the study may be more difficult to complete if subjects from a subgroup are rare or difficult to recruit.

Purposive sampling is even more restrictive and subjective than quota sampling. In this case, the researcher has specific requirements for the sample and picks subjects who meet these strict criteria. For example, the researcher may be interested in the behavior of experts but recognizes that there may be regional differences. The investigator then purposely could select a number of nationally recognized experts in air transport from each of the Association of Air Medical Services regions.

Another case in which purposive sampling could be used is when the sample will be small and 100% cooperation is needed. In such cases, the researchers may ask specific subjects who they know will volunteer and follow through with the study protocol. This approach is also sometimes called *judgmental sampling*, because enrollment depends on the investigator's judgment as to who qualifies. This "judgment" may lead to serious investigator bias in subject selection, which could invalidate the study results.

Because a non-random selection of subjects is much easier to obtain—and sometimes the only way to get subjects—why use more "expensive" random sampling techniques? Random sampling techniques provide a higher quality of research results. First, selecting a sample at random helps to reduce bias from the process of sample selection itself. For example, your transport program may have a different philosophy, may have a different set of protocols, or may just differ in the quality of care provided to patients when compared with other air transport services. As a result, any study performed using subjects "convenient" to your program, or similar programs, may give results that are biased by these factors. As a consequence, the results would be applicable to your program only and would not be generalizable to other programs.

A second reason for using a random sample relates to the statistical analysis of the data at the conclusion of the study. The inferential statistics commonly used in health care research generally include an assumption that the sample under study is truly random. The probability tables used to determine whether your results are different enough to be considered statistically significant were developed using random samples. Consequently, purists may say that your statistical analysis is flawed if your sample was not selected randomly.

What can you do if randomly selecting the sample is not possible? First, the researcher should try to minimize bias in sample selection. Patients should be selected using explicit inclusion/exclusion criteria without first trying to determine whether they will be "good" subjects whose data may be more likely to support the study hypothesis. The final sample, even though convenient, should be as representative of the overall population as possible.

Second, the researcher can try to further diversify the sample by, for example, engaging in multicenter studies, which have a wider range of subjects than do single-site studies.

Even the best selected samples may still be somewhat biased. Thus, in the analysis phase, the researcher must specifically look for bias and institute statistical adjustments, if necessary. For example, in a study comparing drug A with drug B, the effect of an extraneous variable, such as gender, might be causing bias in the results. A statistician can use statistical techniques such as regression or analysis of covariance to adjust for this and other, imbalanced variables.

This issue has discussed methods of identifying a sample for your study. The next installment will address validity and reliability, two other potential sources of bias or inaccuracy in the study findings.

Reference

1. Cohen J. Statistical power analysis for the behavioral science. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

Recommended Texts

Bailey DM. Research for the health professional: a practical guide. 2nd ed. Philadelphia, PA: FA Davis; 1997.

Huley SB, Cummings SR, Browner WS, Grady DG, Hearst N, Newman TB. Designing clinical research: An epidemiologic approach. 2nd ed. Baltimore, MD: Lippincott, Williams and Wilkins; 2001.

Okolo EN. Health research design and methodology. Boca Raton, FL: CRC Press; 1990.

Edward A. Panacek, MD, MPH, is professor of emergency medicine and director of clinical trials at the UC Davis Medical Center in Sacramento, California. Cheryl Bagley Thompson, RN, PhD, is an associate professor and director of the health informatics program at the University of Nebraska Medical Center, College of Nursing in Omaha, Nebraska. She can be reached at cbthompson@unmc.edu.

1067-991X/$32.00

Copyright 2007 Air Medical Journal Associates

doi:10.1016/j.amj.2007.01.001