

Sampling

The following chapter is excerpted from *Designing HIV/AIDS Intervention Studies: An Operations Research Handbook*, Andrew Fisher and James Foreit, 2002, Washington, DC: Population Council. ([More on OR Handbook](#))

Many operations research projects depend on data that have been obtained from samples. The purpose of this chapter is to explain why researchers sample and to describe some basic sampling considerations and strategies. The chapter seeks to enable you to identify situations in which different types of sampling might be appropriate. Although it will prepare you to work with a sampling specialist, it will not make you a sampling specialist.

A sample can be thought of as a model of a larger population. A sample consists of a relatively small number of individuals or other units that are selected from a larger population according to a set of rules. For example, you may **randomly** select 100 PLHA who are part of a total population of 600 PLHA served by the outpatient clinic of Central Hospital. If you have a good model, you may be able to generalize from your sample of PLHA to all PLHA in the outpatient clinic.

Similarly, if you use a sample consisting of all types of women in the country, you may be able to generalize your results to the total population of women. The advantage of studying a sample of cases as opposed to all cases is that the research can be done more quickly, less expensively, and often more accurately than a large **census** (survey of the entire population). In fact, given limited research budgets and typically large population sizes, there is usually no alternative to sampling.

SAMPLE OR CENSUS?

Under what circumstances would you study a sample of population units and under what circumstances might you perform a complete census? If the universe, a term that is defined in the box below, were very small, for example, 20 health centers, one hospital, or 50 peer educators, you would do a census, that is, include all units in your study. You might also do a census of a large universe if the elements were easily and quickly accessible. For example, if the records of all visits to AIDS outpatient clinics were computerized, you might obtain data from all the records rather than from a sample.

There are two general types of samples: **probability (or random)** and **nonprobability samples**. The nature of your operations research study will determine which type of sampling you should use. Large-scale descriptive studies almost always use probability sampling techniques. Intervention studies sometimes use probability sampling but also frequently use nonprobability sampling. Qualitative studies almost always use nonprobability samples.

Probability Sampling

In chapter 7, we discussed experimental and quasi-experimental design techniques for controlling the internal validity of a study. Probability sampling is a technique you can use to maximize external validity or generalizability of the results of the study. Descriptive studies and large operations research intervention studies are frequently designed so that their results are as generalizable as possible.

For example, you might want to evaluate the impact of a national radio campaign on the attitudes of adults toward people living with HIV/AIDS. Such a study would probably involve comparing the results of two sample surveys. The first might be a nationally representative survey of adults conducted before the campaign, and the second a survey of the same group conducted after the campaign. Similarly, if you were studying the effect of an intervention to increase dual protection among married couples, you might introduce the intervention in one district and use a matched district as a control. You would then conduct a survey that was representative of all married couples in the districts and compare the prevalence of dual protection.

SOME BASIC TERMINOLOGY

If information from a person, or subject, is gathered from an interview survey or a questionnaire survey, the subject can be referred to as a **respondent**. A more general term that may be used to describe the unit studied, regardless of the type of study or the type of unit being studied (for example, a person, clinic, village, or patient record) is a **case**.

The **population** or **universe** consists of all the members of a group. The population is also the total collection of units from which you select your sample. A **sample** is the subset of the population that you examine in order to generalize about the total population. **Sampling units** are the **elements** into which a population is divided. For example, if there are 2,000 rural health centers in a country and you select a sample of 285 rural health centers, the sampling unit is the rural health center.

Three factors determine how accurate a sample is as a description of a population:

1. The methods used to select the sample must not bias the sample, that is, the sample must be truly representative of the larger universe. For example, if you wanted a sample of women ages 15–49 and selected only unmarried women who are friends of the interviewers, this would be a biased sample.
2. The characteristics of the sample must be consistent with the characteristics of the population of interest (for example, if the population is unmarried women, the sample must consist of unmarried women).
3. The numerical estimates provided by the sample must accurately represent the true values in the population. For example, if the true number of condom users in a population is 30 percent and your sample estimates that it is between 35 and 45 percent, the sample is not an accurate representation of the population. However, if you estimate that the true number of condom users is between 28 and 32 percent, your sample is a fairly accurate representation of the population.

The essence of probability sampling is that each element of the larger population (that is, each couple, each field worker, or each clinic) has a **known, non-zero probability of being selected**. This is achieved through random selection of units for the sample from a list or sampling frame (see definition below). The **random process** guards against the introduction of bias into the sample by the researcher and against other types of systematic bias. The accuracy of the sampling frame is important for meeting the criterion that the characteristics of the sample be consistent with the characteristics of the total population.

A **sampling frame** is a list of the population from which the sampling units are drawn. In the rural health center example, your sampling frame would be the list of health centers maintained by the Ministry of Health that are located in rural areas. The completeness of a sampling frame is critical to the “representativeness” of a sample chosen from the frame. If the Ministry’s list of clinics were out of date and did not include clinics opened after 1999, the sample would be representative only of clinics opened in 1999 or earlier, not of new clinics. Similarly, the sample frame would also be inaccurate if some health centers that had been closed in 2000 were still included on the list, or if some centers were inadvertently listed more than once.

Finally, the **sample size** must be large enough to deliver the level of accuracy or **precision** required in your estimate of the value in the total population. We will discuss five commonly used probability sampling techniques that prevent bias.

Simple Random Sampling

In simple random sampling, each element of the larger population is assigned a unique number, and a table of random numbers or a lottery technique is used to select elements, one at a time, until the desired sample size is reached. Bias is avoided because the person drawing the sample does not manipulate the lottery or random numbers table to select certain individuals.

Simple random sampling is usually reserved for use with relatively small populations with an easy-to-use sampling frame. For example, if the medical records files (the sampling frame) of 600 outpatients (the universe) are ordered consecutively from 1 to 600, it will be quite easy to draw a simple random sample of 100 outpatients. However, this procedure can be very tedious when drawing large samples.

Systematic Sampling

This is a modification of simple random sampling, which is ordinarily less time-consuming and easier to implement. The estimated number of elements in the larger population is divided by the desired sample size, yielding a **sampling interval** (let us call it n). The sample is then drawn by listing the population elements in an arbitrary order and selecting every n^{th} case, starting with a randomly selected number between 1 and n .

In the text box above, we discussed the problem of sampling rural health centers. In this example, your sampling frame would be a list of rural health centers arranged alphabetically by health center name. If your desired sample size is 285 rural health centers drawn from a universe of 2,000 rural health centers, the sampling interval is $2,000/285 = 7$. You would then choose a randomly selected number between 1 and 7 as your start. If your random number is 3, the first unit selected would be the 3rd rural clinic listed in the sampling frame, the second would be the 10th ($7 + 3$) clinic listed, the third the 17th, and so on until the sampling frame is exhausted. Systematic sampling is useful when the units in your sampling frame are not numbered, when the elements are not numbered serially, or when the sampling frame consists of very long lists.

Stratified Sampling

Populations often consist of **strata**, or groups, that are different from each other and that consist of very different sizes. For example, rural health centers, urban health centers, and hospitals are very different kinds of establishments in most developing countries. Similarly, the proportion of urban and rural residents in a country or of HIV-positive and HIV-negative patients attending prenatal clinics are liable to be very different. To ensure that all relevant strata of the population are represented in your sample, you would use a technique called **stratified sampling**.

Stratification may be used in conjunction with either simple random sampling or systematic sampling. When stratifying, each stratum is treated as a separate population. You would arrange your sample frame by strata, and then draw a random or systematic sample from each. Estimates for each stratum are then combined to produce an estimate for the total population.

You can draw either a **proportionate** or **disproportionate** stratified sample. If it were important that the age distribution of your sample is the same as the age distribution of your population, you would draw a proportionate sample by using the same or **unified sampling** fraction for each group (for example, if your strata were ten-year age groups between ages 15 and 44, you might sample every 100th person aged 15–24, every 100th person aged 25–34, and every 100th person aged 35–44).

Proportionate stratified samples are perhaps the most commonly used type of stratified sampling. However, in HIV/AIDS operations research, you sometimes encounter situations in which strata are so different in size that it is impossible to get a needed minimum sample size. If you use a single sampling fraction, you must draw a disproportionate stratified sample. For example, if your strata are 4,000 rural health centers, 3,000 urban health centers, and 50 hospitals and you want to estimate the proportion of AIDS-related visits in each strata, you would have to use a smaller sampling fraction for hospitals than for health centers.

Cluster Sampling

Cluster sampling is the most commonly used probability sampling technique in the behavioral sciences. Cluster sampling refers to techniques in which samples are selected in two or more stages. Cluster sampling is used when it is not possible to get an adequate sampling frame for the individuals you wish to study, or when a simple random sample technique would result in a list of individuals so dispersed that it would be too costly to visit

each one. The disadvantage of a cluster sample is that it increases sampling error and requires a larger sample size for reliable estimates of population characteristics. If the cost of the larger sample size outweighs the costs associated with unclustered sampling, clustering should not be used.

A **cluster** is a group of sampling units rather than an individual unit. Examples of clusters include all the AIDS patients in a hospital, all the peer educators in a district, all the women in a town, or all the children in a household. You would probably use cluster sampling to study AIDS widows. No list of AIDS widows exists, but you do have a list of households. Your strategy would be to first select a random sample of households. If the clusters contained a small number of individuals—for example, only one or two women of marriageable age per household—then you might interview all of the individual sampling elements included in the cluster. However, if the number of sampling elements per cluster is large (for example, the number of AIDS patients in a hospital), you would select a random sample of elements from within the cluster. This is referred to as **two-stage cluster sampling**.

Multi-stage Cluster Sampling

Sometimes, when populations are extremely complex, it is necessary to go beyond two stages in cluster sampling, a technique referred to as **multi-stage cluster sampling**. For example, if you do not have a list of households for your survey of AIDS widows, you might have to begin with a random sample of villages (called the **primary sampling unit** or **PSU**), and when you arrive at each village, make a list of households (called the **secondary sampling unit**) and draw a random selection of households to visit. When you arrive at a household, you would randomly select a woman to interview, or interview all eligible women. In either case, you would apply a sampling fraction to each village, such as one out of five households or one out of ten eligible women.

Nonprobability Sampling

Nonprobability sampling refers to the selection of a sample that is not based on known probabilities. It is distinguished from probability sampling by the fact that subjective judgments play a role in selecting the sampling elements.

Nonprobability sampling procedures are not valid for obtaining a sample that is truly representative of a larger population. Almost always, nonprobability samples tend to over-select some population elements and under-select others. When the known probabilities of selection are not known, there is no precise way to adjust for such distortions.

Despite these drawbacks, there are many instances in which obtaining a truly representative probability sample may be too difficult or too expensive. In fact, much, if not most, of HIV/AIDS operations research uses some kind of nonprobability sampling. For example, it is usually necessary to use nonprobability samples when studying sex workers and their clients, injecting drug users, and men who have sex with men. The external validity of intervention studies that use nonprobability sampling techniques depends on replication of the study results in different populations.

There are two broad types of nonprobability samples: (1) **convenience** samples, which are selected from whatever cases happen to be available at a given time or place, and (2) **purposive** samples, which consist of units deliberately selected to provide specific information about a population.

An example of a **convenience sample** in HIV/AIDS research might occur when you place an advertisement in a newspaper stating you want to interview men who have sex with men. The men who answered your advertisement would be a convenience sample. Similarly, if you wanted to find out how much condoms cost in pharmacies by visiting the five drugstores nearest to your office, you would be using a convenience sample.

A special type of convenience sample is the **snowball sample**. In this technique, persons who have agreed to be interviewed recommend acquaintances for interviewing. Your convenience sample of men answering an advertisement for men who have sex with men would become a snowball sample if you asked everyone who responded to the original advertisement to recruit other men who had sex with men for your study. It should be noted that internal validity is not compromised in a study that uses a nonprobability sample and then randomly assigns cases to an experimental and control group. However, external validity is compromised. The number of groups for which the results can be considered valid will be smaller than if a probability sample were used. In operations research some studies emphasize internal validity; others emphasize external validity or representativeness. Few are able to emphasize both internal and external validity.

Purposive samples are commonly used in qualitative and experimental operations research studies. If you conduct a focus group to select a name for a condom brand that is especially appealing to adolescent males, you might purposively include in your group only 15- to 19-year-old men. If the brand were to be marketed only in urban pharmacies, the sample might be further restricted to adolescent males who live in urban areas and have purchased a condom in a pharmacy in the past six months. When you conduct experiments, the costs and procedures needed to ensure internal validity often preclude use of a representative sample of the

population. In selecting samples to include in an experiment, you might use a convenience sample (for example, a few clinics that volunteer to participate in the study), or, more often, you might purposively select units because they are either **critical cases** or **typical cases**.

A Ministry of Health might want to determine whether multipurpose, rural community health workers are capable of using the “syndromic” approach to diagnose and treat STIs in men. You might choose a group of health workers whom you think might have difficulty in implementing the new activity (perhaps because of limited literacy or the pressure of other activities). You might make this choice because you reason that if the weakest group of health workers can perform the task successfully, then average and above-average workers will also be able to perform it. Conversely, you might select a group of the most able community health workers to test your intervention, because you believe that if they are unable to implement the intervention successfully, all other community health workers will also be unable to implement the intervention. Either strategy would involve the selection of critical cases. If you wanted to draw a purposive sample of critical cases, you might select a group of workers whose educational level is within the modal range for the entire population of community health workers. The homogeneity in this sample would allow you to draw conclusions about whether or not the “typical worker” would experience problems implementing the intervention.

In contrast to the critical case or typical case approaches, you might want to draw a purposive sample that is somewhat representative of the heterogeneity found among community health workers. This type of sample is called a **quota sample**. In quota sampling you purposively select elements for your sample. There are two broad reasons for quota sampling:

- To ensure that the sample composition is proportionate to the population for variables believed critical to the study. In the case of the health worker experiment, you might draw a sample for which the educational distribution and distribution of years of experience of the workers are the same as for the population of community health workers.
- To obtain a desired sample size. For example, you might decide to visit all patients going to an STI clinic until you have interviews from 400 women and 400 men.

Sample Sizes

Many handbooks contain formulas for estimating sample size because the size of the sample is one of the most important determinants of the accuracy of survey estimates. However, we will not provide formulas for sample size estimation. Formulas differ among sampling strategies (for example, those used in cluster sampling are different from those used in simple random sampling); population size; the type of variable being studied; experimental design, if any; and type of statistical comparison planned. Explaining all of these formulas is beyond the scope of this *Handbook*, and presenting any single formula would be of little relevance to most OR studies. Rather, in the remainder of this chapter we will discuss some of the basic factors that affect sample size, to familiarize the reader with the concepts necessary to work with a sampler or select an appropriate formula from any standard textbook on sampling.

We will begin our discussion of sample size estimation with an important caveat. If your objective is to obtain a probability sample that is representative of a relatively large population, such as in a typical descriptive survey, you need to obtain the assistance of a sampling specialist. A nationally representative

health survey may have sample sizes of 5,000 to 6,000 or more individuals. Sample sizes of this magnitude will allow accurate estimates of several different variables for different subpopulations, but the cost may be hundreds of thousands of dollars.

A sample size appropriate to the needs of a researcher depends on two concepts: **precision** and **confidence level**. Precision is the amount of sampling error that can be tolerated by the researcher. Confidence is the level of certainty that the true value of the variable being studied is captured within the **standard error**, or sampling error. A standard error is simply the difference between the true value of the variable in the population and the estimated value of the variable obtained from the sample.

In calculating sample size, the researcher and program decision maker must first decide how much precision they need in their estimate and how much confidence they need in the result. The greater the precision and confidence required, the larger the sample size needed. For some purposes, an error of ± 10 percent might be tolerable, but for other purposes a standard error greater than 1 percent might not be tolerable. Usually, the degree of precision needed depends on the consequences of accepting a study finding as true when in fact it is not true—in other words, it is an error. If people may die because of an error, a great deal of precision is needed in the estimate. However, if the practical effects of an error are likely to be trivial and easily fixed, less precision may be acceptable.

Another important factor in determining sample size is the amount of resources available for the study. Do you have the resources for your study to afford a 1 percent error, or must you settle for a 5 percent error? While most people would prefer small errors, that requires large samples which, in turn, require large resources. The availability of resources usually determines the upper limit of the sample size used in surveys.

In discussing with a statistician or sample expert how large a sample you will need for your study, it is important to have a fairly good understanding of several key concepts, some of which have been introduced above. One of these concepts is the **standard error**.

The standard error is expressed as a range around a point estimate of a variable in a study. Suppose you interviewed a sample of 200 peer educators and found that 40 percent (the point estimate) had talked to someone about HIV/AIDS that day. It is quite unlikely that exactly 40 percent of your universe of peer educators talked about HIV/AIDS on the day of the interview. It is more likely that the true number is slightly different than your point estimate, by, say, plus or minus 3 percent. The interval extends above and below your point estimate (in this case 40 percent) because half the time the true population value will be below your point estimate, and half the time it will be above your point estimate. Thus, given an error of ± 3 percent, you would say that the number of all peer educators talking about HIV/AIDS on that day is between 37 and 43 percent. If you want greater precision, you need to have a smaller sampling error and therefore a larger sample size. A sample that captures the true population value within ± 3 percent provides a considerably more precise estimate than a sample that captures the true value within an interval of ± 10 percent.

What is somewhat confusing is that this standard error interval is referred to as a **confidence interval**. In contrast, the **confidence level** is the degree of certainty (expressed as a probability) that the researcher has that the true population variable is captured within the confidence interval. Thus, in reporting the result of a survey of HIV prevalence, the researcher might say something like, “The survey estimates HIV prevalence in this region at 15.6 percent. We are 95 percent certain that true regional prevalence is between 13.6 and 17.6

percent,” or more simply, “We can state with 95 percent confidence that HIV prevalence is 15.6 percent, plus or minus 3 percent.”

So far, we have discussed sample size as a way of influencing the precision of an estimate of a single variable or observation. But in operations research intervention studies, researchers are usually interested in comparing two or more observations.

For example, you might ask: Has the proportion of sex workers who used a condom with their last client increased over time? Is the observed difference in increased condom use the result of your educational campaign or is it just due to sampling error? Is the difference between HIV/AIDS knowledge scores of experimental and control groups due to your new teaching approach or is it just chance?

OR studies should determine the sample size needed to detect real differences between variables during the project design phase by using an appropriate sample size formula. These formulas all require minimum information that the researcher must be able to provide the sampler, including the following:

- **The baseline value of the dependent variable.** For example, is the baseline value 10 percent prevalence, \$12 per case treated, or 200 visits per month?
- **The size of the difference between two estimates that you want to find statistically significant.** The smaller the difference between the two estimates, the larger the sample needed. The reason for this is that standard errors must be smaller to detect a small difference than a large difference. For example, attributing a 1 percent change (for example, from 2 to 3 percent) in condom use to an intervention rather than to sampling error implies the need for a confidence interval around each point estimate that is less than ± 0.5 percent (the upper bound of your confi-

dence interval around 2 percent is 2.5 percent; the lower bound of your confidence interval around 3 percent is also 2.5 percent). However, you need a much smaller sample size if you want to attribute a 50 percent difference (for example, from 20 to 30 percent) in condom use to your intervention rather than to sampling error. The confidence intervals only need to be less than ± 20 percent.

- **The significance level.** This refers to the probability that the size of the observed difference between the two variables could have been produced by sampling error or by chance rather than by the intervention. The smaller the significance level, the lower the probability that the result could be the result of chance. Thus, a significance level of .1 means that the probability that the observed difference was produced by chance is 1 in 10. A significance level of .01 means that the probability that the difference was produced by chance is 1 in 100. Traditionally, significance levels are usually set at .05 (or 1 chance in 20). The smaller the probability of finding a difference that is the result of chance or sampling error, the larger the sample size required. Thus, to reduce the probability that the result is due to chance from 1 in 20 to 1 in 100 can, for a given difference, increase required sample sizes from 70 to 90 percent.
- **The confidence level.** This is the probability that the true value is within the specified confidence interval (see above).

Usually, other issues also need to be addressed in the sample size formula. If a cluster sample will be used, it is necessary to adjust sample size upward. If the sample is from a small universe (less than 10,000 units), a multiplier (the finite population correction) can be used to adjust sample size downward.

What To Do: Sampling

1. Decide first whether you want to draw a sample and, if so, whether it should be a probability sample or a nonprobability sample. In making this decision, take into account the objectives of the study, the extent to which the findings need to be representative of a larger population, and such resource factors as cost, time, and personnel.
2. Calculate the size of the sample required for your study. Seek the assistance of a statistician to do this, if possible. The statistician needs to know your estimates of the values of the variables to be tested, the degree of accuracy needed, and so on. It is always better to have a somewhat larger sample than required instead of a smaller one to conduct the research.