Preparing to Make Sampling Choices

In: Sampling Essentials: Practical Guidelines for Making Sampling Choices

By: Johnnie Daniel

Pub. Date: 2012
Access Date: April 9, 2019
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781412952217
Online ISBN: 9781452272047
DOI: https://dx.doi.org/10.4135/9781452272047
Print pages: 1-22

Preparing to Make Sampling Choices

What you will learn in this chapter:

- Milestones in the history of sampling
- Major steps in selecting a sample
- Preparations important for making sampling choices
- Guidelines for preparing to make sampling choices

Introduction

The key to good research is preparation, preparation, and preparation. Hence, the key to making good sampling choices is preparation, preparation, and preparation. **Sampling** may be defined as the selection of a subset of a population for inclusion in a study. If done properly, it can save money, time, and effort, while providing valid, reliable, and useful results. On the other hand, if done poorly, the findings of a study may have little scientific and practical value. In order to increase the likelihood that the findings of a study will have value, preparations should be carried out *before* making sampling choices.

This chapter begins with a brief history of sampling, followed by a description of the major steps in selecting a sample. Preparation is the first of these steps. The preparation should include a careful review of the study's purpose, the nature of the population, the available resources, various research design considerations, and ethical and legal considerations. Guidelines for making these preparations are described in this chapter.

Milestones in the History of Sampling

Although sampling probably has always been part of human history, many of the sampling procedures used today have a relatively short history. Governments have long collected population data for taxation, military purposes, and other objectives. Typically, total enumeration was sought. On the other hand, private pollsters tended to use availability sampling such as straw polling. However, by the end of the 19th century, “scientific” procedures for selecting a sample began to surface. U.S. governmental agencies, in particular, began experimenting with these procedures that later became known as probability sampling. Private pollsters, on the other hand, continued to rely on availability sampling until 1936.

Critical changes in sampling procedures used by private pollsters came about in 1936 and again in 1948. In 1936, the failure of the *Literary Digest* to predict the winner of the presidential election led to a movement away from availability sampling to quota sampling. Using availability sampling of millions of respondents,

the *Literary Digest* was successful in predicting the winner in each U.S. presidential election that was held between 1916 and 1932. However, it failed to do so in 1936. On the other hand, using a new sampling procedure that later came to be known as quota sampling, pollsters George Gallup, Elmo Roper, Paul Cherington, and Richardson Wood were successful in predicting Franklin D. Roosevelt as the winner in that election. This caused pollsters to pay more attention to quota sampling and less attention to availability sampling (Bryson, 1976; Cahalan, 1989; Katz & Cantril, 1937; Squire, 1988). These sampling procedures are described in detail in Chapter 4.

The failure of the *Literary Digest's* prediction of the winner of the 1936 presidential election was primarily due to two factors: coverage bias and nonresponse bias.

- **Coverage bias** is the lack of a one-to-one correspondence between the elements in the target population and the elements encompassed by the respondent selection procedures used in a study. **Sampling frame bias** is the extent to which there are differences between the elements that are listed in the frame and the elements that make up the target population. The sampling frame of the *Literary Digest* consisted of its subscribers, listings in telephone books, and listings of automobile registrants. These lists were not representative of the 1936 voting population.
- **Nonresponse bias** is the extent to which there are significant differences between the respondents and nonrespondents in terms of the variables of interest in a study. The proportion of Republicans among the respondents to the *Literary Digest's* straw poll was higher than the proportion of Republicans among the registered voters at that time. The magazine's straw poll was thereby not representative of voters throughout the country.

After the 1936 election, private pollsters increasingly used quota sampling instead of straw polling in predicting the results of elections. On the other hand, U.S. governmental statisticians and academic statisticians increasingly focused their attention on probability sampling. Probability sampling became a fixture of the U.S. decennial censuses in 1940. However, it took the failure of polls utilizing quota sampling to predict the winner of the 1948 U.S. presidential election to cause private pollsters to adopt this evolving sampling procedure. Using quota sampling, the major polling companies (Gallup, Crossley, and Roper) predicted that Thomas Dewey would beat Harry S. Truman in the presidential election of that year. On the other hand, academic researchers utilizing probability sampling predicted Truman would win. Truman won the election by more than 2 million votes and 114 electoral votes.

The failure of the pollsters using quota sampling in predicting the winner of the 1948 presidential election was due to several factors: basing projections on outdated data, stopping data collection too soon, the impact of interviewer bias, and changes in party identifications that were not factored into the projections.

- *The election projections were based on outdated data.* The quota sampling procedures used by the pollsters were based on the 1940 census, data collected 8 years earlier. As a result of major population changes during those 8 years, 1940 census data did not reflect the 1948 voting population.

- *The pollsters stopped collecting data too soon.* Gallup and Crossley stopped collecting data mid-October. Roper stopped in August. At the time they stopped polling, there were yet many voters undecided. A large proportion of these voters decided to vote for Truman.
- *Interviewer bias also contributed to the problem.* Quota sampling has an inherent problem of interviewer bias. In using this sampling procedure, interviewers have discretion to interview whomever they desire as long as they satisfy quota control requirements of the sampling procedure. As a result, working class voters are more likely to be ignored by interviewers, and in 1948, these voters were more likely to vote Democratic.
- *Changing dynamics of political party identification also contributed to the failure of the pollsters.* The Progressive Party and the Dixiecrat Party also had candidates in the race. The effect of these candidates tended to help the Democratic Party. The Communist Party USA did not run a candidate for president, but endorsed the Progressive Party's candidate. This endorsement deflected anti-communism attacks away from the Democratic Party. Many White southerners left the Democratic Party to support the Dixiecrats. This made the Democratic Party more acceptable to Blacks, and they gave it their support.

The failure of the major polling companies to predict the winner of the 1948 U.S. presidential election motivated them to move away from quota sampling and incorporate probability sampling into their polling procedures. They joined statisticians in the federal government and academia in endorsing probability sampling. Probability sampling became the dominant sampling procedure for estimating population parameters. The major types of probability sampling are described in Chapter 5.

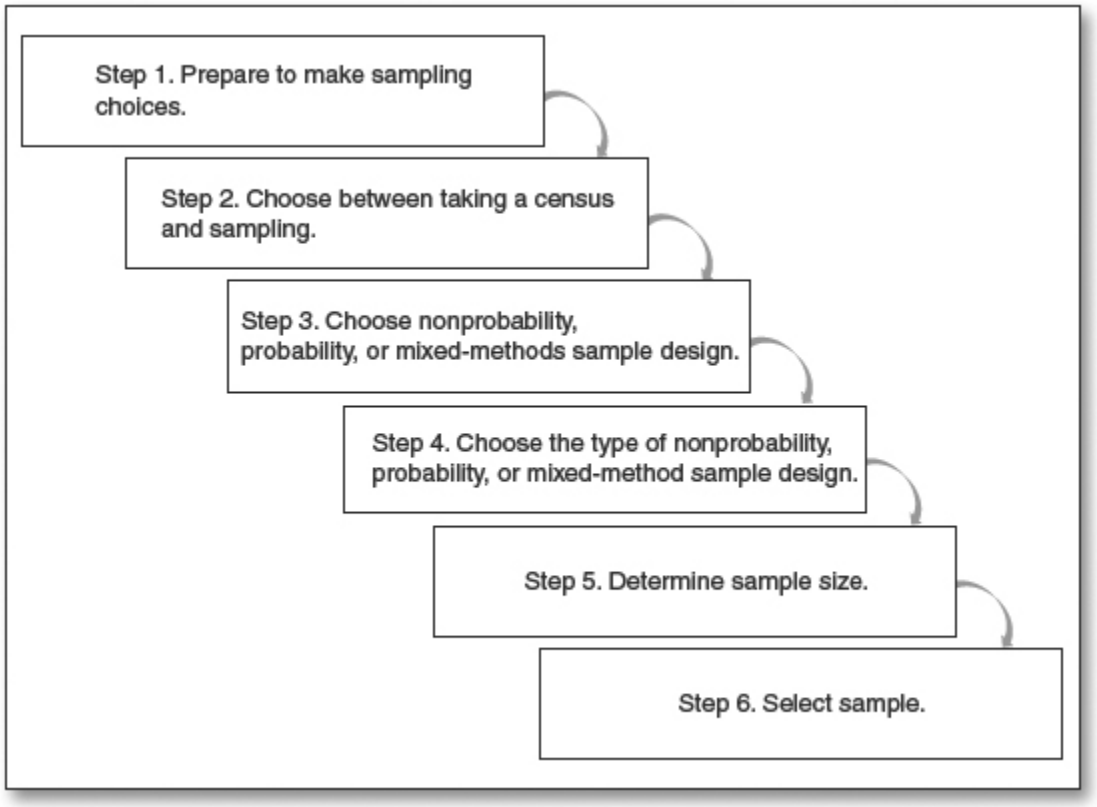
Up to today, sampling procedures continued to evolve. To a certain extent, as modes of collecting data changed, sampling procedures changed. During the period of the 1970s through the 1990s, there was a movement from personal interview surveys to telephone surveys. Variants of random digit dialing (RDD) sampling procedures were developed to meet challenges of telephone surveys. As research methods embraced advances in electronic technology, including the use of online surveys, fax machines, and cell phones, sampling procedures were further modified and adjusted. Today, a wide range of nonprobability and probability sampling procedures are used, making sampling choices more challenging than ever before.

Major Steps in Selecting a Sample

One may identify six major steps in selecting a sample (see Figure 1.1):

Step 1. Prepare to make sampling choices.
Step 2. Choose between taking a census and sampling.
Step 3. Choose nonprobability, probability, or mixed-methods sample design.
Step 4. Choose the type of nonprobability, probability, or mixed-methods sample design.
Step 5. Determine the sample size.
Step 6. Select the sample.

Figure 1.1 Major Steps in Selecting a Sample



Step 1. Prepare to Make Sampling Choices

Specific preparation should be made before making sampling choices. Such preparation should include a careful review of the purpose of one's study, the nature of the population, available resources, research design considerations, and ethical and legal issues considerations. Guidelines for making these preparations are presented in the next section of this chapter.

Step 2. Choose Between Taking a Census and Sampling

The second step involves choosing between selecting the entire target population (taking a census) and selecting a subset of the target population (sampling). In making this choice it is important that one has a good understanding of the differences between random sampling error and systematic error. A description of these differences and guidelines for choosing between taking a census and sampling are described in Chapter 2.

Step 3. Choose Between Nonprobability, Probability, or Mixed-Methods Sample Designs

Once a decision is made to sample, the next step involves choosing between the two major types of sampling: nonprobability sampling and probability sampling. Probability sampling gives every element in the target

population a known and nonzero chance of being selected. Nonprobability sampling does not. Guidelines for choosing between nonprobability sampling and probability sampling are described in Chapter 3. Guidelines for choosing mixed-methods sample designs are included in Chapter 6.

Step 4. Choose the Type of Nonprobability, Probability, or Mixed-Methods Sample Design

The next step involves choosing the specific type of nonprobability or probability sample design to be employed. One may opt to utilize a mixed-methods sample procedure combining different types of nonprobability sampling procedures, different types of probability sampling procedures, or combining nonprobability and probability sampling procedures. The major types of nonprobability sample designs are described in Chapter 4; the major types of probability sample designs are described in Chapter 5. At the end of these two chapters, guidelines are presented to assist in making a sample design choice.

Some sample designs are distinguished by the nature of their sampling units and others by the mixing of more than one sample design type. These types of designs are described in Chapter 6. Designs distinguished by the nature of their sampling units are telephone-based sampling, web-based sampling, address-based sampling, time-based sampling, and space-based sampling. In mixing different types of sample designs, one may mix different types of nonprobability sample designs, mix different types of probability sample designs, or mix nonprobability and probability sample designs. Mixed-methods sample designs are also described in Chapter 6.

Step 5. Determine the Sample Size

Having chosen a specific type of sampling design to be used to select a sample, the next step involves determining of the number of elements to be selected. Chapter 7 describes factors that should be considered in determining sample size and guidelines for doing so.

Step 6. Select the Sample

The final step in sampling involves implementing one's sampling choices. The quality of the resulting sample is dependent substantially on the first step: preparing to make sampling choices. Guidelines for preparing to make sampling choices are presented below.

Guidelines for Preparing to Make Sampling Choices

Specific preparation should be made before making sampling choices. Several guidelines may be proposed. In considering the guidelines listed below and others presented in other chapters, it should be noted that they are not equally important; nor are they absolute. In many cases, their applicability is contingent on

specific conditions that may or not be present. Often, the researcher must balance competing and conflicting guidelines.

Before making sampling choices one should be able to clearly answer such questions as: What are the objectives of the study? How is the target population defined? What is the nature of the population (i.e., its size, heterogeneity, accessibility, spatial distribution, and destructibility)? What resources are available to conduct the study? What type of research design will be implemented? What ethical and legal issues should be taken into account? Considering questions such as these, guidelines for preparing to make sampling choices may be categorized as:

- Objectives of the study
- Definition of the population
- Nature of the population
- Availability of resources
- Research design considerations
- Ethical and legal considerations

Objectives of the Study

> **Guideline 1.1.***Objectives of the study.* Prior to making sampling choices, make sure one has a good understanding of the objectives of the study, the importance of the study, and the special needs of the study, if any.

There should be a good fit between the objectives of a study and the sampling choices a researcher makes. A research study may have only one or a combination of the following objectives: exploration, description, prediction, evaluation, and explanation. **Exploratory research** targets information seeking to better understand a population, theoretical issues, or methodological issues relating to a study. A study with a descriptive objective seeks to describe the parameters of a population, differences between or among population, or relationships among variables. A study with a prediction objective seeks to predict future parameters of a population, differences between or among populations, or relationships among variables. **Evaluation research** seeks to determine the need for an intervention, how the need should be addressed, the ongoing progress of an intervention, and the outcome of an intervention. **Explanatory research** attempts to explain the patterns of population parameters and the relationships among variables.

Different objectives of a study may require different sampling choices. Typically, exploratory research does not require a rigorous sample design. A nonprobability sample with a small sample size may suffice. On the other hand, the requirements of the sample design of a descriptive study may depend on the amount of detail required, the confidence level requirements, and the homogeneity/heterogeneity of the population. The required precision of one's predictions may determine the amount of rigor in the sample design of a prediction study. In part, the sample design for an evaluation research project is dependent on the type of evaluation research one is conducting (formative evaluation, process evaluation, or outcome evaluation). A study with

an explanatory research purpose may require a more rigorous sample design than a study that has one of the other purposes.

There should be a good fit between the importance of a study and the sampling choices that are made. Studies are not equally important. Highly important studies require a much more rigorous sample design than studies that are not as important. A researcher should have a clear understanding of the importance of a study before making sampling choices.

Moreover, there should be a good fit between the special needs of a study, if any, and the sampling choices that are made. The special needs of a study may require the selection of particular population elements, a particular sample size, or a particular situation. In order to ensure that the special needs of a study are met, the researcher possibly should purposefully select particular elements of the population.

Definition of the Population

Before making sampling choices, it is important that one clearly defines the **target population** (the set of elements one desires to apply the findings of the study). An ambiguously defined target population may lead to **population specification bias** (systematic bias resulting from a poor fit between the definition of the target population and the actual population studied). One should have an unambiguous definition of the target population before making sampling choices.

The definition of the target population should clearly identify inclusive and exclusive criteria for participation in the study. That is, it should clearly indicate which elements are included in the target population and which elements are not excluded in it. Inclusive and exclusive criteria should be specified. Inclusion criteria are a set of conditions that must be met for participation in a study. Exclusion criteria are a set of conditions for not allowing participation in a study. Elements may be excluded because they have health problems, their age, they possess characteristics that may confound the study's findings, they may have a language barrier, or they may create a burden for collecting data. Persons living in nursing homes, mental institutions, prisons, or jails may be excluded from the target population even though they satisfy inclusion criteria of the study.

The definition of the target population should specify:

- Nature of the elements
- Sampling units containing the elements to be selected
- Geographic location of the elements
- Time period under consideration

For example, for a study of persons with HIV who live in Washington, DC, one may define the target population as persons with HIV (nature of the elements) living in households (sampling units) in Washington, DC (geographical location), June 2008 (time period). Given this definition, persons with HIV who do not live in Washington, DC, and persons who live in Washington, DC, but reside in a hospice would not be included

in the target population.

The following research note illustrates the use of inclusive and exclusive criteria in defining a target population.

Research Note 1.1: Example of the Use of Inclusive and Exclusive Criteria in Defining a Target Population

Wingood and DiClemente (1998) described the inclusive and exclusive criteria they used in defining their target population in their study of African American women's noncondom use during sexual intercourse as follows:

> Inclusion criteria consisted of being a sexually active, heterosexual African American female, 18–29 years of age, residing in the Bayview-Hunter's Point neighborhood. Exclusion criteria consisted of a history of injection drug use or crack cocaine in the past 3 months. Study participants were recruited using street outreach and media advertisements placed throughout the community Indigenous African American female field recruiters, familiar with the Bayview-Hunter's Point neighborhood, approached and screened women at the local unemployment office, the Social Security office, public laundry facilities, beauty salons, grocery stores, health clinics, and the local (AFDC) office to identify women eligible for participation in the study.

Nature of the Population

> **Guideline 1.2.***Nature of the population.* Prior to making sampling choices, one should have a good understanding of the target population, including its content, size, heterogeneity/homogeneity, accessibility, spatial distribution, and destructibility.

Content of the Population

The content of the population should affect sampling choices. The elements of the population of a research study may be people, things, places, events, situations, or time. The composition of many materials that may be studied in the natural sciences may be assumed to be constant, making any sample essentially identical to any other sample. A sample of blood taken from one's finger may be considered equivalent to a sample of blood taken from one's arm. In such situations, generalizing from a single sampled element may be acceptable. However, different samples of a human population may be extremely unlike each other. Generalizing from a small sample of people may be untenable.

Size of the Population

The size of a population is a critical factor in making sampling choices. One should have knowledge of the size of the population before making sampling choices. More resources and a larger sample size may be necessary to study a large population than to study a small population. Costs, the amount of time needed to collect the study's data, management issues, random sampling error, and systematic error are tied to the size of the population, and thereby affect sampling choices.

Heterogeneity/Homogeneity of the Population

The homogeneity/homogeneity of a population should be considered in making sampling choices. Studies of populations that are relatively homogeneous require smaller samples than studies of populations that are relatively heterogeneous. Before making sampling choices one should determine the homogeneity/heterogeneity of the target population. In conducting a literature review and exploratory research, special attention should be paid to measures of the variability (i.e., standard deviation, variances, etc.) of one's key variables of interest. A pilot study may be in order to acquire such information.

Accessibility of the Population

Populations vary in their accessibility. The accessibility of a population will affect the ability of a researcher to successfully implement a sample design, and should be considered in making sampling choices. Segments of the population may be in remote locations, gated communities or buildings, or other inaccessible locations. Some populations are considered to be “hidden” due to the difficulty to locate them and gain their cooperation. Examples include persons at risk of HIV infection, gang-affiliated adolescents, gays and lesbians who are “in the closet,” injection drug users, sex workers, and the street homeless. Elements of the target population may be inaccessible because they have neither a postal address nor an e-mail address. In preparing to make sampling choices, one should determine the accessibility of the target population. Important segments of the target population that are inaccessible may result in coverage bias.

Spatial Distribution of the Population

The spatial distribution of a population is likely to have a significant impact on the data collection costs of a study and the amount of effort necessary to get the study done. Generally, the more scattered the population, the more resources (i.e., people, time, money, etc.) are necessary to contact and collect data from the population. The sample design should take this factor into account. Acquiring information on the spatial distribution of a population is an important part of preparing to make sampling choices.

Destructibility of the Population

Total enumeration is not an option if making contact and collecting data from a population element destroys it or seriously affects its utility for use in future research. The destructibility of the population elements should be considered in making sampling choices. If destroyed or changed in a significant way, use of the population

for future research may be compromised. In preparing to make sampling choices, one should assess the likelihood of the destructibility of the population.

Availability of Resources

> **Guideline 1.3.***Availability of resources.* Prior to making sampling choices, make a comprehensive assessment of the resources available to conduct the research.

An assessment should be made of available resources to conduct a study before making sampling choices. Typical resources include money, time, personnel, authority, facilities, information sources, equipment, and sampling frame. One should determine the adequacy of availability of resources. If adequate resources are not available, one should acquire what is needed, or make appropriate adjustments to one's sampling choices.

The availability of an appropriate **sampling frame** (i.e., a listing of the target population) is critical in making sampling choices. It may determine whether one should choose nonprobability sampling or probability sampling. Examples of sampling frames include a listing of names of employees, names of customers, addresses, telephone numbers, city directory, and a map. A suitable sampling frame may not exist or may not be accessible to the researcher because of privacy regulations or other reasons. Moreover, it may be very time-consuming to develop an appropriate frame or expensive to purchase one from a vendor.

A good sampling frame would identify all members of the target population only once, and have no other entries, but also include auxiliary information that may be useful in making sampling choices. A good sampling frame would be complete, accurate, up-to-date, reliable, and convenient to use. In preparing to make sampling choices, one should determine the availability of a good sampling frame, and the resources available to create one if one does not exist. Once obtained or developed, an assessment should be made of any sampling frame bias that may exist.

Research Design Considerations

> **Guideline 1.4.***Research design considerations.* Prior to making sampling choices, determine the type of research design, the data collection design, and the data analysis design that will be used.

Sampling choices should be made in conjunction with other choices relating to the research design of a study. Most important are choices relating to the following:

- Type of research design
- Data collection design
- Data analysis design

Type of Research Design

There should be a good fit between the type of research design used in a study and the sampling choices that are made. One should determine the type of research design that will be employed before making sampling choices. Major types of research designs may be classified as follows:

- Qualitative versus quantitative research designs
- Nonexperimental versus experimental research designs
- Cross-sectional versus longitudinal research designs
- Mixed-methods designs

Qualitative Versus Quantitative Research Designs. Sampling choices for qualitative research tend to be different from sampling choices for quantitative research. **Qualitative research** primarily involves the collection and analysis of non-numerical data, with more attention focused on understanding the nature of the elements selected for study than to generalizing to a target population. It is characterized by in-depth inquiry; immersion into the social setting of that being studied; emphasis on the understanding of the participants' perspectives; and comprehensive description of the study's topic. Conversely, **quantitative research** primarily involves the collection and analysis of numerical data, with more attention focused on generalizing to a target population than understanding the nature of the elements selected for study.

Both probability sampling and nonprobability sampling are utilized throughout quantitative research, whereas qualitative research primarily employs nonprobability sampling. Probability sampling might not yield elements of the population that can satisfy the needs of qualitative research. Typically, the qualitative researcher is not interested in estimating population parameters, but rather interested in selecting population elements that are most useful in providing rich information about the topic of the study. These elements may have to be purposefully selected.

Typically, in quantitative research a fixed sample size is set prior to data collection. A researcher might set a target sample size so as to yield a specific margin of error or confidence interval within which is expected to include the true population value. On the other hand, in qualitative research, a sequential approach is more common. Using such an approach, a qualitative researcher would continue to sample until data saturation is reached, that is, as new elements are added to the study, no new information or understanding is forthcoming.

Nonexperimental Research Versus Experimental Research Designs. Sampling choices for experimental research tend to be different from sampling choices for nonexperimental research. In **experimental research,** a researcher controls exposure to the key independent variable of a study. The researcher creates variability in the key independent variable. On the other hand, in **nonexperimental research,** a researcher does not control exposure to the key independent variable of a study. Instead of creating variability in the key independent variable as in experimental research, the researcher measures naturally occurring variability in the variable. Although both probability sampling and nonprobability sampling are used in nonexperimental research and experimental research, the sample designs tend to be more complex and the sample sizes tend to be larger in nonexperimental research studies than in experimental research studies. The primary purpose of experimental research is to make generalizations about cause and effect relationships, and making

generalizations about population parameters is of secondary importance. Experimental research attempts to achieve external validity through replication, whereas nonexperimental research attempts to achieve external validity through its sample design.

Two major subtypes of experimental design research are quasi-experimental designs and true experimental designs, such as randomized clinical trials. These designs differ in terms of the extent to which extraneous variables are controlled, and require different sample designs. In quasi-experimental designs, nonprobability sampling is used, and in true experimental designs, random sampling and random assignment to treatment modalities are used.

Cross-Sectional **Versus Longitudinal Research Designs.** Sampling choices for cross-sectional research tend to be different from sampling choices for longitudinal research. **Cross-sectional research designs** involve the collection of data within one time period. **Longitudinal research** designs the collection of data over multiple time periods. Cross-sectional designs are sometimes referred to as “in-time” studies, and longitudinal designs are referred to as “through-time” studies. In cross-sectional research, all measurements are made at a single point in time or over a relatively short period. There is no need to followup the participants for later data collection. Sampling requirements for cross-sectional research are generally less difficult than sampling requirements for longitudinal research.

Three major terms are used to classify different types of longitudinal designs: *trend study, cohort study*, and *panel study.* A **trend study** examines patterns of changes in variables over time. The terms **cohort study** and **panel study** are often used interchangeably as a study that tracks over time population elements that have a common experience. At times, the terms *cohort study* and *panel study* are used differently. The term *cohort study* may be used to refer to a longitudinal study in which samples are selected over time from the same sampling frame, and the term *panel study* is used to refer to a longitudinal study in which the same sample is studied over time. In this context, depending on the population size and the sample size, in using a cohort study, the same people may not be selected for each data collection event.

Longitudinal designs raise issues of bias due to repeated surveying, attrition, the burden of repeated data collection on respondents, and the proper mix of old and replacement respondents. At times a cross-sectional design is combined with a longitudinal design by systematically adding new elements from the target population to compensate for attrition. A **rotating panel design** may be used to reduce respondent burden. This involves the use of multiple panels of population elements with each being used a fixed number of times and targeting different variables of interest. Given that these designs require different sampling choices, they require different preparations to enhance their effectiveness.

Mixed-Methods Research Designs. A researcher may not limit the study he or she conducts to a single design. A combination of research designs, mixed-methods research designs, may be employed. Much of the literature on mixed-methods designs focus on the mixing of qualitative and quantitative research designs. Other options include the mixing of nonexperimental and experimental research designs, and/or the mixing of cross-sectional and longitudinal research designs. The mixing of different designs has implications for the

sampling choices that are made. A mixed-methods sample design may be advisable, and a larger or smaller sample size may be required than would be required if a single-method design is used. One may mix different nonprobability sample designs, mix different probability sample designs, or mix nonprobability and probability sample designs. Mixed-methods sample designs are described in Chapter 6.

A research design has three major components or sub-designs: the selection of study participants design, the data collection design, and the data analysis design. These sub-designs must fit well together and with the purpose of a study for a study to be effective.

Data Collection Design

There should be a good fit between the data collection choices and the sampling choices that are made. The different modes of collecting data tend to have different sources of systematic error and costs. The preparation of making sampling choices includes estimating eligibility rates, unit nonresponse rates, item nonresponse rates, and data collection costs. All of these factors should be recognized and taken into account when making sampling choices.

Data Analysis Design

There should be a good fit between the data analysis choices and the sampling choices that are made. The data analysis plans of a research project might require only a few basic analyses. On the other hand, the data analysis design may require complex, multivariate procedures. Moreover, statistical procedures vary in terms of their sampling requirements. Some statistical procedures require probability sampling. Generally, the more complex the data analysis design, the larger the required sample size. Sampling choices should be made after determination of the data analysis requirements.

Ethical and Legal Considerations

> **Guideline 1.5.***Ethical and legal considerations.* Prior to making sampling choices, identify any ethical or legal concerns relating to the research project.

There should be a good fit between the ethical and legal concerns and the sampling choices that are made. Concerns relating to informed consent, privacy, anonymity, confidentiality, and professional codes of ethics may make it impractical or impossible to implement certain sample designs. As part of the preparation in making sampling choices, one should make sure one is aware of the relevant ethical and legal regulations.

In order to acquire sufficient information to apply the above guidelines, it may be necessary to conduct formative research including a comprehensive literature review and exploratory research. Research Note 1.2 describes the formative research conducted in a study of men who have sex with other men.

Research Note 1.2: Example of Formative Research in Preparing to Make Sampling Choices: A Study of Men Who Have Sex With Men

In preparation for the National HIV Behavioral Surveillance (NHBS) study of men who have sex with men (MSM), MacKellar et al. (2007) conducted formative research to acquire information for planning the sample design of the study. They described the formative research conducted as follows:

> Formative research was conducted to learn about the venues, times, and methods to recruit MSM. To meet these objectives, staff reviewed advertisements for MSM in online and print media, interviewed key informants, and conducted observations at venues. Key informant interviews were conducted with MSM, MSM researchers, and knowledgeable staff of state and local health departments, prevention planning groups, community-based organizations, service providers, and commercial and social MSM venues. To help ensure that all potential venues were identified, interviews were conducted with key informants of different age groups, race/ethnicities, and sexual orientations.
>
> For each venue identified in these interviews, staff collected information on MSM attendance during specific days and times; the estimated distribution of patrons by race, age group, and sexual orientation; safety issues; and management contact information (if applicable). Staff also asked key informants about participation motivations and disincentives, optimal referrals for prevention and health-care services, and needs for coordination with other venue-based research and prevention efforts. When applicable, staff met with venue owners or managers to solicit their approval to conduct NHBS-MSM on their property.
>
> Finally, staff observed MSM attendance and patron flow patterns at identified venues to learn how recruitment and interview methods might be optimally applied (logistics). If initial formative research did not yield sufficient information on attendance, staff conducted 30- to 60-minute enumerations of male patrons within identified VDTs [venue-specific, day-time periods]. If the proportion of male patrons who were MSM was also unknown or thought to be low (e.g., at street locations, mixed clubs, parks), staff would approach counted men to ascertain their demographic, residence, and sexual-behavior characteristics. These data were then used to estimate the number and proportion of eligible MSM who attended VDTs. VDTs estimated to yield .75% MSM of men approached were

considered MSM venues.

Once appropriate preparation has been made, one would then address the first major choice: whether to conduct a census or whether to sample. Guidelines relating to making this choice are presented in the next chapter.

Summary

Throughout much of human history, total enumeration was considered the only valid way to study populations. Sampling was limited to straw polling, volunteer sampling, and other availability sampling procedures. Quota sampling was introduced in the early 20th century, and took off when pollsters using the technique were successful in predicting the winner of the 1936 U.S. presidential election, whereas pollsters, the *Literary Digest* in particular, utilizing haphazard polling techniques failed to predict the winner. Toward the end of the 19th century, scholars proposed the use of "scientific" sampling. As pollsters turned to quota sampling, governmental researchers and academicians turned to probability sampling in their research. Probability sampling became the preferred technique in 1948 when pollsters utilizing quota sampling failed to predict the winner of the U.S. presidential election that year.

The sampling process has six major steps:

Step 1. Prepare to make sampling choices.
Step 2. Choose between taking a census and sampling.
Step 3. Choose nonprobability, probability, or mixed-methods sample design.
Step 4. Choose the type of nonprobability, probability, or mixed-methods sample design.
Step 5. Determine the sample size.
Step 6. Select the sample.

Proper preparation should be made in the first step so that one has the necessary information to effectively carry out the subsequent steps. One should clearly understand the objectives of the study, definition of the population, content of the population, size of the population, heterogeneity of the population, accessibility of the population, spatial distribution of the population, destructibility of the population, availability of resources, type of research design to be employed, and relevant ethical and legal considerations.

Review Questions

What are the major milestones in the history of sampling?
What should contemporary political pollsters learn from the problems political pollsters had in predicting the winner of the 1936 and the 1948 presidential elections?

What are the major steps in selecting a sample?

What are the major types of research objectives? Why should the objectives of a study be clarified before making sampling choices?

What are the characteristics of a good definition of a target population? Give examples.

In preparing to make sampling choices, why is it important to consider the target population's (a) content, (b) size, (c) heterogeneity, (d) accessibility, (e) spatial distribution, and (f) destructibility?

Prior to making sampling choices, what steps should one take in determining the availability of resources to plan and execute a sample design?

Suppose you were charged with developing a sample design for a study of the problems children who are homeless have in completing their school assignments. Describe in detail the formative research that you would conduct to prepare for constructing your sample design.

What are the major types of research designs, and how do they differ in terms of the preparation that should be done for making sampling choices?

How might the sampling requirements of the following pairs of research designs tend to differ from each other?

- Qualitative versus quantitative research
- Experimental versus nonexperimental research
- Cross-sectional versus longitudinal research

What ethical and legal factors should be considered in making sampling choices?

It is often difficult to obtain a good sampling frame for a population-based study covering a large geographical area. Would a list of driver's licenses serve as a good sampling frame? Why or why not? Once you have answered these questions, consider: Lynch, C. F. et al., “The Driver's License List as a Population-Based Sampling Frame in Iowa” (1994).

Key Terms

Define and give examples of the following concepts:

census
cohort study
coverage bias
cross-sectional research design
evaluation research
experimental research
explanatory research
exploratory research
longitudinal research
nonexperimental research
nonresponse bias

panel study
population specification bias
qualitative research
quantitative research
rotating panel design
sample
sampling
sampling frame
sampling frame bias
target population
trend study

References for Further Study

Bryson, M. C. *The Literary Digest poll: Making of a statistical myth.* 30.(1976)184–185.

Cahalan, D. *The digest poll rides again.* 53.(1989)129–133.

Cowles, M.(2000). Statistics in psychology: An historical perspective.Mahwah, NJ: Lawrence Erlbaum.

Crossley, A. M. *Early days of public opinion research.* 21.(1957)159–164.

Edwin, R. *Gallup polls public opinion for 25 years.* 93.(1960)62–63.

Frankel, M. R. & Frankel, L. R. *Fifty years of survey sampling in the United States.* 51.(1987)S127-S138.

Gallup, G. *The changing climate for public opinion research.* 21.(1957)23–27.

Hansen, M. H. *Some history and reminiscences on survey sampling.* 2.(1987)180–190.

Ibo, S. E. *A gold mine and a tool for democracy: George Gallup, Elmo Roper, and the business of scientific polling, 1935–1955.* 42.(2006)109–134.

Katz, D. & Cantril, H. *Public opinion polls.* 1.(1937)155–179.

Kish, L. *The hundred year wars of survey sampling.* 2.(1995)813–830.

Kruskal, W. & Mosteller, F. *Representative sampling, IV: The history of the concept in statistics, 1895–1939.* 48.(1980)169–195.

Link, H. C. *Some milestones in public opinion research.* 313.(1947)225–234.

Lynch, C. F. , Logsden-Sackett, N. , Edwards, S. L. & Cantor, K. P. *The driver's license list as a population-based sampling frame in Iowa.* 84.(1994)469–472.

Northrop, A. *The rise of the polls: Bloopers amid improving aim.* 3.(1971)1703.

Squire, P. *Why the 1936 Literary Digest poll failed.* 52.(1988)125–133.

Stephan, F. F. *History and uses of modern sampling procedures.* 43.(1948)12–39.

Stephan, F. F.(1949). Development of election forecasting by polling methods. In Edited by: **F.Mosteller**, **H.Hyman**, **P. J.McCarthy**, **E. S.Marks**, & **D. B.Truman** (eds.), The pre-election polls of 1948: Report to the Committee on Analysis of Pre-Election Polls and Forecasts, Social Science Research Council, Bulletin no 60 (pp. 8–14).New York: Social Science Research Council.

Sudman, S. & Blair, E. *Sampling in the twenty-first century.* 27.(1999)269–277.

Wallace, H. A. & McCamy, J. L. *Straw polls and public administration.* 4.(1940)221–223.

Weiss, N. J.(1983). Farewell to the party of Lincoln.Princeton, NJ: Princeton University Press.

http://dx.doi.org/10.4135/9781452272047.n1