SAMPLING

A Practical Guide for Quality Management in Home & Community-Based Waiver Programs

A product of the
National Quality Contractor

developed by:
Human Services Research Institute

And

The MEDSTAT Group, Inc.

developed for:
Centers for Medicare and Medicaid Services

March 2006

Authors:

Ruth Freedman, PhD
Boston University School of Social Work
Boston, Massachusetts

Sarah Taub, MMHS
Human Services Research Institute
Cambridge, Massachusetts

Other Contributors:

Julie F. Silver, MMHS
Elizabeth Pell, LICSW
June Rowe, CAGS
Giuseppina Chiri, MA

Human Services Research Institute
Cambridge, Massachusetts

Beth Jackson
Sara Galantowicz

The MEDSTAT Group, Inc.
Cambridge, Massachusetts

Acknowledgements

The authors express their appreciation to the participants of an initial focus group who offered ideas and suggestions regarding what states need to know about sampling. Their experiences, candor and contributions all helped to shape this guide. Focus group members included: Ellen McClimans, Chris Newman, Ruth Schanke, Dave Engels, Karen Glew, David Goddu, Gail Grossman, Connie Lehr, John Zeeck and Dana Ciccone.

We also thank the Centers for Medicare and Medicaid Services for their support and enthusiasm for this practical response to states' needs and concerns.

This Guide was produced for the Centers for Medicare and Medicaid Services, CMS Contract No. 500-96-006, Task Order No. 2. and CMS Contract No. 500-00-0021, Task Order No. 1.

Table of Contents

Introduction

In the past decade we have witnessed appreciable changes in quality management (QM) for community-based long-term care services and supports provided to elders and people with disabilities. The number of people receiving services and supports in the community, along with the complexity of the systems for delivering those services, has increased dramatically. In response, states have worked to develop QM processes that can both address individual concerns and identify system-wide issues.

At the same time, the Centers for Medicare and Medicaid Services (CMS), which helps fund many of these services through the Medicaid 1915c waiver program, has refocused its quality oversight responsibility for home and community-based services (HCBS) at the federal level to reflect three important principles. These are:

- States have first-line responsibility for assuring the quality of services and supports provided through their HCBS waiver programs, and for assuring the health and welfare of program participants.

- CMS oversight of this responsibility must be continuous over the life of an approved waiver, and requires sustained and on-going communication between the federal government and state waiver staff.

- The focus of quality management efforts should be on meeting the waiver assurances articulated in federal regulations and on continuous quality improvement within individual waiver programs.

These principles have direct implications for the design and practice of quality management at the state level. They require that a state develop the ability to retrieve and analyze information from state and local agencies, providers and participants, and use this information for quality improvement up, down, and across the organizational hierarchy.

In recognition of these three principles, CMS revised its quality review process for Medicaid HCBS waivers through implementation of the Interim Procedural Guidance for Assessing HCBS Waivers, in January 2004.[1] The new procedures shifted the paradigm of federal oversight from a point-in-time, on-site review of waivers to a continuous quality improvement cycle, characterized by two processes:

- On-going dialogue between federal reviewers and state staff; and

[1] CMS plans to replace the Interim Procedural Guidance with an annual report (the proposed CMS 373Q) from the state to CMS on the quality achieved in their waivers.

- Provision of evidence by states that illustrates they have systems in place to identify, in a timely manner, when quality issues occur; to take action to remedy individual quality issues; and to prevent their reoccurrence through appropriate intervention(s).

Inherent in the quality improvement cycle are the three functions of **Discovery, Remediation, and Improvement**, as articulated in the *HCBS Quality Framework*, released by CMS in 2004.[2] Each of these functions relies on accurate and representative information to identify and address quality issues.

Discovery is the first step in managing and improving quality, and provides information (in accordance with the CMS Assurances) on whether program participants are provided appropriate and adequate access to services and supports; that these services and supports are delivered as intended; that participants' health and welfare is achieved; that providers of services and supports are qualified; that the financial integrity of the program is maintained; and that the single state Medicaid agency oversees and is actually involved in the quality management enterprise.

In implementing Discovery approaches, states face questions of what types of information to gather and what data sources can provide the needed information. For example, Discovery around health and welfare might involve aggregated data from case manager supervisory record reviews; from independent case record reviews by the Medicaid agency; from a survey of program participants; from a comparison of plans of care with claims data, and so forth. Each of these discovery methods focuses on a different kind of information from a different source, but all have the potential of providing important information about whether beneficiaries are receiving the necessary services and supports and/or are achieving outcomes consistent with the intent of the program.

Gathering information from each waiver participant or service plan, however, can be costly and time intensive, and is not always necessary. Depending upon the size and scope of the state's waiver program, it is often sufficient and more cost effective to draw representative samples in order to gather information and make inferences about an aspect of program quality overall. In its oversight role, CMS does not prescribe sampling methods for states to use. It does, however, expect that states will use sound and reasonable processes to gather information from which conclusions about quality can be drawn and acted upon.

To that end, this Guide is designed to provide states with practical information about sampling techniques and strategies that they can employ in their quality management work. Its purpose is to offer states information to consider when deciding whether to sample, and to identify issues for consideration once the decision to sample has been made. The information presented here is intended to familiarize the reader with basic concepts and considerations; it should not take the place of seeking technical expertise in sampling methodology. States are strongly advised to consult with a statistician or research methodologist when designing sampling plans.

---

[2] The *HCBS Quality Framework* is available at http://www.cms.hhs.gov/HCBS/05_Quality%20Oversight.asp#TopOfPage

How to Use This Guide

This Guide seeks to provide a user-friendly, step-by-step approach to explaining sampling, identifying alternatives among different sampling techniques, and understanding how to use these techniques for specific purposes in a quality management strategy. Much of this Guide should be useful to agencies that have already made the decision to sample, by providing guidance on implementing that decision.

However, before launching into sampling, states must first assess whether or not sampling is the best means for securing the information they are seeking. Sometimes sampling is not the most appropriate way to gather data. Certain types of information, such as health and safety data, are routinely gathered for all program participants, and states may find it more useful to examine data from the entire population of program beneficiaries when analyzing trends.

To help users determine the purpose of a proposed data collection effort, the Guide begins with a discussion of issues to consider when deciding whether to sample (*To Sample or Not?*). The subsequent sections address various sampling considerations and techniques (*Defining the Target Population, Sampling Methods, Determining Sample Size, Stratification* and *Sources of Error*). Critical terms are defined in call-out boxes throughout the text, and highlighted words are defined in the Glossary found at the end of the Guide.

The Guide is intended for use in quality management for Medicaid HCBS waivers serving any population, and for any type of service or support. Examples of applications in different types of waiver programs are provided throughout the text. Additional references for applying these concepts are provided in the Appendices, which include: a list of Sampling Resources, such as web-based calculators for determining sample size (Appendix A) and a Glossary of terms (Appendix B).

To Sample or Not?

A sample is a subset of a population. Since it is sometimes impractical, or not necessary, to collect information on the universe of individuals or other entities of interest, such as providers or care plans, a smaller segment of the population is selected to estimate the characteristics of the whole. The purpose of sampling is to collect information that can be generalized to the broader population from which the sample is drawn.

When drawing a sample, the goal is to select one that is representative of the population of interest. A sample is considered representative of the population if the characteristics of, or outcomes associated with, the sample (e.g., age, type of disability, percent receiving a level of care determination on time; percent getting their needs addressed) are similar to their distribution in the population. Ideally, sample characteristics/outcomes should mirror their manifestation in the population. For example, if 40 percent of the population lives in group homes, than approximately 40 percent of the sample should, too.

However, before any discussions about sampling design can begin, states should examine the purpose of the data collection effort and determine the best approach for gathering information. This section takes the reader through some basic issues to consider when making the decision whether or not to sample.

Representative Sample

A sample is considered representative of the population if the characteristics of the sample (e.g., age, gender, type of disability) are similar to the distribution of these characteristics in the overall population.

Determining the Need for New Information

States continually gather information from many disparate sources for a variety of different purposes, including quality assurance and improvement. Before an agency decides to sample, it must first determine exactly what information it needs and if this information is already available. Sometimes the needed information, such as use of dental services, can be drawn from an existing source, e.g. claims data. If it cannot, the next consideration is how these new data should be collected. To first assess the need for collecting new data, administrators must answer three fundamental questions:

1. What is the question to be answered?

2. What is the source of information that can answer/address the question?
3. Is it necessary to collect new information, or can existing information be used?

For example, a state agency may have questions about waiver beneficiaries' access to transportation services based on concerns raised by case managers. After discussion among members of state's Quality Committee, the following question emerges: What proportion of participants in the state's aged/disabled waiver actually used the county-run van service in the past 12 months? The committee then identifies two potential sources of information to answer this question. The first is the billing system used by the transportation service, which submits an electronic invoice for each trip that includes information on the number of riders. The second is a proposed survey of waiver participants assessing their use of the transportation service in the past year. After further investigating these options, the committee determines that the invoices are not an adequate source of information because they cannot be used to estimate an unduplicated number of users. As a result, committee members decide they must collect new data.

Collecting New Information

After a state has determined it needs to obtain new information, the next consideration is how best to collect these data. In some cases, it may be sufficient or preferable to collect information from only part of the population of interest and use these data to estimate the outcomes of the overall population. In other cases, sampling the population may not be appropriate. Each of these scenarios is discussed below.

Deciding to Sample

Sampling to collect new information should be considered when:

- It is not necessary to collect information about the entire population to determine if there is a problem or to answer a question.

- It would take too long to gather information systematically about an entire population, particularly if the population is large and/or geographically dispersed;

- Collecting information about a full population would be labor-intensive and require significant staff time and money.

If any of these conditions are met, state officials should consider drawing a sample, and collecting information only on the sample. In the example from the previous section, let's assume that the number of aged/disabled waiver participants is quite large and contacting all of them would be very time-consuming and labor-intensive. The Quality Committee is confident that any transportation problems experienced by a representative sample of waiver participants would be indicative of access issues for the waiver population overall. In this case, they determine that using a sample for their proposed survey is the best method to collect new information.

There are a variety of methods that can be used to select samples from populations. Some of these methods are more likely to yield a representative sample than others. These techniques are discussed in detail in the next section.

Deciding Not to Sample

Sometimes, collecting new information about the full population- called a census - is more appropriate than sampling, for both methodological and pragmatic reasons. For example, collecting new information from an entire population could be useful when:

- The size of the population of interest is so small that it is feasible to collect information from all members of that population.

- There is a legislative or regulatory mandate to collect information about all members of the population. For example, if a state is required to gather data about all service providers for certification and licensure, selecting a sample of providers would not be acceptable.

- Collecting information from all members of the population may be perceived by stakeholders as more legitimate or credible.

- Information on the entire population is readily available (e.g., automated information on critical incidents, automated data on claims and plans of care to assess degree to which authorized services are actually delivered). Sampling always involves some amount of error, so if you have the information you need on the entire population and it is readily available, it makes sense to use it rather than to sample.

Census

A process used to collect information about the full population, as opposed to a sample or subset of the population.

Other Methods for Collecting New Information

Even when states determine that new information is necessary and sampling is the best path, a state may still determine that its resources are not adequate, and drawing a truly representative sample is not a practical option. In these instances, states can consider a range of other approaches to gathering information. States frequently and effectively use other qualitative methods of inquiry, such as:

- Focus Groups – structured discussions used to gather information and insight from a small group of individuals

- Public Forums – larger gatherings used to present information and gather feedback from individuals in specific communities

- Targeted Reviews – examinations of a single issue, with a group of selected individuals

- Root Cause Analysis – in-depth examination from a systemic perspective of a seminal event

- Quality Improvement Committees – appointed bodies whose purpose is to advise a public system

Sampling considerations and methods to use when applying these types of qualitative approaches are beyond the scope of this guide.

Defining the Target Population

One of the first and most important steps in sampling is to define the target population (i.e., population of interest). The target population is the collection of entities you want to study. There needs to be a clear and explicit definition of the whole population or universe before the sample is drawn from it. If the population is not well specified, it is difficult to determine whether the sample is representative of the population and erroneous generalizations or conclusions may result. It is important to note, however, that the term "population" can refer to elements or units other than people. In addition to individuals, states frequently will examine provider organizations, records, critical incident reports, and plans of care, as well as other populations in their quality management systems.

Inclusion and Exclusion Criteria

Clear inclusion and exclusion criteria are necessary to specify the population of interest. These criteria indicate who is eligible (or not eligible) to be selected from the target population into the sample. Inclusion and exclusion criteria specify the characteristics of individuals or entities to be included or excluded, respectively, in the sampling frame (e.g., age, type of disability, type of residential setting, geographic location). These criteria help to clarify the boundaries of the specific target population. Regardless of the inclusion/exclusion criteria used, the rules must be consistently applied and explicitly documented. Doing so allows states to make accurate estimates about the population based on what they have learned from the sample. If these criteria are not carefully outlined, it is possible to make erroneous generalizations. For example, if you decide to include only waiver participants with mobility problems, you will only be able to generalize your findings to that group, not to the entire waiver population (even though persons with mobility problems are a subset of the larger waiver population).

Sometimes the definition of the target population may be based upon certain operational considerations, such as litigation or legislation. For example, if a class action settlement mandates that annual surveys be conducted to assess state agency compliance in providing services to protected "class members," then the target population by definition is all of the individuals who are members of this protected class. Other times there may be practical considerations, such as the lack of interpreters or foreign language surveys, which affect the scope of the target population.

Sampling Frame and Units of Analysis

The actual list of the target population, created by applying the inclusion and exclusion criteria, is referred to as the sampling frame. The sampling frame operationally defines the target population. It is the list of people or entities from which the sample is drawn. In many cases, the sampling frame consists of individuals – for example, a list of program participants currently served by an agency, or a list of people who are on a waiting list for environmental modification services. The unit of analysis is the element about which information is collected and will provide the basis for the analysis. Sometimes the unit of analysis in the sampling frame includes elements other than individuals – for example, a list of home health care service providers, grievance reports, or individual support plans. For different analyses in the same inquiry, there may be different units of analyses.

Sampling Frame

The list of all the units of analysis in the population that meet the sampling criteria.

The sampling frame must be as accurate as possible. If the list of people or entities in the target population is incomplete or outdated, the resulting sample may be biased and therefore not truly representative of the population. For instance, a year-old roster for a waiver with a high death rate (due, for example, to the acuity of the program participants) or short average length of stay may not be very accurate for defining the sampling frame. Therefore, the list must be screened for completeness and accuracy immediately prior to drawing the sample. Common inaccuracies in sampling frames include omissions, duplications, or ineligible cases. Bias due to coverage errors (selecting the wrong people or selecting too few or too many people with certain characteristics) or inaccurate conclusions can result if certain members of the population are:

- mistakenly omitted from the sampling frame;
- listed more than once in the sampling frame; or
- mistakenly included in the sampling frame when they do not meet the inclusion criteria.

Unit of Analysis

The element about which information is being collected and analyzed.

In some situations, a complete list of the population may not exist, or the list may not be available or accessible for sampling (e.g., due to privacy and confidentiality issues, data access problems, or organizational constraints). In many cases, one “master list” is not available, but can be constructed by combining lists from different data sources. Cooperation and/or agreements with sister agencies may be necessary to secure accurate and complete lists. If multiple lists are combined, it is important to screen for duplication. Sometimes an initial outlay of effort is necessary, such as checking for current address, but then processes can be put in place to reduce effort in the future. If these hurdles can not be overcome, it may be necessary to define your population more narrowly.

Sampling Methods

There are two broad types of samples: probability and non-probability. In this section, probability sampling methods are described first, followed by discussion of non-probability sampling methods. In a probability sample, every member of the target population has a known, non-zero probability (or likelihood) of being included in the sample. That is, everyone in the target population has a chance of being selected into the sample, and their chances of selection (likelihood) are greater than zero. A probability sample is considered "representative" of the population, and therefore the findings based on the sample can be generalized to the population overall. Probability samples are essential if the goal of the data collection is to make estimates about the whole population or to use data from the sample to draw conclusions. When drawing probability samples, specific random selection procedures are used that eliminate subjectivity or bias in the sample. "Random" selection in this context does not mean haphazard or coincidental. Rather, it refers to precise procedures based upon probability theory.

Probability Sampling Methods

Probability Sample

A sample drawn according to random selection procedures in which every member of the target population has a known, non-zero chance of being included in the sample.

In order to use probability sampling methods, a complete list of the target population (the sampling frame) is needed. Common probability samples typically fall into one of two types: single stage random samples and multi-stage random samples. Single stage random sampling methods assist states to generalize across the entire target population, and frequently are more cost-efficient. However, when states want to ensure that sub-populations are adequately represented in their sample, they use a multi-stage probability technique.

Some of the common probability sampling methodologies are described below.

Single Stage Sampling Techniques

Simple random sampling:

Using simple random sampling, each unit in the sampling frame (e.g., each individual on the population list) is assigned a number and then numbers are randomly selected using a random numbers table or a computerized random selection program. Numbers are randomly drawn until the desired number of cases for the sample has been reached.

Systematic sampling:

In systematic sampling, selection typically begins at a random place in the population list (sampling frame) to identify the first case to be selected into the sample and then cases are selected at regular intervals from the list – for example, every 6th person is selected, or every 10th person. This type of sampling is typically less cumbersome than simple random sampling, particularly if the population list is long. It is considered as accurate and unbiased as a simple random sample, provided that there is no repetitive pattern or ordering to the sampling frame list. If there is an inherent cycle in the list, linked, for example, to age or residence, selecting every "nth" person might reflect this order bias and result in under- or over-representing certain types of cases on the list.

To determine the appropriate "interval" for your sample (e.g., every "nth" case) divide the population size by the desired sample size. For example, if the sampling frame lists 1000 names, and a sample of 200 is desired, then the sampling interval is 5 (1000/200). That is, every 5th person on the list would be selected into the sample. Determining the appropriate sample size is discussed in more detail below.

Multi-stage Sampling Techniques

Stratified Sampling:

In this technique, the population is first divided into homogeneous strata or sub-samples (grouping of individuals or entities based on characteristics they share), and then simple random sampling or systematic sampling is used to select cases within each stratum. Stratified sampling is used when the state wants to control the relative size of each stratum or sub-sample, instead of leaving this to chance in sampling the full population. A common reason for stratification is to ensure representation of small groups that might otherwise not have a large enough presence in the sample about which to make statistical generalizations. For example, agency administrators may want to compare participant outcomes across five geographic regions of a state and be assured that the sample size within each region is sufficient for making credible generalizations about each region. To use stratified sampling, administrators would first divide cases of the sampling frame into the five state regions (strata) and then randomly sample participants from within each region. This ensures that cases from each stratum are adequately represented in the full sample.

In order to use stratified sampling, there must be sufficient information about the population to decide in advance to which subgroup or stratum each member belongs. The principles of sample size (see next section, *Determining Sample Size*) will apply to each stratum, however, and in order to make accurate analyses and reduce errors, administrators must ensure adequate size of the strata. Additionally, when examining the data in the aggregate, strata may need to be weighted to mathematically account for the disproportionate contributions of the various strata. For a more information regarding analysis of stratified samples, see the *Stratification* section of this guide and consider consulting with a statistical analyst.

Cluster Sampling:

Cluster sampling is typically used in large-scale studies covering broad geographic areas or organizational units, and it involves a multi-stage process. Cluster sampling is used when a complete centralized sampling frame is not available; however, complete sampling frames for each cluster must be available. The first step in drawing a cluster sample is to identify key geographic groups or distinct information clusters (e.g., census tracts, counties, regional offices, differing data sources), then a random sample of these clusters is drawn, and, finally, cases within the randomly selected clusters are randomly selected into the sample.

Cluster sampling is not generally recommended for agency-initiated inquiries as this technique is costly, complex and requires more intensive efforts to control error. While cluster sampling allows agencies to use random selection techniques throughout the process, in the absence of a centralized population list, this technique is considered somewhat less reliable than other “pure” forms of random sampling. There is potential for sampling errors and inaccuracies at each step of this multi-stage process, particularly at the cluster selection stage. Moreover, weighting the results of a cluster sample back to the population is highly complex. Cluster sampling should not be attempted without the services of a seasoned sampling statistician.

Non-Probability Sampling Techniques

In a non-probability sample, the likelihood of selecting any one case from the population into the sample is not known. Random selection procedures are not used in non-probability samples. Instead, cases are selected from the population based upon other criteria such as the judgment of the people doing the study, requirements of other entities, or availability of subjects; there is greater potential for subjectivity or bias in non-probability sample selection. Non-probability samples are often used when a sampling frame is not available, and/or when the time requirements or costs of using probability methods are prohibitive. A non-probability sample may not accurately represent the population, and the generalizability of findings is limited. Despite of these limitations, non-probability samples can be useful and appropriate in certain situations such as descriptive, exploratory, and qualitative studies in which generalizability of findings to broader populations may not be necessary.

Non–Probability Sample

A sample drawn without random selection procedures; the likelihood of selecting any one subject is not known and generalizability to the population is limited.

Some common types of non-probability samples are:

- *Availability or convenience sample*: Sampling those people readily available or convenient to study -- for example, surveying available and willing participants in a day program about their experience with program staff.

- *Purposive sample*: Selecting individuals from the population based upon professional experience, knowledge, or judgment (i.e., purposely handpicking sample members) – for example, purposefully selecting typical or atypical cases for inclusion in the sample or critical cases judged by quality management staff to be important to investigate.

- *Quota sample*: Setting a quota for inclusion of specified numbers of people with certain characteristics (e.g., establishing a sample quota of 25 adults with MR/DD living in community settings, 25 adults with psychiatric conditions in community settings, 25 people with physical disabilities in community settings). Individuals with these characteristics are selected, not necessarily randomly or from a known list, until the specified quota is achieved.

- *Snowball sample*: Making initial contact with known members of the population, and then asking those sample members for referrals to other members of the population. This chain-like informal word of mouth referral process is especially helpful in identifying hard-to-reach populations such as people who are homeless, undocumented immigrants, or people who are socially isolated.

The advantages of these non-probability methods are that they are often less time-consuming and resource intensive than probability methods. While they may be appropriate in certain contexts as described above, the findings based upon these methods are limited in terms of generalizability to the broader population.

Determining Sample Size

State officials frequently grapple with the question: How large should our sample be? The answer is, “it depends.” This section will discuss some of the factors that states should consider when making decisions about sample size. To make this decision, states need to evaluate several factors, balance tradeoffs, and ultimately decide what works best, given their specific information objectives, resources, and constraints.

Relationship between Sample Size and Population Size

One consideration when determining sample size is the size of the population. The table below displays sample sizes that were calculated using different-sized populations, all at 95% confidence level and +/-5% margin of error. (These terms will be explained below.) This table illustrates that once the population size reaches the thousands, the required sample size increases very incrementally.

A common misconception is that samples should be determined based on a certain “percentage” or fraction of the population. This is not true. Looking at an online calculator or a statistical table it is clear that the formulas behind the calculations are not based on percentages of the total population.

Sample Sizes for Different Size Populations

Population	Sample*
300	169
1,000	278
2,000	322
3,000	341
30,000	379
300,000	384

*95% Confidence Level and +/- 5% margin of error

Degree of Accuracy

Whether you decide to use a statistician, an online calculator, or your old statistics textbook to determine sample size, you will need to be familiar with several key terms related to accuracy. Critical concepts include: sampling error, confidence level, and margin of error. As discussed earlier, probability sampling methods that use random selection procedures are considered to have a higher degree of accuracy than non-probability methods. Random sampling allows one to make generalizations about the population based upon the data collected from the sample. However, even when random sampling is used, the sample characteristics (also called sample statistics) are likely to differ somewhat from the true population values (also called population parameters). This discrepancy is referred to as sampling error.

Sampling Error

The amount of discrepancy between the characteristics of the sample and the actual characteristics of the population. Sampling error is due to chance and can be estimated mathematically.

Sampling error occurs simply as a result of the process of drawing a sample. It is a type of error that is due to mathematical chance, or the probability of selecting cases that do not estimate exactly the population parameter. There are two pieces of "good news" about sampling error. First, increasing sample size reduces the amount of sampling error. The larger the sample size, the more likely the sample values will be close to the true population values. Second, if random sampling methods are used, it is possible to estimate mathematically the amount of sampling error – that is, the extent to which the sample may differ from the population.

Confidence Level

A statistical estimate used in random sampling, stated as a percentage, of the degree of certainty that the true population value is within a specified range of values.

The next important concepts to understand are confidence level, confidence interval, and margin of error. These concepts are inter-related and can all be traced back to sampling error. Key decisions will need to be made about these factors, which will then be used to determine sample size.

In a nutshell, states need to decide how much sampling error they are willing to tolerate and their desired confidence level, given their specific objectives. There are statistical tables in many research texts and computerized software programs that calculate the sample sizes necessary to provide population estimates at various levels of precision, by specifying confidence levels and confidence intervals. (A list of web-based and print resources is included in the Appendix of this guide.) Because of the complexity of determining the best approach for a particular research effort, states may want to consult with a statistician or survey methodologist about whether and how to use these resources to determine appropriate sample size, confidence level, and confidence interval. Keeping this caveat in mind, the basic concepts related to accuracy are illustrated below.

Confidence Interval

A statistical estimate of the range of values within which the true population value is likely to fall. Confidence intervals are often denoted by a single number that identifies the margin of error, such as + or - 5%.

Random sampling allows us to estimate statistically the range of values within which the true target population is likely to fall (the "confidence interval") and how certain we can be that the true population value is within that range of values (the "confidence level"). This allows us to make a statement like:

> *Based upon the sample data, we are 95% certain [confidence level] that between 65% and 74% [confidence interval] of E/D waiver participants have been visited by a case manager in the past two months.*

Because sampling error is due to mathematical chance, the sample is just as likely to underestimate as it is to overestimate true population values. The confidence interval (also referred to as "margin of error") is sometimes expressed as "plus or minus" the number of units around the sample statistic. The statement above could also be written as:

> *Based upon the sample data, we are 95% certain [confidence level] that 69.5%, +/- 4.5% [margin of error] of E/D waiver participants have been visited by a case manager in the past two months.*

A common confidence level used in scientific reporting is 95%, and a generally acceptable margin of error is +/-5%. The larger the sample size, the more accurate population estimates will be. However, once a sample reaches a certain size, there are diminishing returns on accuracy. The table below displays the margin of error associated with different samples sizes selected to represent a population of 1500 and based on a confidence level of 95%. As you can see, the margin of error is significantly higher for a small sample than it is for larger samples. As sample size increases from 30 to 200, the margin of error drops dramatically from +/-17.7% (which is unacceptable) to +/-6.5% (much better). However, increasing the sample size from 600 to 800 only changes the margin of error from +/-3.1% to +/-2.4% (both of which indicate high levels of accuracy). It is important to understand these concepts when determining sample size so that states can ensure the level of accuracy selected is appropriate and acceptable for the purpose of the inquiry.

Relationship Between Accuracy and Sample Size

Sample	Margin of Error*
30	+/-17.7%
200	+/-6.5%
400	+/-4.2%
600	+/-3.1%
800	+/-2.4%

*For a population of 1,500 and a 95% confidence level

Degree of Variability in Population

The more variability in the population, the larger the sample size needs to be. For example, if the research question relates to health conditions and medication use and there is an indication that the target population is quite varied in terms of these two characteristics, then a larger sample size will be needed to capture the diversity (heterogeneity) of that population. The opposite is true when population members are quite similar in terms of health conditions or medications. Of course, there are many situations where we do not know in advance how diverse the population is likely to be in terms of various characteristics. In fact, states may often collect data to examine the characteristics of the population. When the variability of the sample is unknown, it is generally better to be conservative, assume large variability within the population, and draw a larger sample to capture potential diversity within the population.[3]

[3] On-line calculators typically assume the greatest variability possible in the population and build this assumption into the sample size calculation. Therefore, the resulting sample size is somewhat larger than if more homogeneity is assumed. However, when the level of variability in the population is unknown, this assumption is the safest and most reasonable.

Number of Variables to Be Examined

Larger sample sizes are advised if you intend to examine large numbers of variables simultaneously. For example, if a state plans to analyze 15 different demographic, clinical, and environmental factors that may predict client satisfaction with services, the analysis will likely involve multivariate statistical techniques that simultaneously investigate how these various factors may in combination affect client satisfaction. The more variables that are investigated in combination, the larger the sample size should be. A statistical analyst or survey methodologist can help advise about whether multivariate analysis is relevant and its implications for sample size.

Non-Response Rate

Even if a state follows proper procedures to select a random sample of cases or individuals, ultimately some of those selected will not respond to the inquiry. "Non-response" of potential participants occurs for a variety of reasons and has implications for sample size. If a state anticipates that there may be appreciable non-response (e.g., due to refusals, inability to make contact, cancellations, unexpected illness, etc.), given their knowledge of the population and past experiences, then it may be wise to draw a larger sample than needed in order to compensate for the potential non-response factor. This process is referred to as "oversampling." For example, a state may decide to randomly select 400 program participants, in hopes of actually ending up with information from a desired sample size of 350. Non-response also varies by the research method chosen. For example, response rates to mail surveys are usually low, and to ensure an adequate sample size the state must over-sample by a much larger factor than they would for face-to-face interviewing.

Non-response also has implications for error. It is helpful if states have access to some background demographic data about the non-respondents in order to analyze whether and how non-respondents may differ from the actual sample respondents. If non-respondents are significantly different from the respondents (e.g., in terms of type of disability, age, type of services, socio-economic factors, language proficiency, etc.), then the final sample has limited generalizability to the population and may be considered skewed.

Budget Resources and Time Constraints

States need to factor in all of these concerns and evaluate the relevance of these issues within their particular operational context. Ultimately, states must balance the tradeoffs of obtaining a sufficient sample size within their existing budgetary resources and time constraints. They must anticipate the costs of drawing a sample and the related tasks of data collection, data entry and management, and data analysis when determining an appropriate and feasible sample size.

As a rule, a sample should be large enough for the state to feel confident in the generalizability and consequent credibility of its results. However, samples large enough to satisfy conventionally acceptable levels of accuracy and precision may require more resources than a state has at its disposal. As such, states must weigh their need for data accuracy and precision against their budget concerns and resource constraints.

Stratification

This section provides additional information about stratified random sampling methods.

Sample Stratification

In stratified sampling, the population is divided first into mutually exclusive subgroups (called strata), and then random samples are drawn from each stratum. This approach helps ensure representation of key subgroups of the population, which is helpful when there are differences in the relative size of groups within the overall population of interest. To use this sampling method, there must be sufficient information for categorizing each member of the population into a stratum (subgroup). This information must be present BEFORE the sample is drawn. Stratification is only possible if the state can divide the list of the target population (i.e., the sampling frame) into non-overlapping homogeneous subgroups from which to draw a random sample. The strata membership (or subgroups) of individuals must be known in advance of selecting the sample members. Moreover, it is also imperative that all strata are mutually exclusive; for example, using age as stratification criteria, only ONE strata can include persons 18-44 years old and only ONE strata can include those 45-64 years old.

Typically a population is stratified based on a key variable upon which comparisons will be made, such as demographic, administrative, or background characteristics. For example, a state agency may decide to stratify its sample by geographic region in order to ensure an adequate sample size within each region, especially if it believes there are differences in service use, service availability, or other resources between regions. Other variables that states often use to stratify a population include service or provider type (e.g., day/residential/supports) and residential setting (e.g., community residence, family home, assisted living facility.)

In some cases a state may decide to stratify the population based upon two or more variables, such as stratifying simultaneously by both region and provider type. First, the population list would be divided into specific regions, and then into provider subgroups within each region, to create a series of mutually exclusive subgroups called “cells.” A sample of individuals would then be randomly selected from each sampling cell. A note of caution -- the more stratifying variables used, the more complex the sampling and analysis procedures, and the larger the sample size required for a given confidence level. If the population is subdivided into too many strata, there is the risk of having too few cases within each of these cells to make statistically valid and reliable comparisons.

Variable

A characteristic of an individual or case. The characteristic must be able to take on more than one value and must be measurable.

Proportionate and Disproportionate Stratified Samples

Samples that are randomly selected within each stratum can be proportionate or disproportionate to the population size. In a proportionate stratified sample, the number of cases selected from each stratum is based upon the subgroup's size relative to the population size. That is, if 40% of the total population resides in Region 1, then residents of Region 1 (the stratum) should comprise 40% of the total sample as well. And, if only 10% of the population resides in Region 2, then Region 2 should comprise only 10% of the sample in a proportionate sample. This is the simplest method of stratification – the number of individuals or elements taken from each stratum or group is proportionate to their distribution in the overall population. While a simple random sample of sufficient size should also result in fairly accurate proportions of each subgroup, proportionate stratified sampling *guarantees* that the subgroups will be proportionate to their known sizes in the population. This helps ensure that key subgroups are accurately represented in the sample. The table below illustrates an example of a proportionate sample, stratified by regional subgroups. In this example, the overall sample of 513 represents a confidence level of 95% and a margin of error of +/-4%. The sample size of each region is based upon the region's size relative to the population. For example, 14% of the overall population lives in the West region, so 14% of the total sample was randomly selected from that region. A proportionate stratified sample is used to ensure that the regional distribution of the sample members matching the regional distribution of the population.

Proportionate Sample

Subgroup	Population		Sample	
	No.	% of Pop.	No.	% of Sample
(1) East	500	14%	72	14%
(2) West	700	20%	103	20%
(3) Central	850	24%	123	24%
(4) North	300	9%	46	9%
(5) South	1150	33%	169	33%
TOTALS	**3500**	**100%**	**513**	**100%**

In a disproportionate stratified sample, the size of each sampled subgroup is not proportionate to its size in the overall population. Some subgroups may be over-sampled or under-sampled relative to their actual proportion in the population. This type of sampling method is typically used when states want to ensure adequate representation of smaller subgroups within a population. For example, states may choose to over-sample or disproportionately select cases from smaller size regions, or from low incidence disability groups, or from ethnic minority groups because proportionate samples would yield too few cases from these small groups. Further, for some types of statistical analyses, a minimum sample size is needed for each subgroup, and disproportionate sampling may help achieve the necessary sample size for each stratum. It is best to consult a statistician about sample size if your state plans to examine subgroups.

The table below illustrates an example of a disproportionate stratified sample. The total sample of 625 will yield an overall confidence level of 95% and a margin of error of +/-4%. In this example, the goal was to obtain a minimum of 125 sample members from each ethnic subgroup, to allow for statistical comparison between subgroups. To ensure adequate representation of each ethnic subgroup, certain subgroups (African-American, Hispanic, and Asian) were over-sampled, while the Caucasian subgroup was under-sampled. For example, 20% of the total sample was drawn from the Asian subgroup (to achieve a minimum number of 125 Asian respondents), although only 8% of the cases in the overall population were from this subgroup. On the other hand, 40% of the sample respondents were Caucasian, compared to 65% in the overall population.

Disproportionate Sample

Subgroup	Population		Sample	
	No.	% of Pop.	No.	% of Sample
Caucasian	6500	65%	250	40%
African-American	1500	15%	125	20%
Hispanic	1200	12%	125	20%
Asian	800	8%	125	20%
TOTALS	**10,000**	**100%**	**625**	**100%**

One cautionary note about reporting data based upon disproportionate samples – if a state plans to report on an estimate about the sample (e.g., proportion with unmet need in transportation), the estimate(s) must be statistically "weighted" back to the population; this is true for estimates based on the entire sample or subgroups (strata). The weighting process accounts for the under- and over-representation of strata that occurred during sampling. States should consult with a statistician or survey methodologist to ensure proper weighting techniques are applied.[4]

[4] Also, if there is interest in estimating statistical differences between subgroups or conducting multivariate analyses (e.g., regression), professional help is recommended in controlling for the "design effect" which results from disproportionate sampling. Many statistical software packages have the ability to weight data and account for design effect.

Sources of Error

Neither a probability sample nor a large sample alone can guarantee accurate or reliable results. Indeed, a large sample obtained through non-probability sampling methods is not considered as representative as a smaller-sized sample obtained through probability methods. Non-sampling factors can influence the accuracy of the sample and the credibility of the findings. There are four types of error that can affect the accuracy of sampled data:

- sampling error
- non-response error (surveys)
- coverage error
- measurement error

States should pay attention to all of these in their attempt to produce accurate, reliable, and credible results.

Sampling Error

As discussed earlier, sampling error is the amount of discrepancy between the characteristics of the sample and the "true" population values. It occurs due to chance and can be estimated statistically. The generalizability of the sample to the population is limited to the extent that sampling error occurs. The best way to minimize sampling error is to use probability sampling methods that employ random selection techniques.

Non-Response Error

Non-response error occurs when an appreciable number of participants initially selected in a survey sample do not participate in the survey and these non-respondents systematically differ, in terms of key characteristics, from participants who do respond. This type of error is a common problem in survey research. Potential respondents may decline to participate for a variety of reasons, including lack of interest, lack of time, concerns about privacy and confidentiality, discomfort with being interviewed, or inability to give informed consent. Another common cause of non-response may be that program participants are away from home at the time of the survey (e.g., at school, doctor's appointment, on vacation, shopping, out with friends, in the hospital, in a nursing home).

The response rate is calculated by dividing the number of individuals who actually participate in a survey (numerator) by the number of eligible participants who were selected into the sample (denominator). For example, if a state randomly selects 300 waiver participants from its target population for participation in a satisfaction survey, and 225 actually participate (i.e., 75 refuse to participate or cannot be located), then the response rate is 225/300 or 75%.

What is a reasonable response rate for a survey? Similar to the sample size question, there is no one answer. A reasonable response rate depends upon the goals of the survey, the nature of the population being surveyed, the degree of potential non-response error that the state and its constituencies are willing to tolerate, how the results will be used, and what other sources of information will be used to supplement or validate survey results.

Some researchers warn that response rates under 60 to 70% are a red flag, and some agency standards require a minimum 75% response rate. Response rates are usually higher if the inquiries are relatively short, well designed, and relevant to potential respondents. Surveys that use in-person interviews tend to yield higher rates of response, while mail surveys lower ones. Surveys that include follow-up or reminder notices to potential eligible respondents are likely to produce higher response rates than surveys that contact potential respondents only once. In order to improve response rates, some studies offer incentives (such as payment) to individuals for their participation. Incentives, however, may not be practical or appropriate for state agency surveys. Other common approaches to minimize non-response error include:

- clearly explaining the survey and its purpose to potential respondents
- ensuring that confidentiality of data will be protected to the extent possible
- sending reminder notices
- following up by telephone

When conducting a study, an analysis of the respondents and non-respondents is highly recommended. Such analyses can be used to identify systematic bias in the results. (See also *Determining Sample Size – Non-Response Rate*.)

Coverage Error

Coverage error occurs when the sampling frame (the list used to draw the sample) is incomplete or inaccurate, and therefore does not include all individuals or cases in the target population. Any discrepancy between the target population and the actual list used for the sampling frame is a source of potential bias. For example, if the target population of a state survey is intended to include adults with disabilities age 18 years and older, but the agency list used to draw the sample only contains adults age 22 and over, then coverage error is a source of bias, as the sample will systematically exclude people age 18 to 22. To avoid coverage error, states need to use up-to-date, accurate sampling lists consistent with the clearly defined target population and inclusion and exclusion criteria. Death is another common source of coverage error – especially in the elderly population. (See also *Defining the Target Population –Sampling Frame and Units of Analysis.*)

Measurement Error

Another non-sampling source of error is measurement error, which refers to inaccuracies or ambiguities in the measurement or collection of data. Measurement error can stem from poorly worded survey questions, ambiguous response options or coding criteria, inadequately trained interviewers, improper administration of a survey, or respondents who cannot provide reliable survey responses. While these sources of error do not relate directly to sampling issues, they are equally important considerations that states need to consider. It is strongly recommended that, whenever possible, states select measures or tools that have been professionally assessed for reliability and validity. And, if such measures are not available, the state can reduce measurement error by careful attention to the construction and wording of survey questionnaires and thorough training and preparation of interviewers.

Appendix A: Resources

Handbooks and Texts

Alreck, P.L. and Settle, R.B. (1995)
The Survey Research Handbook: Guidelines and Strategies for Conducting a Survey, 2nd edition
New York: Irwin Professional Publishing
This survey handbook includes a detailed chapter (with step-by-step guidelines) on designing the sample, reliability and validity, sample size determination, and sample selection methods.

Fink, A. (1995)
The Survey Handbook
Thousand Oaks, CA: SAGE Publications
This nine-volume survey kit helps users prepare surveys and collect data. Volumes in the kit include how to: ask survey questions; conduct self-administered and mail surveys; conduct interviews by telephone and in person; design surveys; sample in surveys; measure survey reliability and validity; analyze survey data; and report on surveys.

Fink, A. (1995)
How to Sample in Surveys
Thousand Oaks, CA: SAGE Publications
This is Volume 6 of *The Survey Handbook* (see above reference), designed to guide the reader in selecting and using appropriate sampling methods. The handbook provides information about probability and non-probability sampling methods and statistical issues related to sampling, including calculation of sample size and acceptable response rate.

Henry, G.T. (1990)
Practical Sampling, Applied Social Research Methods Series, Volume 21
Thousand Oaks, CA: SAGE Publications
This book provides detailed examples of practical sampling designs related to sample selection, sampling frames, sampling techniques, sample size considerations, and post-sampling choices.

Salant, P. and Dillman, D. A. (1994)
How to Conduct Your Own Survey
New York: John Wiley & Sons
This helpful handbook about designing and conducting practical surveys includes a chapter that discusses when to use sampling and how to select a sample.

Web Resources

National Audit Office Publication: A Practical Guide to Sampling
http://www.nao.gov.uk/publications/samplingguide.pdf
This guide provides helpful information about sample design, sampling methods, interpreting and reporting the results. The guide provides case examples and colorful graphics.

The Survey System Sample Size Calculator
http://www.surveysystem.com/resource.htm
A web-based public service of the Creative Research Systems – an on-line "calculator" used to determine how large a sample is needed in order to get results that reflect the target population as precisely as needed.

The Research Methods Knowledge Base
http://www.socialresearchmethods.net/kb/index.htm
This web-based textbook by William M. Trochim at Cornell University addresses topics in a typical introductory undergraduate or graduate course in social research methods including: formulating research questions; sampling (probability and nonprobability); measurement (surveys, scaling, qualitative, unobtrusive); research design (experimental and quasi-experimental); and data analysis. The sampling section is quite basic and uses helpful graphics.

W.K. Kellogg Foundation Evaluation Handbook
http://www.wkkf.org
This handbook provides a framework for thinking about evaluation as a relevant and useful program tool. While this handbook does not specifically address sampling issues, it discusses many important issues that relate to sampling, such as identifying stakeholders, developing evaluation questions, determining data collection methods, collecting data, and analyzing and interpreting data. The handbook can be found on the website in the "Publications and Resources" section under "Toolkits."

Sage Publications Website
http://www.sagepub.com
This publishing company offers reference books on research methods and evaluation (including some titles listed above) for order on their website.

Appendix B: Glossary

Availability or Convenience Sample
A non-probability sampling method of selecting readily available individuals or cases into the sample.

Census
A process used to collect information about the full population, as opposed to a sample or subset of the population.

Cluster Sample
A multi-stage process typically covering broad geographic areas or organizational units, used when a complete centralized sampling frame is not available. Key geographical groups or clusters are identified; then a random sample of these clusters is selected; and then cases within the randomly selected clusters are selected into the sample.

Confidence Interval
A statistical estimate of the range of values within which the true population value is likely to fall. Confidence intervals are often denoted by a single number that identifies the margin of error, such as + or – 5%.

Confidence Level
A statistical estimate used in random sampling, stated as a percentage, of the degree of certainty that the true population value is within a specified range of values.

Coverage Error
A source of bias that occurs when the sampling frame (the list used to draw the sample) is incomplete or inaccurate, and therefore does not include all individuals or cases within the target population.

Disproportionate Stratified Sample
A type of probability sampling method in which the number of cases selected from each stratum of the population is disproportionate to the overall population size – that is, some subgroups may be over-sampled or under-sampled relative to their actual size in the population.

Exclusion Criteria
Rules for defining which individuals or cases are excluded from the sampling frame.

Inclusion Criteria
Rules for defining which individuals or cases are included in the sampling frame.

Measurement Error
A source of bias resulting from inaccuracies or ambiguities in the measurement or collection of data, such as poorly worded survey questions, ambiguous response choices in question items,

inadequately trained interviewers, or respondents who cannot provide reliable survey responses.

Non-Probability Sample
A sample drawn without using random selection procedures. The likelihood of selecting any one case from the population into the sample is not known and is usually different for each person or case in the sample.

Non-Response Error
A source of bias that occurs when an appreciable number of individuals in the sample do not respond/participate and these non-respondents differ in terms of key characteristics from individuals who do respond.

Non-Response Rate
A number that describes the proportion of individuals selected into the sample who do not respond/participate, typically due to the inability to locate the individuals, ineligibility, or their refusal to participate.

Non-Sampling Errors
Types of errors due to flaws in the design of how the sample is drawn or how the data are collected. Non-sampling errors cause bias in one direction or another and cannot be estimated mathematically.

Population
The full universe of individuals or entities from which the sample is drawn.

Population Parameter
A number that represents the true value or occurrence of something in the total population. The theory behind sampling is that the values obtained from a sample will approximate or estimate the population parameters; however, exact population parameters can only be obtained through a complete census.

Power Analysis
A technique used by statisticians to decide how large a sample is needed to make statistically accurate and reliable judgments, as well as how likely the selected statistical tests will be able to detect significant differences.

Probability Sample
A sample drawn according to random selection procedures in which every member of the target population has a known, non-zero chance of being included in the sample.

Proportionate Stratified Sample
A type of probability sampling method in which the number of individuals or cases selected from each stratum of the population is based upon the subgroup’s size relative to the overall population size.

Purposive Sample
A non-probability sampling method that involves selecting “typical” individuals or cases from the population based upon professional experience, knowledge, or judgment.

Quota Sample
A non-probability sampling method that involves setting a quota for inclusion of specified numbers of individuals or cases with certain characteristics, and then selecting cases on an availability basis.

Random Sample Selection
A process, based on scientific probability theory, that ensures individuals or cases in a population have an equal chance of being selected into the sample.

Representative Sample
A sample is considered representative of the population if the characteristics of the sample (e.g., age, gender, type of disability) are similar to the distribution of these characteristics in the overall population.

Response Rate
A number (expressed as a percentage) that describes the proportion of individuals who actually participate in the inquiry (the numerator) divided by the number of eligible respondents who were selected from the population and asked to participate (the denominator).

Sample
A subset of individuals or cases selected to represent a particular population.

Sample Statistic
A number that represents the value or occurrence of something in the sample.

Sampling Error
The amount of discrepancy between the characteristics of the sample and the true population values. Sampling error is due to chance and can be estimated mathematically.

Sampling Frame
The list of all units from which the sample is drawn.

Significant Difference
A term used to describe an observed result that cannot be attributed to sampling error alone. A finding is described as statistically significant if the probability of obtaining such a difference by chance alone is very low (for example, 5 in 100, if the significance level chosen is 95%).

Simple Random Sample
A probability-based sampling method that ensures that each member of the population has an equal probability of being selected into the sample (as if pulling individual names out of a hat).

Snowball Sample
A non-probability sampling method that involves a chain-like referral process. Initial contact is made with known individuals in the population, and then these individuals are asked to refer others for inclusion in the sample.

Strata
Subgroups defined within a population.

Stratified Sample
A multi-stage probability sampling method in which the population is first divided into homogeneous strata or subsamples (grouping individuals or cases based on characteristics they share) and then random samples are selected from each stratum.

Systematic Sample
A probability-based method in which individuals or cases are selected at regular intervals from the sampling frame (e.g., every 10th name).

Target Population
The population of interest in the study; the larger group from which the sample is drawn and which the sample is intended to represent.

Unit of Analysis
The element about which information is being collected and analyzed.

Variable
A characteristic of an individual or case. The characteristic must be able to take on more than one value and must be measurable.