

# OFFICE OF CHILD SUPPORT ENFORCEMENT

An Office of the Administration for Children & Families



## Statistical Sampling

### Techniques for Effective Management of Program Operations (TEMPO)

Published: April 11, 2002

Information About: **State/Local Child Support Agencies**

Topics: **Federal Reporting, Self-assessment Reporting System**

#### Abstract

The Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) of 1996 revised federal audit requirements to focus on data reliability and to assess performance outcomes instead of determining compliance with process steps. PRWORA also amended the Social Security Act (the Act) by requiring each State to conduct an annual review of its Child Support Enforcement (IV-D) program to determine if federal requirements are being met and to provide an annual report to the Secretary of the Department of Health and Human Services on the findings. The annual self-assessment's purpose is to give a State the opportunity to assess whether it is meeting federal requirements for providing child support services and providing the best services possible. It is to be used as a management tool, to help a State evaluate its program and assess its performance.

#### What Constitutes Developing a Sample

Thoughtful planning and careful implementation are two basic characteristics of developing a good self-assessment. The success of a self-assessment review depends to a great degree on the extent to which the sampling methodology has been thought through by the self-assessment evaluator. In the planning stage of the self-assessment process, special emphasis is often given to creating the self-assessment instruments and the analyses of data. A critical concern in planning a review, however, is the determination of what child support cases should make up the universe or population from which sample cases should be selected for the self-assessment review. This is essential because choosing the cases can influence the quality of the data obtained from the assessment.

To determine the type of child support cases to include in the self-assessment review sample, you must first ask yourself the following questions. Your answers to these should then be used as a guide to determine which cases to include.

#### Table of Contents

##### *What Constitutes Developing a Sample*

##### *A "Focused" versus a "Statewide" Sample*

##### *The Focused Sample*

##### *The Statewide Sample*

##### *Defining the Sampling Frame*

##### *Target Population or Universe*

##### *Sampling Frame*

##### *Sampling Element*

##### *Determining the Type of Sampling Methodology to Use*

##### *Probability Sample Designs*

##### *Simple Random Sample*

##### *Systematic Random Sample*

##### *Stratified Sample*

##### *Non-Probability Sample Designs*

##### *Advantages and Disadvantages of Various Sample Designs*

##### *Factors to Consider When Designing a Sample*

##### *Precision*

##### *Accuracy*

##### *Complexity*

##### *The Confidence Level and Confidence Limits: Their Role in Sampling*

##### *What Confidence Limits Are and What They Are Not*

##### *What Constitutes Developing the Sample Size*

##### *Evaluating Sample Results*

##### *Step 1: Compute the Efficiency Rate*

##### *Step 2: Compute the Standard Error*

##### *Step 3: Construct the Confidence Interval*

##### *Examples of State Self-Assessment Sampling Procedures*

<b>North Dakota</b>
<b>Oregon</b>
<b>Conclusion</b>
<b>Bibliography</b>
<b>Appendices</b>
<b>Glossary of Terms</b>
<b>Sampling Table</b>
<b>Random Numbers Table</b>

**Figure 1. Determining Cases to Include in the Sample**

QUESTION	CONSIDERATIONS IN DETERMINING YOUR ANSWER
<b>WHAT IS THE FOCUS OF THE SELF-ASSESSMENT REVIEW?</b>	Determining what you want to find out in your self-assessment review is an essential first step.
<b>WHO OR WHICH CASES ARE THE FOCUS OF THE REVIEW?</b>	For a typical self-assessment review, the object of the study can be categorized in one or two ways: It can represent an individual criterion within the self-assessment review. For example, all child support cases that have been closed within the review period, or all cases that require a review and adjustment may act as an individual criterion. This type of sample is often referred to as a “focused” sample by self-assessment evaluators as well as by Office of Child Support Enforcement (OCSE) staff. Alternatively, it can represent the program’s population of open child support cases on the State’s automated system. This is often referred to as a “statewide” sample by fellow self-assessment evaluators as well as by OCSE staff.
<b>WHEN IS THE REVIEW BEING DONE?</b>	When to conduct the review is dictated by Federal regulations (45 CFR 308.1(d)). Per the regulations, the self-assessment review should occur annually with a review period of one year. You should keep this parameter in mind when developing the sample.
<b>WHY IS THE CHILD SUPPORT PROGRAM CONDUCTING THIS REVIEW?</b>	The most important question is why. The answer should contemplate increasing the performance of the State’s child support program. You may need to reflect upon the State’s goals and objectives for the fiscal year as well as the national strategic objectives. The answer should assist in channeling your thoughts and approach for planning the sampling methodology.

Carefully thinking through the questions and points above—and then scrutinizing their applicability in the self-assessment review methodology—will lead to a more precise assessment producing valuable performance information derived from a sample that truly represents the child support program criteria being reviewed.

**A “Focused” versus a “Statewide” Sample**

Now that we have talked about what constitutes developing a sample, let’s spend some time talking about some self-assessment sampling terminology. When discussing self-assessment reviews, we often hear references to a “focused” or “statewide” sample. However, when self-assessment evaluators turn to the statistical textbooks, there is nothing in the glossary or index pertaining to either of these terms. They are explained below.

**The Focused Sample**

The term “focused sample” is really not a statistical term. It was developed out of necessity by both OCSE and our federal/state self-assessment core workgroup. Essentially, both parties needed a term to describe a simple method of extracting a sample that used the eight self-assessment criteria as a way of differentiating the child support caseload. [1] In other words, a focused sample is just a sample whose target population uses each of the eight self-assessment criterion as eight discrete samples.

**The Statewide Sample**

The term “statewide sample” was also developed out of necessity. Essentially, the self-assessment core workgroup adopted the term from the old program audits conducted by our Office of Audit (OA) prior to the passage of the PRWORA. For the old program audits, the auditors extracted a random sample of cases from States’ continuous caseload in order to determine compliance with the Federal regulations. The OA called this process, of extracting a random sample from a State’s entire caseload without differentiating between subject matter, a statewide sample.

**Defining the Sampling Frame**

For State self-assessment evaluations, the sampling process determines which cases will be included in the evaluation. Sampling makes gathering data on the performance of the child support enforcement program more manageable and more affordable. It enables the characteristics of States’ child support programs to be inferred with minimal errors from information collected on relatively few cases. Given this, the first thing you have to determine is the target population from which you will draw the sample. To determine this, we first need to define the difference between some common statistical terms: “target population or universe,” “sampling frame,” and “sampling element.”

## Target Population or Universe

The “target population” for the self-assessment review is the cases or groups of cases about which information is desired. This is sometimes also referred to as the “universe.” It is the group to which one wishes to generalize or make inferences from the sample. In defining the target population or universe there must also be sample inclusion criteria, such as all open cases on the State’s child support system. There must also be exclusion criteria, such as cases with no action required during the review period. Taking the time to determine the inclusion and exclusion criteria can save you time and effort during the review process.

## Sampling Frame

The “sampling frame” is the list of cases from the target population from which you will draw the sample. It is the actual operational definition of the target population or universe. In essence, it is the designation in concrete terms of which cases will be included in the sample. For example, to extract a sample for a self-assessment review, the universe or target population could be a State’s entire IV-D caseload. The sampling frame is a list of all the cases minus those cases subject to the exclusion criteria. In the terminology we are using, this would be called a statewide sample.

Conversely, if States wish to pull a focused sample, then the sampling frame is simply delineated by the eight self-assessment criteria. In this case, a State would have eight sampling frames and each frame would represent its own particular “focus” or self-assessment criterion.

While reviewing self-assessment reports, we have found that several States have combined methods of sampling, thereby employing the use of both statewide samples and focused samples. Their combinations reflect the capabilities of their automated systems. For example, one State chose to create a sampling frame combining four self-assessment criteria: establishment of paternity and support order, expedited process, securing and enforcing medical support orders, and enforcement of orders. To sample the remaining criteria, the State constructed discrete sampling frames that were focused on the remaining four criteria: case closure, review and adjustment of orders, interstate services, and disbursement of collections. Sometimes there are problems with the sampling frame. For example, the self-assessment evaluator needs to make sure not to overstate or understate the sampling frame. Overstating the frame would result in cases not applicable to the criteria being reviewed. Understating the frame would result in the elimination of cases from the evaluation process and quite likely bias inferences drawn from the cases sampled. The best way to cope with sampling frame definition problems is through careful planning of the sampling methodology. If this does not occur, the sample is likely to be plagued with biases because cases that should be included in the sample frame are not and cases that should not be included are.

## Sampling Element

The “sampling element” refers to the case from which data will be collected during the self-assessment review process. Essentially, an element is that case about which data is collected and that provides the basis of analysis. [2] The sampling element for the evaluation should be clearly defined within the target population or universe. For example, as the self-assessment evaluator, do you want to sample cases by child, non-custodial parent, or custodial parent? Also, are you concerned with cases that do not yet have an order established or only those cases that require enforcement? You may need to go through several steps to reach the ultimate sampling element.

Furthermore, depending upon how you have structured your evaluation process, you may have several sampling elements during the course of the evaluation. For example, if you choose to “focus” your review on the eight self-assessment criteria you could have eight separate and distinct sampling elements. So if you wanted to extract a sample just for the establishment criterion, your sampling element will be all cases on the automated system that were open at the beginning and/or during the review period and require the establishment of paternity and/or a support order.

Now that we have defined some common statistical terms, we are ready to determine the target population from which the sample will be drawn, its sampling frame, and its sampling elements.

## Determining the Type of Sampling Methodology to Use

A critical step in selecting the sample is determining the type of sample design to be used. There are two principal types of sample designs:

1. Probability
2. Non-probability

A probability design relies on the laws of chance for selecting the sample, while a non-probability design relies on human judgment.

### Probability Sample Designs

There are three types of probability sample designs:

1. Simple random sample
2. Systematic random sample
3. Stratified sample

The approaches we will discuss are often used in combination with each other and grow increasingly more complex as we move from a simple random sample to a stratified sample. However, regardless of how complex the sample design becomes, the workhorse of probability sampling is systematic random sample design.

### Simple Random Sample

A fundamental point in selecting a sample is that every element in the sampling frame has a known, nonzero, and equal chance of being included in the sample. Given this, there are two most commonly used methods to extract a simple random sample. The first is a lottery and the second is a random numbers selection procedure. Irrespective of which method you decide to use, every element in the sampling frame should be assigned an identifying number.

To use a lottery method, the case ID numbers can be placed in a container and mixed together. Eventually, someone draws out numbers from the container until the desired sample size is attained.

To use a random numbers selection method, a random number selection tool (usually a piece of software such as MS EXCEL) produces a series of case ID numbers through a random numbers generation process. Each number is unique, has an identifier, and is independent of all the others. To identify the numbers of the cases to be included in the sample, the analyst first chooses a place to start. This can be done by simply closing your eyes and pointing. If you choose this method, decisions rules—such as which direction to move in—should be determined in advance. An example of a random numbers table is shown at the end of this Techniques for Effective Management of Program Operations (TEMPO) monograph (**Random Numbers Table**). In addition, the website **Randomizer.org** (<http://www.randomizer.org>) will randomize numbers for you.

### Systematic Random Sample

The systematic random sample procedures are similar to the simple random sampling procedures. Essentially, the analyst selects a random starting point and then systematically selects cases from the sampling frame at a specified sampling interval.

The starting point and the sampling interval is based on the required sample size. The sampling interval will be represented as (k) within this TEMPO. Let's determine the starting point for the following scenario: the State has 1,000 child support cases needing enforcement. You want a sample of 100 to determine why process has not been served in many of these cases. In order to get 100 cases we need to determine the sampling interval. To determine the sampling interval (k), divide the total number of the sampling frame by the desired sample size. So an equation would represent the following:

$$k = 1000/100$$

$$k = 10$$

This means that the analyst should count down ten cases after starting from the case chosen as the random starting point within the first to the tenth case on the sampling frame and continue to identify every tenth case until the one hundred cases are selected.

### Stratified Sample

A stratified sampling approach should be used when you want to make sure that certain cases are included in the self-assessment evaluation, or if you want certain cases to be sampled at a higher or lower rate than others. To use a stratified sampling approach, the entire sampling frame is divided into subgroups of interest such as: enforcement cases, establishment cases, closed cases, interstate cases. In fact, more complex stratification methods are possible. You may even choose to stratify by gender, race, income, child support order, or other subgroup. Once you have determined your strata, the population is then organized into homogeneous subset/strata (with heterogeneity between subsets). Then, the analyst uses a systematic random sampling process to select cases from each stratum. The strata may resemble the following, as shown in figure 2, with the asterisks representing the selected cases that would be pulled for the sample.

**Figure 2. Sample Systematic Random Sampling Strata<sup>[3]</sup>**

TYPE OF DESIGN	SELECTED EXAMPLES OF DRAWING A SAMPLE		
	STRATUM A	STRATUM B	STRATUM C
STRATIFIED SAMPLE (DIVIDE POPULATION INTO HOMOGENEOUS STRATA AND DRAW RANDOM-TYPE SAMPLE SEPARATELY FOR ALL STRATA.)	A1 A2 A3	B1 B2 B3	C1 C2 C3
PROPORTIONATE : SAME SAMPLING FRACTION IN EACH STRATUM	A4 A5	B4 B5	C4

The premise behind dividing the sample into strata is to identify the cases you want to include in the sample based on the purpose of the self-assessment evaluation. The principal reason for using a stratified sample is to ensure that cases from all of the eight self-assessment criteria are captured in the sample. For example, if there is only a small amount of a particular type of case within the caseload, the systematic or simple random sample design may result in none, or very few, cases being included. This happens simply because there is a very small probability that such cases will be included in the sample.

Given the above, if you have a particular interest in your State's interstate case population, but there are far fewer interstate cases within your caseload than any other type of case, you could extract a disproportionate sample as opposed to a proportionate sample.

Disproportionate sampling (sometimes referred to as "over sampling") varies the proportion of cases in the stratum as opposed to across the strata. For example, you may decide to take one out of every five interstate cases in the interstate stratum while you only take one out of every ten establishment cases in the establishment stratum. By doing this, you are extracting a disproportionate sample. This type of sample contains all the validity and randomization that a proportionate sample has.

### Non-Probability Sample Designs

There are circumstances that require the use of each type of sample design. For example, a lack of resources and inadequate statistical software may require the use of a non-probability sample design. While non-probability sample designs serve a number of useful purposes, it would be unlikely that this sample selection method would be used to evaluate a State's child support program for the purposes of Category I of the self-assessment. A State may, however, wish to employ a non-probability sample design to do an in-depth investigation of a particular group of cases for inclusion as a study piece in Category II of the self-assessment report. For example, if the self-assessment evaluator notices—during the self-assessment case review process—a locate problem with enforcement cases that have a child support order and an arrears balance of \$500 or less, then extracting a sample using a non-probability sample design will provide an adequate picture of the locate problem and, thereby, provide important performance information. There are several methods for extracting a non-probability sample design. The one most often used is the "chunk sample." A chunk sample is a group of people who happen to be available at the time of the study (e.g., people waiting in a waiting room at a hospital or people walking through a mall). Using our locate example above, a chunk sample may be the population of cases within the closest field or county office.

### Advantages and Disadvantages of Various Sample Designs

The designs we discussed above have several advantages and disadvantages. For the purposes of self-assessment reviews, OA recommends using the systematic random sample design. OA recommends this design because it is easy to use and will yield very useful information for the evaluator.

Figure 3 below outlines the advantages and disadvantages of each of the sampling designs. Using this table can help you determine which design best suits the type of self-assessment evaluation you intend to perform, given the administration and the size of your child support program.

**Figure 3. Advantages and Disadvantages of Different Probability Sample Designs [4]**

DESIGN	ADVANTAGES	DISADVANTAGES
<b>SIMPLE RANDOM</b>	Requires little knowledge of population in advance.	May not capture certain groups of interest. May not be very efficient.
<b>SYSTEMATIC</b>	Easy to analyze data.	Periodic ordering of elements in the sample frame may create biases in the data. May not capture certain groups of interest. May not be very efficient.
<b>STRATIFIED</b>	Enables certain groups of interest to be captured. Enables disproportionate sampling and optimal allocation within strata. Highest precision.	Requires knowledge of population in advance. May introduce more complexity in analyzing data and computing sampling errors (standard errors).

### Factors to Consider When Designing a Sample

There are three primary factors to consider when designing a sample:

1. Precision
2. Accuracy
3. Complexity

These three factors work together to ensure that when the sample elements (that is, the child support cases) are extracted and analyzed, they will provide solid performance information. For example, statistical *precision* is directly reflected in how *accurately* the sample reflects the caseload.

## Precision

Sampling is all about precision. Precision refers to how close the estimates derived from a sample are to the true population. It refers to the tolerable *deviation* of values around the estimate and this can be expressed in terms of units of standard errors. These units of standard error are also known as the “standard deviation.”

Standard error/standard deviation indicates the extent to which the sample estimates are distributed around the population parameter, i.e. mean. <sup>[5]</sup> The standard error is actually a function of probability theory. This theory states that a certain proportion of sample estimates will fall within specified increments—each equal to one standard error—from the population parameter. <sup>[6]</sup> The incremental arrangement of sample estimates is referred to as a distribution. The most familiar and commonly used distribution to illustrate the presentation of sample results is the normal distribution. The normal distribution indicates that 68 percent of all sample estimates will fall within plus or minus one standard deviation of the population mean. Therefore, approximately 34% (.3413) of the sample estimates will fall within one standard error increment above the population parameter and 34% will fall within one standard error below the parameter. <sup>[7]</sup> This standard error is considered to be random error and is a function of both the size of the standard error and the sample size. For example, if the standard error does not change and the sample size is increased, then the standard error becomes smaller. The logic surrounding this states that, given a random selection procedure, a large sample is more likely to provide a precise estimate of a population than a small sample.

Probability theory also dictates that approximately 95% of samples will fall within plus or minus two standard errors of the true value, and 99.9% of the sample will fall within plus or minus three standard errors. <sup>[8]</sup> Given this, the proportion of samples that would fall between one, two, or three standard errors of the population parameter is constant for any random sampling procedure. For example, 68% of the sample means (of the same size) will fall within the range of plus or minus one standard error (1.00), 95% will fall within plus or minus 1.96 standard errors, and 99% within plus or minus 2.58 standard errors.

Random error occurs when there is some difference between the sample statistic and the target population as it is expressed in the sample frame. This can occur even when proper random sampling procedures are used. This error occurs because of chance factors that influence the outcomes of the sample selection procedures. Usually this type of error is found in smaller samples—thus, the smaller the sample, the greater chance there is for random error. Further, larger samples tend to have less random error simply because the sample is bigger and there is a greater chance of capturing all the nuances of the target population. Given that the standard error is also a function of the sample size, the standard error will always be reduced by half if the sample size is quadrupled. <sup>[9]</sup> For example, if you have a sample of 500 child support cases with a standard error of 5%, to reduce the standard error to 2.5%, you must increase the sample size to 2000 cases.

## Accuracy

Accuracy refers to how closely the estimates from the sample are to the true population as a function of systematic error. Accuracy relates most specifically to something called “systematic error,” which is also referred to as “bias.” This occurs when there is a flaw in the actual sampling procedure so that not all elements in the population had an equal and independent chance of being included in the sample. Essentially, systematic error is the result of a flawed sampling frame. The only way to correct systematic error is to revise the sampling frame using the what, who, when, and why formulae discussed at the beginning of this TEMPO. (see **Determining Cases to Include in the Sample**)

## Complexity

Complexity is important because it forces the evaluator to think about the amount of information that must be gathered in advance of doing the self-assessment evaluation. For example, the evaluator must think about what he or she is trying to assess. For self-assessment, the State is trying to assess compliance with the standard for each criterion.

## The Confidence Level and Confidence Limits: Their Role in Sampling

When designing a sample strategy for a self-assessment evaluation, you will note the reference to a requirement that States' sample sizes must have a minimum confidence level of 90%. Many self-assessment analysts have asked: *What is a 90% confidence level and how do I incorporate this into a sample design?* The 90% confidence level implies that you are 90% sure that the sample *mean* represents the population *mean*. The *mean* is the sample average.

Essentially, confidence levels/limits and intervals work together to provide the evaluator with information to make informed decisions about how large the sample size should be.

Applying a confidence level to a sample design helps the self-assessment evaluator cope with a large standard error and imprecise estimates. An imprecise estimate refers to not knowing exactly what the target population's *mean* is. For example, if you have information that leads you to believe that your establishment caseload has a lot of variation in it, (simply because you cannot get an accurate count of the number of establishment cases), you can

improve the quality of the sample by drawing a larger sample. Increasing the sample size can, however, reach a point of diminishing returns—whether it does this depends upon the amount of variation a measurement has in the population. As we stated earlier in this TEMPO, as a general rule the sample size must be quadrupled to reduce the standard error by half.

The confidence level also involves “confidence intervals.” Essentially, confidence intervals are two estimates that consist of an upper and lower value (also referred to as upper and lower bound) with the mean falling somewhere between the upper and lower value. There is an inverse relationship between the level of confidence and the precision or width of a confidence interval. The greater the confidence, the wider the limits and the less precision. For example, if you used confidence limits to guess the mean age of non-custodial parents in your caseload and you use very wide limits, such as 16 to 75 years, you will have greater confidence that the intervals include everyone in the caseload but very little precision because the intervals are so wide that they fail to provide precise information. Conversely, narrow limits give you precision but reduce your confidence. In other words, confidence increases as the margin of error increases.

**What Confidence Limits Are and What They Are Not**

If you say that the 95% confidence limits for the estimated mean age of non-custodial parents in the United States are 26 to 36, this does not mean that there is a 95% chance that the true mean lies somewhere in that range. The true mean may or may not lie within that range and we have no way to tell. What we can say, according to H. Russell Bernard in *Social Research Methods*, [10] is that:

1. if we take a very large number of random samples, and
2. if we calculate the mean, and the standard error for each sample, and
3. if we then calculate the confidence intervals for each sample mean, based on a standard error of 1.96, then
4. 95% of these confidence intervals will contain the true mean.

This rubric is derived from the “Central Limit Theorem.” According to this theorem, if you take many samples of a population, and if the samples are big enough, then the mean and the standard deviation of the sample mean will approximate the true mean. Further, the standard deviation of the population and the distribution of sample means will approximate a normal distribution. Given this, when we are doing self-assessment evaluations and we do not have the time or resources to take 10 or 50 samples, we have to derive efficiency rates for each reviewed criteria and make program decisions based on the findings of one sample.

**What Constitutes Developing the Sample Size**

A frequently asked question in self-assessment is *How large should the sample be?* The answer to this question is not as straightforward as some may wish to believe, but OA has provided the following guidelines:

**Figure 4. OA Guidelines for Determining Sample Size**

1. If a state uses a "focused sample", a minimum of 100 cases per criterion should be selected.
2. If a state uses a "statewide" sample, a minimum of 500 cases should be selected. Further, at least 50 cases should be selected for the most infrequently occurring criterion in the IV-D caseload.
3. Results of the sample should be evaluated using the confidence interval method and the results must be projected statewide.

Confidence levels and confidence intervals provide not only a statement of accuracy, but they also provide the basis upon which to determine the appropriate sample size for an evaluation. There are varieties of formulae associated with determining the necessary sample size and the Office of Audit has provided States with the above guidelines to be used as a starting point. Because there are so many formulae and all of them provide reasonable sample estimates, this TEMPO will not designate one formula. Instead, we will discuss the rationale and the logic behind determining the sample size. The figure below provides criteria and logic for estimating the sample size as it outlines the general steps for selecting the sample and determining the necessary sample size.

**Figure 5. Criteria and Logic for Estimating Sample Size [11]**

STEP	DESCRIPTION

1	Identify the portion of the caseload to be evaluated or assessed. This could be the entire caseload or just a portion of it.	
2	Select the sub groups in target population. This step is only necessary if you choose a stratified sample design.	
3A	Indicate what you expect the population value to be.	Determining which one to use depends upon whether you are sampling as a proportion to size or for the mean, which assumes that the estimate of interest can be specified or obtained. In other words, you are able to determine the number in the target population/universe.
3B	Estimate the standard deviation of the estimate.	
4	Decide on a desired level of confidence (federal minimum for self-assessment is 90%).	
5	Decide on a tolerable range of error in the estimate (precision).	
6	Compute the sample size based on the above assumptions.	

There are tables available that assist with sample size estimation. Usually, these tables assume a 95% confidence level and they can be found in any statistics text book. A sample estimate table is also presented in the appendix of this TEMPO (see **Sampling Table.**). This will help the evaluator estimate the sample size based on the amount of tolerable sampling error.

### Evaluating Sample Results

On page 19, we stated that confidence levels and confidence intervals provide not only a statement of accuracy, but they also provide the basis to determine the appropriate sample size for an evaluation. OA suggests using the confidence interval method to evaluate the sample results of the self-assessment evaluation. By doing this, the evaluator can be sure that the sample size accurately reflects the State's child support caseload. To evaluate the sample results, each State should follow the procedures shown below in figure 6.

**Figure 6. Procedures for Evaluating Sample Results**

STEP	PROCEDURE
1	Compute the Efficiency Rate
2	Compute the Standard Error
3	Construct the Confidence Interval

#### Step 1: Compute the Efficiency Rate

To compute the efficiency rate, an unbiased estimate should be used. This unbiased estimate can be thought of as an efficiency score for each of the eight criterion reviewed. We call it an unbiased estimate because it is computed using interval data whereby no prejudice can be brought into the calculation. To calculate the efficiency rate, compute an unbiased estimate for each criterion using the following formula:

The formula for the efficiency rate states that:

1. The number of errors should be subtracted by the number of cases reviewed and this is represented by the letter (x)
2. The sum of (x) should be divided by the number of cases reviewed and this is represented by the letter (y).
3. The sum of (x)/(y) gives you the efficiency rate represented by the letter (p).

$$(Number\ of\ Cases\ Reviewed - Number\ of\ Errors) / Number\ of\ Cases\ Reviewed = Efficiency\ Rate$$

or

$$x/y = p$$

#### Step 2: Compute the Standard Error

Once you have computed your unbiased estimate, otherwise known as the efficiency score, the standard error must be calculated. To calculate the standard error the following formula should be used:

The formula for standard error states the following:



**se** = standard error.

**p** = the efficiency score.

**f** = the percentage of the population contained in the sample. For example, if there are 350,000 cases in the population/universe, the sample represents 300 cases. To obtain the percentage, divide 300/350,000.

**n** = the number of cases in the sample

$$se = (\sqrt{p(1-p)(1-f)}) / (n - 1)$$

Given the above, to compute the standard error, we must first solve  $(1-p)(1-f)$ . Once we have multiplied  $(1-p)(1-f)$  we must then multiply its sum by  $p$ . Once we have multiplied by  $p$ , we must divide its sum by the sample size minus 1. From here, we simply calculate the square root. After we calculate the square root, we construct the confidence interval.

### Step 3: Construct the Confidence Interval

To construct the confidence interval, we must turn our confidence level into a standard deviation unit. For example, we ask States to sample using a confidence level of 90% for self-assessment. This 90% needs to be turned into a standard deviation unit to construct the confidence interval. The chart in Figure 7 below converts confidence levels, expressed as percentages, to standard deviation units based on a normal distribution curve.

**Figure 7. Conversion of Confidence Level to Standard Deviation**

CONFIDENCE LEVEL	STANDARD DEVIATION UNITS
99.9%	3.2905
99.5%	2.8070
99.0%	2.5758
98.0%	2.3263
95.5%	2.0000
95.0%	1.9600
90.0%	1.6449

Federal regulations stipulate that States should use a minimum confidence level of 90%, which converts to a standard deviation unit of 1.64. However, upon review of many self-assessment reports, OA found that many States are sampling at a 95% confidence level, which converts to a standard deviation unit of 1.96. Further, the universal standard for extracting random samples is typically 95% confidence level.

Once you have found your standard deviation unit for your sample, you are ready to construct the confidence interval. To construct the confidence interval, the following three steps should be followed:

#### **Figure 8. Steps for Constructing the Confidence Level**

1. *Multiply your computed standard error by your standard deviation unit. To represent 90%, multiply standard error by 1.64.*
2. *Compute the upper bound of the confidence interval. To do this, add the standard error and the confidence level computed in step 1 above to your efficiency rate.*
3. *Compute the lower bound of the confidence interval. To do this, subtract the value computed in step 1 above to your efficiency rate.*

As we stated earlier in this TEMPO, if you have a large interval between your scores, you have very little error but not a lot of precision. Since self-assessment's primary goal is to provide management information, it should be the goal of the evaluator to have both a small amount of error and the desired degree of precision. Without precision, program managers cannot make sound policy decisions. See **Sampling Table** in the Appendix for an at-a-glance method of calculating error and determining sample size at a 95% confidence level.

Evaluating the sample results should be applied to both statewide samples as well as focused samples. Our example above is computed such that it can be applied to either method. For those States that employ focused samples, the sample results should be computed for each sample pulled from each

sample frame, and the confidence interval should be calculated for each sample representing its discrete criterion.

## Examples of State Self-Assessment Sampling Procedures

As we mentioned earlier in this TEMPO, there are several formulae that could be used to extract a random sample and there are several methods by which States may present their sampling methods. Given this, we will use this section to present some selected State sampling procedures and demonstrate how these States include their sampling procedures in their annual self-assessment reports.

We selected two States' sampling procedures to present in this TEMPO. These States, North Dakota and Oregon, demonstrated sound sampling procedures that had the attributes defined earlier in Factors to Consider When Designing a Sample (page 14): precision, accuracy, and complexity. Both selected States, used focused samples, thereby delineating separate sampling frames for each of the eight criteria.

### North Dakota

The State of North Dakota places an emphasis on its statistical procedures by including their general sampling methodology in the introduction of their report. First, as mentioned above, the State uses a focused sample. Therefore, the State extracts eight samples, each representing one of the eight self-assessment criteria. We must also note that North Dakota's automated system has the capability to determine the population sizes of each of the eight self-assessment criteria. This makes it possible to extract a focused sample. Each of the eight samples was developed to represent estimates at a 90% confidence level. Applying a 90% confidence level to their sample selection procedures provided not only compliance with federally mandated sample selection criterion, but it also assured accuracy up to 90%.

Next, North Dakota wanted to make sure that accuracy did not overtake precision. To ensure that the State had a reasonable balance between precision and accuracy, it set its precision at + 5%. The State wanted to ensure that the confidence interval did not exceed 10% of the confidence interval's upper limit and lower limit. This allowed North Dakota to apply its results from its self-assessment review to its total population of cases, thereby permitting informed policy and management decisions based on its self-assessment review findings.

After using a focused sampling approach to extract a random sample, North Dakota then evaluated each sample extracted from their caseload. In doing this, they utilized the sample evaluation method we described above and applied it to each evaluated criterion. The State took the process a step further by applying an efficiency estimate by criterion to each region in their IV-D program. Therefore, the State had an efficiency estimate and a confidence interval by region as well as an efficiency estimate and confidence interval statewide for each criterion. The state then presented the information in a table that resembled that shown in Figure 9 below.

**Figure 9. North Dakota - Efficiency Estimate/Confidence Interval**

REGION	CASES REVIEWED			EFFICIENCY ESTIMATE	90% CONFIDENCE INTERVAL			
	TOTAL CASES REV'D	ACTION CASES	ERROR CASES		UPPER LIMIT	LOWER LIMIT	RANGE	RANGE/2
WILLISTON	9	8	1	90	99.2	56.1	43.1	21.6
MINOT	28	25	3	89	96.7	73.9	22.8	11.4
DEVILS LAKE	13	13	0	100	100	76.9	23.1	11.6
GRAND FORKS	16	15	1	93	99.6	72.3	27.3	13.7
FARGO	49	49	0	100	100	93.0	7.0	3.5
JAMESTOWN	20	20	0	100	100	84.0	16.0	8.0
BISMARCK	47	46	1	98	99.9	89.5	10.4	5.2
DICKINSON	18	17	1	94	99.6	74.8	24.8	12.4
STATEWIDE		193	7	97	98.3	93.4	4.9	2.5

This method allowed for performance-based decisions at both the statewide level as well as the regional office level.

Finally, North Dakota included the sampling parameters within their assessment report. For example, for each reviewed criterion, the State noted the following:

- Sample size
- Population size
- Brief explanation of how their automated system documented the sampling frame under study

The State also included a pie chart portraying the efficiency score for action cases verses error cases and the total sample size for the specified criterion.

## Oregon

The State of Oregon highlights their sample procedures in the Methodology section of the self-assessment report. In this section, the State explains that it used a focused sample to obtain the required 90% confidence level. To create discrete sample frames of each self-assessment criterion, the State had to first obtain population sizes for these criteria. Once population sizes were obtained, the State applied the following statistical equation to achieve their 90% confidence level:

$$n[(z \ 940/2)^2 \times p(q)] / E^2$$

The formula for Oregon's statistical equation to achieve their confidence level states the following:

**n** = the sample size

**z** = the z score

**940** = 1-confidence interval

**p** = probability

**q** = 1-p

**E** = tolerable error rate

The State then outlined its parameters for acceptability. For example, tolerable error was accepted at 5% and a presumed probability of 50 – 50 was used. This meant that there was a 50% chance that the desired results would occur and a 50% chance that the desired result would not occur. Oregon charted the number of cases required for both a 90% confidence level as well as a 95% confidence level.

Oregon utilized the what, who, when, and why formulae presented in the first section of this TEMPO (**Determining Cases to Include in the Sample**) to define which cases should be included in sampling frame. For example:

- **What?** Case closure
- **Who?** Any case closed
- **When?** Any case closed during the review period even if the case was subsequently reopened
- **Why?** To determine whether the State met the benchmark standard of 90% compliance

Oregon identified their case closed population of approximately 10,000 cases and calculated the sample size based on the formula mentioned above. The State calculated a sample size of 264 cases to review. However, the State anticipated a high exclusion rate and therefore over-sampled to compensate for this. Given this, Oregon randomly selected 371 cases and of the 371 cases, one quarter of these cases was excluded. Because the State excluded 139 of its cases, a second sample was necessary to meet the 264 case standard. Given this, a second sample of 117 cases was pulled for review.

The Oregon report also included the rationale for how they defined their eight program areas in the self-assessment report. The State explained that they broadly defined each of the eight program areas. Further, they were cognizant of the probability that some cases would be included which should have been excluded. This meant that the discrete population was actually smaller than what they identified. This allowed the State to capture all cases (for a particular criterion) they needed.

## Conclusion

The regulations prescribing the self-assessment approach (45 CFR 308.1) ask States to utilize a statistically valid sample to evaluate the caseload. Given this, there are two conditions that must exist for a sample to be considered statistically valid, according to *Sawyers Internal Auditing*. These conditions are shown in figure 10 below.

### **Figure 10. Conditions for a Statistical Sample**

1. *The sampling units must be randomly selected.*
2. *The sampling units must be quantitatively evaluated through the application of probability theory.*

The absence of either requirement defines the approach as non-statistical. <sup>[12]</sup> Given this, self-assessment samples should be defined using probability theory and, furthermore, they should be evaluated using the method outlined in this TEMPO. Evaluating the sample according to the method presented provides self-assessment analysts with assurance that their evaluation was accurate.

As we stated in the introduction of this TEMPO, determining what child support cases should be involved in the sample can influence the quality of the data. To ensure that we have the right cases—and therefore, quality data in the sample—we should consider the **what, who, when, and why** points. These points should be used as a guide to determine which cases should be included in the review sample.

As important as the what, who, when, and why points is what type of sample to extract. This TEMPO has discussed three probability sample designs:

1. Simple random sample
2. Systematic random sample
3. Stratified sample

Each of the three designs has its advantages and disadvantages and they should be considered closely.

The question we most often hear from States is *how big should a sample be?* As we discussed in this TEMPO, there are some things you can do to ensure that the sample you take provides robust information. First, the evaluator needs to ensure accuracy by making sure that every element in the designated population has an equal chance of being selected. Second, the evaluator needs to ensure precision. The way to ensure precision is by making the sample larger. However, the sample can only increase to a point before experiencing diminishing returns. Both accuracy and precision play a large role in determining the sample size.

To ensure accuracy, the evaluators should always ask themselves a series of questions such as:

1. Is the population of interest homogenous or heterogeneous?
2. Are there subgroups within the population? If there is more than one subgroup the evaluator needs to make sure that each subgroup is represented in the sample.
3. What are you trying to find out?

As we explained in this TEMPO, ensuring precision is complex but it is easier to fix errors related to precision than errors resulting from inaccuracy. Ensuring precision is really about probability theory and sample distributions. Small samples have a greater chance of random error, whereas larger samples tend to have less random error.

Credibility of the self-assessment review comes from the sampling methodology employed. An accurate and precise sample allows the evaluator to present the results of the self-assessment review with confidence. Furthermore, this provides program managers with confidence in making procedural and policy decisions based on the self-assessment review.

## Bibliography

Aday, Lu Ann. *Designing and Conducting Health Surveys*. San Francisco: Jossey Bass Inc., Publishers, 1996.

Babbie, Earl. *The Practice of Social Research*. Blemont, CA: Wadsworth Publishing CO (8<sup>th</sup> edition), 1998.

Bernard, H. Russell. *Social Research Methods*. London: Sage Publication 2000.

Sawyer, Lawrence. *Sawyers Internal Auditing*. Altamonte Springs, FL: Institute of Internal Auditors, 1988.

## Glossary of Terms

### Confidence Levels and Confidence Intervals

Confidence levels and confidence intervals are two key components of sampling error estimates. Confidence levels and confidence intervals allow us to express the accuracy of our sample statistics in terms of a level of confidence that the statistics fall within a specified interval from the parameter.

### Element

The sampling element refers to the unit or case from which information or data will be collected during the self-assessment review process.

#### Formula for precision

**Figure A.1**

$$[d = z 1 - \alpha/2 (SE p)]$$

The formula for precision states the following:

**d** = desired precision

**z 1 -  $\alpha/2$**  = Standard Error

**P** = estimate of proportion in the population

**SE** = Standard Error

#### Mean

The average. A measure that describes the center of a distribution of values.

#### Parameter

A summary description of a given variable in a population. For example, the mean, or average number of child support cases with a support order on the child support automated system.

#### Population/Universe

The target population for the self-assessment review is the cases or groups of cases about which information is desired. This is sometimes also referred to as the "universe."

#### Sampling Error

Probability sampling provides statistics exactly equal to the parameter they are to estimate. However, probability theory allows us to estimate the degree of error to be expected for a given sample design. <sup>[13]</sup>

#### Sampling Frame

The list of cases from the target population from which the sample will be drawn.

#### Standard Deviation

The square root of the variance. It describes the variability in a population or a sample.

#### Statistic

A summary description of a variable in a sample. For example, the mean of child support cases with a child support order in a sample.

#### Variable

A set of mutually exclusive attributes, such as gender, age, or case type.

#### Variance

A measure that describes the dispersion of values about the mean.

### Sampling Table

**How to use this table for estimating a sample based on the amount of error:** Find the intersection between the sample size and the approximate percentage distribution/efficiency score. The number at the intersection represents the sampling error calculated at a 95% confidence level. The error is expressed in percentage points of plus or minus.

**Example :** In a self-assessment statewide sample of 500 child support cases, 80 of the cases were in compliance while 20 percent were not in compliance. According to the table below, the sampling error is estimated at plus or minus 3.6. The confidence interval is therefore between 76.4 and 83.6. We could then say that we are 95% sure that the total population of child support cases that are in compliance is somewhere within that interval. If the State determines that this interval is satisfactory for creating management decisions for self-assessment evaluations, then a sample of 500 cases is adequate.

**Figure B.1. Sampling Table [14]**

SAMPLE SIZE	PERCENTAGE DISTRIBUTION (EFFICIENCY SCORE)				
	50/50	60/40	70/30	80/20	90/10
100	10	9.8	9.2	8	6
200	7.1	6.9	6.5	5.7	4.2
300	5.8	5.7	5.3	4.6	3.5
400	5	4.9	4.6	4	3
500	4.5	4.4	4.1	3.6	2.7
600	4.1	4	3.7	3.3	2.4
700	3.8	3.7	3.5	3	2.3
800	3.5	3.5	3.2	2.8	2.1
900	3.3	3.3	3.1	2.7	2
1000	3.2	3.1	2.9	2.5	1.9
1100	3	3	2.8	2.4	1.8
1200	2.9	2.8	2.6	2.3	1.7
1300	2.8	2.7	2.5	2.2	1.7
1400	2.7	2.6	2.4	2.1	1.6
1500	2.6	2.5	2.4	2.1	1.5
1600	2.5	2.4	2.3	2	1.5
1700	2.4	2.4	2.2	1.9	1.5
1800	2.4	2.3	2.2	1.9	1.4
1900	2.3	2.2	2.1	1.8	1.4
2000	2.2	2.2	2	1.8	1.3

**Random Numbers Table**

**Figure C.1. Random Numbers Table [15]**

39634	62349	74088	65564	16379	19713	39153	69459	17986	24537
14595	35050	40469	27478	44526	67331	93365	54526	22356	93208
30734	71571	83722	79712	25775	65178	07763	82928	31131	30196
64628	89126	91254	24090	25752	03091	39411	73146	06089	15630
42831	95113	43511	42082	15140	34733	68076	18292	69486	80468
80583	70361	41047	26792	78466	03395	17635	09697	82447	31405
00209	90404	99457	72570	42194	49043	24330	14939	09865	45906
05409	20830	01911	60767	55248	79253	12317	84120	77772	50103
95836	22530	91785	80210	34361	52228	33869	94332	83868	61672
65358	70469	87149	89509	72176	18103	55169	79954	72002	20582
72249	04037	36192	40221	14918	53437	60571	40995	55006	10694
41692	40581	93050	48734	34652	41577	04631	49184	39295	81776
61885	50796	96822	82002	07973	52925	75467	86013	98072	91942
48917	48129	48624	48248	91465	54898	61220	18721	67387	66575
88378	84299	12193	03785	49314	39761	99132	28775	45276	91816
77800	25734	09801	92087	02955	12872	89848	48579	06028	13827

24028	03405	01178	06316	81916	40170	53665	87202	88638	47121
86558	84750	43994	01760	96205	27937	45416	71964	52261	30781
78545	49201	05329	14182	10971	90472	44682	39304	19819	55799
14969	64623	82780	35686	30941	14622	04126	25498	95452	63937
58697	31973	06303	94202	62287	56164	79157	98375	24558	99241
38449	46438	91579	01907	72146	05764	22400	94490	49833	09258
62134	87244	73348	80114	78490	64735	31010	66975	28652	36166
72749	13347	65030	26128	49067	27904	49953	74674	94617	13317
81638	36566	42709	33717	59943	12027	46547	61303	46699	76243
46574	79670	10342	89543	75030	23428	29541	32501	89422	87474
11873	57196	32209	67663	07990	12288	59245	83638	23642	61715
13862	72778	09949	23096	01791	19472	14634	31690	36602	62943
08312	27886	82321	28666	72998	22514	51054	22940	31842	54245
11071	44430	94664	91294	35163	05494	32882	23904	41340	61185
82509	11842	86963	50307	07510	32545	90717	46856	86079	13769
07426	67341	80314	58910	93948	85738	69444	09370	58194	28207
57696	25592	91221	95386	15857	84645	89659	80535	93233	82798
08074	89810	48521	90740	02687	83117	74920	25954	99629	78978
20128	53721	01518	40699	20849	04710	38989	91322	56057	58573
00190	27157	83208	79446	92987	61357	38752	55424	94518	45205
23798	55425	32454	34611	39605	39981	74691	40836	30812	38563
85306	57995	68222	39055	43890	36956	84861	63624	04961	55439
99719	36036	74274	53901	34643	06157	89500	57514	93977	42403
95970	81452	48873	00784	58347	40269	11880	43395	28249	38743
56651	91460	92462	98566	72062	18556	55052	47614	80044	60015
71499	80220	35750	67337	47556	55272	55249	79100	34014	17037
66660	78443	47545	70736	65419	77489	70831	73237	14970	23129
35483	84563	79956	88618	54619	24853	59783	47537	88822	47227
09262	25041	57862	19203	86103	02800	23198	70639	43757	52064

[1] The eight review criteria are: case closure, review and adjustment, paternity and order establishment, enforcement, medical support enforcement, interstate, disbursement of collections, expedited processes.

[2] Babbie, Earl, *The Practice of Social Research* (Blemont, CA: Wadsworth Publishing CO, 8<sup>th</sup> edition, 1998), p. 202

[3] Adapted from *Designing and Conducting Health Surveys* by Lu Ann Aday (San Francisco: Jossey Bass Inc., Publishers, 1996), p. 119

[4] *Ibid.*, p. 122

[5] Babbie, p. 208

[6] *Ibid.*

[7] *Ibid.*, p. 202

[8] *Ibid.*

[9] Ibid., p.209

[10] Bernard, H. Russell. *Social Research Methods* (London: Sage Publication 2000), p. 169

[11] Aday, p. 147

[12] Sawyer, Lawrence. *Sawyers Internal Auditing* (Altamonte Springs, FL: Institute of Internal Auditors, 1988), p. 416.

[13] Babbie, p. 202

[14] Ibid.

[15] Adapted from Table of Random Numbers at <http://mnstats.morris.umn.edu/introstat/public/instruction/ranbox/random...>  
(<http://mnstats.morris.umn.edu/introstat/public/instruction/ranbox/randomnumbersII.html>)

---