

1500 CORRECTION FOR UNDERSAMPLING AND EXCESSIVE OVERSAMPLING

When using systematic random sampling, imprecise caseload projections or an unexpected drop rate will result in the State or Tribe not obtaining its target sample size. If the actual universe is larger than the estimated size, oversampling may occur. If the actual universe is smaller than the estimated size, undersampling may occur. A State agency and Tribal grantee must correct for undersampling to the extent necessary to meet sample size requirements for TANF reporting and a State agency must correct for undersampling to the extent necessary to meet sample size requirements for reporting of separate State programs. A State agency or Tribal grantee has the option as to whether or not to correct for excessive oversampling. However, we encourage States and Tribes to select larger than the minimum required annual sample size in order to increase the precision of statistics that are estimated from the sample data.

In correcting the TANF or SSP sample size, care must be taken to assure that the statistical principles of "randomness" and measurability are not violated. The selection of additional families for the TANF and SSP samples or deletion of units from the samples must be done in a manner that assures all cases in the population have a known, non-zero probability of selection into the final sample. In addition, techniques of stratification should not be employed in such a way that small additional strata are created for which computed estimates may be unreliable, resulting in a loss of precision in population estimates.

The procedures that a State (Tribe) uses to correct for excessive oversampling or correct for undersampling will depend partly on the procedures the State (Tribe) used to select its original sample cases. States and Tribes may choose from a wide variety of sampling methods. State agencies or Tribal grantees that select their TANF samples or State agencies that select their SSP samples using the systematic sampling method can use the procedures in Sections 1510, 1520, or 1530 of this manual to adjust sample sizes. State agencies or Tribal grantees that select their TANF samples or State agencies that select their SSP samples using the simple random sampling method can use the procedures in Sections 1540 to adjust sample sizes. For State agencies or Tribal grantees that use another method to select their TANF samples, ACF Regional Office staff will be happy to provide technical guidance on procedures to correct for excessive oversampling or undersampling to ensure that the principles of probability sampling are retained.

Monthly sample sizes should be monitored throughout the reporting period and correction should be made only when it becomes clear that target samples will not be met. It is good practice to re-estimate caseloads at the end of each quarterly reporting period. Waiting to the end of the annual period to make necessary corrections could create difficulties in

collecting the information and adversely affect the State's (Tribe's) ability to submit data in a timely manner.

The following procedures allow State agencies and Tribal grantees to make corrections in all months starting with the first month of the reporting period. A consideration for a State in selecting this method is that, in certain circumstances, it may be difficult to obtain accurate information for past months. This method does not involve the creation of additional strata.

1510 Standard Method to Correction for Undersampling or Oversampling when Sample Selected Using Systematic Random Sampling

1510.1 Correction for Oversampling

1. Using the procedure described in Section 1430, Step 1 (page 32), re-estimate the caseload size, adding on the expected number of cases to be dropped as listed-in-error, and compute a revised sample interval.

For each month in which the sample cases have already been selected:

2. Divide the size of the monthly sample frame by the revised sample interval (Step 1) to obtain the revised estimate of the number of sample cases that should have been selected.
3. Subtract the number of cases obtained in Step 2 from the number of sample cases that have been selected. This is the number of sample cases to be eliminated.
4. Divide the number of sample cases that have been selected by the number of cases to be eliminated (Step 3) to obtain the secondary sample interval to be used in identifying the cases to be eliminated.
5. Use a random start and apply the secondary sample interval obtained in Step 4 to select cases from the list of sample cases already selected. The cases so identified are to be eliminated regardless of whether or not data had already been collected.

For months in the annual period for which sample cases have not yet been selected:

6. Use the corrected sample interval for the period obtained in Step 1 to select

sample cases from the monthly frames.

1510.2 Correcting for Undersampling

1. Using the procedure described in Section 1430, Step 1, (page 32) re-estimate the caseload size, adding on the expected number of cases to be dropped as listed-in-error, and compute a revised sample interval.

For each month in which the sample cases have already been selected:

2. Divide the size of the monthly sample frame by the revised sample interval (Step 1) to obtain the revised estimate of the number of sample cases that should have been selected.
3. Subtract the number of sample cases already selected from the number obtained in Step 2. This is the number of additional sample cases to be selected from the monthly frame.
4. Divide the total monthly sample frame size by the number identified in Step 3 to obtain the secondary sample interval to be used in selecting additional cases from the monthly sample frame.
5. Use a random start and apply the secondary sample interval obtained in Step 4 to the monthly sample frame from which cases have already been selected. (If correction for undersampling is required only for the third and/or fourth quarters of the annual period, the State has the option of applying the secondary interval either to the first month of the sample period (October) or the first month of the applicable quarter (April or July)). Add the specific cases identified to the cases already selected for the same month as the month of the sample frame from which they were selected. If a case previously selected in the sample is again selected and identified for the same month as previously selected, an alternate case is to be selected by using a table of random numbers.

For months in the annual period for which sample cases have not yet been selected:

6. Use the corrected sample interval for the period obtained in Step 1 to select sample cases from the monthly frames.

1520 Alternate Method of Correcting for Undersampling or Oversampling When Sample Selected Using Systematic Random Sampling

An alternate method involves no adjustment for the months for which cases were already selected, however it does result in stratification of the sample by time. The alternative method entails the computation of a new sample interval that will either (1) undersample the remaining months of the 12-month sample period to meet sample size requirements if the earlier months had been oversampled, or (2) oversample the remaining months of the annual period to meet sample size requirements if the earlier months had been undersampled.

Because two different sample intervals will have been used, results of cases selected by each sample interval cannot be directly added to obtain State-wide (Tribal-wide) estimates as the proportions of the monthly frames sampled are different, i.e., the total sample is not a self-weighting sample. 1/ The alternate method will require all data to be weighted at the end of the 12-month period. The procedure involves inflating the various frequencies (e.g., number of families with an adult working, the number of families with a minor parent head of household, cases with earned income, etc.) in cases obtained using each sample interval, to their representation in the caseload and dividing the result by the caseload. 2/ This gives the weighted rate for the State (Tribe). In order to make each of the frequencies (number of families with an adults working, the number of families with a minor parent head of household, etc.) comparable with those of other States (Tribes), it is necessary to multiply the weighted rate by the total sample size. The equation for this procedure is as follows:

$$\text{Weighted State Rate} = \frac{\sum (x_m) (SI_m)}{\sum (n_m) (SI_m)}$$

-
- 1/ It should be noted that a self-weighting sample, except for rounding, must possess the following characteristic:

$$\frac{\text{Sample cases selected in a specific month}}{\text{Total sample cases selected in sample period}} = \frac{\text{Cases in sample frame for same month}}{\text{Total cases in all sample frames in sample period}}$$

- 2/ "Caseload", for the purpose, is defined as the completed sample size multiplied by the sample interval.

where:

Σ = the sum of . . . all strata ("stratum" is defined as part of the annual period using the same sample interval);

m = the m^{th} stratum (m is the stratum index);

x_m = "characteristic of interest" in the m^{th} stratum;

n_m = completed sample size in the m^{th} stratum; and

SI_m = sample interval used in the m^{th} stratum.

For example, assume that a State originally had estimated that its caseload would average 80,000 cases for the annual sample period. Assuming a 5 percent drop rate, the State used a sample interval of 303. Actual experience after 10 months resulted in the State revising its average caseload to 75,000, making no change in its drop rate. If the State made no corrections, the final completed sample size for the period would be short approximately 188 cases.

Assume that the State decides to obtain the additional 188 cases by using a revised sample interval of 219 for the last 2 months of the sample period. Also assume for the first 10 months of the sample period ($m = 1$) that the --

Number of cases completed (n_1) = 2,350

Number of cases with "characteristic of interest" (x_1) = 112

and for the last 2 months of the sample period ($m = 2$) that the --

Number of cases completed (n_2) = 650

Number of cases with "characteristic of interest" (x_2) = 37

Using the definition of "caseload" as defined earlier, i.e., sample cases completed multiplied by the sample interval, the weighted proportion of the case with the characteristic of interest would be computed as follows:

$$= \frac{(112 \times 303) + (37 \times 219)}{(2,350 \times 303) + (650 \times 219)} = \frac{42,039}{854,400}$$

$$= .0492$$

The State case proportion for the "characteristic of interest" would be .0492. The reported number of cases with the characteristic of interest for the 12-month period, for comparability with other States, would be 148, i.e., .0492 x 3,000.

Note that each frequency of occurrence or proportion of the total sample must be calculated in the same way, e.g., number of families with an adults working, the number of families with a minor parent head of household, the number of child only cases, the number of cases with earned income, etc. Caseload weights are to be used in computing State-wide (Tribal-wide) characteristics.

Note, it is important that the appropriate code be entered on the coding schedule to identify the stratum from which the case was selected.

1530 Correcting for Undersampling Using a Reserve Sample Pool

Correcting for undersampling using the sample interval (see Section 1520, page 44) involves resampling the original frame using a new sample interval. A State (Tribe) may find this to be difficult and/or costly. The same result can be achieved by selecting a reserve sample pool at the time of original sample selection. The designated reserve sample cases are to be used only if correction for undersampling is required. Properly selected reserve pool cases retain the self-weighting property of the final sample. However, careful attention to the controls is necessary to ensure that cases are properly selected. Any number of cases may be designated as a reserve pool -- a good number could be 10 or 15 percent of the required sample size.

The State (Tribal) sampling plan must describe in detail the procedures for setting up a reserve sample pool. If a random number generator is used, the type of generator and seed number is to be specified.

1531 Procedure for Setting Up a Reserve Sample Pool

Procedures for setting up a reserve sample pool are similar to those outlined in Section 1430 (page 32). To illustrate the procedures, the example in Section 1430, is used, i.e., a State (Tribe) uses the systematic random sampling method, elects the standard sample size, has an estimated average monthly caseload of 42,600 and estimates a 5

percent drop rate for the 12-month sample period. In addition, the State (Tribe) specifies 15 percent of its selected sample as reserve pool cases each month.

1. Determine Average Monthly Sample Size

Divide the number of sample cases for which data is to be collected in the annual sample period by $(1 - 0.15)$ to obtain the estimated total number of sample cases to be selected. In our example, according to Section 1430, Step 2, (page 33) the number of sample cases (completed and dropped cases) is 3,158. The number of cases to be selected would be $3158 \div (0.85)$, or 3,715, or an average of 309 cases per month. The average number to be placed in a reserve pool each month is 15% of 309 cases, or 46 cases (rounded down).

Note that the reserve pool is only to be used to correct for undersampling; it is not to be used to replace dropped cases.

2. Select Monthly Sample

Using the monthly sample size from Step 1, 309 cases, and the procedures outlined in Section 1430, Steps 5, 6 and 7, (page 34) compute the sample interval, determine a random start and select monthly sample cases from the sample frame.

3. Compute Secondary Interval for Selection of Reserve Pool Cases

Compute a secondary sample interval to be applied to the list of sample cases selected each month. This is obtained by dividing the estimated average monthly sample size by the average estimated number of cases designated for the reserve pool. In our illustration, the sample interval is $309 \div 46$, or 6.72.

4. Select and Identify Monthly Reserve Pool Cases

Since the interval obtained in Step 3 above is not a whole number, the acceptable method is to round up to the next higher number. (Note that rounding up is recommended to ensure that the basic sample will have a sufficient number of cases.) In our example, 1 out of every 7 cases on the monthly list of selected sample cases would be identified for the reserve pool, using a random start number between 1 and 7 inclusive. It is important in selecting monthly reserve pool cases to apply the same sample interval to the entire list of selected cases each month. This is an important part of the sample design and should not be violated in order to obtain a specific number of reserve pool cases each month.

5. Submission of Sample Cases Selected

The estimated average caseload, the specified percentage of monthly selected sample cases for the reserve pool, the computed sample intervals, manually generated random start and seed numbers to be used in the 12-month sample period for selection of total sample cases and reserve pool cases should be sent to the ACF Regional Administrator thirty (30) calendar days before the October sample selection. The monthly list of selected sample cases, with reserve pool cases identified, and computer generated random start and seed numbers should be submitted within 10 days of the date of selection specified in the State sampling plan.

If random numbers are used to identify cases for the reserve pool, it is absolutely essential that the total number of sample cases selected each month is known.

The following procedures are to be used and repeated each month if random numbers are used.

1. Determine the total number of sample cases selected. In our example, assume that the number selected for October was 309.
2. Multiply the number obtained in Step 1 by the percentage of selected sample cases specified for the reserve pool. In our illustration, the number of cases to be placed in a reserve pool for October is $309 \times .15$, or 46 cases (rounded down). Note that the same percentage must be applied each month.
3. Randomly select and identify reserve pool cases. In our example, 46 random numbers between 1 and 309, inclusive, would be selected.

If a Table of Random Numbers is used (see Appendix A, page 103), a photocopy of the page(s) used, showing start number, direction, and all selected numbers circled, is to be submitted each month along with the total list of selected sample cases. Sample cases corresponding to the circled random numbers are to be identified on the total list.

1532 Procedure for Obtaining Cases from a Reserve Sample Pool

States (Tribes) with reserve pools must use the same procedures in correcting for undersampling as outlined in Section 1520 (page 44) and 1530 (page 47). A revised estimate of the number of sample cases that should have been selected (excluding reserve pool cases) is to be computed. The difference between the number that should have been

selected and the number that was selected is the number of additional sample cases that will need to be selected from the reserve pool.

If a State (Tribe) uses a disproportionate stratified sample design, the State must maintain a separate reserve sample pool for each stratum. If a State with a disproportionate stratified sample design undersamples, the State must use the allocation procedures specified in its sampling plan to determine in which stratum (or strata) the State has undersampled. The State must correct for undersampling in each stratum in which undersampling occurred.

The same primary sample interval as determined in Section 1430, Step 5 (page 34), Step 2, must be used to select sample cases for months in the annual period for which such cases have not yet been selected. However, a new secondary sample interval to be applied to the monthly lists of selected sample cases must be used in identifying cases for the reserve pool.

The revised secondary sample interval is the product of the original secondary sample interval and the number of cases in the reserve sample pool prior to selection of cases from the reserve sample pool divided by the number of cases remaining in the reserve sample pool after selection of cases from the reserve sample pool.

1540 Correction for Undersampling or Oversampling When Sample Was Selected Using Simple Random Sampling

As described in Sections 1510 and 1520 (pages 42 and 44), there are two basic approaches to correcting the annual samples. The first approach is to correct the sample for both the months for which the sample has already been selected and the months for which the sample has not been selected. This approach provides an annual sample with approximately one-twelfth of the sample selected each month. The second approach is to make the entire adjustment in the months for which the sample has not yet been selected. Monthly samples selected using simple random sampling are less likely to need large adjustments for undersampling than for samples selected using systematic random sampling. This is true because under simple random sampling a fix number of sampling units is selected each month regardless of the monthly caseload. Under systematic random sampling, a fixed proportion of the caseload is selected each month. However, caseloads can vary from month to month. This variation of the monthly caseload results in variation in the monthly sample size.

1540.1 Correcting for Undersampling

If a small correction (e.g., less than 50 cases) is needed to ensure the State agency or

Tribal grantee will meet its minimum required annual sample size and no month is substantially short of the approximate one-twelfth of the annual sample, then the State or Tribe should correct for undersampling by adjusting the sample size in months for which the sample has not yet been selected. If the sample for all months have been selected, then the adjustments should be made for the months in the last quarter of the fiscal year. On the other hand, if a large correction is needed for months in which the sample have already been selected, the State or Tribe should consider making adjustments to all monthly samples. To correct a monthly sample for undersampling, use the following procedures:

1. Retrieve the Original Monthly Sample Frame

As required under Section 1460, States and Tribes must to retain their original monthly sample frames. The State or Tribe should locate the original month sample frame for use in selecting the additional sample cases.

2. Review Original Determinations of Total Monthly Caseload and Average Monthly Sample Size

Review the original application of the sample selection procedures from Section 1440 Steps 2, 3, and 4 (Page 38) to identify the reason for undersampling (e.g., under estimated the number of listed-in-error cases).

3. Determine the Number of Additional Sample Cases Needed

Subtract the number of completed cases for the sample month from the required number of sample cases for the month to determine the short fall. Allowing for some additional listed-in-error cases (using the procedures in Section 1440, Step 4 (Page 38)), determine the number of additional sample cases to be selected from the original sample frame.

4. Select the Additional Sample Cases

Using the same procedures as in Section 1440, Step 5 (page 38), select the addition sample cases from the sample frame and forward the sample selection list to the ACF Region Office.

1540.2 Correcting for Oversampling

States are not required to correct for excessive oversampling. If correction is desired, the procedures to correct for excessive oversampling are similar to correcting for

undersampling. However, instead of using the original monthly sample frame, the State or Tribe would use its monthly sample selection list and apply the following procedures:

1. Determine the Number of Cases on the Original Monthly Sample Selection List

This is the total number of sample cases, including cases that were listed-in-error.

2. Determine the Number of Excess Cases

In determining the number of excess cases, make allowances for the number of listed-in-error cases in the original sample. A proportion of these cases will be selected as cases to be removed. For example, if the original monthly sample had 325 cases of which 25 were listed-in-error, then the listed-in-error cases represent about 7.7% of the total sample and the total number of case of the sample frame. In reducing the sample so that there are at least 250 completed cases, the State or Tribe could expect about 20 listed-in-error cases. Therefore, the number of excess cases is $325 - 270$ or 55 cases.

3. Select the Excess Cases to be Removed from the Sample

Using the same process as in Section 1440, Step 5 (page 38), select the number of excess cases from the original sample and forward the sample selection list of excess cases to the ACF Regional Office.