

**APPENDIX A**  
**SAMPLE SELECTION**

# APPENDIX A

## Sample Selection

Appendix A explains the procedures for selecting the samples for nonmonetary determinations review. Explanations of **options** have been included; each state should select the option best suited to their particular operation. The option preferred by the National Office (NO) will be indicated with reasons for the recommendation.

### WHAT DOES SAMPLING REQUIRE?

The sampling methodology for nonmonetary determinations contains five distinct steps.

- 1) Identify, find, or gather data elements for sampling the universe files;
- 2) Extract or collect data to create the universe files;
- 3) Determine which transactions to select for the sample;
- 4) Select the cases to review;
- 5) Create output reports and files of the selected cases.

How these five steps are accomplished is the SWA's choice. Not all states have the same level of automation, and varying file structures may lend themselves to different sampling approaches.

### Step 1 - Gather Data for the Sampling the Universe

#### ***Collect Required Data***

The first step is to gather or have access to **the universe (ALL)** of the particular transactions to be reviewed). **It is**



**essential that every transaction or item meeting the criteria be included.** This means that all possible sources or locations of the transactions must be searched. For example, since nonmonetary determinations generated in various units such as BPC and BAM (formerly BQC) are included by definition, check to be certain they will all be included in the universe.



Be sure to check that only valid transactions are included. Refer to the definitions in the UI Reports Handbook No. 401 and in Chapter V of this Handbook to determine which records should be included.

For instance, determinations which are generated for the sole purpose of establishing an overpayment amount are not valid in terms of the definitions.

The SWA ADP staff is responsible for creating the universe files which contain the requested information. The NO has developed specifications of the minimum data needed for each sampling of the universe. The SWA ADP staff is responsible for creating the programs and/or utilities to extract or gather the requested data elements.

Each sample being reviewed will be selected from a universe which includes the nonmonetary determinations which were dated within the three months in the preceding quarter, which is referred to as the review quarter. The date the determination was issued determines in which quarter's universe that record will be included.

### ***Schedule Data Capture***

**In building the universe files, nonmonetary determinations should be captured as they occur.**

This is important as the desired transaction may be superseded by a subsequent transaction and the desired information may no longer be readily available. This may be especially true in highly automated states where data fields are often overlaid with the most recent information. Some states may be able to reconstruct events by using daily transaction logs maintained in their data processing environment. It is still better to capture the transactions as they occur during the time period to be reviewed.



**HINT:** Once the elements for the universe files have been identified and the extraction program created, the BTQ reviewer should examine a small cross section of the records to verify that the data elements are correct and the proper time frames are being followed.

After the BTQ reviewer has approved the data elements and time frame, the SWA ADP staff should establish procedures for building the universe files, selecting the sample cases and saving the universe files.

## Step 2 - Collect Data to Create the Universe

### ***Data Collection***

The SWA ADP staff must create a file of the transactions that make up the universe. The SWA can either write a program to create the universe file or use commercial software. The resulting transaction files can then be used as input into a sample selection program such as PICKNMBR, which is described in Appendix A of ET Handbook 401.

### ***Save the universe Files***

States **MUST** save the universe files from which samples are selected for data validation purposes for one year.

## Step 3 - Determine Records for the Sample

### ***Perform Calculations***

The third step is to determine which records to select for the sample. The formulas used to determine which records to select must be the formulas provided by the NO, or alternative formulas approved by the NO.

To perform the calculations, three numbers are needed:

1. **Total Records in the universe** - Once the universe has been created, a count of all the transactions in the universe must be performed. This count is represented by "P" in the calculations.
2. **Number of Records to Sample**. The number of cases to sample for nonmonetary determinations depends on the **total** number of nonmonetary determinations reported by the state on the ETA 9052, Nonmonetary Determinations Time Lapse Report, in the preceding calendar year. States reporting 100,000 or more nonmonetary determinations will sample 50 separation issues and 50 non-separation issues each quarter for quality review. States reporting fewer than 100,000 determinations in the preceding calendar year will sample 30 separation issues and 30 nonseparation issues each quarter for quality review. States may sample larger numbers if they choose, but all of the determinations sampled must be reviewed and entered into the database in order to preserve the validity of the sample. Before running the sampling routine, the universe of nonmonetary determinations must be sorted by separation issues and nonseparation issues so that an independent sample can be drawn from each.
3. **Random Number**. This is the third critical number necessary to perform the sample calculations. It is represented as "R" in the formulas. Random numbers are distributed by the National Office for each calendar year, may be generated on the state 's ADP system, or may be obtained from any statistics manual.

## FORMULAS TO IDENTIFY RECORDS FOR QUALITY SAMPLES

The following are the steps needed to determine which records to select for the sample. These steps must be repeated for each sample that will be selected.

- **A count of the total number of transactions in the universe must be performed.** The SWA ADP staff can supply this number. This number is represented by "P" in the calculations. Note that for nonmonetary determinations, a sort must be performed to divide the transactions file into its component universes of separation issues and nonseparation issues before proceeding.
- **Determine the number of cases to sample.** Based on the number of nonmonetary determinations reported by the state in the previous calendar year, determine the number of cases to sample for each. The letter "N" represents sample size in the calculations.

- **Obtain a Random Number.** In the calculations, "R" represents the random number, which can be obtained from the National Office, from a statistics handbook, or from the ADP system. The random number must be a decimal between 0 and 1 and must be at least three digits (for example, .729). For states with large universes, the random number must contain four digits, if the sampling interval is greater than 1,000.

**After the above mentioned numbers are identified, several calculations must be performed.**

First, determine the sampling interval (K), by dividing the sample size into the universe size. If the result of this calculation is not a whole number, round the result to the nearest integer.

$$K = P/N \text{ (round to the nearest integer)}$$

Second, determine the starting point (I) within the universe. This is accomplished by multiplying the sampling interval (K) by the random number (R) and rounding to the nearest integer.

$$I = (R * K) \text{ (round to the nearest integer)}$$

Next, "N" cases must be selected. This is accomplished by selecting pairs of cases (J) until all the cases have been identified. First, the number of pairs must be determined by:

**If N is even, J = 0, 1, 2, ... (1/2N - 1)**

**If N is odd, J = 0, 1, 2, ... 2(N - 1) - 1, the remaining case is calculated separately.**

Once the number of pairs is determined, the cases are selected by using the following formulas:

$$I + JK \text{ and } (P - JK) - I + 1$$

The remaining (odd) case is calculated by:

$$I + 2(N - 1)K$$

Alternatively, the cases can be selected systematically beginning with the starting point, I; subsequent cases will be selected by I + JK, with J = 1, 2, . . . , (N-1). The procedure is described in Chapter III of ET Handbook 395, 4<sup>th</sup> Edition, Benefit Accuracy Measurement State Operations Handbook.

## **CALCULATIONS**

Example 1: P = 43, N = 5, R = .261

$$K = 43/5 = 8.6 = 9 \text{ (rounded)}$$

$$I = (.261 * 9) = 2.349 = 2 \text{ (rounded)}$$

Since  $N = 5$ ,

$$J = 2(5 - 1) - 1 = 1$$

The following records would be selected:

	$\frac{I + JK}{}$	$\frac{(P - JK) - I + 1}{}$
when $J = 0$	$2 + (0 * 9) = 2$	$(43 - 0 * 9) - 2 + 1 = 42$
when $J = 1$	$2 + (1 * 9) = 11$	$(43 - 1 * 9) - 2 + 1 = 33$

the remaining case is calculated by:  $I + 2(N - 1)K$

$$2 + 2(5 - 1)9 = 20$$

Records 2, 11, 20, 33, and 42 would be selected for the sample.

Example 2:  $P = 244$ ,  $N = 10$ ,  $R = .743$

$$K = 244/10 = 24.4 = 24 \text{ (rounded)}$$

$$I = (.743 * 24) = 17.832 = 18 \text{ (rounded)}$$

The following records would be selected:

18, 42, 66, 90, 114, 131, 155, 179, 203, 227.

### ***Other Automated Approach***

The SWA may choose to use another automated method of identifying which records will constitute the sample. However, it is imperative that the formulas described on pages A-5 and A-6 or an alternative method approved by the National Office be used to ensure that the sample selection is non-biased.

## **Steps 4 & 5 - Select Cases and Create Sample Files**

The last steps of the sampling process involve the creation of files containing the selected sample cases. These steps use the universe file from step two and the calculated record numbers from step three to create the sample file. The calculated record numbers from step three identify which records from the universe file will comprise this file.

**The sample case file must contain, at a minimum, the skeleton fields identified in ET Handbook 402, 3<sup>rd</sup> Edition, Change 2, plus the claimant's social security number.**

**The PICKNMBR program can be used for every sample process.** The calculations performed by this program are designed to ensure a non-biased systematic sample. This program is not dependent on the method used to select sample cases.

Appendix A of Change 7 to ET Handbook 401 references two programs which were developed by the Department for the Tax Performance System (TPS) samples: PICKNMBR and SAMPSOnn. The TPS SAMPSOnn programs, specifically the TPS Status Determination sample selection program, can serve as a general model for the development of a sample selection program specific to the Benefits Quality program. It would require extensive modifications for use in selecting nonmonetary determinations or lower authority appeals samples. States are discouraged from undertaking this task.

States are reminded that, as stated in Change 7 to ET Handbook 401, "States are responsible for creating the programs/utilities necessary to extract the data elements for the universe files for each [benefit quality] sample." The Department cannot provide programming resources to write or modify sample selection routines.

The PICKNMBR program can be used if the universe is:

- not stored using the NO format; or if
- kept as transactions occur.

**Detailed descriptions of PICKNMBR and the National Office COBOL sampling programs are provided in ET Handbook 401, Appendix A.**