

APPENDIX A

Sample Selection

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

WHAT DOES SAMPLING REQUIRE?

The sampling methodology for lower authority appeals 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; and
5. Create output reports and files of the selected cases.

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

The state may use an automated or manual method to create the universe files and select the samples. Throughout this section of the Appendix, both methods will be discussed. Various details and tips on how to determine the best approach for your state is also provided.

Step 1 - Gather Data for the Sampling the Universe

Collect Required Data

The first step is to gather or have access to the sample 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 appeal decisions may be generated and stored in various units, 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 to determine which records should be included. ETA 5130 Benefit Appeals Report and ETA 9054 Appeals Time Lapse are the reports to review.

For instance, episodic claims programs such as Extended Benefits, Disaster Unemployment Assistance, Trade Readjustment Allowances along with disposed by other than decision, such as a dismissal, are not valid in terms of the definitions for the sample universe.

Depending upon the automation level of the Appeals Unit, the creation of the universe may require the extraction of data from a computer (mainframe or PC) or the keeping of a manual list or searching file drawers.

If an automated sampling method is chosen, the state's Information Technology (IT) 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 IT 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 lower authority appeals which were dated within the three months in the preceding quarter, which is referred to as the review quarter. The date the appeal decision was issued determines in which quarter's universe that record will be included.

Schedule Data Capture

In building the universe files, lower authority appeals 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.

In manual record storage systems, the transaction source information is tracked by a filing system. The universe to be sampled should be isolated in some manner so that the integrity of the data will not be compromised with it is time to select the sample.

HINT: *Once the elements for the universe files have been identified and the extraction program created, the quality 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 quality reviewer has approved the data elements and time frame, the IT 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

Save the universe files. Each state has the option to collect the data for the universe files any way they choose. If the necessary file is automated, the IT 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 later in this appendix.

For files not automated, the state must collect the information manually. This can be accomplished many ways but a simple method can be the creation of a list. For example, as lower authority appeals are completed, each hearings officer will record the docket number and the date of the appeal was heard. Throughout the quarter, the list is updated. At the end of the quarter, the list is assembled as the universe and the number of completed appeals are sequentially numbered and counted.

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 lower authority appeals quality depends on the total number of lower authority appeals decisions reported by the state in the preceding calendar year from ETA 9054 Appeals Time Lapse report, cell C1. States reporting 40,000 or more decisions completed from the previous calendar year will sample 40 decisions each quarter for quality review. States reporting fewer than 40,000 decisions in the preceding calendar year will sample 20 decisions each quarter for quality review. State may sample larger numbers if they choose, but all the decisions sampled must be reviewed and entered into the Unemployment Insurance Report (UIR) database in order to preserve the validity of the sample.

Note: Interval sampling can be implemented with no difficulty in states whose automated systems allow appeals disposed of by withdrawal, dismissal, or “no show” to be identified by code and excluded from the universe. If these types of disposals are indistinguishable and cannot be excluded from the universe before sampling, there are three alternatives to pursue.

- The first, and possibility easiest, method would be to add a marker with which to indicate those lower authority appeals in which an actual hearing led to a decision. The marked appeals would constitute the universe for pulling a random sample by interval sampling.
 - As a second alternative, IT could be asked to randomize the lower authority appeals file transactions within the quarter being reviewed. Randomizing routines are readily available, and once the file is randomized, the quality reviewers can begin drawing their sample from the top of the list, examining each record to be certain it meets criteria before saving it to the sample list. As many records as necessary can be pulled and examined until the required number is obtained, and the result will be a random sample for valid appeals decisions.
 - A third approach requires running the sampling routine three times. First, draw a random sample of twenty decisions from the universe, removing them. Then draw a second sample of 5 decisions from the remaining universe, removing them. And finally, draw a third sample of 3 decisions for the universe remaining. If all twenty of the first sample are valid hearings decisions, the next two samples will not be needed. If only 15 of the first sample are valid, move to the second sample of 5. If all of them are valid, stop. But if only 4 of them are valid, the third sample will be needed. The drawback here is that to preserve sample integrity, all of the valid cases from the third sample must be reviewed and entered, even if only one is needed.
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 IT 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. If automated, the state's IT staff can supply this number. This number is represented by "P" in the calculations.
- Determine the number of cases to sample. Based on the number of lower authority appeal decisions 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 NO, from a statistics handbook, or from the IT 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.

Balanced Systematic Sampling

CALCULATIONS

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:

$$\text{If } N \text{ is even, } J = 0, 1, 2, \dots (\frac{1}{2} N - 1)$$

$$\text{If } N \text{ is odd, } J = 0, 1, 2, \dots \frac{1}{2}(N - 1) - 1, \text{ the remaining case is calculated separately.}$$

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

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

The remaining (odd) case is calculated by:

$$I + \frac{1}{2}(N - 1)K$$

CALCULATIONS EXAMPLES

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 $N = 5$,

$$J = \frac{1}{2} (5 - 1) - 1 = 1$$

The following records would be selected:

	<u>$I + JK$</u>	<u>$(P - JK) - I + 1$</u>
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: $I + \frac{1}{2} (N - 1)K$

$$2 + \frac{1}{2} (5 - 1)9 = 20$$

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:

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

Systematic Sampling

Under certain circumstances, balanced systematic sampling can result in the selection of duplicate sample cases, especially if the population is small. To avoid duplicates, systematic sampling can be used as an alternative selection method.

First, a skip interval is computed by dividing the number of records in the sampling frame (P) by the number of records to be sampled (N). The first sample case selected is determined by multiplying the skip interval by the random start number (R), which is obtained as described for *balanced systematic sampling*. The product of the skip interval and the random start number is rounded to the nearest integer. If the rounded integer is zero, the case corresponding to the rounded skip interval is selected as the first case in the sample.

CALCULATIONS EXAMPLES

Number of Records in the Sampling Frame (P) = 118

Random Start Number (R) = .261.

Total Number of Cases to be Sampled (N) = 20.

Skip interval (K) = $118 / 20 = 5.9$

Initial case selected (I) = $.261 \times 5.9 = 1.54 = 2$ (rounded)

Record 2 in the sampling frame is the first record selected for the sample. Subsequent cases are selected using systematic sampling.

1. Select the initial sample case as described above.
2. Select the next (N-1) cases by adding multiples of the skip interval (K), rounded to the nearest integer, to the case number of the initial selection (I): $I + \text{round}(JK)$, where $J = 1, 2, \dots, (N - 1)$.

In the example, cases 2, 8, 14, 20, 26, 32, 37, 43, 49, 55, 61, 67, 73, 79, 85, 91, 96, 102, 108, and 114 will be selected from the sampling frame of 118 records.

If the last case designated for selection by the sampling algorithm is greater than the size of the sampling frame (P), the case will be selected from the beginning of the sampling frame. That is, the sampling frame will be considered to be circular. For example, if the last case selected is $P + 1$, the 1st case in the sampling frame will be selected.

The general rule is:

if $(I + \text{round}(JK)) > P$, select case H, where $H = [(I + \text{round}(JK)) - P]$ and $1 \leq H < I$.

Other Automated Approach

The state may choose to use another automated method of identifying which records will constitute the sample. However, it is imperative that the formulas described on the previous pages 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. The skeleton fields for lower authority appeals are the case identification, the decision date, and the case docket number.

States are responsible for creating the programs/utilities necessary to extract the data elements for the universe files for each quarterly sample. USDOL cannot provide programming resources to write or modify sample selection routines.

The NO sampling system consists of two COBOL programs which were developed for the Tax Performance System (TPS) samples: PICKNMBR and SAMPS0nn. The TPS SAMPS0nn 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 Appeals Quality program. It would require extensive modifications for use in selecting lower authority appeals samples. States are discouraged from undertaking this task.

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.

The PICKNMBR program can be used if the universe is:

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

PICKNMBR Processing – The narrative below describes the processing steps that are performed in the PICKNMBR program. These processing steps are also illustrated in the flowchart format.

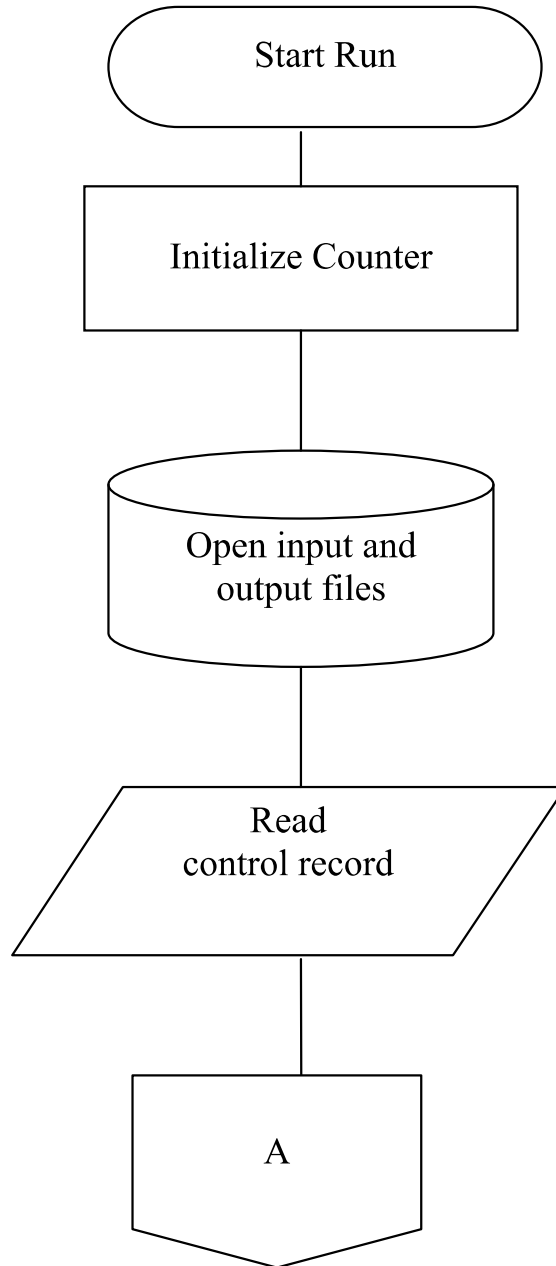
- 0000-DRIVER-ROUTINE – this section is the main routine for the program. This routine calls all of the other routines.
- 0010-LISTING-HEADING; 0020-LISTING-HEADING – these sections control printing of the report page and column header information, line count, and page advancement
- 0011-CS011, 0031-CS031, 0041-CS041, 0042-CS042, 0043-CS043, 0051-CS051, 0061-CS061 – these sections identify lower authority appeals and corresponding year/quarter fields for the activity being processed.
- 0100-OPEN-ROUTINE – this section opens the input file CNTRL-DATA, and output files SELECT-NUMBERS, PICKNUM-LIST and reads the CNTRL-DATA file.
- 0110-CNTL-OPTION – this section determines which function is being processed.
- 0120-CNTL-ERROR – this section validates the three CNTRL-DATA file fields

CNTRL-RANDOM-ALF, TRANS-REC-CNTRL-ALF, and SAMPLED-NMBR for non-numeric values. These fields must be numeric for the program to execute. To assist in the validation, a STOP-FLAG field is incremented by a certain amount. As a result, a comparison is made between the incremented STOP-FLAG and a value in the range 0 through 7. Within this range, different error messages will be displayed depending upon the error detected, and then the program is terminated.

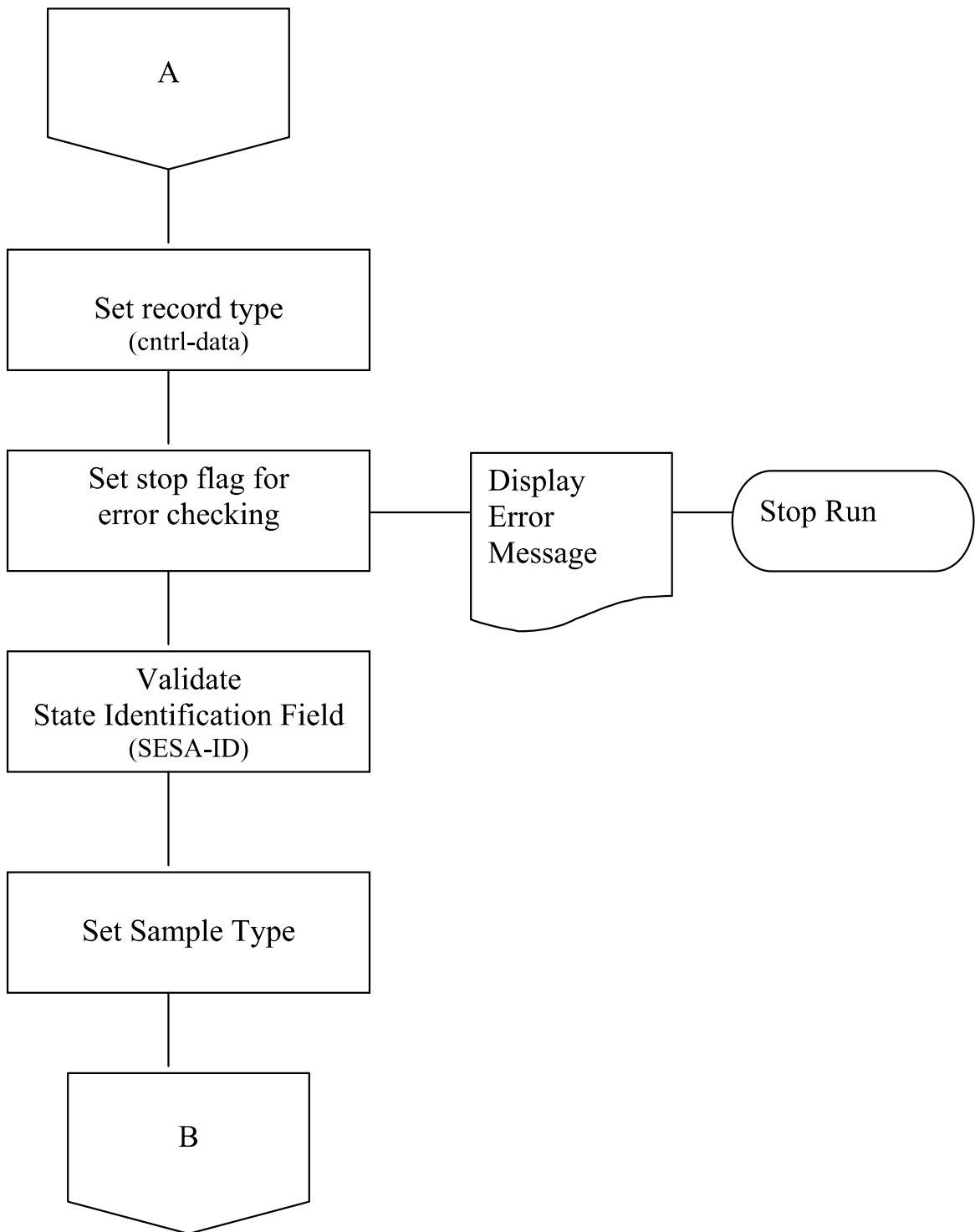
- 0130-FIPS-TABLE – this section searches SESA-ID in the FIPS table to find the exact state name associated with its abbreviation.
- 0140-SPL-TABLE – this section searches the SAMPLE-TYPE field of the CNTRL-DATA file for a corresponding match in the sample table (SPL-TYPE-DATA). If a match occurs, the sample type abbreviation is replaced by the exact sample type description to be utilized in the output report formats.
- 0200-CALC-SKIP-INTERVAL – this section calculates the SKIP-INTERVAL (K) utilizing the following $K=P/N$. P: the total number of records, N: sample size.
- 0300-INITIAL-CASE – this section calculates the initial sample case number (I). It is determined by truncating the result of $I = R * K + 0.5$. The INITIAL-CASE field (I) is defined as a 5-position numeric integer. The right side of the equation $I = R * K + 0.5$ yields a real number, thus allowing (I) to truncate the result of the calculation. R: CNTRL-RANDOM-ALF, and K: SKIP-INTERVAL.
- 0310-CHECK-ODD-EVEN – this section determines whether the number of records (N) to be selected for the sample (SAMPLED-NMBR) is either odd or even. If (N) is odd, the 0320-ODD-RTN procedure is executed.
- 0320-ODD-RTN – this routine calculates the additional number that was described in the random function formula. The equation is as follows: $ONE-MORE-REC = I + \frac{1}{2} (N - 1) * K$.
- 0330-CREATE-REC – this routine writes the calculated record numbers to an output data file (SELECT-NUMBERS) and an output print file (PICKNUM-LIST). A record counter (MATCH-CNTR) is incremented by one each time a record is added to the files.
- 0400-REMAINING-NUMBER – this section performs the calculations to determine the remaining numbers for the sample. As the balanced systematic sample of the random function formula, if N is even, $N/2$ pairs of records are selected. If N is odd, the iteration number is as follows: 0, 1, 2, $\frac{1}{2} (N - 1) - 1$ and $\frac{1}{2} (N - 1) - 1$ pairs of records are selected.
- 0500-CALC-SKIPINTERVAL – this section will calculate the SKIP-INTERVAL-B for a transaction sample size less than 200 using the equation $K = P/N$ (K not rounded).

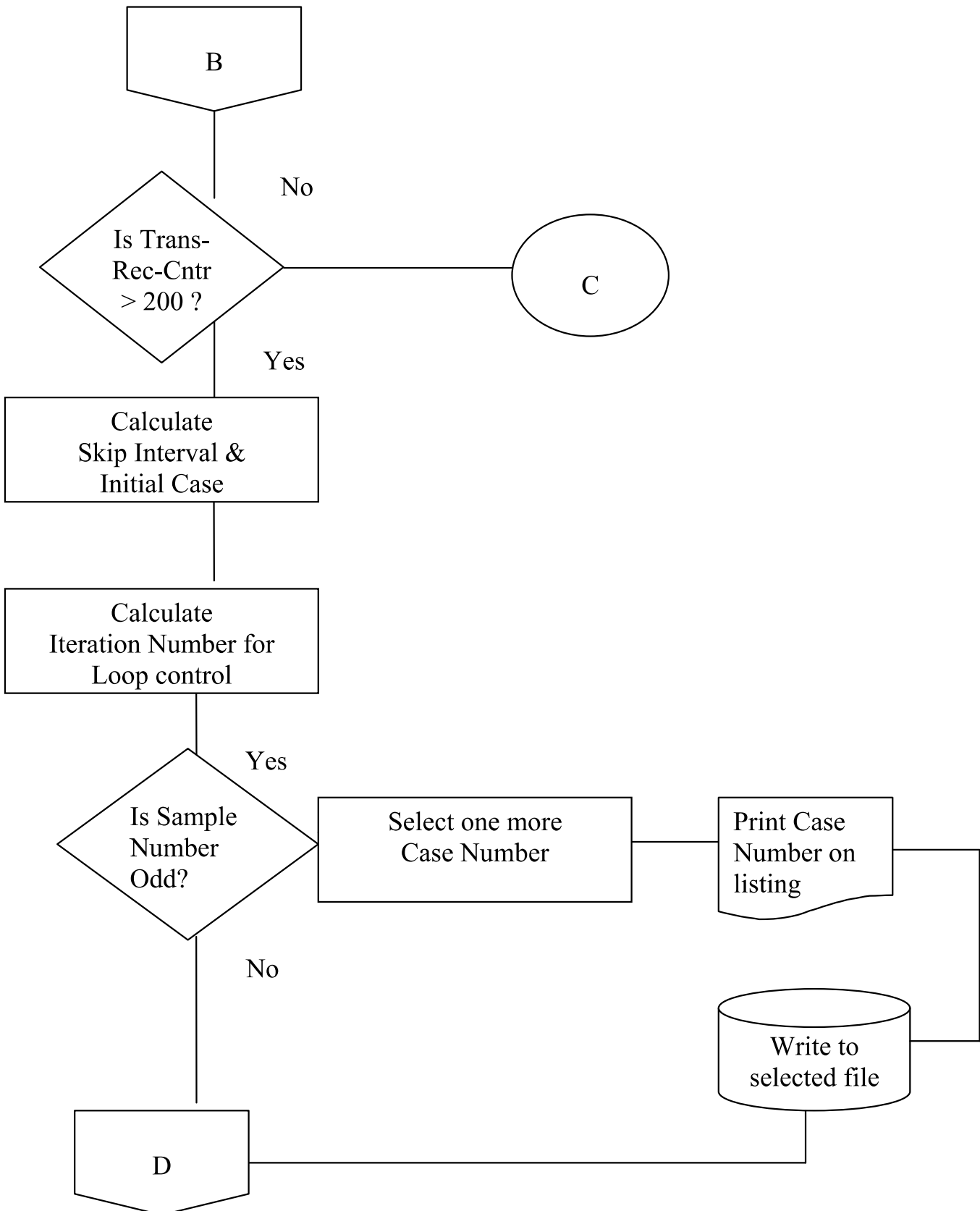
- 0600-CALC-INITIAL-CASE – this section calculates the initial sample case number (I) for transaction sample size less than 200. It is determined by truncating the result of $I = R * K + 0.5$.
- 0700-SELECTED-NUMBERS – this section performs the calculations to determine the numbers to select for the sample. This procedure is based on the transaction sample size less than 200. The second record is calculated by adding the skip interval (not rounded) to the initial case (truncated) and rounding the result. The remaining numbers are calculated by adding the skip interval to the previous (not rounded) number and then rounding the result. This process is continued until all the records have been calculated.
- 0800-CHK-SPL-NBR – this section verifies that the number of records written to the SELECT-NUMBERS files equals the number of records to be sampled (SAMPLED-NMBR) in the CNTRL-DATA file. If these fields are not equal, the error message is displayed.
- 0900-TRAILER-LIST – this section prints the information that was used to perform the calculations and select the record numbers.
- 9999-CLOSE-FILE – this section closes the files CNTRL-DATA, SELECTED-NUMBERS, and PICKNUM-LIST.

PICKNMBR

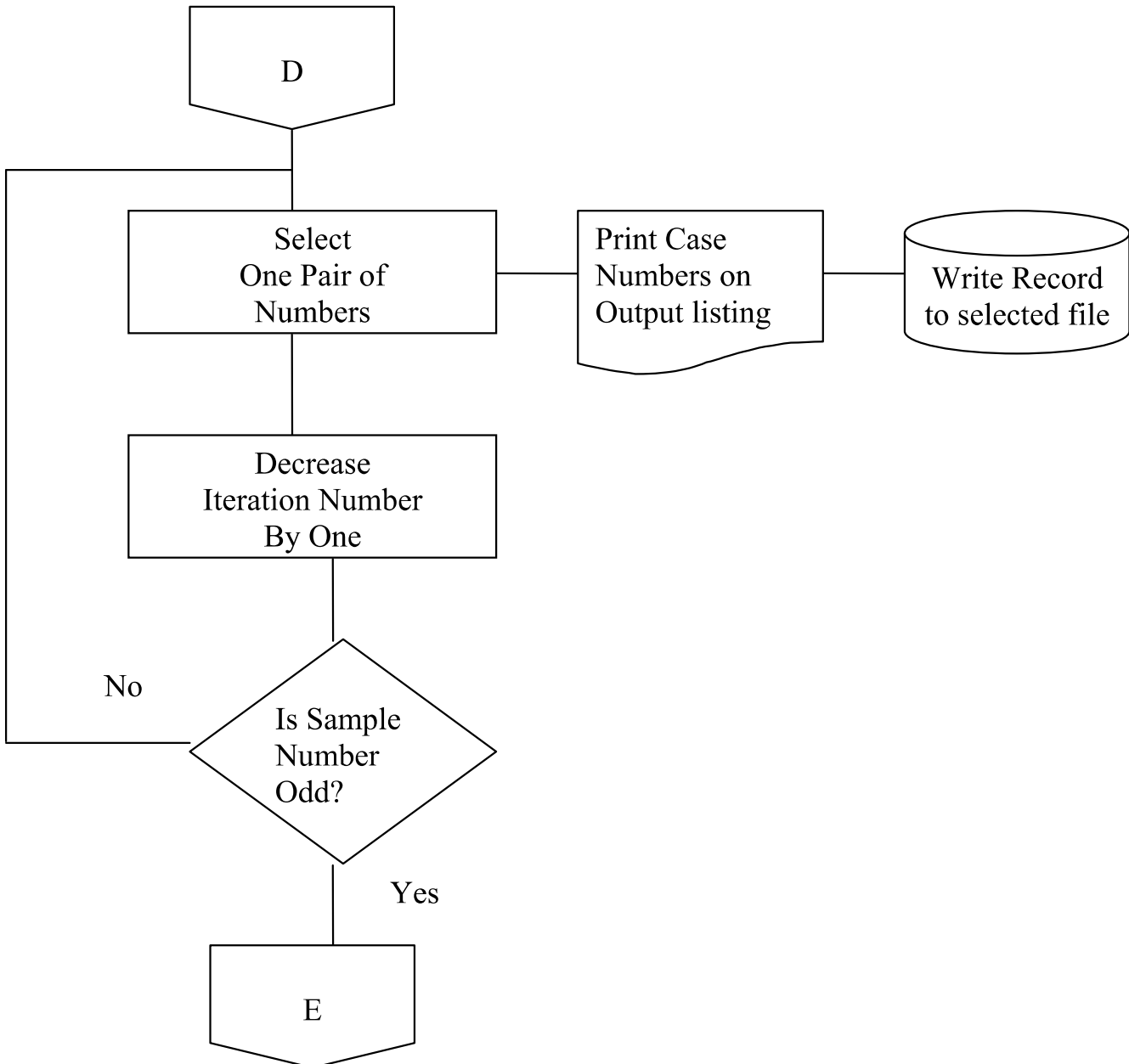


PICKNMBR





PICKNMBR



PICKNMBR

