Sampling Guidelines CRA

Introduction

This section provides sampling guidelines to assist examiners in selecting a sample of loans for review for CRA.

General Sampling Guidelines

Based on loan sampling, examiners should estimate three sets of proportions in connection with CRA examinations of small institutions:

- Loans inside and outside an institution's assessment area(s);
- Loans in low-, moderate-, middle-, and upper-income geographies in an assessment area; and
- Loans to low-, moderate-, middle-, and upper-income borrowers within an assessment area and/or loans to small businesses/farms of different sizes within an assessment area.

Examiners should analyze the results based on the performance context and other information obtained during the examination.

Small institutions (including intermediate small banks) are not required to collect data for CRA examination purposes. However, some small institutions may choose to provide data regarding their loans, including the census tract locations and borrower incomes, similar to the data requirement for large institutions. Some institutions may even provide a summary of their distribution of loans. In this case, if the examiner is able to verify the institution's information, the examiner may use the data supplied by the institution and will not need to perform sampling to evaluate the institution's CRA performance.

These sampling procedures may also be utilized at large institutions if data have not been collected for some reason. For example, if a large institution has chosen not to collect consumer loan information, yet it comprises a substantial portion of lending, an examiner may choose to review consumer loans. In this case, these sampling guidelines would apply.

Statistical Sampling at Small Institutions

- Determine major product lines from which to select a sample, taking into account factors such as the institution's business strategy and its areas of expertise. As an initial matter, examiners may select for review from among the same categories of loans that are to be used when reviewing large institutions (for example, home mortgages, small business and small farm loans, and consumer loans).
- Determine the universe of loans for each category.
  NOTE: The universe of loans is defined as the total number of loans, both originated and purchased by the institution, for a major product category.
  In order to determine the number of loans for the sample (known as the sample size), examiners should determine the total number of loans in the universe.

  This universe can be defined as any of the following:

  - Total number of loans since the previous examination;
  - Total number of loans in the previous year; or
  - Total number of loans in the previous six months.

- At a minimum, the universe of loans should cover at least the activity in the six months prior to the start of the examination. However, for institutions required to report HMDA and CRA data, all loan data should be analyzed for the most recent 24 to 30 months, and partial year-to-date when a minimum of six months of data is available for the most recent calendar year.
- Determine the number of loans to be sampled for each product category by using the Sample Size Table. The table indicates the sample size based on the universe of loans for each product and the desired confidence and precision levels.
- Initially, examiners should select samples based on a 90% confidence interval, with a level of precision of plus or minus 15 percentage points. This means that there is a 90% chance that the results from the sample will be within 15 percentage points of the true proportion, for whichever criterion is being evaluated.
- For loan products or institutions that require further investigation or are undergoing greater scrutiny for any reason, a larger sample may be necessary because examiners may need results with a higher degree of reliability. Examiners may use the 90% confidence level with a level of precision of plus or minus 10 percentage points when a larger sample is necessary. Examiners should use their judgment to determine which sample size to use based on the initial scoping of the examination and subsequent findings on site.

Sample Selection from Automated Download or Loan Trial

Once the number of loans to be sampled is determined, examiners should select loans from the financial institution's automated download of the loan portfolio or a loan trial, as opposed to making a general request for loans and allowing bank management to select the loans for review. Examiners should first determine if an automated download is available. Procedures for selecting loan samples from an automated download are explained in the next section. If an automated download is not available, the sample should be selected through the financial institution's loan trial. The use of random selection methods (whether from an electronic or a hard copy list) should increase substantially the objectivity of the examination, as compared to those based on judgmental sampling. This is because the use of random selection methods removes the potential for bias in results associated with the loan selection process.

NOTE: There may be some difficulties with manipulating the electronic files or loan trial because of the need to understand the file structure and/or the codes used by the bank. Examiners should also request information about the loan codes from financial institution management.

Sample Selection from Automated Download

The following provides guidance for selecting a stratified random sample of loans for CRA review when the financial institution's loans are provided in an electronic format. The different types of loans being sampled at a financial institution constitute the strata. A separate random sample will be selected from each stratum (loan type). Because the loans are in an electronic file, this is a fairly straightforward process: the loans can be electronically arranged and sorted in specified ways for sampling purposes.

Step 1. Once an electronic file of the financial institution's loans is obtained, and the coding structure is known, the process of selecting a sample is relatively easy. The loans are grouped into the various loan universes that will be sampled (e.g., home purchase loans, small business loans). Each loan universe could be a separate electronic file or a subfile of the main file. Within each loan universe, a random sample of loans will be selected.

NOTE: There should be sufficient information in the file to identify the different loan types, though an examiner may need assistance from bank management to interpret the codes.

Step 2. For each universe, place the loans in random order. The random ordering is achieved by assigning a random number between zero and one (i.e., a decimal number) to each loan in the file, and then sorting the loans in ascending order by the random number. Excel software contains a random number generation tool that examiners should use to randomly select loans for review during an examination. This "random number generation" tool can be accessed by following these steps:

- Select the following from the Excel Menu:
  - Developer (top menu tab)
  - Add-Ins (menu ribbon)
  - Analysis Tool Pak (check box and click "OK")
  - Data (top menu tab)
  - Data Analysis (menu ribbon)
  - Random Number Generation (highlight and click "OK")
- Respond to the items on the Random Number Generation screen as follows:
  - Number of Variables (Leave blank)
  - Number of Random Numbers (Leave blank)
  - Distribution (Select "Uniform" from list)
  - Parameters (Leave default set at 0 and 1)
  - Random Seed (Leave blank)
  - Output Options (Click "Output Range")
    - (Click the box for "Output Range")
    - (Select the range (output location) for the random numbers by highlighting the column on the spreadsheet where the random numbers should be placed. Use the "Shift" key and the down arrow to highlight the column.)
    - (Click box at end of selected range or press enter)
    - (Click "OK")
- The random numbers are automatically assigned and placed in the designated statistical column. Sort the loans in ascending order by the random number as follows:

- Click a cell in the column you would like to sort by
- Click "Sort Ascending"

Step 3. Once the loans for a given type are placed in random order, determine the number of loans to select for review. The sample needed for the examination is simply taken from the top of the list. If an "out-of-scope" loan is identified in the sample, the examiner selects the next loan from the list to replace it. From within each loan type, print out all of the loans that will be reviewed. This information should be forwarded to bank management so that the loan documentation files will be ready at the commencement of the examination.

NOTE: An "out-of-scope" loan would be any loan that is not in the target universe. For instance, a loan that was originated prior to the last examination date; or a business loan that is greater than $1,000,000; or a small farm loan greater than $500,000.

Step 4. If an examiner decides that additional loans for a given stratum (type) are needed based on the results of the initial sample of loans, additional loans should be selected. As with replacements for "out-of-scope" loans, the examiner would take the additional sample of loans from the list, picking up where he or she left off when selecting the initial sample.

Sample Selection from a Loan Trial

For the case of sampling from a loan trial, the selection of a simple random sample from each stratum (type) would be rather difficult to do. Instead, another type of statistical sample, referred to as a systematic sample, should be selected from each stratum. As is true for a simple random sample, each loan in a given stratum will have the same chance of being selected for review. The basic idea of systematic sampling is straightforward:

- Select every nth loan (for example, every fifth loan) after selecting the initial loan at random from among the first "k" loans (for example, a random selection from among the first five loans).

The logistics of preparing a "universe" of loans from the loan trial for sampling, however, can be somewhat problematic, as will subsequently be discussed.

The different types of loans being sampled at a bank constitute strata. A separate systematic sample will be selected from each stratum (loan type). There may be a separate loan trial (or group of two or more trials) for each stratum. Alternatively, the different types of loans may be mixed together on a single loan trial. These two possible situations are addressed separately as Case 1 and Case 2, as follows:

Case 1: The bank provides separate loan trials for each of the major types of loans being sampled.

Assume there is only one loan trial for each type of loan. If there are two or more trials for a type, the lists can just be merged to form one list.

Step 1. Review the loan trial and cross out any loans that can be identified as being "out-of-scope" for the review (e.g., the date of the loan is not within the time period specified for review).

Step 2. Once the "out-of-scope" loans have been crossed off, number the remaining loans consecutively from 1 to N, where N is the total number of "in-scope" loans on the list.

NOTE: Some of the loans in the remaining list may still be "out-of-scope," but cannot be identified prior to the review process.

Step 3. Determine a sampling (or skip) interval "k" by dividing N by the target sample size n, and rounding the result off to the nearest integer. For example, if there were N=123 loans on the trial for a given loan type, and the examiner decides to select n=10 of them for review, the selection interval "k" would be 12 (i.e., 123/10, rounded to the nearest integer).

NOTE: To compensate for "out-of-scope" loans, it is at this point that the examiner may wish to "oversample." This procedure is explained in Step 5.

Step 4. Identify the "random start" (initial selection) for the systematic selection, which is a randomly selected loan from among the first "k" loans listed in the trial. To do this, refer to the list of random numbers provided. Take the first random number available from this list (for example, the first one that has not yet been used). This will be a decimal number between 0 and 1. Multiply the random number by the selection interval "k" and round the result up to the next integer. This integer will identify the "random start." Suppose that for the example introduced above with k=12, the next available random number from your list is 0.34309. The product of 12 and 0.34309 is 4.11708. Therefore, the random start (first selection) would be loan number 5 (since you always round up). Once you have the random start, the other selections are identified by repeatedly adding the skip interval "k" to the previous selection number. For example, in the case discussed above with a random start of 5 and a selection interval of

k=12, the other selections would be loans numbered 17, 29, 41, 53, 65, 77, 89, 101, and 113.

Step 5. The systematic selection process should continue until the end of the list is reached, even if this adds one or two selections to the target sample size (this could happen because of the rounding of the skip interval "k" to an integer). If any of the loans selected are identified during the review process as "out-of- scope," it should be dropped from the sample *without being replaced*. For example, *do not* replace an "out-of-scope" loan with the next one on the list. The deletion from the sample of "out-of-scope" loans may reduce the sample size below the target. If such a reduction is minimal, and if the examiner feels that the remaining sample is adequate, nothing more needs to be done. However, if the reduction in the sample size is of concern, there are two methods that can be used to compensate for "out-of-scope" loans.

- **Method 1.** The preferred method is to "oversample" to allow for anticipated deletions. For example, if the examiner expects, based on previous experience, that 10% of the loans in the loan trial are actually "out-of-scope," the sample can be increased by 10% to account for this. In the example discussed above with a target sample size of 10, the examiner would select a systematic sample of 11 to compensate for an anticipated "out-of-scope" rate of 10%. This would dictate a selection interval "k" of 11 rather than 12. If all 11 loans turn out to be "in-scope," they should all be retained for the sample.

- **Method 2.** Select a supplemental sample after a review of the initial sample. In this case, the size of the supplemental sample would be determined to provide the correct number of additional sample loans. The selection procedures used for the supplemental loan sample would be the same as those used for the initial sample, after removing the initial selections. For example, if in the case discussed above for selecting ten loans, suppose that three of them turn out to be "out-of- scope," leaving a deficit of three. A supplemental sample of three or four loans could then be selected. It would be wise to select four, since the initial sample suggests that the "out-of-scope" rate is 30%. To select four additional loans, the remaining loans (i.e., 123-10) would be renumbered from 1 to 113. The new selection interval "k" would be 28 (i.e., 113/4, rounded). To identify the random start for the supplemental sample, use the next random number from the list of numbers provided at the end of this section. Multiply this random number by 28, and round the product up to the next integer to identify the first selection.

Case 2: The bank provides only one loan trial, and all loan types are mixed throughout the list.

- This case is more complicated and more prone to errors than Case 1, even though the basic idea is the same. The first step is identical to that for Case 1— Review the loan trial and cross out those that are known to be "out-of-scope."

- Next, go through the list and classify each loan by type. This might best be done using different colored highlighters to identify the loans by the different types. Then, the loans of a given type (same color) are numbered consecutively from 1 to N, and selection from these would be carried out the same way as it was in Case 1 (Steps 3–5), including any possible "oversampling" or sample supplementation.

Information to be Gathered for Each Loan Record

- Once the loans for each sample have been identified, record relevant loan information into a spreadsheet. Data for each loan should include, at a minimum:
  - Institution's internal loan ID number;
  - Loan type;
  - Loan dollar amount;
  - Location – In cases where the census tract of the loan is not readily available, examiners are expected to geocode the loans (refer to "Geocoding Loan Locations" below);
  - For the home related and consumer loans sampled, the borrower income that was used to approve the loan; and
  - For the small business and small farm loans sampled, the business's or farm's revenue. When income information used in the analysis of the applicant is not readily apparent and tax returns are present, this information can usually be obtained from the "gross receipts;" or "gross farm rental income" and "gross income" fields, respectively.

- When data is missing, attempt to obtain this information through discussions with institution personnel. Obtaining information through these discussions can significantly reduce the number of records in the sample with "missing data" and thereby increase the validity of each sample.

Geocoding Loan Locations

If the institution has not already geocoded each loan by determining the Metropolitan Statistical Area (MSA) (if applicable), state, county, and census tract, the examiners will need to

determine this for the loans in the sample.
MSA /Census tract information is available through the Internet.

Calculating Proportion Estimates

- Calculate the following proportion estimates as itemized in the examination procedures:
  - Percentage of the number of loans (by product type) inside and outside the assessment area(s); and
  - Percentage of the dollar amount of loans (by product type) inside and outside the assessment area(s).
- In accordance with the CRA examination procedures, examiners should tabulate the following statistics based on *only those loan records from the sample that are within the assessment area(s)* for each product category:
  - Number and percentage of loan originations (by product category, if applicable) in low-, moderate-, middle-, and upper-income geographies;
  - Dollar amount and percentage of loan originations (by product category, if applicable) in low-, moderate-, middle-, and upper-income geographies;
  - Number and percentage of loan originations (by product category, if applicable) to low-, moderate-, middle-, and upper-income borrowers;
  - Dollar amount and percentage of loan originations (by product category, if applicable) to low-, moderate-, middle-, and upper-income borrowers;
  - Number and percentage of loan originations to small businesses/farms of different sizes (by revenue); and
  - Dollar amount and percentage of loan originations to small businesses/farms of different sizes (by revenue).

Examiners should follow the guidelines in the CRA examination procedures for analyzing the results from sampling and, ultimately, assign a rating to the institution's lending performance.

Sampling Guidelines for Compliance

Sample Size Table

90% Confidence Interval

Number of Originations of Purchases	*Sample Size*	
	10% Precision	**15% Precision**
10	9	8
50	34	24
100	50	31
200	67	37
300	76	39
400	81	40
500	84	41
600	86	42
700	88	42
800	89	42
900	91	43
1,000	92	43
1,250	93	43
1,500	94	43
1,750	95	44
2,000	96	44
2,250	96	44
2,500	97	44
2,750	97	44
3,000	97	44
3,500	98	44
4,000	98	44
4,500	98	44
5,000	99	44

Random Numbers

0.38200	0.11692	0.54509	0.93725
0.10068	0.76571	0.03854	0.57637
0.59648	0.80123	0.99829	0.71999
0.89911	0.75829	0.05795	0.95322
0.88461	0.16840	0.78448	0.21992
0.95846	0.17753	0.22163	0.14505
0.01450	0.68123	0.55965	0.82025
0.40742	0.32841	0.59865	0.06351
0.86325	0.15769	0.81817	0.16541
0.13858	0.12033	0.16309	0.95004
0.24503	0.09137	0.57833	0.06900
0.04547	0.47011	0.10123	0.56258
0.03238	0.35252	0.37065	0.22309
0.16413	0.46727	0.11570	0.23917
0.21961	0.88348	0.51720	0.43397
0.01709	0.71441	0.01358	0.91305
0.28504	0.40043	0.08927	0.40794
0.34309	0.05341	0.66530	0.25922
0.55364	0.71868	0.53526	0.80044
0.35737	0.07834	0.60927	0.08249
0.37184	0.74892	0.40172	0.44646
0.35560	0.54076	0.28321	0.44240
0.91031	0.09787	0.45540	0.53639
0.46602	0.09870	0.20899	0.47337
0.42616	0.16886	0.47710	0.11240
0.30390	0.97241	0.56795	0.02460
0.97571	0.32383	0.66408	0.76263
0.80667	0.07096	0.60707	0.73745
0.99124	0.66311	0.57186	0.90796
0.25626	0.17273	0.33290	0.62673
0.95169	0.92596	0.47371	0.73421
0.05344	0.89953	0.10895	0.25645
0.70504	0.54039	0.54143	0.74697
0.81652	0.64925	0.19565	0.82986
0.97250	0.57942	0.97919	0.23957
0.46632	0.46541	0.08670	0.56990
0.30021	0.93738	0.79952	0.79647
0.75021	0.15278	0.82580	0.27607
0.35148	0.04978	0.93988	0.42332
0.77566	0.91806	0.45799	0.53157
0.07434	0.56850	0.50566	0.13248
0.19843	0.55220	0.50697	0.32572
0.06406	0.19382	0.71819	0.93927