

2011 FDIC Survey of Banks' Efforts to Serve the Unbanked and Underbanked

Appendix B



December 2012

Appendix B – Survey Methodology

1. Sample Design

The goal of this study is to obtain information from the headquarters (HQs) of all FDIC-insured financial institutions regarding strategies for serving the unbanked and underbanked and products and services offered to these populations. The target universe (or population) consists of main offices associated with all FDIC-insured bank charters, excluding institutions that do not have standard retail branching operations.¹ The basic sample design was a stratified random sample, with three size strata defined by asset value. These three strata, also referred to as tiers, were defined as follows:

Tier 1: Top 25 largest banks by assets

Tier 2: Banks with \$1 billion or more in assets outside the top 25

Tier 3: Banks with less than \$1 billion in assets

For sample design development purposes, we used the FDIC's Summary of Deposits (SOD) database for June, 2010. From this database, we defined a preliminary universe of 7,691 FDIC-insured institutions (bank charters) having standard retail branching operations. There were 94,313 branch offices associated with these 7,691 FDIC-insured institutions. Table 1 presents the number of bank HQs and branches by Tier size categories.²

The HQs sample was designed so that survey-based estimates of universe proportions for Tier 2 and Tier 3, such as the proportion of HQs in Tier 2 (or in Tier 3) that would say yes to a yes-no question, could be computed with acceptable precision (± 5 percentage points at the 90 percent level of confidence). For Tier 1 (the 25 largest banks), a 100 percent sample was selected because of the importance of obtaining survey information from these large banks.

¹ The current data universe includes banks with a) brick and mortar branches (code 11) and b) banks in retail locations (code 12). Examples of institutions excluded from the analysis are: Banks with branches with deposits above \$900 million, Industrial Loan Companies, and Trust Companies.

² In addition to the HQs sample, the sampling plan also consisted of a second-stage sample of 2,000 branches, selected randomly from those branches associated with HQs that participated in the HQs survey and also gave us permission to survey their branches. However, since the focus of this report is the HQs survey, this appendix will describe the statistical methodology of the HQs survey, and will not refer further to the branch survey, aside from the presentation of branch counts given in Table 1.

Table B1: Universe Counts (HQs and Branches) by Asset Size

	Initial Number of HQs	Number of Branches
Tier 1 Top 25 banks by assets	25	39,707
Tier 2 Banks with \$1 Billion or more in assets outside the Top 25	562	25,401
Tier 3 Banks with less than \$1 Billion in assets	7,104	29,205
Universe/Sample Frame	7,691	94,313

To achieve the precision targets for the Tier 2 and Tier 3 strata, the target number of completed questionnaires for each of these two strata (shown in the third column of Table 2) was calculated based on the following formula:³

$$n = \{((t^2 * p * q) / d^2)\} / \{1 + (1/N) * ((t^2 * p * q) / d^2) - 1\}, \quad (1)$$

where

- n = the target number of completed questionnaires being derived,
- p = the approximate value of the universe proportion (with 0.5 used for planning purposes),
- q = 1 - p,
- N = the number of banks in the HQs universe for the stratum,
- d = the target level of precision (0.05), and
- t = the appropriate constant (1.645) for computing a 90 percent confidence interval based on the normal probability distribution.

The Tier 2 and Tier 3 target numbers of completes were computed from Equation 1, using the initial universe counts (number of HQs) shown in the first column of Table 2. These sample sizes were also computed based on the final universe counts (second column of Table 2), which were obtained from FDIC's Call Report data as of June 2011, with additional screening criteria used to filter out non-standard retail institutions. As it turned out, the target numbers of completes for Tier 2 and Tier 3, based on the final universe counts, were the same as those based

³ See William Cochran, *Sampling Techniques*, Third Edition, 1977, formula 4.1, page 75.

on the preliminary universe counts, except that the Tier 3 target number of completes was 260, rather than 261. Due to the closeness of these numbers, the target number of completed questionnaires for Tier 3 was left as 261.

Table 2 also provides, in the last two columns, the survey sample sizes and the number of completed HQs Internet-based questionnaires. As indicated previously, all 25 Tier 1 banks were included in the sample. Also, a strong effort was made to obtain a response rate of 100 percent among Tier 1 banks, which was achieved. (As such, there is no sampling error or nonresponse bias associated with any of the Tier 1 results.)

For the Tier 2 and Tier 3 strata, the sample sizes (shown in column 4, Table 2) were based on the target number of completed questionnaires and an assumed response rate of 65 percent. Therefore, the sample sizes for Tier 2 and Tier 3 were computed by dividing the target number of completes for each stratum by 0.65. For example, for Tier 2, the sample size of 282 was computed by dividing 183 by 0.65.

As it turned out, the response rates for Tiers 2 and 3 far exceeded the minimum target of 65 percent. The response rate for Tier 2 was 80 percent, and the response rate for Tier 3 was 79 percent, for an overall response rate across the three tiers of 80 percent (567/709). The number of completed online questionnaires is shown for each tier in the last column in Table 2.

The designated initial samples, as specified in Table 2, were selected randomly from the updated Tier 2 and Tier 3 universe lists. In addition, there were plans in place for the possibility of having to supplement the Tier 2 or Tier 3 samples with additional randomly selected banks from a “reserve sample,” in the event that the response rates were below 65 percent. However, since the response rates for Tier 2 and Tier 3 were substantially higher than 65 percent, there was no need to supplement the sample.

2. Data Collection Procedures

Gallup, Inc. received from the FDIC updated lists of HQs containing the name, asset size, and mailing address of all FDIC-insured HQs. These lists were used as the source of the universe (sampling frame) for the selection of the HQs sample. The sample was selected as a stratified random sample of 709 HQs from the list provided. The sampling unit for the survey of HQs, as mentioned above, was main offices associated with bank charters, excluding charters that do not have standard retail operations. The respondent sought was the bank officer expected to be most knowledgeable about strategies and practices for serving the unbanked and underbanked households.

Data collection was conducted entirely online, although a mixed-mode approach was used for inviting and prompting participation by HQs management. The initial survey request was sent by mail, by means of a letter from the FDIC introducing the study, accompanied by a letter from Gallup outlining the procedures for logging onto the website to respond to the survey. The website required a unique user name and password to respond in order to monitor response behavior and ensure privacy.

A series of prompts by mail, email (when available), and telephone were carefully spaced out in order to boost response rates. All prompts encouraged the representative of the financial institution to go online to Gallup’s secure website to respond to the survey. If an institution did not have access to the Internet, a paper version of the questionnaire would have been made available. However, providing a paper version was not necessary for any of the HQs respondents.

3. Survey Estimation Procedures

Survey estimates for HQs were based on weighted survey responses, where the weights were derived from HQs probabilities of selection, and adjusted for response levels. Weighting and estimation formulas are given in the following subsections. A summary description of how significant differences in sample estimates of proportions and means were computed is also provided.

Table B2: Universe and Sample Counts for the HQs Survey, by Strata

Sampling Strata	Initial Number of HQs (1)	Final Number of HQs (2)	Target Number of Completed Questionnaires (3)	Sample Size of HQs (4)	Actual Number of Completed Questionnaires (5)
Tier 1 Top 25 banks by asset size (equated to total assets greater than \$38 billion)	25	25	25	25	25
Tier 2 Banks with at least \$1 billion in assets and less than \$38 billion in assets	562	559	183	282	225
Tier 3 Banks with less than \$1 billion in assets	7,104	6,745	261	402	317
Total	7,691	7,329	469	709	567

Weighting the Sample Data

The sample of HQs, as described above, was a single-stage stratified random sample. As such, the selection probabilities for all units within a stratum will be equal to the stratum sampling rate.

Let h denote the stratum and the index i denote a bank (i.e., HQs) within the stratum. If we select n_{ho} (the initial sample size for stratum h) out of the N_h banks in the sampling frame (or list) for stratum h , then the selection probability for all banks in stratum h is given by

$$\pi_{hi} = n_{ho} / N_h$$

The probability or base weight (w_{1hi}) assigned to all sampled banks (HQs) in stratum h is the inverse of the selection probability:

$$w_{1hi} = N_h / n_{ho}$$

The base weights were adjusted to account for the nonresponding banks. This weight adjustment was obtained by redistributing the weights of the nonresponding banks in a stratum across the responding banks in the stratum. (Some consideration was given to possibly define nonresponse adjustment cells by other variables in addition to the size tier; however, mostly because of the high response rates, it was decided to use only the size tiers to define nonresponse adjustment cells.)

Suppose that n_h of the n_{ho} banks sampled from stratum h turned out to be eligible for the survey, and that r_h of these banks participated in the survey. Then the adjustment for nonresponse in stratum h , w_{2h} , was calculated as the ratio of the number of eligible sampled banks (both respondents and nonrespondents) in the stratum to the number of banks that participated in the survey:

$$w_{2h} = n_h / r_h$$

The final weight (w_{hi}), adjusted for nonresponse, for each bank in stratum h was computed as follows:

$$w_{hi} = (w_{1hi})(w_{2h}) = (N_h / n_{ho})(n_h / r_h) \quad (1)$$

As it turned out, all banks selected from stratum h were classified as eligible for the survey. Therefore, $n_{ho} = n_h$, and the final weight for each bank in stratum h was simply N_h / r_h . Furthermore, the sum of the weights of all of the r_h respondents in stratum h equals the number of banks in the sample frame for stratum h , N_h .

Preparing Survey Estimates and Frequency Tables

Most of the estimates derived from the survey results were estimates of universe proportions or universe means. An example of a statistic for the HQs universe that was estimated was the proportion of banks that give top priority to reach out to unbanked and underbanked households. Defining y_{hi} as “1” if the i^{th} respondent bank in stratum h does assign top priority to this effort, and assigning “0” to y_{hi} otherwise, the universe proportion was estimated as follows:

$$p = \sum w_{hi} y_{hi} / \sum w_{hi}, \quad (2)$$

where the sum is taken over all of the sample HQs in the domain for which the proportion is being estimated. Note that the estimated proportion is simply the sum of the weights of all responding HQs in the domain that do assign top priority to reaching out to unbanked and underbanked individuals, divided by the sum of the weights of all responding HQs in the domain.

For deriving frequency tables, estimates of universe cell frequencies for HQs were estimated using the formula in the numerator of Equation (2) above.

For estimates of universe means, Equation (2) above was used, where the value of the y -variable was the value of the statistic (like a fee amount) for which the universe mean was being computed (rather than being the 0-1 indicator variable used for estimating proportions).

Comparing Sample Proportions or Means for Significance

For various universe domains, either defined by strata or by other variables, including variables based on survey responses, estimates of proportions or means were made to determine whether the differences of the sample estimates were “significant,” indicating that it would be unlikely that the corresponding universe proportions (or means) would be equal. The criterion used to identify significant differences was the 10 percent level of significance. Sample differences had to be significant at this level to be reported here as identifying universe differences. The significance determinations were made using the Survey-Reg procedure in SAS software v9. 3.