

METHODOLOGICAL ISSUES IN SURVEYING THE NONBANKED POPULATION IN URBAN AREAS¹

by Constance R. Dunham, Office of the Comptroller of the Currency,
Fritz J. Scheuren, Ernst & Young, LLP, and Douglas H. Willson, Macro International Inc.
Constance R. Dunham, OCC, Economics Dept. 2-1, 250 E Street, S.W., Washington D.C., 20219
Reprinted with permission.

Key Words: Banking; Financial access; Low-income populations; Sensitive survey topics.

I. Introduction

The Office of the Comptroller of the Currency (OCC) is responsible for regulating and supervising the national banking system, which comprises all banks with a federal charter and about 55 percent of the total assets of the U.S. banking system. The OCC's major objectives include ensuring bank safety and soundness and seeking to assure fair access to financial services for all Americans. As part of the latter effort, the OCC has sponsored a Survey of Financial Activities and Attitudes (Survey) to better understand why millions of adults in the United States rarely, if ever, conduct their regular financial activities, particularly savings and transaction activities, through banks.

We define "nonbanked" individuals as those who conduct their financial activities primarily outside the banking system, including those who may have a small, inactive bank account but who primarily patronize nonbanks when conducting their regular financial activities. Thus, our "nonbanked" target group includes more individuals than those conventionally termed the "unbanked," the 13 percent of American households (containing over 30 million adults) who have no transactions accounts at any depository institution [1].

Early on in the development of the Survey, the OCC convened a Forum on approaches to servicing the nonbanked. Participants included knowledgeable representatives of banks, check cashing and bill-payment firms, technology providers, consumer organizations, financial sector consultants, academic experts, and policy makers. The discussions revealed that survey information on the nonbanked would be useful to many; that surprisingly few financial organizations conduct surveys; and that few of those studies are publicly available [2].

Accordingly, the OCC sought to supplement the sparse information available on the nonbanked, and to provide several public benefits. The Survey will generate new data on financial activities of the nonbanked, the costs they thereby incur, their attitudes toward banks, and any prior experience with banks. This information may

help reduce the uncertainty that banks face in considering how to serve this potential market, and may help inform bank efforts to design more appropriate products and more effective outreach methods to increase financial access. Public availability of the OCC's survey results, the survey instrument, and the survey methodology may also help reduce costs of private-sector market research, product development, and outreach efforts.

Moreover, the findings from the Survey may help the OCC carry out its supervisory responsibilities more effectively. By better understanding the factors limiting financial access to, and use of, bank-provided financial services, the agency can better propose ways for banks to consider removing barriers; identify banking practices that may inadvertently limit financial access; and assess the extent to which banking practices meet local financial service needs.

Several factors posed special challenges to the development of a survey design that would achieve good coverage and high response rates. These included characteristics of the Survey's target population, the sensitive topics addressed by the Survey (personal financial matters), and the paucity of prior research on the topic. This paper discusses the ways in which we met those challenges in developing the survey design.

II. Development of the Survey

A review of the literature revealed that little research has been carried out on the nonbanked or on overlapping populations (the unbanked, financial transactions of low- and moderate-income households, customers of alternative financial service providers, etc.) [3].

OCC researchers recognized the need to supplement their knowledge of conventional banking markets in order to improve the quality of the Survey instrument and its relevance to the nonbanked. This effort entailed: (1) analyzing national survey data on overlapping populations to better identify our target population; (2) identifying key issues raised by participants in the OCC Forum; (3) conducting ethnographic work (in English and Spanish) to identify the questions, common financial terms, and answer options most relevant for the target population; (4) conducting several pre-tests and a set of cognitive

¹ Citation: Constance R. Dunham, Fritz J. Scheuren, and Douglas J. Willson (1998), "Methodological Issues in Surveying the Nonbanked Population in Urban Areas," in *Proceedings of the Survey Research Methods Section, American Statistical Association* (1998), pp. 611-16.

interviews, all with interviewer de-briefings, to test core Survey questions; (5) conducting a nationwide pilot survey to further test core Survey questions; and (6) providing drafts of the Survey instrument to bankers, community representatives, researchers, policy makers, and survey experts for review [4].

One national survey data set we analyzed was the Federal Reserve's Survey of Consumer Finances (SCF), the source of the estimate of 13 million unbanked U.S. households, which comprise a substantial subset of the nonbanked. We used 1995 SCF data to develop a rough proxy measure of nonbanked status and socio-economic characteristics of the nonbanked [5].

Based on this rough proxy measure, about 90 percent of the nonbanked are low- and moderate-income households. The proportions of nonbanked individuals were particularly high for Hispanics (55 percent) and blacks (49 percent). Even controlling for income, wealth, and other characteristics, the percent nonbanked was relatively high for those two groups, suggesting that they may face higher barriers. A relatively low proportion of non-Hispanic whites were nonbanked (15 percent), but non-Hispanic whites are so numerous (75 percent of all U.S. households) that their nonbanked portion comprises about half of our target population. Together, these considerations supported our choice of the sampling strata discussed below [6]. Since Spanish is the primary language for a large portion of Hispanics in the United States, the Survey is being conducted in English and Spanish, at the respondent's discretion [7].

The OCC's nationwide pilot survey was administered to 927 randomly-selected adults in 1997 to test core questions of the Survey, as part of an ongoing national omnibus survey conducted by University of Maryland's Survey Research Center (SRC). SRC's survey methodology entails random digit dialing (RDD) telephone interviews in English. This methodology limited pilot survey respondents to English-speaking individuals with telephones, so we expected that the pilot survey would obtain lower coverage of our target population than surveys (such as the SCF) that are administered in person and in English or Spanish. For example, we expected under-representation of low-income individuals (less likely to have telephones) and Hispanic households (less likely to speak English).

Indeed, while the estimated proportion of unbanked households in the 1995 SCF was 12.9 percent, the estimated proportion of unbanked individuals (an upper bound of the percentage of unbanked households) in the 1997 OCC pilot survey was only 10.1 percent. Thus, while the primary purpose of the pilot survey was to test core Survey questions, this outcome further supported the OCC's aim of developing a survey design appropriate to

the target population, one likely to achieve good coverage and high response rates.

III. The Survey Design

The survey design was developed around several key issues: the financial access policy issues studied, the geographical areas surveyed, the target population, and the survey mode.

Policy issues studied. The Survey is intended to generate new data on how nonbanked individuals conduct financial activities (i.e., receiving and making payments, saving, making remittances, using credit cards, and borrowing); the institutions they use (supermarkets, post offices, check cashing outlets, depository institutions, and bill-paying and wire transfer services); their attitudes toward banks and other financial service providers; any prior experience with banks; and costs they incur in carrying out financial activities.

The Survey will also obtain this information for individuals who are similarly situated (in terms of socio-economic characteristics and geographic location) but who primarily obtain services from banks or other depository institutions (and whom we term "bank-serviced"). Learning why these low-income households have become and have remained bank customers may reveal much about why their nonbanked counterparts have not. Patterns of terminations and item refusals in our pilot survey showed that questions about dollar values of assets and liabilities are particularly sensitive. Accordingly, the Survey instrument contains no questions on dollar asset values.

Geographical areas surveyed. The target population for the Survey are low- and moderate-income residents of Los Angeles County and New York City. While a national probability sample would have been desirable, cost and our limited knowledge of the demographics of the nonbanked precluded us from taking that approach. A very large initial screening sample would have been required to reach a sufficient number of nonbanked respondents, as we learned from the pilot RDD survey. In addition, we lacked sufficient data to identify and oversample primary sampling units (PSUs) with large expected proportions of the nonbanked.

Limiting the geographical scope of the Survey has several key advantages. It permits us to better interpret the results in light of local effects, such as local regulations or institutional practices. It permits us to collect a large enough sample size per urban area (1,000) to generate statistically significant results in each. It also resembles the geographic scope of market research conducted by many banks, which may make it easier for some of them to adapt the Survey to other urban areas.

We selected New York and Los Angeles as survey sites based on information developed in the literature review and obtained from OCC compliance examiners. This showed that New York City and Los Angeles County are among the top few areas in the U.S. with large clusters of innovative bank practices intended to reach the traditionally nonbanked [8]. In those two areas the Survey may provide useful baseline data, to which the results of future surveys may be compared. This comparison could reveal which bank efforts have been most effective in reaching and servicing the nonbanked.

Target population group. The Survey focuses on respondents' financial activities and their attitudes toward financial service providers. The unit of analysis is the individual, and no responses are permitted by proxy. Interviewers randomly select one adult per dwelling unit. This approach provides information on a wider range of nonbanked individuals than would selection of only the financially most knowledgeable individual within the unit, by including those who are least well-connected to the banking sector.

For each of the two urban areas, we developed a multistage stratified design. The first stage sample is of 1990 Census tracts stratified by income (low or moderate) and race and ethnicity (predominantly white, predominantly black, and predominantly Hispanic), plus an integrated low- and moderate-income stratum. Tracts were ineligible if a large portion of the residents spoke a primary language other than English or Spanish, or if residents were predominantly full-time students [9]. A total of 21 Census tracts were sampled per urban area (i.e., three tracts per stratum, except in Los Angeles, where the only two eligible non-Hispanic white, low-income tracts were selected with certainty and four moderate-income Hispanic tracts were sampled).

Based on our analysis of SCF data, we expect to interview approximately 50 percent nonbanked and 50 percent bank-serviced individuals. Because of the uncertainty, we initially released half the sample in each urban area, mainly in low-income strata where we expect to find high rates of the nonbanked. If our expectations are met, we will release the second half of the sample so as to achieve an equal allocation of sample per stratum. Otherwise, we will over-sample areas with higher interim percentages of nonbanked in order to achieve the 50 percent target.

Survey mode. The second stage sample is of dwelling unit addresses in the sampled Census tracts. Sampled addresses are designated as in-person or telephone interviews, based on whether or not the address can be linked to a telephone. If a telephone link exists, the address is part of the telephone sample. If not (due to unlisted telephone numbers, gaps in the coverage of

telephone lists, or dwelling units without a telephone), it is part of the in-person sample [10]. About 50 percent of the sample was drawn for each mode.

Other modal designs were considered and rejected. A single-mode in-person survey would have been too costly. A single-mode telephone survey would have had inadequate coverage, especially of low-income individuals, who are less likely to have telephones [11]. The low average level of literacy in our target population argued against a mail survey [12]. The dual-mode (telephone/in-person) approach helps us achieve coverage and response rate goals at reasonable cost.

A dual-mode approach helps us achieve our 75 percent response-rate target in other ways as well. Respondents can be reached through alternative modes if it will increase the likelihood of obtaining the interview [13]. Some respondents call a toll-free number to be interviewed. Respondents who do not answer their listed telephones are visited in person. All respondents may reschedule an interview to a more convenient time. By adjusting the mode to the respondent, we expect to be able to achieve our response rate and coverage goals more efficiently, and to reduce perceived respondent burden.

IV. Sample Design

Sampling frame. We employed a single, area sampling frame. We developed a comprehensive list of addresses for Census tracts in which we are surveying. We then carried out a telephone directory match on the entire address frame for those tracts.

Consideration had been given to using a dual frame, with RDD used to select telephone interview cases. We decided against this for several reasons. First, a single frame approach enabled us to send an advance letter from the Comptroller of the Currency to every sample household. Advance letters from federal government sponsoring agencies can raise response rates [14]. This suggests that an advance letter should be sent to every sample household, to avoid biasing the sample. With a dual frame approach, we could not send out advance letters to all respondents, since we would not have addresses linked to telephone numbers for a part of the RDD sample. This could result in a lower response rate for that part of the sample.

The area frame closely links sample households to specific Census tracts, while an RDD identifies telephone numbers that can be only loosely linked to specific Census tracts. RDD generates enough out-of-sample area telephone numbers that RDD interviews would need to start with a question about the respondent's location in order to screen out the sizeable number of geographically-ineligible individuals. That question could increase

respondents' concerns about the confidentiality of their answers. Since personal finance is a highly sensitive topic, heightened concern over confidentiality could impair response rates. In contrast, the area frame generates sufficiently few out-of-sample interviews (caused, for example, by reassignment of a telephone number; by respondents moving outside the survey area but keeping their telephone number; or errors in directory listings), that we need ask about respondent location only at the end of a telephone interview, which does not affect the response rate.

Finally, a single frame survey generates results that are easier to analyze and interpret than those from a dual frame survey. This is a clear advantage for the Survey, through which the OCC aims to demonstrate a viable method by which banks can collect and assess information on the nonbanked.

Representative sample. We chose a multi-stage stratified random sample design. A purposive selection of several neighborhoods per city might have enabled us to better interpret local area effects, but research on the nonbanked is still too sparse to reliably guide purposive selection of specific survey sites. If we chose neighborhoods based on inaccurate prior views, we might obtain more precise but less useful results. A multi-stage stratified design, in contrast, generates results that are unbiased, if somewhat less precise.

We decided to accept the loss in precision in favor of obtaining unbiased results. Having unbiased results will enhance our confidence in policy recommendations that are based on the statistically significant findings.

We employed Census tracts as PSUs and selected them with probabilities proportional to an estimate of the size of the nonbanked population within each tract. Samples of equal size were selected from the telephone (i.e., listed) and non-telephone (i.e., unlisted telephone and non-telephone) strata. To estimate the variances for the multistage stratified design, we selected 1,000 samples of Census tracts, estimated the variance for a sample proportion conditional on the selected tracts for each sample, and then accounted for the effect of clustering on each sample using the relationship between variances of a single-stage cluster sample and a simple random sample of the same size.

The variances from the multistage stratified design depend in part on ρ , the intracluster correlation coefficient, which measures the heterogeneity within a cluster (e.g., a Census tract) relative to the heterogeneity within the population as a whole. In most situations, ρ is positive because clusters are relatively homogeneous when compared to the population. Such cluster samples are less efficient than simple random samples, and their variance of estimates will be larger than variances of

estimates from a simple random sample of the same size. On the other hand, the more heterogeneous each cluster the more similar is its variance to that developed from a simple random sample.

To obtain some idea about probable values of ρ , we used 1990 Census data for the relevant tracts in New York City and Los Angeles County to construct an estimate for the population. For the proportion of households that had interest, dividend, and net rental income, in the appropriate tracts and in each stratum, ρ was approximately 0.02 [15].

With $\rho = 0.02$, random selection of tracts will generate estimates in Los Angeles within 5.5 percent, and in New York within 4.5 percent, of true values with 95 percent confidence. (In addition, to account for random selection of eligible adults within dwelling units, estimated variances should be increased a further 20 to 30 percent for either a purposive or multi-stage stratified random sample design.)

We consider the decrease in precision for these estimates that is due to clustering an acceptable cost associated with gaining representation of the larger group of tracts underlying the cluster sample.

V. Challenges in Developing the Sampling Frame

PSUs. We chose Census tracts as our PSU, since the coverage of telephone lists at the block level is highly variable. Some blocks had many households but no listed telephone numbers. (We do not know if this is a greater problem in low-income areas.)

Base lists. Most area probability samples require development of an up-to-date list of residential addresses. This usually requires having an enumeration team count and list dwelling units in each survey site. The efficiency of this process is greatly improved if the team starts with a comprehensive base list that excludes geographically ineligible units.

One commonly-used type of base list are the consumer databases compiled by firms such as MetroMail and Donnelly/ First Source. These lists contain data on the great majority of U.S. households [16]. However, non-telephone and unlisted telephone households are not included in these base lists before augmentation. In large urban areas, the coverage of these lists is under 50 percent in many Census tracts, even after augmentation. If we used lists that had such poor coverage in our survey sites, enumerators would have had much more work to develop up-to-date residential addresses.

For our purposes, a better source was the U.S. Postal Service Computerized Delivery Sequence (CDS) file, which contains every physical mailing address that receives mail delivery. Extracts from the CDS file can be

obtained from a variety of direct mail vendors. This list distinguishes business and residential units and includes zip+4, which can be used to closely match the addresses to Census tracts. We developed a base list of addresses, mapped the areas using GIS software, and produced an enumeration manual with street maps, address lists, and instructions. Enumerators updated the base list by walking each street within each PSU of our sample.

Telephone-address links. Another challenge came in matching telephone numbers to dwelling unit addresses in multi-unit buildings. Residents of multi-unit buildings do not usually report their unit number to the telephone company for the listing. Thus, while we could match a telephone number to a building, we could not match it to a specific dwelling unit. However, we needed the dwelling unit number to draw the sample and send each respondent an advance letter. To solve this problem, we gave enumerators lists of multi-unit building addresses and names of individuals with telephones. They verified exact dwelling unit addresses from tenant listings or through the building manager.

VI. Conclusions

We encountered a number of challenges in developing a survey design that would be consistent with the topic and target population of the OCC Survey of Financial Activities and Attitudes. Our solutions should improve response rates, coverage, and reliability of the data. These solutions may also facilitate subsequent market research efforts undertaken by banks and others to improve financial access for traditionally nonbanked individuals.

Notes and References:

- [1] Kennickell, A.B. et al. (1997) "Family Finances in the U.S.: Recent Evidence from the Survey of Consumer Finances," *Federal Reserve Bulletin* (January) p. 7.
- [2] The Forum proceedings, OCC, *Financial Access in the 21st Century* (1997), Washington, D.C., can be ordered from the OCC Communications Department at 202-874-4700 or downloaded from: <http://www.occ.treas.gov/occfinaac.pdf>.
- [3] See Kennickell, A.B. et al., *op. cit.*; Caskey, J.P. (1997) *Lower Income Americans, Higher Cost Financial Services*, Filene Research Inst.: Madison, WI; Bond, P. and R. Townsend (1996) "Formal and Informal Financing in a Chicago Ethnic Neighborhood, Federal Reserve Bank of Chicago *Economic Perspectives*, (July/Aug.); Caskey, J.P. (1994) *Fringe Banking*, Russell Sage Foundation, New York.
- [4] Advice from Marguerite Robinson (Harvard Institute for International Development) regarding the importance of preliminary ethnographic work is gratefully acknowledged, as is her input to the ethnographic work carried out for the OCC by Aguirre International. Advice from Deborah Stone (then at the Bureau of Labor Statistics, now at SAMHSA/CSAP) on survey instrument design and implementation, and her thoughtful review of numerous survey instrument drafts, are also gratefully acknowledged.
- [5] We estimated the number of nonbanked households based on an observation by J.P. Caskey that many nonbank customers have bank accounts but lack sufficient balances to cover their income checks (*Fringe Banking, op. cit.*, p. 73.). Our proxy is the sum of all households without a transaction account at a depository institution and households with aggregate deposit balances of under \$200.
- [6] Other socio-economic factors we considered had weaker independent effects on nonbanked status. These factors included age, net worth, metro/non-metro location, and education level.

- [7] About 78 percent of Hispanic households in the U.S. speak Spanish at home; about half of these report that they do not speak English “very well.” U.S. Dept. of Commerce, Bureau of the Census (1993) *We the American ...Hispanics*. Washington, D.C.
- [8] In a third such area, Chicago, ongoing survey work on similar populations is being conducted. See Bond, P. and R. Townsend, *op. cit.*
- [9] The number of nonbanked Asian-Americans may be under-represented in surveys, such as the SCF, which are conducted only in English and Spanish. Nevertheless, Asian-Americans were not chosen as a stratum in the OCC sampling design because of their small expected numbers among the nonbanked and because the multiple Asian languages would greatly increase the level of effort for this exploratory research.
- [10] Schwarz, N. (1997) "Questionnaire Design: The Rocky Road from Concepts to Answers," in *Survey Measurement and Process Quality*, Lyberg L. et al. (eds.), John Wiley & Sons, Inc., New York.
- [11] Special runs of Current Population Survey data (three-year average, 1994-1996) were carried out for the OCC by the Census Bureau. The data show that 6 percent of U.S. households had no telephones. This percentage is considerably higher for low-income households (under \$15,000 income) who are non-Hispanic white (10 percent), Hispanic (25 percent) or black (23 percent). See also Keeter, S. (1995) “Estimating Telephone Noncoverage Bias with a Telephone Survey,” *Public Opinion Quarterly*, Vol. 59: pp. 196-217; and Thornberry Jr, O.T. and J.T. Massey (1988) “Trends in United States telephone coverage across time and subgroups,” Ch. 3 in Groves, M. et al. (eds.), *Telephone Survey Methodology*, John Wiley & Sons.
- [12] National Center for Educational Statistics (1993) *Adult Literacy in America*. Washington D.C., App. Table 2.3, pp. 144-45.
- [13] Dillman, D.A. and J. Tarnai, Ch. 32 in *Telephone Survey Methodology*, *op. cit.*, pp. 509-28.
- [14] See Dillman, D.A., Gallegos, J.G., and Frey, J.H. (1976) “Reducing refusal rates for telephone interviews,” *Public Opinion Quarterly*, vol.40, No. 1, pp. 66-78.

- [15] We assumed an intracluster correlation coefficient (ρ) of about 0.02 based on Census data on the proportion of households in the appropriate tracts (P_{ij}) and each stratum j (P_j) in each urban area that had interest, dividend, and net rental income in 1989. For a fixed number of M households per tract, the intracluster correlation coefficient for the population in each stratum can be defined as:

$$\rho_j = \frac{M \sum (p_{ij} - P_j)^2 - \sum p_{ij}(1 - p_{ij})}{NMP_j(1 - P_j)}$$

We approximate ρ for each stratum using the average number of households in each tract within stratum j , \underline{M}_j , in place of M . Values of ρ averaged .023 for Los Angeles County strata (ranging from .014 to .037) and .033 for New York City strata (ranging from .020 to .054). Values of ρ for the sample will probably be lower, as households are sub-sampled within each tract.

- [16] These consumer databases are based on sources such as white pages telephone directories, department of motor vehicle records, and real estate transactions. The addresses are geocoded using U.S. Postal Service products, to attach a Census tract and zip+4.