LIST-ASSISTED RDD TELEPHONE SURVEYS

F.J. Potter, J.J. McNeill, S.R. Williams, and M.A. Waitman
Frank Potter, RTI, PO Box 12194, RTP, NC USA

Key Words: Telephone Surveys; Housing Costs

1. Introduction

The Research Triangle Institute (RTI) developed and field tested a survey methodology and instrument for use in a national longitudinal telephone survey of renters under a contract with the Department of Housing and Urban Development (HUD). A list-assisted random-digit-dialing (RDD) telephone survey design was designed and implemented in selected metropolitan and nonmetropolitan counties of HUD Regions 6 and 10 during early 1991. List-assisted RDD telephone samples are available from various vendors. RTI's design differs from these as to the methods used to achieve survey efficiency, such as, exclusion rules and stratification methods used. RTI also conducted a methodological investigation into effectiveness and bias potential of various design strategies. In this paper, we discuss the research goals that led to the design, the sampling design, the sampling frame and its stratification, sample selection, and the results of the survey design and related methodological investigations.

2. Research Goals

HUD needed information on the current rental costs for one- and two-bedroom rental units to develop estimates on the Fair Market Rent for HUD's Section 8 Housing Assistance Payments Programs. The information desired included the current contract rent and the costs for utilities paid for by the renter. Additionally, the design was to be suitable for efficient longitudinal surveys. Telephone survey was the choice of data collection modes because of the limited information sought and the quick turn-around time afforded by telephone surveys.

3. Method Overview

RTI previously had experience with the Mitofsky-Waksberg (M-W) RDD sample design (Waksberg 1978) for similar surveys of renters in three urban areas. That experience indicated that between 35 and 45 percent of the telephone numbers attempted in an M-W cluster design will identify residences. RTI sought to develop a more efficient RDD survey design using a stratified list-assisted RDD telephone design (Potter et al. 1991). This design is based on information about the count of published residential telephone numbers in each 100-block of potential telephone numbers, where a 100-block is defined by the area code, the exchange, and the first two of the last four digits. The information on these counts was obtained from Nielsen Media Research, Inc. (Nielsen). Nielsen's Total Telephone Frame (TTF) was supplemented with active exchanges containing no published residential telephone numbers (active exchanges according to the most recent BELLCORE file; see Section 4).

The counts of published telephone numbers was used to assign 100-blocks to sampling strata (called density strata) formed using the cumulative square-root of the frequency (Cochran 1963). We also developed cost and variance models and estimated an optimal sample size for each density stratum.

Design efficiency was further enhanced by obtaining information on the likely usage of sample 10-digit numbers by reverse-matching them to lists of residential and business telephone numbers. The prescreening information was used to construct secondary strata. The number of unproductive telephone attempts was thus reduced by the use of disproportionate sampling and improved scheduling of work.

The 100-blocks with no published residential numbers (called zero-blocks) represent the majority of the 100-blocks and were expected to contain few residential households. The methodology of telephone sample vendors reviewed during the study excluded these zero-blocks in addition to 100-blocks with only one or two listings. RTI's design excluded only the zero-blocks. The exclusion of the zero-blocks substantially improved the efficiency of the design, but introduces a potential for undercoverage bias (in addition to that related to exclusion of households with no telephone). RTI conducted a methodological investigation to determine the proportion of residential households in zero-blocks and the characteristics of these households. The feasibility of using automated (computerized) dialing in the survey was also investigated.

RTI used computer-assisted telephone interviewing (CATI) for data collection. CATI was used to monitor survey status and facilitated the control of sampling rates by stratum during data collection, but the use of CATI is not essential to the method.

4. The Basic Sampling Frame

The source of information about eligible 100-blocks of telephone numbers is the Nielsen's TTF. Nielsen constructs the TTF using information from Donnelley Marketing Information Services, Inc. on published residential telephone numbers (subsequently called Donnelley's list). This information is merged by area code and exchange with a file of more than 32,500 working exchanges obtained from BELLCORE. Three groups of exchanges are formed: (1) exchanges on both Donnelley's list and BELLCORE's list; (2) exchanges on Donnelley's list only; and (3) exchanges on BELLCORE's list only.

Nielsen considers exchanges only on Donnelley's list to be either errors or exchanges no longer in service. These exchanges are excluded from the frame. They consider exchanges only on BELLCORE's list to be either for nonresidential usage (for example, business or government usage) or new exchanges with no households assigned.

Nielsen also excludes these exchanges from the TTF, but they listed these for RTI along with information about their location; service rates and telephone exchanges are assigned to geographic areas that are linked to a specific city or town in that area, referred to as the Rate Center City for that area. These 100-blocks were included in our pool of zero-blocks.

For exchanges on both Donnelley's list and BELLCORE's list, Nielsen forms 100-blocks of telephone numbers and, using the Donnelley list again, categorizes the 100-blocks of telephone numbers by the count of published residential numbers in each block. This count, along with past RTI experience with using the TTF, provides an indication of the potential "hit rate" of residential households in each 100-block. The actual "hit rate" in a 100-block may be different because the count for the 100-block does not contain unpublished residential numbers or newly established residential telephone numbers, but does contain recently terminated telephone numbers and multiple listings.

Nielsen creates sampling frames three times per year for the purpose of conducting national household telephone surveys (Nielsen Media Research 1988). The TTF consists of all 100-blocks with one or more published residential telephone numbers identified from this process. Nielsen has determined that approximately three percent of the telephone households are missed in this process because Nielsen excludes the zero-blocks.

For the initial sampling frame for these telephone surveys, RTI augmented the TTF with the zero-blocks. Hence, the initial sampling frame used by RTI contained all possible 100-blocks and therefore provided complete coverage on all possible telephone numbers. After conducting the methodological studies, however, RTI recommended that the final design for the HUD longitudinal surveys exclude the zero-blocks.

5. Stratification

The RTI design utilized three stratification factors: (1) geographic location; (2) density (published telephone numbers in a 100-block); and (3) prescreening usage indication (reverse-matching). The density strata within four geographic strata were developed using information on the estimated telephone household "hit rate" in 100-blocks of telephone numbers. That is, four stratified random samples of 100-blocks were selected (one in each geographic stratum), and then a single telephone number was randomly selected within each selected 100-block. The sample telephone numbers were then stratified by information from TeleMatch, a telephone number service in Springfield, VA, that classified each telephone number as either a residential or business telephone number or a nonmatch (not contained in the TeleMatch file).

HUD identified for these surveys specific sets of counties defining metropolitan and nonmetropolitan areas within HUD Regions 6 and 10 (HUD Region 6 includes Arkansas, Louisiana, New Mexico, Oklahoma, and Texas and HUD Region 10 includes Idaho, Oregon, and Washington). Nielsen assigns each exchange with one or more published telephone numbers to a county (based on the location of the majority of the published telephone numbers in the exchange) to allow for geographic stratification. To ensure full coverage of the survey area, RTI obtained a listing of exchanges assigned by Nielsen to ineligible counties that have one or more published telephone numbers in an eligible county. Some of these exchanges, depending upon number of listings in target counties, were reassigned to the eligible counties. Also, some of the exchanges with no published telephone numbers were assigned to the geographic strata (based on the location of the Rate Center City).

After 100-blocks were assigned to geographic strata, the counts of published telephone numbers were used to form sampling strata of 100-blocks. RTI used the "cumulative square root of the frequency (f) rule" to define the boundaries for these strata, where the frequency count was the estimated number of households (Cochran 1963, p. 130). The estimated household count was based on the count of telephone numbers and the estimated percentage of telephone numbers that identify a household (the "hit rate") in each category of 100-blocks. The estimated percentage of telephone numbers that identify a household in the various density strata was based on RTI's experience with a recent national survey. This recent national survey used a sample of 9,203 telephone numbers, also selected from the Nielsen TTF (RTI 1991). We estimated, for example, that the telephone numbers in zero-blocks would have a residence hit rate of 1.4 percent. The estimated hit rates for the other categories of 100-blocks ranged from 1 to 100 percent. To form the sampling strata, RTI treated the 100-blocks with no listings as a separate stratum. Within each geographic area, the sum of the square root of the remaining frequencies was divided by the candidate number of strata. The density strata boundaries were defined by this interval. Three density strata were thus defined in each geographic area (low-, medium-, and high-density).

RTI contracted with TeleMatch of Springfield, VA to reverse-match the sample of telephone numbers with lists of residential and business telephone numbers. A matching telephone number was classified as either a business or a residential number and so recorded along with the name and address associated with the number. This information was used to form secondary sampling strata. Because most businesses will not have a residential unit, the sampled numbers identified as businesses can be subsampled to improve the efficiency of the sample. Also, a substantial gain in efficiency may be possible because of the opportunity to subsample some categories of nonmatches, especially those numbers that were generated for 100-blocks with few or no listings.

In addition, this information allows for better scheduling of the telephone contact attempts. The telephone numbers identified as business numbers can be worked during normal workday hours, and the telephone numbers identified as residential can be worked during evening hours and on weekends.

6. Sample Size and Allocation

The estimated sample size was based on the density stratification, and the sample was selected prior to reverse-matching. The allocation of the sample was reviewed based on the reverse-matching information about the occurrence of residential and business numbers in the sample. Because the strata containing fewer published telephone numbers per 100-block will have lower hit rates of households than will the strata with more published telephone numbers per 100-block, the cost to identify an eligible telephone household will be higher in these blocks. RTI developed cost and variance models for estimating a 45th percentile (the parameter of interest for the renter surveys) to accommodate the cost differential and to control the precision (Potter et al. 1991). The sample size was estimated using these models and the strata definitions following a Neyman-type optimal sample allocation that takes into account the desired precision and estimated costs for screening and interviewing (Cochran 1963, pp. 95-97).

The initial variance model was developed using an asymptotic estimator of the variance for the 45th percentile (Francisco and Fuller 1986). The asymptotic variance estimator requires the estimation of the density function value of the rental costs evaluated at the 45th percentile. A valid survey-based estimate of this parameter, however, was not available for the variance modeling used to design these surveys. As an alternative approach, a second, simpler variance model was developed based on this initial model. In this alternative, the variance is calculated for estimating a proportion (which follows a binomial distribution) using a stratified random sample design. The requisite sample size was calculated as that required to achieve a 95-percent confidence interval half-width of 0.05 around an estimated proportion, p, for $p = 0.45$ (that is, a 95-percent confidence interval of 0.40 to 0.50).

RTI used a three-stratum design, which excluded the zero blocks, because of the cost differential and because the potential for substantial bias was suspected to be minimal (the bias was subsequently investigated). The density strata are shown in Table 1. In addition to the variance properties, one purpose of this stratification is to control the sample and the number of completed interviews because the hit rates and eligibility rates are estimates and therefore are subject to uncertainty. That is, the stratum-specific hit rate for identifying residential telephone households and eligibility rate among identified households can be monitored and, if more (or less) of the sample is needed in a stratum, additional sample telephone numbers for an individual stratum released (or withheld) without jeopardizing the validity of the sample.

7. Sample Selection Procedure

RTI selected a sample of 100-blocks in each sampling stratum and randomly appended the final two digits to form a complete telephone number. To control the allocation of the sample geographically, the frame was sorted by State, county, and exchange within each sampling stratum. Because of this ordering, the sample within each stratum is implicitly stratified by these factors. The sample of telephone numbers was selected using the probability with minimal replacement sequential sample selection procedure (Chromy 1979).

8. Results

The key methodological findings of this study were the percentage of households identified using the density stratification; the effectiveness of the reverse-matching prescreening to identify potential residential telephone numbers, and the estimation of the proportion of telephone households in 100-blocks with no published telephone numbers.

RTI selected a total sample size of 23,211 telephone numbers across the four geographic strata. A total of 95.7 percent of the telephone numbers were successfully resolved (classified as either residential or business telephone numbers or nonworking). Unresolved numbers include those that resulted in rings but were not answered, that were consistently busy over repeated attempts, or that resulted in a wrong connection. Approximately 55 percent of the RDD numbers in HUD Region 6 identified a household telephone number. For Region 10, 59.6 percent of the numbers in the metropolitan counties were assigned to residential telephones, and 49.2 percent of the numbers in the nonmetropolitan counties reached a household (Table 2). The observed identification rate for households was approximately 23 percent higher than the expected household identification rate in the metropolitan counties and between 11 and 16 percent higher than the expected household identification rate in the nonmetropolitan counties. The expected household identification rate is based on the count of published residential numbers as a percentage of the total number of potential telephone numbers.

As noted earlier, RTI reverse-matched the sample of telephone numbers to a file of business and household telephone numbers. Each sampled number was assigned an indicator of business or household usage or indication that the telephone number did not find a match in the file (a nonmatch). The results of the survey showed that approximately 85 to 90 percent of the telephone numbers classified as a household by the reverse-matching identified a household and between 11.5 and 15 percent of the telephone numbers classified as a business identified a household (Table 3). Moreover, approximately 40 percent of the telephone numbers classified by the reverse-matching as a nonmatch (neither a business nor a household) identified a household in the metropolitan counties, and between approximately 26 and 30 percent of the telephone numbers classified as a nonmatch identified a household in the nonmetropolitan counties. Therefore, reverse-matching is an effective method for classifying telephone numbers to efficiently identify households.

The percentages of telephone numbers that were households for the telephone numbers classified as business were between approximately 11 and 15 percent. Some vendors of list-assisted RDD telephone samples exclude all business telephone numbers. The results of these surveys indicate that the arbitrary exclusion of these telephone numbers may lead to potentially biased estimates.

RTI also conducted two methodological investigations in this contract. The first evaluated the 100-blocks of telephone numbers that were excluded from the sampling frame (the zero-blocks). The second investigation evaluated the use of a computer to dial a telephone number and automatically detect, code, and terminate the call for certain results or to transfer the call to an interviewer if a person answers the telephone.

From 4,000 RDD numbers in the zero-blocks (1,000 in each geographic area), RTI determined that 0.7 percent of the telephone numbers in the zero-blocks were households (Table 4), whereas 54.5 percent of the numbers in the 100-blocks with one or more published telephone numbers were household numbers. The percentage of households identified was 2.4 percent for the metropolitan stratum in Region 10, but between 0.5 and 0.7 percent for the other geographic strata. The percentage of all telephone households in the geographic area that were identified in the zero-blocks ranged between 1.2 percent in the metropolitan counties of HUD Region 6 to 4.5 percent in the metropolitan counties of HUD Region 10. Therefore, the zero-blocks can include a significant proportion of the telephone households in a geographic area, especially if the frame is not current. RTI computed estimates of the 45th percentile of the gross rental costs both including and excluding these eligible households. Only relatively minor changes were shown in these estimates.

RTI also determined that a majority of the households found in the zero-blocks had newly listed telephone numbers. These telephone numbers were assigned generally within 6 months of the survey date and after the date when the sampling frame was constructed by Nielsen. Data collection for the methodological investigation was conducted in March 1991, whereas primary data collection was conducted in January and February 1991. The Nielsen TTF was constructed in September 1990. Therefore, RTI expects that fewer households would be found if future surveys are performed within a few months of the construction date for the sampling frame. Some bias may be present in the estimates, but the bias is expected to be relatively minor.

The second methodological study was to test a computerized system to screen telephone numbers. This system (developed by Nielsen) uses a computer to automatically detect, code, and terminate calls resulting in a phone company's recorded message, data phone signals, no rings, busy signal, or no answer after five rings. If a person answers the telephone, a trained interviewer will ask "Have I reached a place of business? (Yes/No)" and then say "Please excuse the call." RTI provided Nielsen with 4,000 telephone numbers (1,000 in each geographic stratum) from the zero-blocks for the test. All telephone calls were conducted by Nielsen in a 2-hour period on March 20 1991. RTI also called the same 4,000 telephone numbers at approximately the same time using the same interviewers and questionnaire that were used for the main survey. RTI's attempts were conducted between March 15 and April 4, 1991.

The computerized system (with only one attempt) identified 3,366 (84.2 percent) of the 4,000 telephone numbers as either a household, a business, or a nonworking number. The classification assigned by RTI's interviewers agreed with the computerized screening system's classification for 3,132 (93.1 percent) of the 3,366 telephone numbers (Table 5). The most striking result is that 2,938 (99.3 percent) of the 2,958 telephone numbers classified by the computerized screening system as nonworking were also classified as nonworking by RTI's interviewers. Therefore, a computerized screening system can result in substantial savings because the nonworking numbers can be identified by the computerized screening system with minimal error and cost.

9. Discussion

RDD telephone surveys are widely used for governmental and commercial sponsored survey research. The M-W RDD sample design is frequently used in federally sponsored telephone surveys and, when properly implemented, will result in complete coverage of all telephone households and unbiased estimates. List-assisted RDD telephone sample designs are widely used for marketing research surveys and is now being explored for use in federally-sponsored surveys (Casady and Lepkowski, in press). Some vendors of list-assisted RDD telephone samples have developed sample designs and selection procedures that enhance the success rate for identifying a telephone household by excluding telephone numbers assigned to a business, and thereby, introduce the potential for biased estimates. The purpose of this research was to develop and evaluate a cost-efficient RDD telephone survey design that capitalizes on the information available from lists of residential and business telephone numbers while providing the option to have complete coverage of all telephone households and unbiased estimates.

In relation to the M-W RDD sample design (Waksberg 1978), the sample design described in this paper results in known selection probabilities for each potential telephone number and improved efficiency when zero-blocks are excluded. The exclusion of the zero-blocks can introduce the potential for biased estimates and this sample design offers a method to efficiently investigate the potential for bias. In relation to list-assisted RDD sample design used by some telephone sample vendors, the sample design incorporates multiple levels of stratification to enhance the efficiency of the design while avoiding arbitrary exclusion of telephone numbers.

REFERENCES

Casady, R. J. and J. M. Lepkowski (in press). "Optimal Allocation for Stratified Telephone Survey Designs," American Statistical Association Proceedings of the Section on Survey Research Methods.

Chromy, J.R. (1979). "Sequential Sample Selection Methods," American Statistical Association Proceedings of the Section on Survey Research Methods, 401-406.

Cochran, W. G. (1963). Sampling Techniques, New York: John Wiley & Sons.

Francisco, C. A. and W.A. Fuller, (1986). "Estimates of the Distribution Function with a Complex Survey," American Statistical Association Proceedings of the Section on Survey Research Methods, 37-45.

Nielsen Media Research, (1988). "The Nielsen Total Telephone Frame Database," Northbrook, Illinois: Dun & Bradstreet Corporation.

Potter, F. J., S. R. Williams, and J. J. McNeill (1991). "Regional Telephone Surveys to Estimate Annual Adjustment Factors for Section 8 Housing Assistance Payments Program: Design Report." Submitted to U.S. Department of Housing and Urban Development under contract HC-5796, Task 4.

Research Triangle Institute (1991). "Nationwide Personal Transportation Survey: Final Operations Report." Submitted to U.S. Department of Transportation under contract DTFH 61-88-C00030.

Waksberg, J. (1978). "Sampling Methods for Random Digit Dialing," Journal of the American Statistical Association, 73, 40-46.

Table 1. Definition of Density Strata
(Published Telephone Numbers [listings] per 100-Block)

HUD Region	Density Strata	Metropolitan Strata	Nonmetropolitan Strata
6	Low	1 - 41 listings	1 - 46 listings
	Medium	42 - 58 listings	47 - 65 listings
	High	59 - 100 listings	66 - 100 listings
10	Low	1 - 45 listings	1 - 47 listings
	Medium	46 - 61 listings	48 - 65 listings
	High	62 - 100 listings	66 - 100 listings

Table 2. Expected and Observed Residential Telephone Households As Percentage of All Potential Telephone Numbers By Density Strata

HUD Region	Geographic Strata	Density Strata	Listings Categories	Expected[a] (Percent)	Relative Observed Percent)	Difference (Percent)
6	Metropolitan	Total		44.3	54.5	23.1
		Low	1 - 41	25.3	36.0	42.1
		Medium	42 - 58	50.1	63.2	26.0
		High	59 - 100	66.2	71.1	7.4
	Nonmetropolitan	Total		47.2	54.7	15.9
		Low	1 - 46	20.9	28.1	34.1
		Medium	47 - 65	57.0	67.6	18.7
		High	66 - 100	72.7	76.5	5.2
10	Metropolitan	Total		48.3	59.6	23.4
		Low	1 - 45	27.4	41.6	52.1
		Medium	46 - 61	54.0	65.5	21.2
		High	62 - 100	67.9	75.1	10.6
	Nonmetropolitan	Total		44.4	49.2	10.9
		Low	1 - 47	18.4	23.8	28.9
		Medium	48 - 65	57.6	63.4	10.0
		High	66 - 100	72.0	74.7	3.7

1 Expected percentage based on count of published numbers in 100-blocks.

Table 3. Reverse-matched Sampled Numbers Identifying Residential, Nonresidential, and Nonworking Telephone Numbers (Unweighted Percentage of Reverse-matched Sample)

	Unresolved (Percent)	Percent of Resolved Numbers		
		Residential	Nonresidential	Nonworking
Total (All Areas)				
Residential	3.4	87.3	1.9	10.8
Nonresidential	1.6	12.9	74.5	12.6
Nonmatch	5.5	33.9	12.7	53.4
Region 6				
Metropolitan Counties				
Residential	1.9	84.6	2.0	13.4
Nonresidential	0.6	11.7	74.0	14.3
Nonmatch	5.0	40.9	13.1	46.0
Nonmetropolitan Counties				
Residential	3.9	90.2	1.7	8.1
Nonresidential	2.9	12.5	73.0	14.5
Nonmatch	5.0	29.7	10.4	59.9
Region 10				
Metropolitan Counties				
Residential	2.8	86.1	2.5	11.4
Nonresidential	0.8	12.6	75.6	11.8
Nonmatch	6.0	39.1	17.0	43.9
Nonmetropolitan Counties				
Residential	4.2	87.0	1.7	11.3
Nonresidential	1.3	15.0	75.8	9.2
Nonmatch	5.5	26.7	11.1	62.3

Table 4. Percentage of Sampled Numbers Identifying a Residential or Nonresidential Telephone Number or a Nonworking Number in 100-blocks with No Published Telephone Numbers

HUD Region	Geographic Strata	Residential	Nonresidential	Nonworking
All Areas		0.7	5.1	94.1
6	Metropolitan	0.5 (1.2)[a]	11.5	88.0
	Nonmetropolitan	0.7 (4.1)	1.7	97.6
10	Metropolitan	2.4 (4.5)	11.9	85.7
	Nonmetropolitan	0.6 (3.2)	1.7	97.7

[a] Percentage of all telephone households in 100-blocks with no published telephone numbers

Table 5. Summary of Nielsen's Semiautomated Telephone Number Screening

	Nielsen's Count	Number Matching RTI's Result	Percent Matching
Total[a]	3,366	3,132	93.1
Nonworking	2,958	2,938	99.3
Residential	40	17	42.5
Nonresidential	368	177	48.1

[a] Excludes 634 telephone numbers that Nielsen classified as no answer, busy, or unknown.