

PRE-IDENTIFICATION OF NONWORKING AND BUSINESS TELEPHONE NUMBERS IN LIST-ASSISTED RANDOM-DIGIT-DIALING SAMPLES

Michael P. Battaglia, Abt Associates Inc.; Amy Starer, GENESYS Sampling Systems;
Jerry Oberkofler, GENESYS Sampling Systems;
Elizabeth R. Zell, Centers for Disease Control and Prevention

Contact Author: Michael P. Battaglia, Abt Associates Inc., 55 Wheeler St., Cambridge, MA 02138

KEY WORDS: *Working number rate, cost-efficiency, coverage bias*

INTRODUCTION

List-assisted random-digit-dialing (RDD) samples can be drawn using equal probability of selection methods from a sampling frame consisting of hundred series banks containing 1+ directory listed residential telephone numbers. Connor and Heeringa (1992) examine list-assisted RDD sampling. Brick, Waksberg, Kulp, and Starer (1994) examine the noncoverage bias resulting from the exclusion of "zero banks" from the sampling frame.

We look at computerized procedures for increasing working number rates, and therefore the cost-efficiency of list-assisted RDD samples, and discusses the potential biases that could be introduced by procedures designed to keep interviewer dialing costs down. The procedures for pre-screening RDD samples to identify and remove a portion of the business and nonworking numbers, before the sample numbers are dialed, are discussed. Although computerized procedures for increasing the cost-efficiency of list-assisted RDD samples have been in place for several years, other than some work by Biemer and Atkin (1994), there has not been much research on this topic. If the pre-screening procedures can substantially increase working number rates without falsely identifying residential numbers as business or nonworking numbers, coverage bias will not be a concern.

The National Immunization Survey(NIS) offers an opportunity to examine the pre-screening procedures on a continuing, large-scale basis. The NIS uses stratified quarterly list-assisted RDD telephone samples generated by the GENESYS sampling system (1994). The 78 strata cover the entire U.S. and consist of Immunization Action Plan (IAP) areas. IAP areas may be entire states, a single county, a multiple county area, or an individual city. The total sample size for each quarter consists of roughly 475,000 sample telephone numbers.

As part of the process of constructing the sampling frame for each IAP area, current estimates of total

households in a prefix area are summed to the IAP-level. Taking the ratio of the estimated number of total households in the IAP area to total telephone numbers in the retained banks in the IAP area yields an approximate expected residential number rate for the IAP area. The expected working number rates vary considerably by IAP area.

TECHNOLOGIES FOR IDENTIFYING BUSINESS AND NON-WORKING NUMBERS

The ten-digit numbers that make up an RDD sample can be categorized in a way that simplifies the explanation of pre-identification processes:

Residential Numbers

- *Dedicated Numbers:* These are telephone numbers that are only used for normal residential phone service.
- *Dual Purpose Numbers:* These are telephone lines that are used for primary residential service, but not exclusively.

Non-Residential Numbers

- *Dedicated Numbers:* These are telephone numbers that are used for only one purpose-Business FAX or Modem applications.
- *Dual Purpose Numbers:* These are business telephone lines that are used primarily for business purposes, but are also used for FAX/Modem applications on a periodic basis.
- *Unassigned Numbers*

GENESYS-ID utilizes both database and computer telephony technology. Once the RDD sample has been generated, it is ready to be processed through the three distinct phases of GENESYS-ID.

The Pre-Test (or Pre-Dialer) Phase

During this phase, the generated RDD numbers are processed against two databases in order to remove numbers that do not need to be tested for their working/nonworking status.

First, the generated RDD telephone numbers are processed against the GENESYS-Plus business database. RDD telephone numbers found on this database are purged. The remaining ten-digit RDD numbers are passed against the Donnelley Listed Household Database. Sample numbers found in this database are temporarily set aside.

The Test (Dialer) Phase

During this phase, the remaining RDD numbers (i.e., those not identified as business or listed residential numbers in the pre-test phase) are processed through the dialer. System features include:

- *Restricted Hours of Operation.* The GENESYS-ID system operates 9 a.m.- 5 p.m. (local time) Monday through Friday. This limits intrusiveness, as these hours represent the time of day when people are least likely to be at home.
- *Ring Detection.* The dialer software is designed to listen for tri-tone FAX tones, and modem tones. In order to minimize intrusiveness, the dialer software has also been modified to control ringing in the event a tone is not detected immediately.
- *Multiple Passes.* This enhancement allows the numbers to be processed through the test phase more than once. That is, all the numbers are tested once, and those not identified as non-working, are tested again.

The Post-Test (Post-Dialer) Phase

During this first phase, the RDD sample is reconstructed. The RDD numbers identified as listed households in the pre-test phase are recombined with those RDD numbers not identified as non-working in the test phase.

RESULTS OF THE PRE-SCREENING PROCESS

Table 1 shows the results of the pre-screening process by IAP area for approximately 1,000,000 sample telephone numbers. The first column of the table indicates the percent of the total sample for each IAP area that was pre-identified as either nonworking or business numbers. The average rate for the 78 IAP areas is 16.8% removed. The next two columns of the table show the split between nonworking and business number identification. The 78 IAP area average is 12.3% identified as nonworking and 4.5% identified as business.

The fourth column of Table 2 shows the percent of all nonworking and business numbers removed from the sample. The 78-IAP area average is 34.1%. There is, however, considerable variation by IAP area.

EXPECTED RESIDENTIAL NUMBER WORKING NUMBER RATE AFTER PRE-IDENTIFICATION

Before the calling of the sample, it is possible to compute an expected residential working number rate based on the percent of all nonworking and business numbers removed from the sample. The fifth column of Table 2 shows the results of this calculation. The 78 IAP area average is an expected working number rate of 61.3%. We again see considerable variation across the IAP areas.

As a measure of the effectiveness of the pre-identification process, the before versus after pre-identification percent change in the expected working number rate was computed and is shown in the last column of Table 2. The 78 IAP area average percent increase in the expected working number rate is 18.6%.

WAYS IN WHICH THE PRE-SCREENING MIGHT FALSELY REMOVE RESIDENTIAL NUMBERS

During the business purging phase, *dual purpose* residential telephone numbers can be erroneously purged if they are listed on the yellow page database *but not on the white page database.*

Misidentification during the dialing phase can occur due to hardware failure or telephone line problems.

The dialing system classifies non-working numbers by two types: Type A is the Tri-tone non-working intercept, and Type B is a No Ring back. There are two classes of Type B calls. The first returns a non-working result based on receiving 40 seconds of silence. The second class of Type B calls returns a non-working result after receiving six seconds of continuous "noise." Most Type B results are legitimate non-working numbers, but can occur due to the problems stated above.

RESIDENTIAL MISIDENTIFICATION ESTIMATES

During the third quarter of 1994 we had the NIS interviewers dial 5,149 sample numbers that had been identified as nonworking or business numbers by the GENESYS-ID process. These numbers are from the first 43 IAP areas, and thus the western part of the U.S. was excluded. The total sample size for these 43 IAP areas was 21,000 telephone numbers. Based on the final disposition codes assigned to these sample numbers, 3.1% (160) were determined to be residential numbers. The expected residential working number rate for the 43 IAP areas, prior to the pre-identification process, is 51.7%. With 21,500 total sample numbers, 11,115 should be residential numbers. Therefore, we estimate that 1.4% (160/11,115) of the total residential numbers were misidentified by GENESYS-ID.

SUMMARY AND CONCLUSIONS

Working number rates can be expected to increase by a substantial amount after the pre-identification process. The conditions under which misidentification can occur are enumerated. Initial evidence indicates that the proportion of residential numbers incorrectly removed from the sample is only about 1.4%. The misidentification rate is influenced by the time lag between the pre-identification phase and the actual dialing of the sample. In the NIS, the interviewers dial the sample within two weeks of the completion of the pre-identification phase.

References

- Biemer, P., and Akin, D. (1994). The efficiency of list-assisted RDD sampling schemes for single and dual frame surveys. Paper presented at the Annual Meetings of the American Statistical Association.
- Brick, J.M., Waksberg, J., Kulp, D., and Starer, A. (1994). Bias in list-assisted telephone samples. Paper presented at the Annual Conference of the American Association for Public Opinion Research.
- Connor, J., and Herringa, S. (1992). Evaluation of two cost efficient RDD designs. Paper presented at the Annual Conference of the American Association for Public Opinion Research.
- GENESYS Sampling System (1994). Methodology. Internal GENESYS report.

Table 1. Expected Working Number Rate by IAP Area for an EPSEM RDD Sample After Pre-Screening

IAP Area Name	(1) Percent of Total Sample Removed	(2) Percent of Total Sample Removed: Business	(3) Percent of Total Sample Removed: Nonworking	(4) Percent of Nonworking and Business Numbers Removed	(5) Expected Residential Working Number Rate After Removal	(6) Percent Change in Working Number Rate
Connecticut	12.7%	4.4%	8.3%	24.2%	54.3%	14.5%
MA-Rest of state	15.8%	4.8%	11.1%	33.2%	62.3%	18.8%
MA-City of Boston	19.4%	4.2%	15.2%	32.5%	49.9%	24.1%
Maine	27.2%	4.0%	23.1%	54.3%	68.6%	37.4%
New Hampshire	20.7%	5.0%	15.7%	44.3%	67.2%	26.1%
Rhode Island	18.1%	5.1%	13.0%	39.8%	66.6%	22.1%
Vermont	27.6%	4.2%	23.5%	58.6%	73.1%	38.1%
NJ-Rest of state	15.9%	4.8%	11.1%	29.7%	55.2%	18.9%
NJ-Newark	24.4%	3.8%	20.6%	36.6%	44.1%	32.3%
NY-Rest of state	19.7%	4.7%	15.0%	39.1%	61.8%	24.5%
NY-New York City	19.6%	3.9%	15.7%	38.6%	61.3%	24.4%
District of Columbia	24.5%	3.2%	21.3%	35.7%	41.6%	32.5%
Delaware	21.1%	4.5%	16.6%	45.1%	67.5%	26.7%
MD-Rest of state	19.3%	4.5%	14.8%	39.7%	63.7%	23.9%
MD-Baltimore City	28.5%	3.7%	24.9%	55.6%	68.2%	39.9%
PA-Rest of state	17.2%	4.6%	12.6%	39.3%	67.9%	20.8%
PA-Philadelphia	25.7%	3.1%	22.5%	50.0%	65.4%	34.6%
Virginia	19.4%	4.0%	15.4%	40.9%	65.2%	24.1%
West Virginia	17.0%	4.2%	12.9%	45.5%	75.4%	20.5%
AL-Rest of state	11.0%	4.7%	6.3%	29.1%	69.8%	12.4%
AL-Jefferson County	12.7%	4.3%	8.4%	24.7%	55.7%	14.5%
FL-Rest of state	17.0%	5.1%	11.9%	35.5%	62.8%	20.5%
FL-Duval County	24.0%	4.4%	19.7%	50.7%	69.3%	31.6%
FL-Dade County	24.9%	5.1%	19.7%	44.6%	58.8%	33.2%
GA-Rest of state	17.4%	4.3%	13.1%	39.1%	67.2%	21.1%
GA-Fulton/Dekalb Counties	18.5%	4.5%	14.0%	30.8%	49.0%	22.7%
Kentucky	14.2%	4.2%	10.0%	32.3%	65.2%	16.6%

Table 1 (continued)

IAP Area Name	(1) Percent of Total Sample Removed	(2) Percent of Total Sample Removed: Business	(3) Percent of Total Sample Removed: Nonworking	(4) Percent of Nonworking and Business Numbers Removed	(5) Expected Residential Working Number Rate After Removal	(6) Percent Change in Working Number Rate
Mississippi	10.7%	5.0%	5.7%	27.9%	69.0%	12.0%
North Carolina	10.3%	4.0%	6.3%	22.1%	59.5%	11.5%
South Carolina	12.0%	4.1%	7.9%	25.6%	60.3%	13.6%
TN-Rest of state	13.0%	4.3%	8.7%	31.2%	67.1%	14.9%
TN-Shelby County	13.2%	3.5%	9.7%	26.9%	58.6%	15.2%
TN-Davidson County	13.6%	4.9%	8.7%	26.9%	57.2%	15.7%
IL-Rest of state	19.4%	4.2%	15.2%	38.9%	62.2%	24.1%
IL-Chicago	25.5%	3.4%	22.0%	48.7%	64.0%	34.2%
IN-Rest of state	18.4%	4.0%	14.4%	39.6%	65.6%	22.5%
IN-Marion County	32.7%	3.1%	29.6%	59.7%	67.2%	48.6%
MI-Rest of state	17.7%	4.5%	13.3%	39.4%	66.9%	21.5%
MI-Detroit	36.1%	2.4%	33.7%	58.7%	60.3%	56.5%
Minnesota	12.9%	4.2%	8.7%	27.7%	61.4%	14.8%
OH-Rest of state	19.1%	3.7%	15.4%	37.8%	61.2%	23.6%
OH-Cuyahoga County	24.4%	3.7%	20.7%	48.3%	65.4%	32.3%
OH-Franklin County	21.7%	4.2%	17.5%	44.7%	65.8%	27.7%
WI-Rest of state	13.7%	4.7%	9.0%	31.0%	64.6%	15.9%
WI-Milwaukee County	21.7%	4.0%	17.7%	44.2%	65.1%	27.7%
Arkansas	11.7%	5.5%	6.2%	29.9%	69.0%	13.3%
LA-Rest of state	13.0%	5.1%	7.9%	29.3%	63.9%	14.9%
LA-Orleans Parish	12.1%	4.3%	7.8%	24.4%	57.4%	13.8%
New Mexico	9.9%	5.3%	4.6%	24.3%	65.8%	11.0%
Oklahoma	13.6%	5.6%	7.9%	30.7%	64.5%	15.7%
TX-Rest of state	15.4%	4.6%	10.8%	32.4%	62.1%	18.2%
TX-Dallas County	15.1%	4.6%	10.6%	27.7%	53.5%	17.8%
TX-El Paso County	10.0%	5.5%	4.5%	26.3%	68.8%	11.1%
TX-Houston	15.2%	4.5%	10.8%	28.2%	54.4%	17.9%

Table 1 (continued)

IAP Area Name	(1) Percent of Total Sample Removed	(2) Percent of Total Sample Removed: Business	(3) Percent of Total Sample Removed: Nonworking	(4) Percent of Nonworking and Business Numbers Removed	(5) Expected Residential Working Number Rate After Removal	(6) Percent Change in Working Number Rate
TX-Bexar County	15.4%	5.1%	10.4%	33.8%	64.4%	18.2%
Iowa	15.6%	4.3%	11.3%	31.6%	60.0%	18.5%
Kansas	13.9%	4.7%	9.2%	30.2%	62.7%	16.1%
Missouri	14.2%	4.7%	9.5%	32.1%	64.9%	16.6%
Nebraska	22.1%	3.9%	18.1%	39.4%	56.4%	28.4%
Colorado	11.5%	5.0%	6.5%	24.8%	60.7%	13.0%
Montana	12.2%	5.4%	6.8%	25.1%	58.6%	13.9%
North Dakota	13.8%	5.0%	8.7%	26.8%	56.2%	16.0%
South Dakota	18.6%	4.8%	13.8%	39.6%	65.1%	22.9%
Utah	11.1%	4.7%	6.4%	24.6%	61.7%	12.5%
Wyoming	13.1%	5.5%	7.6%	26.5%	58.1%	15.1%
AZ-Rest of state	10.3%	4.8%	5.5%	25.4%	66.3%	11.5%
AZ-Maricopa County	11.6%	4.9%	6.7%	26.3%	63.3%	13.1%
CA-Rest of state	14.0%	4.7%	9.3%	27.4%	56.9%	16.3%
CA-Los Angeles County	11.8%	4.4%	7.4%	22.2%	53.0%	13.4%
CA-Santa Clara County	13.7%	5.1%	8.7%	24.0%	49.7%	15.9%
CA-San Diego County	10.6%	5.3%	5.3%	22.4%	59.0%	11.9%
Hawaii	14.0%	5.1%	8.9%	24.8%	50.5%	16.3%
Nevada	8.0%	4.6%	3.4%	15.2%	51.5%	8.7%
Alaska	14.0%	4.5%	9.4%	22.3%	43.3%	16.3%
Idaho	14.8%	5.3%	9.5%	31.1%	61.5%	17.4%
Oregon	13.5%	5.2%	8.4%	28.0%	59.9%	15.6%
WA-Rest of state	11.9%	4.3%	7.6%	26.6%	62.7%	13.5%
WA-King County	12.3%	5.3%	7.1%	24.4%	56.5%	14.0%
IAP area average	16.8%	4.5%	12.3%	34.1%	61.3%	18.8%