INCREASED EFFORTS IN RDD SURVEYS

J. Michael Brick, David Martin, Patricia Warren, Westat; and J. Wivagg, DataSource
Westat, 1650 Research Boulevard, Rockville, Maryland 20850-3195

Key Words: Response Rates, Telephone Surveys, Call Attempts

1.   Introduction

The Urban Institute conducts the National Survey of America's Families (NSAF) to assess the New Federalism by surveying children and adults in 12 states and in the balance of the nation. The 12 states account for about 50 percent of the country's population. The initial round of the NSAF took place in 1997, with followup rounds in 1999 and 2002. The major objective of the study is assessing the effects of the devolution of responsibility for major social programs. A consortium of foundations, led by the Annie E. Casey Foundation, funds NSAF. Westat is responsible for the sample design, data collection, and weighting. The survey is a dual frame design with a random digit dialing (RDD) sample and an area probability sample. In this paper we consider only the RDD sample. See Cunningham, Martin, and Brick (2003) and the Urban Institute's URL for more details on the survey and its methodology (www.urban.org).

Recent literature on random digit dial (RDD) telephone surveys suggests that increased levels of effort are required to contact households due to changes in technology. For example, see Curtin et al. 2000. Answering machines, caller-ID devices, TeleZappers and privacy managers are examples of the types of technologies that are being used at an increasing rate to shield the households from unwanted calls. In addition, it is speculated that more telephone numbers are devoted to purposes such as fax machines and computers and may never be answered no matter how many times the numbers are dialed. This paper uses Round 1 (1997) and Round 3 (2002) NSAF data to examine some of these issues. The next section gives some basic background information needed to help understand the data collection efforts in the two rounds.

2.   Background

As noted earlier, the three rounds of data collection for NSAF were done in 1997 (Round 1), 1999 (Round 2) and 2002 (Round 3). In all three efforts, the content and the method of interviewing for the RDD sample were very similar. The instruments for each round are at (www.urban.org).

The Round 1 and Round 3 RDD sample designs were also very similar. For these two rounds, independent, list-assisted RDD samples were selected from the study areas and from the balance of the nation. The sample design for Round 2, on the other hand, differed from this design. The Round 2 sample included a substantial subsample from Round 1 and the telephone numbers were differentially sampled depending on the Round 1 outcome. In addition, a supplemental sample of new telephone numbers was included in Round 2 to provide complete coverage and reach the desired sample size for that round. Because the Round 2 design is so different from the other two rounds, we only consider Round 1 and Round 3 in this paper.

While the basic sample designs for Round 1 and Round 3 were similar, there were some important differences in the rounds that affect any comparisons of levels of effort. One change was in the composition of the study areas and the sample distribution. In Round 1, Milwaukee and the balance of Wisconsin were separate study areas with their own RDD samples, but in Round 3 Milwaukee was merged with the balance of Wisconsin to form a Wisconsin study area with one RDD sample. Another difference was that the sample allocated to the balance of the nation was larger in Round 3 than in Round 1.

Other important differences in procedures were implemented over the three years between Round 1 and Round 3. Many of these are discussed below, but two that have very important implications for our analysis are mentioned here. The first procedure is the use of incentives. In Round 1, an advance letter was sent to each telephone number linked to a mailing address, but monetary incentives were reserved for converting households and sample persons who refused. In Round 3, a $2 pre-paid incentive was included for telephone numbers linked to a mailing address. If a person was sampled for the extended interview, they were promised $10. It is also worth noting that the ability to find an address for a telephone number has increased over time. In Round 1 only 38 percent of the sampled telephone numbers had mailable addresses, while in Round 3 the percent increased to 72 percent. Of course, the mailable addresses may not reach the household for the sampled telephone number due to errors of various sorts.

Previous research suggests that 15 to 20 percent of the mailable addresses may not be accurate.

A second procedure that affects comparisons is subsampling households that refused to participate in the screening interview. In Round 3, households that refused to respond to the screening items were subsampled and only those subsampled were subject to refusal conversion efforts. About four percent of all telephone numbers were screeners that refused and were not subsampled for refusal conversion. This accounts for about 11 percent of all the residential numbers in Round 3. The subsampling refusal procedure was not used in Round 1. We comment on both of these procedures in more detail as we examine the outcomes.

The tabulations and comparisons of the outcomes from Round 1 and Round 3 presented below are all based on the actual counts and have not been weighted. The advantage of using the raw counts is that it corresponds to the actual levels of effort expended in the operations at the time, which is the focus of the paper. The main disadvantage is that changes in the design such as those described above are more likely to contribute to differences between the rounds. For example, the change in the composition of the sample

(Wisconsin and balance of the nation samples) would not be as relevant in weighted analyses.

We begin by giving some basic comparisons of the RDD samples selected for the two rounds. Table 1 gives the number of telephone numbers sampled by residential status for each round. As shown in the table, the number of telephone numbers sampled was much greater in Round 3, partially because the residency rate of the sampled numbers in list-assisted RDD samples decreased rather dramatically in the five-year period. For the NSAF samples, the percent of telephone numbers sampled that were residential (excluding the telephone numbers with a residency rate that could not be determined) decreased by nine percentage points, a decrease of 91 percent. Another notable difference was the increase in the percent of sampled telephone numbers that had an unknown residential status. This group increased less than three percentage points, but this change is large compared to the percent of numbers in the category. The other factor that must be taken into consideration is that the number of telephone numbers in the list-assisted frame also changed drastically. The relative difference in the frame increased by 30 percent from Round 1 to Round 3. This increase makes it difficult to interpret the percentage change in the unknown residential numbers.

Table 1.   Disposition of sampled numbers, by residency status

Disposition	Round 1	Round 3	Difference	Relative difference
Number sampled	483,260	556,651	73,391	15.2%
Residential	46.0%	36.9%	-9.1%	-19.9%
Nonworking	32.9	40.2	7.3	22.2
Nonresidential	15.0	14.0	-0.9	-6.1
Unknown residency status	6.1	8.8	2.8	45.1

NOTE: Numbers may not add due to rounding.

The much lower residency rate in Round 3 as compared to Round 1 might suggest that the level of effort needed to reach households must have increased substantially in these five years. However, survey researchers also used new technologies to deal with the increased proportion of sampled numbers that are not residential. In both rounds, the sampling vendor we used, Marketing Systems Group (MSG), purged the sample of telephone numbers of those numbers only listed in the Yellow Pages and then autodialed other numbers to eliminate nonworking and some business numbers from the sample. The MSG purging technology (called ID in 1997 and ID*plus* in 2002) advanced so that 38 percent of the sampled telephone numbers were purged in Round 3, while only 18 percent were purged in Round 1. Of the sampled

telephone numbers that were not purged, the percentage that were residential actually increased by about three percentage points from Round 1 to Round 3. Thus, the lower residential rate in Round 3 shown in Table 1 did not increase the percentage of numbers that had to be dialed in the survey because of this technological innovation. In the subsequent sections the purged telephone numbers are not included.

3.   Findings

This section analyzes and compares the levels of effort from the Round 1 and Round 3 surveys. The analysis is restricted to the screening interview where the initial contact is made and any sampling of persons within the household is conducted. In both rounds the

content of the screener was basically the same. The first analysis examines the number of call attempts to complete the screener. The second analysis looks at telephone numbers that ring and are never answered in an effort to find out if more telephone numbers are used for technological purposes. The third analysis looks at levels of nonresponse over time and how other procedural changes affect these outcomes.

Call Attempts

The number of call attempts to finalize a case is a common measure of the level of effort in a survey. Table 2 gives the mean number of call attempts for the two rounds of data collection by the residency status of the number. Overall, the number of call attempts increased by 10 percent between rounds, but this increase is somewhat deceptive because of some of the procedural differences mentioned above. In particular, the Round 3 refusal cases that were not retained due to subsampling are included in the tabulation, even though calls were not made to these numbers after the initial refusal. As mentioned above, this is an important factor and it complicates the analysis. Furthermore, the

decrease in the mean number of call attempts to numbers with unknown residency status is directly attributable to procedural changes. Cunningham, Martin, and Brick (2003) describe these changes and the rationale for them.

When we examine the mean number of call attempts for the residential numbers, the problem associated with not accounting for the refusal subsampling becomes even more apparent. In Table 3, the mean numbers of attempts for all numbers classified as residential are given by the final disposition of the screener. Most of the categories are obvious, with perhaps the exception of Max Calls. These are households that answer the phone at least once, never refuse, but never are available to complete the interview despite repeated call attempts. The table suggests that for every disposition, except other nonresponse, has a smaller number of call attempts in Round 3 than in Round 1. Since the other nonresponse group is very small (less than 0.1% of all dispositions), it is obvious that understanding the effect of the level of effort requires dealing with the refusal subsampling.

Table 2.   Mean number of call attempts, by residency status

Final Disposition	Round 1	Round 3	Difference	Relative difference
Total	6.2	6.8	0.6	9.7%
Residential	6.9	7.1	0.2	2.9
Nonworking	3.0	3.1	0.1	3.3
Nonresidential	3.4	4.4	1.0	29.4
Unknown residency status	15.9	12.0	-3.9	-24.5

NOTE: Numbers may not add due to rounding.

Table 3.   Mean number of call attempts for residential numbers, by final disposition

Disposition type	Round 1	Round 3	Difference	Relative difference
Complete	4.8	4.5	-0.3	-6.3%
Language problem	14.6	13.5	-1.1	-7.5
Max call	37.9	24.3	-13.6	-35.9
Refusal	13.6	10.3	-3.3	-24.3
Other nonresponse	3.5	11.8	8.3	237.1

NOTE: Numbers may not add due to rounding.

Table 4 divides the residential numbers into those that never had a refusal and those that had one or more refusal to clarify the effect of refusal subsampling. In the top of the table, the number of call attempts for the never refused numbers are relatively consistent across rounds. The only notable difference is that more attempts were made in Round 1 than in Round 3 before classifying the number as a Max Call. A procedural decision was made in Round 1 to set the number of call attempts before finalizing a case as a Max Call to a very

large number. In Round 3, the call attempt limit was reduced from this level. Later we show that the percentage of cases classified as Max Calls more than doubled from Round 1 to Round 3. Since this group of numbers required more attempts, the calling in Round 3 actually increased. Furthermore, even though the level of effort to resolve the Max Calls was greater in Round 1, the literature shows that calls once the number of attempts exceeds 20 there is little increase in the number of completed interviews.

Table 4. Mean number of call attempts for residential numbers, by refusal status

Disposition type	Round 1	Round 3	Difference	Relative difference
Never refused cases				
Complete	3.8	3.8	0.0	0.0%
Language problem	13.7	12.6	-1.1	-8.0
Max call	37.9	24.4	-13.5	-35.6
Ever refused cases				
Complete	7.4	10.0	2.6	35.1
Language Problem	15.6	15.0	-0.6	-3.8
Refusal	13.6	10.3	-3.3	-24.3
Not subsampled refusal	13.6	14.1	0.5	3.7

NOTE: Numbers may not add due to rounding.

The lower portion of Table 4 gives the means for screeners with at least one refusal. The table shows that it took 2.6 more call attempts to obtain a completed screener in Round 3 than in Round 1 (an increase of 35%) if there was a refusal. The row showing the lower mean number of attempts for the refusal cases is again a function of the refusal subsampling. The last row of the table gives the comparable group by excluding those cases that were not retained in the refusal subsampling. It shows that the average number of call attempts was nearly the same for the two rounds for the refusal cases, giving quite a different picture than from the earlier table. Overall, the mean number of call attempts in Round 3 is thus greater than the number in Round 1.

Unanswered Numbers

The next analysis deals with telephone numbers classified as unknown residency status numbers in Table 1. These are those numbers that are dialed numerous times but are never answered. These numbers are partitioned into two groups: Never Answered (NA) numbers are those that ring and are never answered across all call attempts; Answering Machine (AM) numbers are those that are never answered by a person but an answering machine is encountered in one or more of the call attempts. The procedure we followed was to not depend on an interviewer's interpretation of the answering machine message to classify the number as residential or not.

Table 5 gives the percent of all numbers attempted that were classified as NA and AM in each round of NSAF. The table shows that percent of unanswered numbers increased substantially and both the NA and the AM classifications grew, but NA numbers account for most of the unanswered numbers. The literature suggests (e.g., see Oldendick 1993; Oldendick and Link 1994) that answering machines are not a major problem if enough calls are made to the number. The Round 1 results are consistent with this finding, but by Round 3 the percent of AM numbers increased, and the percentage in this category is becoming more problematic.

One of the factors hypothesized to account for part of the large increase in the number of sampled telephone numbers that are not residential (see Table 1) is households dedicating lines for fax or computer use. The increase in the percent of unanswered numbers suggests this possibility. However, as noted earlier the frame of all numbers also changed during this period. Although we have no direct evidence, we would expect a higher percentage of call attempts to such dedicated numbers to be either 'ring no answer' or busy outcomes (if the fax or computer was in use). To investigate this, we examined the NA call attempts in the two rounds. Since both rounds had at least nine call attempts for every NA (in Round 1 all had at least 14 attempts and in Round 3 all had at least nine), we only look at the results of the first nine call attempts for comparability.

Table 5. Percent of dialed numbers that were never answered, by round

Never answered	Round 1	Round 3	Difference	Relative difference
Total	6.1%	8.9%	2.8%	45.9%
NA	5.2	7.3	2.1	40.4
AM	0.9	1.6	0.7	77.8

NOTE: Numbers may not add due to rounding.

When we reached a busy number in NSAF we automatically set an appointment in the call scheduler

to dial that number again 15 minutes later. Up to four consecutive busy outcomes are allowed with this

procedure. Our usual practice is to count these busy dialings as one call attempt (this is the method used in the earlier tables). For this analysis we count each dialing separately. Thus, in the nine call attempts we could have up to 36 busy dialings. Table 6 presents some characteristics of the distribution of the busy dialings for the NAs in the two rounds. The mean number of busy dialings to the NAs doubled from Round 1 to Round 3 (1.4 to 2.9). The percentage of telephone numbers in which half or more of the dialings were busy also doubled. Thus, these data are consistent

with the hypothesis of increased use of numbers for other purposes. The implications are two-fold. First, these numbers require a large number of dialing attempts, do not result in completed interviews, and end up increasing the cost of data collection. Second, some of the numbers with a NA disposition are usually allocated as being residential for computing response rates. If many of these are dedicated for computer or fax use but are counted as residential, then we may be underestimating response rates in RDD surveys.

Table 6.   Distribution of busy dialings for NA telephone numbers

At least …busy dialings	Round 1	Round 3	Difference	Relative difference
1	17.0%	24.6%	7.6%	44.7%
2	10.1	20.5	10.4	103.0
3	8.5	18.8	10.3	121.2
18	3.2	6.8	3.6	112.5
30	1.4	3.3	1.9	135.7
Mean	1.4	2.9	1.5	107.1

NOTE: Numbers may not add due to rounding.

Response Issues

Households that refuse account for the vast majority of all screener nonresponse in NSAF and nearly all RDD surveys. In Round 1, 89 percent of nonresponse was due to refusals with Max Calls and language problems accounting for 6.5 percent and 4.0 percent, respectively. Similarly in Round 3, refusals were 86 percent of all nonresponse (including the cases that were subsampled), Max Calls were 10.0 percent, and language problems were 3.9 percent. As noted earlier, the Max Calls require the greatest number of call attempts and the increase of 177 percent in these numbers does have cost and response rate implications. However, because the percent of nonresponse due to

refusals is so large, this category of nonresponse is the main problem and is considered now.

There are two ways of examining the effect of refusals over time using NSAF data without having to deal with refusal subsampling in Round 3. The first is to examine the percentage of residential numbers that ever refused (initial refusal rate). The second is to examine the percentage of refusals that interviewers classified as hostile (most of these are classified as hostile on the first contact and thus are not affected by subsampling). Table 7 gives the percent of residential numbers that ever refused and the percent of refusals interviewers classified as hostile in the two rounds. Perhaps the most striking result in the table is that the percentages do not vary much across rounds.

Table 7.   Percent of residential numbers that refused and percent of all refusals that were hostile.

Screener outcomes	Round 1	Round 3	Difference	Relative difference
Initial refusal	45.4%	47.3%	1.9%	4.2%
Hostile refusals (of all refusals)	1.0	0.8	-0.2	-20.0

NOTE: Numbers may not add due to rounding.

The refusal rates in Table 7 may appear unexpected, especially since the screener response rate was 77 percent for Round 1 and 65 percent for Round 3 using a weighted version of AAPOR (2000) definition RR3. This relatively large decrease in the response rate is generally not consistent with a constant initial refusal rate. The difference relates to the manner in which

incentives were used in the two rounds. In Round 3, the incentive was front-loaded in the advance letter because evidence suggested doing this might lower the initial refusal rate. In Round 1, incentives were used to convert those households that refused, thus boosting refusal conversion rates. As a result, the refusal conversion rate in Round 1 was 49 percent and in

Round 3 was only 38 percent (for those subsampled for conversion). If the prepaid incentives had not been used in Round 3, it is likely (see the experimental data on incentives in Round 3 in Cantor et al. 2003) that the initial refusal rate for Round 3 would have been six percentage points higher. This result is more consistent with the overall screener response rate and the common perception of the research community that refusals are becoming an even greater problem in RDD surveys over time.

4.   Summary

The findings above show that while much has changed in only five years, it is difficult to evaluate the effect that would have been obtained if the same procedures were used uniformly over time in the same survey. The NSAF is a good example in the sense that the interview itself had only small changes, but the procedures used were constantly revised to attempt to keep up with changes in the RDD survey environment.

The data show that despite the increasing percentage of the frame that is not residential, revisions in methods used by researchers to purge the frame of nonresidential numbers more than keep pace with this change. We did find the mean number of call attempts increased from Round 1 to Round 3, despite changes in calling protocols that limited unproductive calls to numbers that virtually never resulted in a completed interview. The comparisons for residential numbers were somewhat difficult to analyze due to the introduction of refusal subsampling in Round 3. The main points that emerged were: the mean number of attempts for households that did not refuse were constant over time; the mean number of attempts went up substantially for those households that completed after refusing at least once; the mean number of attempts did not change much for the households that ended up in the refusal category; and, the overall mean number of attempts increased also because the number of Max Call cases increased.

We also looked at the NA cases to try to assess if there was any support to the notion that more telephone numbers are being devoted to fax and computer uses. By looking the busy dialings to these numbers, we found nearly twice as many appeared to be dedicated to uses other than regular incoming and outgoing calls. We concluded that this change requires more dialings, but does not increase the completion rate. We also suspect it may be artificially depressing response rates in RDD surveys.

The third topic we examined was initial refusal rates and found that the procedural changes made this difficult to evaluate. While NSAF response rates declined over the five years, the initial refusal rates remained about the same because incentives were used differently. We believe that the Round 3 procedure enabled us to achieve higher response rates than would have been possible for the same cost using the Round 1 methods.

In general, the statement that increased efforts are necessary to achieve good results in RDD surveys is not very relevant. The environment for doing RDD surveys changes rapidly and methods to contact and obtain completed interviews must be continually revised to address those changes. The NSAF is an example of attempts to do exactly that and provide valuable data to analysts.

References

AAPOR (2000). *Standard definitions: Final dispositions of case codes and outcome rates for surveys*. Ann Arbor, Michigan: AAPOR.

Cantor, D., Cunningham, P., Triplett, T., and Steinbach, R. (2003). *Comparing Incentives at Initial and Refusal Conversion Stages On a Screening Interview for a Random Digit Dial Survey*. Presentation at AAPOR in Nashville, TN.

Cunningham, P., Martin, D., and Brick, J.M. (2003). *Scheduler Experiment*. Presentation at AAPOR in Nashville, TN.

Curtin, R., Presser, S. and Singer, E. (2000). The Effects of Response Rates Changes on the Index of Consumer Sentiment. *Public Opinion Quarterly*, 64, pp. 413-428.

Oldendick, R. W. (1993). The Effect of Answering Machines on the Representativeness of Samples in Telephone Surveys. *Journal of Official Statistics,* 9, pp. 663-672.

Oldendick, R.W., and Link, M. (1994). The Answering Machine Generation. Who Are They and What Problem Do They Pose for Survey Research? *Public Opinion Quarterly*, 58, pp. 254-273.