Comments on the Two Articles (1)

Michael P. Battaglia*

Tags: survey practice

Survey Practice

Vol. 2, Issue 1, 2009

Comments on the Two Articles (1)

List-assisted random-digit-dialing sampling retains 100-banks with one or more directory-listed residential numbers (1+ banks). It assumes that a very small percentage of residential telephone numbers are excluded from the sampling frame. Today, with the rapid growth of cellular-only households, the assumption can be restated as a very small percentage of residential landline telephone numbers being excluded from the frame.

With the decline in the proportion of landline households that have a directory-listed telephone number and the bundling by cable television providers of telephone service with Internet access and cable television, it is important to examine coverage of voice-use residential landline telephone numbers in the list-assisted sampling frame, and to identify currently excluded strata that may contain a substantial proportion of voice-use residential numbers. The two key strata where such numbers may exist are 100-banks with zero directory-listed numbers and “remaining POTS” 100-banks.

When a sample of telephone numbers is called, we end up with known households, likely households (e.g., ambiguous answering machine message), nonresidential numbers, and undetermined (unresolved) numbers. The largest category of undetermined numbers is ring no answer to all call attempts.

What is known regarding the total number of residential landline telephone numbers in the U.S.? The August 2008 *Trends in Telephone Service* report by the Federal Communication Commission indicates that there were 89.5 million primary residential wirelines in 2006 and that there were another 10.5 million non-primary residential wirelines. Some landline households have two or more voice-use telephone lines while others maintain one or more additional lines for devices such as facsimile machines, home security systems, etc. Both totals have been declining in recent years and so in 2008 the total number of voice-use residential lines may range from somewhat less than 89.5 million to somewhat less than 100.0 million.

* Institution: Abt Associates Inc.

The current issue and a recent issue of *Survey Practice* contain articles by Fahimi et al. and Boyle et al. examining the RDD coverage issue. Fahimi et al. sampled telephone numbers from the three strata described above. Those numbers were called a maximum of 9 times using Genesys-CSS. The undetermined numbers were then reverse-matched against some number of commercial address data bases. For most samples a small percentage of the reverse matches yield incorrect address information. Base sampling weights were applied and the estimates in Table 1 indicate that around 20% of voice-use residential telephone numbers are outside the 1+ listed stratum. Totals are not presented and sampling variances are not given. It is not clear whether Table 2 includes the results of the commercial data base matching. It would make sense not to include that component of the process if the purpose of the table is to indicate the residential working number rates that will be experienced in the three strata. Looking at the undetermined row of Table 2 it appears that about the same percentage of sample numbers in each stratum ended up unresolved. This is important because the coverage estimates in Table 1 assume that none of the undetermined numbers are voice-use residential telephone numbers.

Boyle et al. present considerably more detail in their article. They defined the same three strata but found differences in the frame totals when comparing SSI to MSG. They ended up adding a fourth stratum from the non-Telcordia banks in order to end up with total frame counts that are within 1.6 percent (9.0 million versus 9.2 million). Examining Figure 1 we would expect the RDD 1+ and RDD Zero rows to be in close agreement. This holds for the RDD 1+ frame counts where the difference is 3.5 percent, but the RDD Zero frame counts differ by 36 percent. Although the two studies were not conducted during the same time frame, it seems very likely that the differences are due to definitional differences in the construction of the sampling frames. This may reduce the comparability of the results by stratum.

Table 1 Universe Counts for Listed and Unlisted Hundreds Banks by Vendor.

	Marketing Systems Group	Survey Sampling Inc.
RDD 1+	2,920,039	2,816,471
RDD Zero	4,009,944	2,562,600
OPOTS	2,099,950	2,286,330
Non-Telcordia		1,506,480
TOTAL	9,029,933	9,171,881

Boyle et al. used up to a maximum of 11 call attempts apparently using predictive dialers. For numbers where contact was made (completed screener), up to 9 additional call attempts were made. Looking at the RDD 1+ column of Figure 2 the 10,000 sample numbers are initially divided into four categories: bad numbers (57.9%), permanent no answer (2.9%), presumed good numbers (22.4%), and completed screener (17.0%).

Table 2 Sample Disposition of Interviewer Screening of Household Status.

Sample Disposition	Percent within Stratum				
	RDD 1+	OPOTS	RDD Zero	Non-Telcordia	Total
Total Numbers Dialed	10,000	10,000	10,000	7,175	37,175
Bad Numbers					
Not in service	48.4%	71.4%	77.0%	97.4%	71.1%
Business/gov/non-res	4.5	7.2	6.9	0.1	5.0
Fax/modem	4.8	8.3	3.0	1.1	4.5
Cell phone	0.1	0.1	0.2	0.4	0.2
Child/teen phone	0.1	0.0	0.0	0.0	0.0
Permanent no answer	2.9	1.1	1.3	0.0	1.4
Presumed Good Numbers					
No answer	4.5	1.8	1.9	0.2	2.2
Answering machine/Voicemail	2.8	1.5	1.3	0.1	1.5
Busy	0.5	3.0	2.3	0.1	1.6
Callback	7.1	0.6	0.6	0.1	2.3
Refusal	6.1	0.4	0.4	0.1	1.9
Language	0.6	0.1	0.1	0.0	0.2
Health/hearing	0.8	0.1	0.1	0.0	0.3
Completed Screener (Percent)	17.0	4.3	5.1	0.3	7.2
Completed Screener (Number)	(1,700)	(433)	(508)	(20)	(2,661)

Figure 3 indicates that the additional call attempts on the 1,700 completed screener numbers yields an estimate for the RDD 1+ column that 72.7% are residential numbers. Continuing to Figure 4 for the 1+ RDD column the ratio (0.404) of the sum of presumed good numbers and completed screeners to the sum of presumed good numbers, completed screeners and bad numbers is applied to the permanent no answers (285) to estimate the number that are presumed good numbers (115), and an estimated total of 4,044 (3,929 + 115) presumed good numbers and completed screeners.

Table 3 Outcome of Interviewer Screening for Household Status.

Household status	Percent within Stratum				
	RDD 1+	OPOTS	RDD Zero	Non-Telcordia	Total
Total Numbers Completed Screen	1,700	433	508	20	2,661
Private residence	72.7%	7.4%	5.3%	50.0%	49.0%
Business	24.8	85.7	89.2	45.0	47.2
Dormitory or group home	0.7	0.9	0.6	0.0	0.7
Non-residential	0.8	5.5	4.7	5.0	2.4
Refused	0.9	0.5	0.2	0.0	0.7

Table 4 Estimated Household Distribution Based On Sample Disposition And Interviewer Resolved Cases.

	1+RDD	OPOTS	Zero RDD	Non-Telcordia	Total
Total numbers dialed	10,000	10,000	10,000	7,175	37,175
Bad numbers (out of frame)	5786	8706	8707	7108	30,307
Presumed Good Numbers	3929	1186	1165	66	6346
Permanent no answers	285	108	128	1	522
Ratio of Presumed Good Numbers/(Bad Numbers + Presumed Good Numbers)	.404	.120	.118	.009	.173
Estimated good numbers out of permanent no answers	115	13	15	0	143
Total potential HH numbers (Presumed Good #s + Estimated Good #s)	4044	1199	1180	66	6489
Interviewer Resolved numbers	1700	433	508	20	2661
Number of households	1236	32	27	10	1305
Percent households	72.7%	7.4%	5.3%	50.0%	49.0%
Unresolved numbers	2344	766	672	46	3828
Estimated hh in unresolved numbers	1704	57	36	23	1820
Actual + estimated households	2940	89	63	33	3125
Estimated Household Rate	29.4%	0.89%	0.63%	0.46%	8.4%

The bottom half of Figure 4 implements a second set of calculations. For the RDD 1+ column the Figure 3 estimate that 72.7% of the 1,700 completed screeners are residential numbers yields 1,236 residential numbers ($1,700 \times 0.727$). For the remaining 2,344 presumed good numbers ($4,044 - 1,700$) the 72.7% estimate is applied to obtain an estimate of 1,704 residential numbers. The total estimated number of residential numbers for the RDD 1+ column is therefore 2,940 ($1,236 + 1,704$) or 29.4%. A similar set of calculations is used in the other columns of Figure 4.

Turning to Figure 5, the estimated percentage of residential numbers for each stratum is applied to the total number of telephone numbers in each stratum to yield a total of 87.1 million residential telephone numbers in the U.S. and that only 5.0% of all residential telephone numbers are outside the 1+ listed banks. The estimate of 87.1 million residential numbers is at the lower end of the FCC range and putting aside sampling variability, seems to point to some underestimation of the total number of voice-use residential telephone numbers in the U.S. As with the Fahimi et al. article no sampling variances are presented. It appears that base sampling weights were not calculated, limiting the analytic utility of the survey data, but this does not cause any problems for the specific estimates presented in the article.

Table 5 Estimated Distribution of Households by Stratum.

	RDD 1+	OPOTS	RDD Zero	Non-Telcordia	Total
Total Numbers by Stratum	281,647,100	228,633,000	256,260,000	150,648,000	917,188,100
Household rate	29.4%	.89%	.63%	.46%	
Household numbers	82,804,200	2,034,800	1,614,400	692,981	87,146,400
Percent numbers with HH in all banks	95.0%	2.3%	1.9%	0.8%	

Putting aside the differences caused by the definition of the sampling frames, we can see that the two articles used different estimation methodologies. Fahimi et al. attempted to reduce the undetermined rate to a low level using commercial data base reverse matching. This approach assumes that the reverse matching only yields a very small percentage of false matches. Also, in the calling process the classification of likely residential numbers as residential or undetermined can have an impact on the estimates.

Boyle et al. attempted to allocate the undetermined and also the presumed good numbers (some which appear to be likely household numbers) in a two-step process in Figure 4. This approach assumes that the residential rates within each stratum (Private residence row of Figure 3) apply to the undetermined sample in those strata. That assumption can hold for 1+ bank RDD sample, although in some RDD samples it overestimates the percentage of undetermined numbers that are residential. We however do not know how well that assumption holds for the other three strata used by Boyle et al. One

could consider applying the Fahimi et al. methodology to the Boyle et al. sample using archived commercial address data bases from the time frame of their study, but to make the results more comparable one would also need to ensure that the classification of known residential, likely residential and undetermined telephone numbers is the same between the two studies.

Where do we go from here? Putting aside the different findings of the two studies, it seems very likely that for some state and local RDD samples the coverage of voice-use residential telephone numbers in the traditional 1+ listed 100-bank sample design has declined over time. Neither study sheds any light on the magnitude of the bias from the exclusion of landline residential numbers but similar to unit nonresponse bias and bias from cell-only households, it is probably going to be close to zero for some survey variables and very large for other survey variables. We therefore need to consider ways to supplement the traditional 1+ listed 100-bank frame with one or more strata in an effort to increase coverage of landline residential numbers while at the same time employing dual frame designs to also cover cell-only households. Fahimi et al. suggest one approach to a new list assisted RDD frame by switching to 1+ listed 1,000 banks. Depending on the geographic area covered by an RDD survey, one might also need to consider a design that also incorporates a “remaining POTS” stratum.

At some point in time cell phone coverage may increase to a level where we no longer need to sample landline telephone numbers. Until we reach that point designing RDD samples will get more complex and this may push some surveys over to address based sampling (ABS), which relies on an address sampling frame but must deal with a host of issues related to mode of data collection and within household respondent selection.

Disclosure: Abt SRBI is a subsidiary of Abt Associates Inc.