Survey Practice This Month

Thursday, January 29, 2009, 10:25:32 AM | Editor

The September 2008 issue of Survey Practice had an article by Mansour Fahimi and colleagues that described a bias in the commonly implemented 100-series list-assisted RDD sampling. In this issue, John Boyle and colleagues show that the bias may not be as large as described in the first article. The authors comment on each other's papers and three experts (Michael Battaglia, Clyde Tucker, and Michael Link) also comment on the papers.

The editors' comments - we think that a third organization should conduct a similar project, or at least the differences in the data collection procedures need to be isolated from the estimation procedures. One of the experts makes a similar suggestion because this topic is so important for telephone surveys.

This is the first issue that Survey Practice focuses on one topic. The editors have discussed other possible single topic issues such as the use of non-probability sampling in surveys. If you would like us to consider a topic for an issue of SP, please let us know.

John Kennedy Diane O'Rourke
Andy Peytchev David Moore

survprac@indiana.edu

Comments: 0

Comments

Zero Banks: Coverage Error in List Assisted RDD Samples

Thursday, January 29, 2009, 10:24:54 AM | Editor

John Boyle, Abt SRBI Inc.
Michael Bucuvalas
Linda Piekarski, Survey Sampling Inc.
Andy Weiss, Abt SRBI Inc.

List-assisted random digit dialing (RDD) is the sampling procedure that is normally used in constructing samples of telephone households. This is a truncated design because it only includes telephone hundreds banks with one or more listed numbers. However, this design has become widely accepted after a 1995 study found that only 3.7% of working household telephone numbers fell in the unlisted banks with no significant demographic biases.

A recent study has re-examined the coverage of 100-series banks with one or more listed telephone numbers for landline households. Fahimi and his colleagues concluded that "the coverage loss for designs based on the

1+ listed banks is closer to 20% than 4%" today. Such a coverage error would call into question the acceptability of the current RDD sampling procedures for landline households, and in combination with cell phone coverage issues, the very future of telephone surveys.

The current study attempts to replicate the Fahimi study with sample from a second vendor and a somewhat different process for classifying households and non-households. Based on a national RDD sample of 10,000 numbers from 1+ listed banks and 27,175 numbers from unlisted banks, we find that 95% of landline households are still located in 1+ listed banks. These findings would seem to support the continued viability of list-assisted RDD sampling in the design and conduct of telephone surveys.

Background on List-Assisted RDD Sampling

Telephone surveys became the dominant mode of data collection for general population surveys in the United States in the 1980's. The number of U.S. households with no (landline) telephones fell to about 10% by the early 1970's. The publication of the Mitofsky-Waksberg RDD sampling procedure in 1978 (Waksberg 1978) established an accepted standard for the sampling of telephone households.

During the 1970's and 1980's, market research companies frequently conducted telephone surveys using list-assisted RDD sampling. By restricting the sampling frame to the banks with listed numbers, the efficiency of the sampling procedure is equal to or greater than the Mitofsky-Waksberg method. Moreover, the list-assisted RDD procedures provided an efficient method of drawing an element sample of telephone households. However, adoption of list-assisted RDD sampling was inhibited because of concerns about the unknown coverage of listed banks and the potential sampling bias associated with the excluded population of households.

In 1995, a seminal study of the coverage and bias of list-assisted RDD sampling was conducted. After excluding categories that were not available for general residential usage, they divided the remaining frame of telephone numbers from the Bellcore file into two strata. The first stratum consisted of all telephone numbers in 100-banks that have at least one listed residential telephone number. The second stratum was the zero-listed stratum containing telephone numbers in 100-banks that have no listed, residential telephone numbers." (Brick, et al.,1995).

The investigators drew a single-stage, epsem sample of 10,000 telephone numbers from the zero-listed stratum. These numbers were dialed by interviewers to determine whether they were residential. Out of the 10,000 telephone numbers in the zero banks, only 135 were found to be residential. This was a residential hit rate of 1.4% in the zero-listed stratum. Based on the estimated proportions of residential telephone numbers in the zero-listed banks and the estimate proportion of residential telephone numbers in the listed stratum from other studies, they estimated that 3.7% of all telephone households were not covered when the sample was restricted to the listed stratum. The authors concluded: "The results from this research indicate that the truncated, list-assisted RDD sampling method is efficient and the estimates from the design are not subject to important coverage bias." Consequently, truncated list-assisted RDD sampling procedure emerged as the standard sampling method for telephone surveys of the general population.

Unlisted Blocks Emerge as a Potentially Serious Problem for RDD Sampling

In 2008, a very different set of findings about the coverage of list-assisted RDD samples were reported by

Fahimi and his colleagues. Marketing Systems Group (MSG) drew samples of telephone numbers from "three strata that collectively constitute the entire pool of available landline telephone numbers." One stratum (1+ listed banks) includes all telephone numbers in 100-series banks that have at least one listed number. This is directly equivalent to the "listed stratum" in the earlier study. A second stratum (zero listed banks) consisted of telephone numbers in 100-series banks that have no listed number but are part of telephone exchanges (NXXs) with at least one listed number. Finally, a third stratum of the remaining telephone numbers in plain old telephone service (OPOTS) 100-series and mixed-use banks from exchanges with no listed numbers.

A sample of approximately 20,000 numbers was drawn from the 1+ banks stratum and nearly 10,000 each from the zero banks and OPOTS strata. These numbers were dialed up to 9 times to determine their household status using MSG's CSS-attended screening service. Those numbers whose status was undetermined after nine dialing attempts were processed through two additional database matches. After the nine calling attempts, they reported that approximately 7% of the sample remained undetermined. After the additional matching processes, they reported that less than 3% of the sample remained undetermined.

The estimated hit rate for residential households was 30.8% in the 1+ listed banks, 4.0% in the zero listed banks, and 2.7% in the remaining OPOTS. When the household hit rate in each of the three strata was applied to the number of telephone numbers in each stratum, the authors concluded that the percentage of residential numbers in 1+ banks has dropped from 96% in 1995 to 80% in 2008. Most of the coverage loss (14% of residential numbers) was found in the zero banks where there were no listed numbers in the 100-series but one or more listed numbers in the exchanges. The remaining OPOTS numbers accounted for another 4% of households.

Based on these estimates, the authors concluded in 2008: "These changes have greatly reduced the utility of 100-series banks for constructing RDD sampling frames. Consequently, continuing to sample from a frame that contains only 1+ listed 100-banks entails a much larger coverage loss than suggested by previous studies." "Telephone samples that ignore cell phones and use the standard 1+ listed design can exclude over 30% of the population. The potential for substantial coverage bias in this situation cannot be ignored."

Current Study: A Second Look at Unlisted Hundred Banks

We undertook the current study to replicate the findings of the Fahimi study with a different sample, while also exploring whether households reached in unlisted banks might be represented in listed banks, as well as describing the characteristics of households reached in unlisted banks compared to listed banks. We attempted to replicate the sample design used by Fahimi, Kulp and Brick using a second sample vendor. They used Marketing Systems Group (MSG) as the sample vendor for their survey. We used Survey Sampling, Inc. (SSI) for our study. Both organizations draw their samples based on information from the Telcordia (formerly Bell Core) TPM Data Source, and the same list compiler for determining listed numbers, so we would anticipate general agreement on the definition of the strata and the size of the population within stratum. However, some counts might vary depending on when the last updates of the sampling frame were done. It is also possible that differences in timing and proprietary validations rules could cause some variation in the codes used to define eligible NXXs and thousand banks.

Telephone numbers in the United States consist of ten numbers. The first three numbers are the area code or NPA. The next three numbers are designed as the NXX, which are often called the exchange, prefix or central office number. The N ranges from 2 to 9, while the X ranges from 0 to 9. The last four numbers are designated

as the thousand bank (Xxxx). The last three numbers are designated as the hundreds bank (XXxx). The current practice of list- assisted random digit dialing in the United States is to: (a) construct a frame of all NPA-NXX numbers which are available for residential household numbers; (b) restrict the frame to hundreds series banks with one or more listed numbers; and (c) randomly select a sample of these hundreds banks with listed numbers and append a two-digit random number to complete the ten digit telephone number.

For this study, we initially constructed a sampling frame that included all valid NXX's or thousand banks available for residential numbers in the Telcordia data base. At the time the sample was drawn, there were a total of 766,540 thousand banks or 7,665,400 hundred banks in the sampling frame. The frame was stratified into 1+ listed hundreds banks, zero listed hundreds banks, and the remainder OPOTS. The 1+ listed hundreds bank frame was ordered by State and County, area code, exchange and hundred bank and a systematic epsem sample of 10,000 telephone numbers was generated. Known business numbers were pre-identified but were not removed from the frame or sample. The zero listed and OPOTS frames were ordered by area code, exchange and hundred bank and systematic epsem samples of 10,000 telephone numbers were generated from each frame.

A comparison of the universe counts from the MSG sample and the SSI sample revealed similar counts for the listed hundred banks (2.9 versus 2.8 million). However, the MSG counts from the RDD zero hundred banks and OPOTS (6.1 million) were substantially higher than the SSI counts (4.8 million). The difference, however, could be accounted for by the inclusion of non-Telcordia banks in the MSG sample. Non-Telcordia banks are banks in "pooled" prefixes for which there is no 1000-block record on the Telcordia file. Thousand Block Pooling designates a pool of prefixes that are assigned a thousand telephone lines at a time by the Pool Administrator to potentially different companies. Any 1000-block in a pooled exchange that did not appear on the Telcordia file was presumed to be unassigned or not currently in use and was therefore excluded from the SSI frames. However, rather than under-represent any unlisted banks, we added a fourth stratum from the non-Telcordia banks to our study. An additional 10,000 numbers were selected from this stratum in the same manner as from the other two zero listed strata. With the inclusion of the non-Telcordia banks in the SSI sample, the total number of listed and unlisted banks were roughly equivalent for the MSG (9.03 million) and SSI (9.17 million) sampling frames (Figure 1). The 1.5% difference in the total SSI sampling frame (larger) and the 3.7% difference in 1+ banks (smaller) is small and probably the result of the timing of the updates from the two sampling frames. However, if there were a bias, then it we would expect to find more households in the 1+ banks in the MSG sample, where 1+ banks represent a slightly larger proportion of the total sampling frame.

A sample of 10,000 numbers was drawn by SSI for each of the four non-overlapping sample strata: 1+ listed, zero listed, OPOTS, and non-Telcordia. As a result of a selection error, 2825 of the numbers drawn in the non-Telcordia sample were found to be invalid and dropped from the sample for that stratum. Hence, a total sample of 37,175 telephone numbers across the four strata was drawn and fielded.

These numbers were dialed by interviewers at Abt SRBI Inc. using a predictive dialer, which should be equivalent to the Genesys CSS process. In order to classify the status of as many of these numbers as possible, we increased the contact attempts to reach a household and interview a respondent compared to the earlier study. A total of 11 contact attempts were made to reach an individual with whom to conduct screening for household status.

This is where we expanded on Fahimi's procedures by adding a brief interview regarding the nature of the telephone numbers reached. The interview explicitly confirmed whether the number reached was residential,

business, or some other category. The informant interview then went on to collect some additional information about the nature of the phone number and the household. This brief interview allowed us to expand on the information collected by Fahimi, and to delineate a major difference in the estimated distribution of household phone numbers between "listed" and "unlisted" telephone bank strata.

When contact was made at a sampled number, the maximum number of attempts was increased from 11 to 20 in order to complete the informant screening interview. The samples were drawn from frames that represented the May 2008 Telcordia file and June 2008 list-assisted frame. The samples were later matched to the most recent Telcordia file and list-assisted frame in order to append current Telcordia and list frame information for analysis. The survey was conducted between July and October 2008.

Findings on Household Coverage in Listed and Unlisted Banks

The vast majority of all numbers dialed were classified as non-residential or "bad numbers" prior to the interviewer administered screening question. This ranges from 58% bad numbers in the 1+ listed stratum to nearly 90% bad numbers in zero banks and the remaining OPOTS strata. Not surprisingly, virtually all of the numbers (99%) dialed in the non-Telcordia banks were bad numbers (Figure 2).

There were a total of 522 numbers (1.4%) that were "no answer" on each of 11 attempts over the course of several weeks of interviewing. The number of permanent no-answers was somewhat higher in the listed banks (285) than the zero banks (108), the OPOTS (128), and the non-Telcordia strata (1). These "permanent no answers" are likely to be a mix of unassigned numbers, unattended numbers (e.g., public phones, seasonal or unoccupied locations) and systematically unanswered numbers (e.g., screening by Caller ID).

We allocated the permanent no-answers by stratum, proportionally to the ratio of "presumed good numbers" and known "bad numbers". Excluding the permanent no-answers from the total sample, the proportion of presumed good numbers was 40% in the listed stratum, compared to 12% in the zero banks and OPOTS strata and less than 1% in the non-Telcordia stratum. We used this proportion to allocate the permanent no answers between the bad numbers and the presumed good numbers by stratum. The estimated good numbers from the permanent no answers were added to the other presumed good numbers to yield an estimated number of potential residential numbers per stratum.

The presumed good numbers were dialed up to 20 times in order to classify them as household or non-household. The basis of the classification was an interviewer administered question to anyone answering the phone: "Have I reached a private residence?" The responses were: Yes, private residence or household; No, business; No, dormitory or group home; No, other; or Refused. A case was treated as interviewer resolved when an interviewer obtained a response to this question from a live informant at the number.

Among the interviewer resolved cases in the listed banks, the vast majority (73%) were private residences, while only 25% were businesses. By contrast, the vast majority of interviewer resolved numbers were businesses in both the zero banks (89%) and the OPOTS banks (86%). Only 5.3% of interviewer resolved numbers in the zero banks and 7.4% of resolved numbers in OPOTS were households. By contrast, while there were very few potential household numbers in the non-Telcordia stratum, 50% of the resolved cases were households. Most of the "other, non-household" responses in all four strata could be classified as business (e.g., police station, hospital, military base, conference line, etc.) or group home (e.g., college, school), while none of them were households (Figure 3).

That the difference between the listed and unlisted strata is large is not surprising here, especially for the zero banks. These are frequently banks for which active phone numbers exist but they are not listed as residential (and probably not listed at all). They are primarily business numbers, which are in fact sold in blocks of 100 (or 1000) by telephone companies as "direct inward dial" (DID) to give individual direct numbers to workers served by a company telephone system. In large part, such numbers are not answered with a company (or even departmental) name, but as an individual. They will only be identifiable as business numbers if directly asked, as we found by asking the type of phone in the interviewer resolution process.

There were a total of 6,346 presumed good numbers out of the total sample of 37,175. In addition, another 143 out of the permanent no answers were allocated as estimated good numbers. Hence, there was a total of 6,489 potential household numbers out of the initial sample of 37,175 numbers that needed to be resolved by interviewer screening.

Among these potential residential numbers, we were able to positively resolve the household status on the basis of the interviewer administered question in more than 2 out of 5 cases. The resolution rate among potential residential numbers was approximately the same for the listed banks (42%) and the zero banks (43%), but somewhat lower in the OPOTS (36%) and non-Telcordia (30%) numbers. The unresolved numbers after interviewer screening represented 23.4% of the numbers in the 1+ RDD sample, but only 6.7% of the numbers in the zero banks, 7.7% of the numbers in the OPOTS, and .6% of the numbers in the non-Telcordia sample. In total, 5.5% of the numbers were unresolved in the unlisted banks after interviewer screening.

The standard approach to estimating the number of eligible households in the sample for purposes of response rate calculation is to apply the eligibility rate in the resolved cases to the unresolved cases. In the case of the listed banks, 72.7% of the 1700 resolved numbers were determined to be households. When this eligibility rate is applied to the 2,344 unresolved cases in this stratum, we would estimate that 1,704 would be eligible households. Combining the 1,236 known households and the 1,704 estimated eligible households in the unresolved sample yields an estimated household rate of 29.4% in the 1+ listed household stratum (Figure 4). This household rate is consistent with the 28% to 29% residential hit rates reported in the 2007 National Household Education Survey and the 2006 National Immunization Survey, which used national 1+ list-assisted telephone surveys (Fahimi, et al. 2008).

Using the identical procedure, we apply the 5.3% household rate in the 508 resolved cases in the zero banks sample to the 672 unresolved cases. This yields an estimated 36 households in addition to 27 resolved households. The total number of actual and estimated households in the zero banks is 63 out of 10,000 numbers. Applying the same procedure to the OPOTS, we find a total of 89 actual and estimated households out of the 10,000 numbers dialed. Finally, the total number of actual and estimated households in the non-Telcordia sample is 33 out of 7,175. Since we used the identical procedures for estimating the number of households in the zero banks, OPOTS and non-Telcordia strata that produced the expected rate in the listed banks, we believe these estimates are credible.

When the estimated household rate of 29.4% is applied to the population of 281,647,100 telephone numbers in hundreds banks in the listed stratum, it yields an estimated 82,804,247 eligible household numbers.

When the estimated household rate of .63% in the zero banks is applied to the 256,260,000 telephone numbers in hundreds banks in the stratum, it yields an estimated 1,614,438 eligible household numbers in those banks.

When the estimated household rate of .89% in the OPOTS banks is applied to the 228,633,000 telephone numbers in hundreds banks in that stratum, it yields an estimated 2,034,834 eligible household numbers in those banks. Finally, the same procedure yields an estimated 692,981 eligible banks in the non-Telcordia banks.

Thus, our findings suggest that 82.8 million household numbers out of a total of a potential 87.1 million household telephone numbers are located in 1+ listed hundreds banks. Consequently, these findings suggest that approximately 95.0% of working residential telephone numbers in the United States are found in 100 series banks with one or more listed number. By contrast, 5.0% of working residential landline telephone numbers are located in zero banks, OPOTS and non-Telcordia banks, and hence would be excluded from any sampling frame based on listed hundreds banks (Figure 5).

Database Matching for Resolved and Unresolved Numbers

The Fahimi study conducted nine dialings of sampled numbers as the first step in determining the household status. It is not clear from the paper whether a formal screening assessment similar to our screening question was administered or whether interviewers only classified numbers as businesses/non-residential on the basis of telephone responses, (e.g., "This is Acme Construction. How can I help you?"). It is also not clear how answering machines and voice mail were handled for classification purposes during the dialing. However, the cases that were unresolved after the initial dialings were subsequently compared to two commercial data bases to improve the resolution rate.

Although we believe that an extended interviewer administered screening is the more reliable approach to determining household status, we also submitted our resolved and unresolved telephone numbers to a commercial database match process for comparison purposes. The matching process that we used was Allant's Prevalence Reverse Append (PRA). They report over 640 million feed records per month from more than twenty sources, including telecommunications carriers, caller ID providers, directory compilers, major consumer marketing companies, data compilers and directory assistance providers. They claim 60 million records not found in white pages or Directory Assistance sourced databases. The unresolved and resolved telephone numbers from all four strata were submitted to this database search for consumer name.

Among the 2,661 interviewer resolved households, approximately 38.0% were found to have complete or partial secondary matches on consumer names in the PRA search. There were 988 consumer matches in the resolved listed bank numbers (58.1%). This compares to 1,236 households identified by interviewers in those numbers. There were 11 consumer matches in the resolved zero bank numbers (2.2%), compared to 27 households identified by interviewers in those numbers. There were 8 consumer matches in the resolved OPOTS numbers (1.8%), compared to 32 households identified by interviewers in those numbers. Finally, there were 4 consumer matches in the resolved non-Telcordia numbers (20.0%), compared to 10 households identified by interviewers in those numbers (Figure 6). In short, the database match yielded fewer households compared to interviewer resolved cases in all strata, but the ratio of interviewer to database identified households was much higher in the unlisted banks.

Among the 3,685 unresolved households (not counting the permanent no answers estimated to be good numbers in Figure 4), approximately 36.7% were found to have complete or partial secondary matches on consumer names in the PRA search. The proportion of telephone numbers with consumer matches is actually slightly higher in unresolved cases (59.8%) compared to resolved cases (58.1%) in the listed banks. By contrast, the proportion of matched consumer names is higher in resolved than unresolved bases for numbers from

zero banks (2.2%-1.8%), OPOTS (1.8%-.5%), and non-Telcordia (20.0%-6.5%) strata. So, while the proportion of households among resolved numbers appears to be a relatively good predictor of the proportion of households among the unresolved households, based on consumer name matching comparisons between the two samples, it may somewhat overestimate the number of households in unlisted banks compared to listed banks.

Since our interviewer administered screening achieved a higher household rate for resolved numbers and a higher estimated rate for unresolved numbers than the database matching, we believe it is a more reliable indicator. Indeed, since many of the sources for the database matches come from published directories, there is a bias against households in the unlisted banks. If we had used this approach rather than interviewer administered screening or to estimate the rate in unresolved numbers, then the difference between our estimates of the number of households in unlisted banks compared to Fahimi would be even greater.

Discussion

This project was undertaken to confirm and expand the findings of Fahimi and his colleagues that approximately twenty percent of residential landline household numbers were not covered by the current practice of using hundred bank series with one or more listed numbers. We hoped to test whether transferred numbers (or phantom numbers) or other household numbers in listed banks might mitigate the apparent problem. We also wished to examine the differences in households in listed and unlisted banks to determine the amount of bias associated with the exclusion of unlisted banks.

However, our findings suggest a much smaller coverage error (five percent) from the exclusion of unlisted hundreds banks from RDD landline sampling frames than reported by Fahimi and his colleagues. Our five percent non-coverage is only slightly higher than the proportion of households (3.7%) found in unlisted banks in the 1995 study. Although we used different sources for the sampling frames, our population counts for the total number of banks and the banks with 1 or more listed numbers are almost identical. Our findings on the household rate in the listed banks are similar and consistent with the literature. We used the same procedure for estimating the number of households in the unlisted banks that we used to estimate the household rate in the listed banks. Consequently, we believe that it is more likely that the difference in the estimates of the number of households in unlisted banks between the two studies is a result of the procedures for estimating residential households, than the sampling frame or sample.

These findings would seem to support the continued viability of list-assisted RDD sampling in the design and conduct of telephone surveys, at least in terms of coverage error. However, given the difference in the findings of these two large-scale studies, and the absence of any "smoking gun" that would explain the differences, additional research on this issue is needed. In the meantime, however, we believe that it is not necessary to abandon listed hundreds banks for listed thousand banks, or list-assisted RDD sampling altogether, until this difference is resolved to the satisfaction of the telephone research community.

References

Blumberg, Stephen and Julian Luke (2007). "Coverage Bias in Traditional Telephone Surveys of Low-Income and Young Adults." Public Opinion Quarterly, 71: 734-749.

Brick, J. Michael, Joseph Waksberg, Dale Kulp, and Amy Starer (1995). "Bias in List-Assisted Telephone

Samples." Public Opinion Quarterly, 59: 218-235.

Brick, J. Michael and Clyde Tucker (2007). "Mitofsky-Waksberg: Learning from the Past." Public Opinion Quarterly, 71: 703-716.

Cassady, Robert J. and James M. Lepkowski (1993). "Stratified Telephone Survey Designs", Survey Methodology 19:103-113.

Fahimi, Mansour, Dale Kulp and J. Michael Brick (2008). "Bias in RDD Sampling: A 21st Century Digital World Reassessment." Paper presented at the American Association for Public Opinion Research Annual Meeting, 2008.

Fahimi, Mansour, Dale Kulp and J. Michael Brick (2008). "Bias in List-Assisted 100-Series RDD Sampling." Survey Practice, September 28, 2008.

Keeter, Scott, Courtney Kennedy, April Clark, Trevor Thompson, and Mike Mokrzycki (2007). "What's Missing from National Landline RDD Surveys? The Impact of the Growing Cell-Only Population." Public Opinion Quarterly, 71: 772-792.
Lepkowski, J, "Telephone Sampling Methods in the United States", Chapter 5 in Telephone Survey Methodology, Eds. Robert Groves, et al., John Wiley and Sons, 1988.

Piekarski, Linda B., "Choosing Between Directory Listed and Random Digit Sampling in Light of New Demographic Findings." Paper presented at the American Association for Public Opinion Research Meeting, 1989.

Tucker, Clyde, James M. Lepkowski, and Linda Piekarski (2002). "The Current Efficiency of List-Assisted Telephone Sampling Designs." Public Opinion Quarterly, 66:321-338.

Thornberry, Owen T. and Massey, James T. (1988). "Trends in United States Telephone Coverage Across Time and Subgroups." *Telephone Survey Methodology* (pp. 25-49). New York: Wiley.

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

Comments: 0

Comments

Comments on Boyle et al Article

Thursday, January 29, 2009, 10:24:06 AM | Editor

Mansour Fahimi, Dale Kulp - Marketing Systems Group
J. Michael Brick - Westat

Given the growing number of challenges undermining the RDD sampling methodology, it is important to

conduct independent studies that can identify potential problems so that effective remedies can be devised by the research community. In that sense, we welcome this study that has aimed to quantify the coverage problems associated with RDD-based samples. However, results produced by this work are subject to one critical problem and an assortment of secondary issues. In that order, we start with the main problem and its implications.

Critical Problem

The main finding of this paper - that the 1+listed frame covers about 95% of landline households - is based on a contingent result that there are only a total of 87,146,400 residential telephone numbers in the nation. This estimate is far too low, as explained below.

•I. Beyond seasonal homes and those occupied on an occasional basis, there are about 112.4 million full-time occupied housing units in the US. Of these units, at the time of this study, about 2.5% had no telephone and some 17.5 % were estimated to be cell-only. This leaves about 89.9 million households that have landlines. It is worth noting that an undetermined fraction of cell-only households continue to rely on residential landlines for non-voice applications such as fax/modem or security functions.

•II. According to Brock-Roth et al (2001), about 7% of households have more than one landline. Also, the authors of this study report 15% of their study respondents had more than one residential telephone line. This translates into about 8%, since their sample is of telephone numbers and the resulting estimate should be divided by a factor of about 2 to account for the multiple chances of selection. Consequently, the estimated total number of residential telephone numbers should be at least 97.1 = 89.9 × 1.08 million, not the reported estimate of 87.1 million.

•III. Using the authors' estimate that the 1+listed frame covers 82.8 million residential telephone numbers results in an estimate of coverage rate for the 1+listed frame of about 85% = 82.8/97.1. While this rate is slightly higher than the 80% reported by Fahimi et al, it falls well short of their reported 95%.

•IV. If, on the other hand, one includes residential lines in seasonal homes and those disguised in cell-only households noted above, then the total number of residential lines could quite easily exceed 105 million. As such, the actual coverage rate for the 1+listed frame becomes even closer to 80%, if not lower.

In light of the above, the estimate of coverage rate for the 1+listed frame should be much closer to that reported by Fahimi et al than what is suggested in this paper. It is worth noting that results from two follow-up studies conducted by MSG support the initial findings that the coverage rate for the 1+listed frame has declined drastically. Moreover, these studies support findings presented at the 2008 AAPOR Annual Conference that the 0-listed frame includes millions of residential number assignments. While we believe the coverage rate of the 1+listed frame is the critical issue undermining the utility of the conventional RDD sampling methodology, the issues of coverage and hit rates in the other (0-listed) frames are interesting and worthy of further investigations as well. Such results, however, do not detract from the coverage problem of the conventional 1+listed frame.

Secondary Issues

In addition to the above fundamental problem, this study is subject to a number of secondary issues that may

pose further concerns about the reliability of the reported results. In particular, such issues include concerns about the frame construction methodology, ambiguity regarding the unit of measurement, questions about the employed survey instrument, and the process used for resolution of undetermined cases as briefly discussed below.

•I. As pointed out correctly by the authors, both MSG and SSI utilize the same primary source for identifying listed households. However, it is incorrectly stated that MSG relies on the TPM file from Telcordia for frame construction. Unlike SSI, MSG has been relying on a much more comprehensive (expensive) Telcordia product known as the LERG (Local Exchange Reference Guide) because of observed updating and currency issues associated with the TPM file. This is why, as referenced in this paper, the MSG frame has included many more 100-series banks with approximately 150 million potential telephone numbers than what the authors have considered valid. Interestingly enough, the authors report that almost one million households are covered in such banks. MSG includes these banks because, by definition, all numbers in the associated exchanges are either active or potentially available for residential assignment. This is the conservative approach to frame definition since Telcordia data are always somewhat out-of-date and fail to include many 1000-series blocks that are currently in use.

•II. While there is an important distinction between the two parameters - number of telephone households and count of residential telephone numbers - the authors do not clearly delineate which of the two corresponding units of measure they intended to utilize in this study. As such, it is unclear if any information has been collected from the respondents about the number of phone lines in each household. Without such inquiries it is impossible to bridge the gap between these two key parameters. Regardless of whether the authors have opted for telephone households or residential telephone numbers as their unit of measurement, their estimates fall well short of the actual totals. Without further information it would not be possible to determine whether the source of the resulting underestimations is due to sampling design, data collection process, estimation producers, or a combination thereof.

•III. Several objectives are enumerated for the survey instrument including resolution of "phantom numbers" due to number portability by verifying whether respondents had been reached on the number dialed; eliciting information regarding the listed/unlisted status of numbers; and whether households with landline numbers in the 0-listed frame are reachable through the 1+listed frame. However, very little is reported about the actual scope of the questionnaire, its exact contents, or study results related to these objectives. Ironically, the paper incorrectly states that MSG's interviewers did not confirm the dialed numbers nor determine whether they are residential, business, or of some other category. Actually, these are standard questions as part of our GENESYS-*CSS* attended screening process.

•IV. It is correct that for calculation of response rates survey researchers often allocate the remaining undetermined cases based on the observed distribution of the resolved cases. However, the problem at hand is one of estimating coverage rates and not approximating response rates. This is why MSG has taken additional steps to resolve the remaining undetermined cases by searching in various commercial databases to trace as many such telephone numbers as possible. Accordingly, we have managed to find a final disposition for nearly one half of our undetermined cases to reduce the net rate to about 3% as compared to the 7% the authors have proportionally allocated based on a distribution that most likely includes a different mixture of working and nonworking numbers. To put this in perspective, a 4% difference in resolution rate represents about 36 million telephone numbers across the 0- and 1+listed frames. Imputing the final status of these numbers through simple extrapolations alone can easily explain the observed differences in coverage rates for the

corresponding 100-series banks.

•V. Lastly, and as a point of curiosity, the authors initially list one of their research objectives to be a description of the characteristics of households with landline telephone numbers in the 0-listed banks. This is an important investigation that we are pursuing as the next phase of our study - research that is expected to require significant resources because of the low residential hit rates in such banks. However, later in the text the authors report their study has only focused on estimating the extent of coverage error and not on examination of the characteristics of households in the 0-listed banks. It is not clear why an objective that has not been carried out is even mentioned in the first place.

References

Brock-Roth, S., J. Montaquila, and J. M. Brick (2001). "Effects of Telephone Technologies and Call Screening Devices on Sampling, Weighting, and Cooperation in a Random Digit Dialing (RDD) Survey." *Proceedings of the American Statistical Association, Survey Research Methods Section*, Atlanta, Georgia.

Reply to Fahimi et al Comments

Thursday, January 29, 2009, 10:23:05 AM | Editor

John Boyle, Abt SRBI Inc.
Michael Bucuvalas
Linda Piekarski, Survey Sampling Inc.
Andy Weiss, Abt SRBI Inc.

We are extremely pleased Survey Practice provides a forum for authors of papers with conflicting findings to have the opportunity to exchange observations and comments that may facilitate future research, if not an outright resolution of differences. Fahimi and his colleagues say that the results of our paper "are subject to one critical problem and an assortment of secondary issues." We will attempt to address these critiques below. It is worth noting that Survey Practice submissions are deliberately restricted in length, so some of the items Fahimi found curious are addressed in a longer article that has been submitted for journal publication.

Critical Problem

Fahimi states that: "The main finding of this paper — that the 1+ listed frame covers about 95% of landline households — is based on a contingent result that there are only 87,146,400 residential telephone numbers in the nation. This result is far too low …."

We regret that any readers have misinterpreted our attempt to estimate the proportion of residential voice lines included in the 1+ listed frame as an attempt to estimate the total number of landline households, particularly for validation purposes. Our goal was transparency; we recognize our findings conflict with Fahimi's data and felt a detailed disclosure would help us all to better understand the difference between the two

approaches. In order to estimate the proportion of total residential voice landlines in the 1+ banks, we applied our survey estimate of the proportion of landlines in each sample stratum to the universe of banks (and by extension numbers) in each stratum. Then, by adding the total estimated landlines across strata for the denominator of total landlines, we could calculate the proportion (95%) from the 1+ listed banks.

We believe the approach we have outlined in our paper provides for an unbiased estimate of the **proportion** of residential voice landlines in listed and unlisted banks. The basic approach is to conduct rigorous interviewer screening of numbers to identify the proportion of **resolved numbers** that can be classified as residential voice lines. Then, the proportion of residential voice lines in the resolved numbers by stratum is applied to the remaining **unresolved numbers** in the stratum to generate an **estimated number of residential voice landlines** by stratum. However, let us consider Fahimi's contention that the estimated number of voice landlines represented in the denominator of our equation (approximately 87.1 million) undermines the validity our estimates of the proportions by stratum.

We accept Fahimi's estimates that there are approximately 112.4 million households in the United States and approximately 2.5% of these households have no telephones while another 17.5% of these households have cell phones only. Hence, the number of households in the U.S. with residential voice landlines would be approximately 89.9 million (112.4 x .80).

We also agree that the number of residential landlines will be larger than the number of residential households as a result of those households with more than one landline used for voice. However, we believe the current rate of multiple landlines used for voice is much smaller than estimated by Fahimi. Although Fahimi correctly cites 7% of households as having multiple telephone numbers in the Brock-Roth paper, he fails to mention that only 59% of those households had additional residential landline numbers (see Table 1 in that paper). So, the proportion of households with more than one residential landline number appears to have been approximately four percent in that study. It is also important to note that the Brock-Roth paper cited by Fahimi was based on 1999 NHES data. It would certainly not be surprising to find that the proportion of households with multiple voice landlines had declined from 1999 as a result of the growth of cell phones (cell only households) and the decrease in dial-up internet connections. In our study, while we found that 14.7% of households reported more than one phone, most of these were reported as wireless numbers. Actual residential multiple landlines for voice was about 2.4% in our study. So, we would project that the number of residential voice landlines in the United States, today, is probably about 92 million (112.4 x .80 x 1.024) rather than the 97.1 million estimated by Fahimi.

Although it was not our intent to focus on the total estimated number of residential voice landlines, if this was to be the critical test of the validity of the findings, then we would look for validation by some external standard. For example, the FCC Report "Local Telephone Competition: Status as of December 31, 2007" provides the following numbers:

129.7 million incumbent LEC switched access lines,

28.7 million CLEC switched access lines

158.4 total switched access lines.

The FCC also estimates the Percentage of Lines Provided to Residential Customers = 59%, yielding the result

158,436,758 * .59 = 93,477,687 residential switched access lines (83% of all households). When you subtract 966,701 for US Territories and Protectorates (counts withheld for CLECs but probably insignificant since suppressed), this translates to 92,907,333 Residential access lines, including households with multiple landline numbers. This falls in the same range of our earlier estimate of 92 million residential voice landlines.

Although we believe that the survey estimate of 87.1 million residential voice landlines is not dramatically different than the FCC based estimate of 92.9 million, it is worth considering why the survey estimate might be lower. First, our estimate of households with voice landlines is based on the resolution of the sample of telephone numbers by an interviewer administered screener. However, in addition to the 2.5% of households with no telephones and 17.5% with cell phones only, the NCHS estimates that there are another 13.3% of households who are **cell phone mostly,** having received all or almost all calls on wireless telephones despite having a landline telephone in the home. Since the cell phone mostly households are less likely or unlikely to answer their landlines, we would anticipate that there will be bias against cell phone mostly households in the interviewer resolved numbers, although they are certainly in the sampling frame.

Second, in order to estimate the total number of residential landlines based on unresolved as well as resolved cases, we assume the residential hit rate in resolved cases would be the same for unresolved cases. We believe that — in the absence of evidence to the contrary — this is a reasonable approach to generating an unbiased estimate across strata. Even if the true household rate is actually lower or higher in the unresolved cases, the result is unbiased so long as it does not behave differently by strata. Although this approach is fine for generating estimates of proportions, if the assumption is incorrect it would affect the total number of estimated households in the individual strata and total. If a disproportionate amount of cell phone mostly cases cannot be interviewer resolved because they never answer their landline, then the household rate would be higher in the unresolved compared to resolved cases, yielding an underestimate of the total number of residential voice landlines. While this may account for some of the difference between our estimated 87.1 million residential voice landlines and an estimated 92.9 million from the FCC figures, it should have no impact on our estimates of proportion by stratum since there is no reason to assume a differential effect by stratum.

While we hoped the use of actual counts from the universe and the samples would facilitate understanding of the basis of our estimates, we acknowledge that the use of reciprocal weights to estimate proportions would have produced the same results without the controversy. Fahimi and his colleagues have the advantage of not providing counts, as distinct from proportions in their Survey Practice article. However, in their AAPOR presentation they applied their estimated residential rate from their survey to the population of their three strata to yield residential estimates of 16,151,285 in the zero listed, 89,902,643 in the 1+ listed, and 5,615,877 in the remaining POTS. Hence, while we believe the total number of residential landline numbers should be about 92.9 million, the Fahmi study yields a total count of 111.7 million residential landline numbers. So, the apples to apples comparison of the projected number of residential voice landlines for the two studies is 87.1 million in our study and 111.7 million in the Fahimi study, compared to an FCC based estimate of 92.9 million. In addition, Fahimi leaves 4.0% of zero-listed numbers, 2.1% of 1+ listed numbers, and 3.7% of remaining POTS as undetermined, so instead of allocating these numbers between residential and non-residential these cases are effectively treated as non-residential for the estimates. If they were allocated then the Fahimi estimate for total residential voice landline would be even larger.

Secondary Issues

- **I.** We apologize if we incorrectly inferred the process used by MSG, but we had to guess since it was not

documented in either the Survey Practice paper or AAPOR presentation. We would note that what is on the LERG file eventually gets to the TPM file, so these blocks are never permanently excluded. Also given 1000-blocks pooling, considering "all numbers in associated exchanges [as] potentially available for residential assignment" may be a dangerous assumption since many exchanges today mix POTS and non-POTS services. Nevertheless, we included this stratum in our analysis (labeled non-Telcordia), and found that it included less than 1% of eligible residence lines. Most importantly, it represents no impact on the study conclusion because it is already included in the total 5% estimated non-coverage of residential voice landlines for 1+ block samples.

•II. As noted above, while the distinction between households and residential landlines is absolutely valid as a unit of measure, since only 2.4% of households with landlines have more than one voice landline, the difference is less critical than it might otherwise be to our estimates. We apologize for any confusion we may have caused by the terminology, which we will correct in any longer publications.

•III. The length restrictions for Survey Practice publications limited the degree of detail that we could report for these other objectives. We are providing a copy of the survey questionnaire used to estimate the size, nature and characteristics of households in listed and unlisted banks. We apologize for inferring that MSG interviewers did not determine the household status of numbers dialed by asking respondents directly whether they had reached a household, business, or something else, but they did not provide any documentation of exactly what was done.

•IV. As we described in our paper, we also explored a database matching step. We found that the database matching yielded a lower incidence of residential lines for the unresolved cases than resulted from using a simple allocation based on the resolved cases. We chose the allocation because we felt it was based on the superior information gained from the interviewer resolved cases and also provided a higher (more conservative) allocation of residential lines to the unlisted segments.

•V. We describe all of objectives of the study in the introduction to this paper. However, restriction on the length of this paper meant that some study objectives, such as the characteristics of households in listed and unlisted banks, could not be addressed in this paper. We hope to publish and present this data in other forums.

Comments: 0

Comments

Comments on the Two Articles (1)

Thursday, January 29, 2009, 10:22:30 AM | Editor

Michael P. Battaglia
Abt Associates Inc.

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.

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.

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%).

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.

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 x .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.

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.

Comments: 0

Comments

Comments on the Articles (2) - A Failure to Communicate

Thursday, January 29, 2009, 10:19:29 AM | Editor

Clyde Tucker
Bureau of Labor Statistics

Both the Fahimi et al. and Boyle et al. papers provide reasonable approaches to measuring the same thing - landline coverage in the listed residential 100-banks from Telcordia. The fact that the two papers reach different conclusions is puzzling since both appear to start from about the same place. In order to sort out this controversy, we need to look at each component in the process separately.

Although the two sets of authors stratify the frame in different ways, both end up with relatively equal counts of listed 100-banks (just under 3 million). When the 100-banks from the other strata in the frame are added to those counts, both papers report totals of about 9 million or so 100-banks. Moreover, after doing a little checking, I have concluded that both groups of researchers have counts of total households (about 112 million) and non-landline households (those cell-only or with no phone service) that are roughly the same (close to 20% of households or 22 million). Since they both start equal, why don't they end up that way? I believe the problem, as concluded by Boyle et al., is a non-statistical one, it is all in the administrative process that goes into identifying a residential number.

Let's begin with the listed residential banks. Fahimi et al. report a residential hit rate of 30.8%, and Boyle et al. have the rate at 29.4%. Based on the strata size in each paper, Fahimi et al. estimate the number of residential numbers in this part of the frame to be about 89 million and Boyle et al. get 82 million. This seems to result

from the different hit rates each estimates and the slightly different count in listed 100-banks. Splitting the difference on the hit rates gives about .30 which produces a residential number estimate of 85.5 million in that part of the frame. Neither paper deals explicitly with the number of households with multiple voice landlines, where the range seems to be between 3% and 7%. However, whichever line is reached, it counts as a household hit in terms of raw (not weighted) counts.

Fahimi et al. and Boyle et al. approach the estimation of residential numbers in the other part of the frame differently. The former uses reverse matching to commercial databases for undetermined numbers, and the latter estimates the number by extrapolating from their empirical results to the undetermined in that part of the frame. The results are quite different in the two papers. Fahimi et al. conclude that almost 20% of residential lines are not in listed banks, while Boyle et al. estimates only 5% are outside listed 100-banks. This is quite a difference. How do we explain it?

I think we have to look beyond the pure statistics to the actual survey procedures. Given that there are now cell-mostly households that rarely use their landline and call screening devices which reinforce concerns about privacy, more burden is put on survey researchers to identify residential phones. Boyle et al. discuss their procedures for making call attempts and verifying residences once the phone is answered, but Fahimi et al. are relatively silent on that score. In any case, identifying residential numbers is difficult no matter what procedure is followed; and, as response rates continue to decline, the exact nature of these procedures will take on more importance. Yet, there are other problems to overcome. With the expansion of the number of telephone service companies, getting information on individual numbers is almost impossible. This is clearly a greater problem in the non-listed banks, but it is still a problem in listed banks where the residential hit rate has declined with the expansion of area codes and the less systematic assignment of numbers (estimates of residential hit rates obviously vary even here). Although making many call attempts and verifying if the number is at a residence after someone answers is a step in the right direction, what about all those numbers that are ring-no-answer? The jury is still out on the adequacy of matching to commercial or other databases as Fahimi et al. do to resolve the problem. Of course, all of these issues do not even consider the rules used by a survey organization to assign final dispositions.

The only remedy is for the different survey organizations to start directly talking to one another. Joint studies should be conducted where all procedures used are transparent and state-of-the-art. The results of that research should be presented jointly in journals and at conferences. This does not ensure we have the right estimates, but it is a big step in the right direction.

Comments: 0

Comments

Comments on the Articles (3) - Three Key Takeaways from the Zero Bank Debate

Thursday, January 29, 2009, 10:18:43 AM | Editor

Michael Link
The Nielsen Company

Considering the important and timely work by Fahimi et al. and Boyle et al. on the issue of residential numbers

in zero listed telephone banks, there are, I believe, three key take-a-ways from this exchange.

First, science proceeds through a series of tests and retests and a single study does not a definitive conclusion make. While not a novel observation, the iterative approach to hypothesis testing is one of the oldest precepts of science, it does from time-to-time help for us to be reminded of this. Advancement of our understanding in critical and changing areas necessitates multiple points of inquiry. Where we stand now with this particular issue is at a vexing crossroads, the direction of which has major implications for a large segment of the survey research community. This is one time when "more research is needed" is not a simple tagline at the end of an article, but a call for real and necessary action.

Second, it reminds us that some of the foundations upon which we build our industry are not ours, but borrowed from others. In this instance I'm referring to the use of the U.S. telephone system database as the critical frame upon which many of our studies are built. We sometimes forget that this seemingly simple system of number assignment is not structured with surveys in mind and numbers are not allocated in ways to necessarily make it easier for us as survey researchers. Rather structural changes in the telecommunication system are occurring at an increasing rate to meet the demand for numbers and as a result of, as Fahimi et al. note, the breakup of the larger telecommunications system into a series of smaller, yet interconnected networks. This lesson is one we should heed not just for telephone sampling, but for many of the newer approaches being championed as well, including online strategies and address-based approaches. It is crucial that we maintain a critical eye on the origins and characteristics of our sampling frames to ensure we understand the limits of these sources.

Third, the studies highlight the continued trouble in which our industry finds itself with respect to landline-based random digit dialed (RDD) surveys (as an aside, it is important to separate out the concept of "telephone surveys" from "RDD landline surveys" as the former will continue to be with us for quite some time as a critical component of multimode and alternative survey designs, while the latter's days may be numbered). This is a discussion which cannot be had without including the implications of cell phone only (as well as cell phone mostly) households - to do so misses much of the real problem. When we consider the cell phone issue in the context of the zero banks debate we are left with one of two unpleasant alternatives: if Boyle et al. are correct, then the situation for landline RDD is bad; if Fahimi et al. are correct the situation is horrible, indeed untenable. That's not to say there are no RDD solutions to be had. Fahimi et al. propose some potential approaches for the landline portion of the sample, which combined with an RDD-based cell phone sampling could provide a solution (a costly and inefficient one, but a solution nonetheless). The notion, however, that landline-based, list-assisted RDD is a viable solution in terms of coverage error without placing that in the context of all residential households, which includes cell phone only homes, significantly misses the mark and may lead astray those who don't consider, as Paul Harvey might put it, "the rest of the story."

FIGURE 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

FIGURE 2: Sample Disposition of Interviewer Screening of Household Status					
	Percent within Stratum				
Sample Disposition	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: Outcome of Interviewer Screening for Household Status					
	Percent within Stratum				
Household status	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

FIGURE 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%

FIGURE 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%	

FIGURE 6: Commercial Database Name and Address Matches

Database matches among	RDD 1+	OPOTS	RDD Zero	Non-Telcordia	Total
Permanent no answers	(285)	(108)	(128)	(1)	(522)
Percent Matching	21.8%	0.9%	1.6%	0.0%	12.5%
Interviewer resolved numbers	(1700)	(433)	(508)	(20)	(2661)
Percent Matching	58.1%	1.8%	2.2%	20.0%	38.0%
Unresolved good numbers (excluding estimated good numbers)	(2229)	(753)	(657)	(46)	(3685)
Percent Matching	59.8%	0.5%	1.8%	6.5%	36.7%