

Comments on Boyle et al. Article

Mansour Fahimi^{*}, Dale Kulp[†], J. Michael Brick[‡]

Tags: survey practice

DOI: [10.29115/SP-2009-0006](https://doi.org/10.29115/SP-2009-0006)

Survey Practice

Vol. 2, Issue 1, 2009

Comments on Boyle et al. Article

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

^{*} **Institution:** Marketing Systems Group

[†] **Institution:** Marketing Systems Group

[‡] **Institution:** Westat

numbers should be at least $97.1 = 89.9 \times 1.08$ 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.

1. 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.

2. 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.
3. 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.
4. 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.

5. 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." In *Proceedings of the American Statistical Association, Survey Research Methods Section*. Atlanta, Georgia.