# Bias in List-Assisted 100-Series RDD Sampling

Mansour Fahimi[*], J Michael Brick[†]

Tags: survey practice

List-assisted Random Digit Dial (RDD) sampling methodology was developed decades ago when local telephone exchanges relied on 100-series telephone banks as physical building blocks. In recent years, however, the telecommunication industry has undergone a number of fundamental changes including a complete transition from analog to digital call routing and departure from an AT&T-dominated infrastructure to what is provided by regional independent operating companies as well as a growing number of alternative landline service providers. Combined with the decline in the proportion of directory-listed households and dilution of the residential landline assignment density due to a sharp increase in the number of residential exchanges, these changes have all but eliminated the utility of 100-series banks for frame construction and sampling purposes.

In spite of the above drastic changes, the sampling frame construction methodology for RDD samples has changed very little (if any) over the years. This note provides an overview of a research conducted to reexamine the underlying assumptions that were conducive to list-assisted RDD sampling against the ground realities of today. Specifically, the extent of undercoverage bias in traditional RDD samples is quantified while alternative methods of frame construction are introduced that aim to restore some of the lost coverage. Proposed alternatives are evaluated in light of cost implications that can result from adoption of more inclusive sampling frames, since such expansions will inevitably require additional resources for sample designs and survey administrations.

## INTRODUCTION

A major breakthrough in telephone survey research methodology was introduced when the Mitofsky-Waksberg (1970) technique of RDD sampling was simplified to include only 100-series banks with at least one listed telephone number. As such, a two-stage cluster sampling methodology that entailed both operational and technical complexities was replaced by a

* **Institution:** Dale Kulp – Marketing Systems Group
† **Institution:** Westat

single-stage *epsem* sampling method that could produce survey estimates with smaller sampling variances. Of note, these impressive gains were exercised at the expense of accepting a modest coverage bias that could be easily tolerated when time and cost saving considerations were kept in balance. Brick et al. (1995) had estimated that only 3.7 percent of all telephone households were not covered when the frame was confined to listed 100-series banks.

Coupled with solid theoretical underpinnings with respect to design and estimation issues, as examined by a number of researchers including Casady and Lepkowski (1993), list-assisted RDD samples have served at the nexus of telephone surveys during the past three decades. In recent years, a number of studies have reassessed the efficiency of such sample designs in light of the changes in US telephony. For instance, Tucker et al. (2002) have concluded that now more than before list-assisted methods are important since it is becoming exceedingly more difficult to identify residential numbers. Nonetheless, the question that is the genesis of this research has remained unanswered: "*How large is the coverage bias in current list-assisted RDD samples?*" Other than anecdotal projections based on surveys designed for completely different analytical objectives, there has been no specific research in recent years to reassess the magnitude of this bias.

### RESEARCH METHODOLOGY AND RESULTS

In order to provide a current estimate of the coverage bias in list-assisted RDD samples, Marketing Systems Group (MSG) selected a sample of 40,000 telephone numbers from three strata that collectively constitute the entire pool of available landline telephone numbers. The first stratum (Zero-Listed Banks) consisted of telephone numbers in 100-series banks that had no listed numbers but were part of telephone exchanges (NXXs) with at least one listed number; the second stratum (1+Listed Banks) consisted of telephone numbers in 100-series banks that had at least one listed number (complement of the first); and the third stratum (Remaining POTS) included telephone numbers in POTS 100-series and mixed-use banks with no listed numbers.

All sample telephone numbers were called a maximum of 9 times using MSG's GENESYS-*CSS* attended screening service to obtain an initial disposition for each number. Subsequently, the pool of 2,722 numbers remaining *CSS*-undetermined (no answer or busy) were cross-referenced against available databases to determine a final disposition for each sample telephone number. Finally, the entire sample was weighted to reflect the employed stratified design before proper estimates could be developed for percent residential and other categories in each of the three strata. The following table provides a summary of the final disposition status for each stratum.

While the weighted results from the above table can be used to estimate the percentage of residential numbers (coverage rate) in each stratum, the following table provides results that can be used to estimate percent residential

hit rates by stratum.

Prime among the results summarized in the above tables is that the extent of coverage bias in list-assisted RDD samples that exclude zero-listed banks is no longer as little as 3.7 percent. As shown in Table 1, indeed, this rate has now peaked to about 20 percent, representing a non-ignorable and most likely a nonrandom subset of US households. More specifically, these results suggest that 75 percent (or 14.5 percentage points) of this undercoverage rate is attributed to residences whose telephone numbers are now in zero-listed banks. As mentioned earlier, this is a direct byproduct of the significant increase in the number of residential exchanges during the past decade.

**Table 1**  Weighted coverage rates by stratum.

| Disposition | Zero-Listed | 1+Listed | Remaining POTS | Total |
|---|---|---|---|---|
| Residential | 14.5% | 80.5% | 5.0% | 100% |
| Nonresidential | 48.6% | 25.2% | 26.2% | 100% |
| Undetermined | 49.1% | 30.5% | 20.4% | 100% |

**Table 2**  Weighted hit rates by stratum.

| Disposition | Zero-Listed | 1+Listed | Remaining POTS |
|---|---|---|---|
| Residential | 4.0% | 30.8% | 2.7% |
| Nonresidential | 90.2% | 64.2% | 92.7% |
| Undetermined | 5.8% | 5.0% | 4.6% |
| Total | 100% | 100% | 100% |

### ALTERNATIVE FRAME CONSTRUCTION METHODOLOGIES

List-assisted RDD samples selected from listed 100-series banks no longer provide a representative sample or one that could be remedied through applications of post-stratification adjustment techniques. Actually, coupled with the fact that more than 16 percent of the US households are now reachable only via cell phones, it can be deduced that traditional RDD samples *at best* cover less than 70 percent of all US households. To make the situation even more complicated, it is estimated that a growing percent of households – currently estimated at about 15 percent (Blumberg and Luke 2008) – are mostly reachable via cell phones. These *cell-only* and *cell-mostly* households present yet another formidable source of coverage bias for list-assisted RDD samples.

Is using the current method of RDD for sample selection no longer a practical option? Clearly, results from this and related studies seem to suggest that if frame construction and sampling methodologies stay the same results from such samples can no longer withstand scientific scrutiny. Already, there are researchers who have questioned the future utility of RDD-based surveys in

certain settings (Link 2006). In what follows, a few simple alternatives for frame construction are introduced that can eliminate some of the coverage bias currently undermining the utility of RDD samples selected from the listed 100-series banks.

It is our submission that future RDD frames have to be developed using 1000-series telephone numbers as their basic building blocks. With that, the listed status of each block will have to be determined based on whether the associated 1000-series block contains any listed numbers or not; this way, many of the 100-series banks that are currently unlisted can be included as part of a listed 1000-series block. Actually, in the extreme case, a 1000-series block can be comprised of 9 zero-listed 100-series banks and only one listed 100-series bank. This is how a transition to listed 1000-series blocks can entail additional screening resources to identify residential numbers for the benefit of reducing the coverage bias. Based on our research, frames developed from all 1+listed 1000-series blocks are expected to increase the residential coverage rate from 80 to about 90 percent. On the negative side, household incidence (hit) rates are expected to decrease by about 10 percent.

Also, it no longer seems feasible to limit the sampling frames to include only traditional exchanges, since the landline coverage rate even when using listed 1000-series blocks for frame construction is expected to be at best 90 percent; therefore, future frames should be supplemented with the remaining POTS exchanges deemed to have residential assignments. This too, however, will further dilute the sampling frame as the rate of residential number assignments in such exchanges is currently very low. Lastly, it is becoming an obvious necessity for future RDD samples to include proper mixtures of cellular phone numbers to compensate for the *cell-only* and *cell-mostly* households that are not included in the landline frame. Given the growing number of such households, however, it is impractical to suggest standard methods for this supplementation at the present time.

## SUMMARY AND CONCLUSIONS

Digital transition of the telephone network infrastructure has all but invalidated the utility of the 100-series banks. The unfolding changes in US telephony have introduced new sources of coverage bias in traditional RDD samples with magnitudes that are no longer ignorable. Predominantly, the source of this coverage gap is due to a combination of a decrease in residential number assignment density and an increase in alternative dial-tone providers such as cable that have much lower listed rates. Recapturing this coverage will require developing sampling frames that are more inclusive even though this will entail lower residential hit rates and additional costs for screening efforts.

Given the fluidity of the current situation, it is important to implement tracking mechanisms that can assess and report the ongoing changes in the structure of telephone frames. In parallel, it will be necessary to introduce new

screening procedures that can reverse the cost drain associated with decreased hit rates resulting from expanded RDD frames. Moreover, it is highly advisable for the research community to investigate and shed more light on the emerging peculiarities associated with these changes. For instance, it will be revealing to know why the *time-to-listing* of residential numbers among alternative providers is so long and whether such low listed rates are due to number porting.

It should be noted that MSG has since conducted a second study based on a sample of 10,000 telephone numbers and obtained results that completely corroborate what is presented in this paper. Moreover, similar results have been reported for the NHES 2007 with estimated residential, nonresidential, and unknown rates of 26.9, 66.9, and 6.2 percent, respectively. Additionally, disposition results from the 2005–2006 NIS report estimates of 23.6 percent residential, 59.9 percent nonresidential, and 16.5 percent unknown. Reallocating the unknown category for each survey in proportions to their respective residential and nonresidential rates will result in estimated residential hit rates of 28.7 and 28.3 percent for NHES and NIS, respectively. Both of these estimates are very much in line with the residential hit rate of about 30 percent reported here.

A longer version of this article was subsequently published in [Public Opinion Quarterly](.).

**REFERENCES**

Blumberg, S.J. and J.V. Luke. 2008. Wireless-mostly households: estimates from the 2007 national health interview survey. Paper presented at the 63rd AAPOR conference, New Orleans, LA.

Brick, J.M., J. Waksberg, D. Kulp and A. Starer. 1995. Bias in list-assisted telephone samples. *Public Opinion Quarterly 59*: 218–235.

Casady, R.J. and J.M. Lepkowski. 1993. Stratified telephone survey designs. *Survey Methodology 19*: 103–113

Link, M.W. and M. Kresnow. 2006. The future of random-digit-dial surveys for injury prevention and violence research. *American Journal of Preventive Medicine 31*(5): 444–450.

Mitofsky, W. 1970. Sampling of telephone households. Unpublished CBS News memorandum.

Tucker, C., J.M. Lepkowski and L. Piekarski. 2002. The current efficiency of list-assisted telephone sampling designs. *Public Opinion Quarterly 66*: 321–338.