ARTICLES

Addressing the Cell Phone-Only Problem: Cell Phone Sampling Versus Address Based Sampling

Michael W. Link¹, Gail Daily¹, Charles D. Shuttles¹, H. Christine Bourquin¹, L. Tracie Yancey¹

¹ The Nielsen Company

Keywords: survey practice

<https://doi.org/10.29115/SP-2009-0007>

Survey Practice

Vol. 2, Issue 2, 2009

Addressing the Cell Phone-Only Problem: Cell Phone Sampling Versus Address Based Sampling

Developing cost effective methods for reaching households which no longer have a landline but do have access to a cell phone, so called cell phone only households, is a critical item on the agenda of most data collection organizations. To date, two methodologies have emerged as potential means for addressing this issue. The first involves sampling telephone numbers from known cell phone exchanges and calling these numbers or combining these with a sample of landline numbers in a dual frame design. An alternative approach involves sampling of addresses rather than telephone numbers. Address based sampling (ABS) is a new technique built upon the use of large scale address databases. These addresses can be reverse-matched to commercially available databases to identify a relatively large proportion of telephone numbers, facilitating the use of mixed-mode approaches. Here we delineate and compare the advantages and limitations of these two approaches, including discussion of sampling and weighting approaches, operational considerations, timeliness, and cost.

Twilight for Landline Random Digit Dialing

For nearly three decades, landline-based random digit dialing (RDD) enjoyed preeminence among survey methodologies, facilitating high quality, quick turn-around, computer-assisted interviewing that met the needs of most researchers and clients. Yet, with the dawn of the 21st century, issues first noted in the late 1990s became true problems. First, the sheer number of telephone numbers increased leading to significant declines in the "hit rate" for residential numbers in telephone samples, increasing the cost and difficulty of identifying working household numbers. Next was the decline in response rates for landline RDD surveys, dropping approximately 2–3 percent per year since the mid-1990s and increasing concerns about the representativeness of the data collected using RDD methods. Third, the growth and popularity of cell phones has irrevocably altered the landscape of telephone survey research, with nearly one-in-five US households being cell-phone only, a proportion which is substantially higher among groups such as renters and younger adults (Blumberg and Luke 2008). Finally, new research indicates that the common practice of excluding zero listed landline banks to improve operational

efficiency in RDD surveys may not be as benign as previously thought, with a higher than expected proportion of unlisted, residential landline numbers now being identified in these banks (Fahimi, Kulp, and Brick 2008). All told, coverage of the landline telephone frame has decreased to pre-1970s levels.

Researchers, data users, and clients who have come to depend on landline-based RDD find themselves at a crossroads. The problems with landline RDD methodology will not be fixed through the use of incentives, advance mailings, additional telephone calls, oversampling, or any of the many techniques designed to improve participation with this methodology. What is needed is a complete re-engineering or even re-imagining of how researchers go about the task of sampling to conduct cost-effective surveys of the general public. Two potential options have emerged: (1) combining cell phone exchanges with landline exchanges in a dual frame approach, or (2) turning away from dependence on telephones as the primary sampling unit and moving instead to sampling of addresses. We consider some of the major arguments for and against these two approaches.

Dual Frame Sampling with Cell Phone Numbers

One potential solution is to sample from known banks of cell phone numbers, combining this sample with the landline sample in a dual frame approach. The combination of the two frames should significantly improve coverage, but would still exclude households with no telephone access and those with unlisted landline numbers in banks not typically sampled by survey researchers. The dual frame approach has a number of other potential benefits. First, it allows researchers to continue to use computer assisted telephone interviewing (CATI) as the primary vehicle for conducting telephone surveys, facilitating use of complex questionnaires and the conduct of quick turn-around surveys. In terms of sample costs, the per unit cost for cell phone numbers is only slightly more than that of traditional landline samples.

Unfortunately, sampling and contacting households by cell phone faces a number of challenges, some severe. First, the sampling frame of known cell phone numbers is an inefficient one, containing a large, but unknown, number of cell phone numbers which are either not in service or in service but rarely used or answered. As a result, researchers must either make a large number of calls per case or sample a larger number of cell phones in order to reach an individual. Cell phones may also not be in use in the geographic area from which they were sampled. Potentially more problematic, the cell phone frame is rather barren in terms of additional information about the number, such as associated address, name, projected demographic characteristics, etc. As a result, frame stratification is limited and some of the common features of modern telephone surveys, such as the ability to send advance letters to homes before a telephone call, are not possible at all.

The dual frame approach also poses a number of operational challenges. Initial contact with cell phone households is limited to telephone contact, limiting the use of other modes during the contact/recruitment phase of a study. Participation in cell phone surveys is already quite low and can only be expected to decline further over time. Additionally, the cell phone must also be viewed as a new mode, with potential uses and constraints that differ from traditional landline interviewing. Some of the areas which require further clarification through research include:

- the optimal questionnaire length;
- the level of “cognitive engagement” of respondents interviewed by cell phone, particularly when multitasking or being interviewed while driving or shopping;
- lack of a common set of disposition codes to cover situations which may be unique to cell phone interviewing (although work in this area is progressing);
- clarification of response rate calculations, in particular the determination of the percentage of uncontacted numbers which should be estimated as eligible households and, therefore, be included in the denominator of a response rate; and,
- the applicability of within household randomization with cell phone interviewing, that is if the devices should be treated as individual or household devices, and how to handle selection within a cell phone only household in which an eligible sample member does not have their own cell phone and, therefore, would be excluded from inclusion in a study.

One particularly vexing issue is the lack of universe estimates or population parameters against which to weight survey data. This is especially problematic at the sub-national levels (state, county, city, etc.).

Surveys conducted via sampling of cell phone exchanges are also much more expensive to conduct than the costs associated with a landline survey. Studies have reported the costs to be nearly twice as high when screening for cell phone-only households was not conducted and nearly four times as great when such screening was used to identify cell phone-only households (Keeter et al. 2008; M. Link et al. 2007).

Cell phone interviewing also involves certain legal and ethical considerations that do not apply to traditional landline interviewing. For instance, the Telephone Consumer Protection Act and the FCC’s implementation (71 Federal Reg 21634, April 26, 2006) prohibit machine-based dialing of cell phone numbers without prior consent from the respondent. These numbers need to be dialed by hand, thereby making them more expensive for many

survey organizations and impractical for organizations dealing with very large samples sizes. With cell phone interviews it is imperative to ensure that respondents are in a safe location or situation before proceeding with the interview. Finally, many US cellular calling plans require the cell phone subscriber to pay for incoming calls, which raises a number of ethical and legal issues associated with soliciting cell phone subscribers without appropriate financial compensation.

In sum, it seems clear from the flurry of research in this area that many survey researchers who currently conduct landline RDD surveys are hopeful that a dual frame landline/cell phone approach can be developed to deal with the growing coverage crisis in landline RDD surveys. It is clear, however, that such an approach faces an array of obstacles. The approach may have short-term appeal, but its long-term prospects are still unclear.

Address Based Sampling (ABS)

As an alternative, researchers have looked to a completely different approach: address based sampling, that is, the use of addresses as the primary sampling unit drawn from a computerized frame of address listings (M. W. Link et al. 2006, 2008). In particular, the Delivery Sequence File (DSF) used by the U.S. Postal Service (USPS) has proven most promising. The DSF is a computerized file that contains all delivery point addresses serviced by the USPS, with the exception of general delivery (United States Postal Service 2005). Each delivery point is a separate record that conforms to all USPS addressing standards, thereby facilitating the drawing of samples from any geography within the US.

When considering all types of addresses, the DSF provides approximately 98% coverage of residential households, thereby providing a means of sampling landline and cell phone only households, as well as providing access to households with no telephone, newly emerging VoIP-only based computer phones, and those in less efficient landline telephone banks (e.g. zero listed banks). Because addresses are in a fixed location, telephone portability is not an issue.

Another important benefit is the rich amount of information that can be matched to an address, facilitating more complex sample designs and providing information for enhanced contacting and recruiting approaches. A majority of addresses can be matched to a landline telephone number via commercial databases, thereby facilitating multiple modes of contact. Survey sample vendors can typically provide case-level variables such as household name, Spanish surname indicator, estimated age of head of household, as well as geocoding and attachment of Census tract information such as the percentage of racial/ethnic groups within a particular geography, median household income of the area, and in some cases even email addresses. These variables can be used in a number of ways to enhance the survey design, such as through

sample stratification on key variables, advance mailings to households, and tailoring of materials, contact scripts, or incentives based on household characteristics such as likely age, race, or ethnicity of the head of household.

ABS facilitates a range of potential survey designs, including single mode mail surveys to all sampled addresses; or a mail invitation to complete a mail or web survey; or, a dual mode design with mail surveys to all households (or just those with no matched telephone number) and telephone follow-up (or first contact) for those with an identifiable telephone numbers; or a more complex mix of mail, Web, interactive voice response (IVR), and outbound or inbound telephone. This gives researchers greater flexibility to match survey mode with the goals and target population for their study.

Weighting and post-survey adjustments can follow traditional survey procedures as population totals or universe estimates are readily available at the level of most commonly used geographies (national, state, county, etc.). Further, because of the near universal coverage the weighted data should more representative of the larger population from which the sample was drawn provided there is little or no systematic bias due to nonresponse.

In terms of cost, an equal number of sampled addresses are about twice as expensive as an equal sample of telephone numbers, although this can vary based on the sample vendor, number of cases sampled, and amount of additional data appended to each sampled case. Because of the efficiency of the frame (i.e., there are relatively few non-residential addresses), far fewer addresses (than telephone numbers) are required to reach a residential household.

Address-based approaches do, however, have some drawbacks. DSF information cannot be obtained directly from the USPS, but must be purchased through a nonexclusive license agreement with private vendors. The quality and completeness of the address information obtained from these vendors varies widely depending on how often the company updates the listings, the degree to which the listings are augmented with information from other databases, and whether the company purges the records of householders who request that their information not be released (M. W. Link et al. 2006). The DSF contains post office (PO) box and multi-drop addresses (multiple persons associated with the same address), which may be problematic for in-person and telephone surveys where a street address is required to locate the household or an associated telephone number. Such addresses may be less problematic for surveys which use mail as the recruitment mode (such as with mail or Web surveys). Households with multiple mailing addresses (for example, a street address and a residential PO box) can introduce selection multiplicities if both addresses are utilized.

From an operational perspective, ABS can limit the ability of a research organization to conduct quick turnaround studies. While a majority of the sampled addresses can be matched to a telephone number, the remaining sample must be contacted/recruited first by mail regardless of the actual survey mode used for data collection. This process takes time. As an alternative, an organization could conduct on-going pre-recruitment efforts with these “unmatched” cases (i.e., those with no matched telephone number), obtaining telephone contact information from respondents and providing a ready bank of numbers from which to sample for this portion of addresses. This is, however, a relatively expensive and somewhat complex proposition.

If limited to mail-only, many surveys would also need to be adjusted in terms of length and complexity, as longer, more complex surveys are not readily feasible with a paper-and-pencil approach. Use of a Web survey option and/or a call-in number to a CATI interviewer can alleviate this problem, however, only with households with Web access and very few respondents are likely to call in to complete a survey with the latter design.

As is the case with cell phones, use of address based sampling also requires modification of standard disposition codes (even the American Association for Public Opinion Research’s [AAPOR] current disposition list for mail surveys is inadequate as the list only applies to mail surveys where the respondent is known) and a reassessment of response rate calculations as there is currently no agreed upon industry standard for determining the percentage of noncontacted addresses (excluding those for which a post office return was received) to include in the denominator of a response rate.

Conclusions

While it would likely be a mistake to declare that the sun has set completely on landline RDD methodology, it is clear that the approach has serious, seemingly non-recoverable problems in terms of coverage and declining response rates. This is not to say that “telephone surveys” are nearing their end, but rather the reliance on the landline telephone frame as the sole basis for drawing samples for conducting surveys of the general population is in jeopardy. Given the considerable cost of conducting in-person interviews, that mode is likely to remain reserved for only the best funded projects. At the other extreme, use of online surveys based on non-probability, opt-in sample designs is likely to remain a niche, but growing methodology for the foreseeable future. Use of dual frame landline/cell phone studies or address based sample approaches may become more common methodologies in the future. Both have their advantages and disadvantages — some which can be resolved with time and additional research, others which may be intractable.

In the end, the use of a dual-telephone frame approach or an address based approach to sampling comes down to how well each fits the requirements of the research at hand, in terms of cost, quality, and timeliness. What may work well for one research endeavor may not match the needs of another. As a short-

term solution, sampling of cell phone exchanges may provide a stop-gap for those conducting smaller to moderate-sized surveys until a more stable, longer-term methodology is refined. An ABS approach is perhaps the most promising foundation upon which such a methodology (or set of methodologies) can be built, providing a stable sampling base, a rich source of characteristic and geographic data for facilitating sophisticated designs, and an opportunity to utilize multiple modes for contacting and conducting surveys with households.

REFERENCES

- Blumberg, S., and J. Luke. 2008. "Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, January-June 2008." 2008. <http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200812.htm>.
- Fahimi, M., D. Kulp, and M. Brick. 2008. "Bias in List-Assisted 100-Series Rdd Sampling." *Survey Practice*. <https://www.surveypractice.org/article/2922>.
- Keeter, S., M. Dimock, C. Kennedy, J. Best, and J. Horrigan. 2008. "Costs and Benefits of Full Dual Frame Telephone Survey Designs." In *Paper Presented at the 63rd Annual Conference of the American Association for Public Opinion Research*. New Orleans, LA.
- Link, M., M.P. Battaglia, M. Frankel, L. Osborne, and A. Mokdad. 2007. "Reaching the Cell Phone Generation: Comparison of U.S. Cell Phone Survey Results with an Ongoing Landline Telephone Survey." *Public Opinion Quarterly* 71: 814–39.
- Link, M.W., M.P. Battaglia, M.R. Frankel, L. Osborn, and A.H. Mokdad. 2006. "Address-Based versus Random-Digit Dialed Surveys: Comparison of Key Health and Risk Indicators." *American Journal of Epidemiology* 164: 1019–25.
- . 2008. "Comparison of Address Based Sampling (ABS) versus Random-Digit Dialing (RDD) for General Population Surveys." *Public Opinion Quarterly*.
- United States Postal Service. 2005. "Delivery Sequence File." 2005. <http://www.usps.com/ncsc/addressservices/addressqualityservices/deliverysequence.htm>.