Predicting the Coverage of Address-Based Sampling Frames Prior to Sample Selection

Joseph P. McMichael[1], Jamie L. Ridenhour[1], Bonnie E. Shook-Sa[1], Katherine B. Morton[1], Vincent G. Iannacchione[2]
[1] RTI International, 3040 East Cornwallis Road, Research Triangle Park, NC 27709
[2]RTI International, 701 13th Street, N.W., Suite 750, Washington, DC 20005

Abstract
In 2009, RTI conducted a field study for the National Survey on Drug Use and Health (NSDUH) aimed at investigating the cost implications and coverage properties of a sampling frame based on address-based sampling (ABS) in area-segments with adequate ABS coverage and field enumeration (FE) elsewhere. Data for this study were based on a probability sample of 200 segments and 3,878 eligible dwelling units from the NSDUH. We found that accurately predicting ABS coverage at the area-segment level prior to sample selection simultaneously lowers costs and improves frame coverage.

Key Words: ABS, coverage, in-person surveys

1. Introduction & Background

The current sampling frame for the National Survey on Drug Use and Health (NSDUH) relies on field enumeration (FE) supplemented with the half-open interval (HOI) procedure (Kish 1965). Listers are sent to area segments, which are comprised of one or more adjacent census blocks that in combination meet or exceed the minimum required number of DUs, to enumerate all the dwelling units (DUs) in segment. Listers also enumerate group quarter areas such as dormitories. Because of the costs associated with FE, the size of the area segments is small: usually about 100 DUs in a rural[1] area and 150 DUs in an urban area. The HOI procedure is used during screening and interviewing to pick up new construction and/or units otherwise missed by listers.

Due to the costs associated with FE, several national in-person surveys are looking at address-based sampling (ABS) instead of or in conjunction with FE due to the lower costs associated with ABS. A study may choose to use ABS in conjunction with FE because although ABS is much cheaper than FE, coverage is not as desirable in areas with group quarters and rural areas (Iannacchione et al. 2007, O'Muircheartaigh et al. 2007). An approach that would utilize the good qualities of both frames is a hybrid ABS/FE frame using ABS in segments where coverage is adequate and retaining FE in segments were ABS coverage is low. To use ABS in combination with FE, there must be a viable method for determining what proportion of the sample segments should be field enumerated and what proportion will use ABS. This determination should be made based

[1] For a segment to be classified as rural, all of the census blocks in the segment have to be rural. If one or more of the segment's blocks are urban, the segment is also urban.

on the coverage properties and costs associated with using ABSE and FE in a sample segment. Accurately predicting coverage is of paramount importance to using a hybrid ABS/FE frame. This paper discusses an examination of a candidate frame for the NSDUH and how coverage can and should be predicted for sample segments *a priori*.

1.1 Field Enumeration Costs

There are several costs associated with field enumeration. Some costs are unique to FE including: training listers on how to enumerate area segments and the subsequent fieldwork, producing lister maps, field support during listing, and processing listed data. Whether FE or ABS is used for surveys, a frame supplementation procedure should be used because coverage of both frames is not complete. There is a cost associated with field interviewers (FIs) implementing the FE frame supplementation procedure, the HOI, during screening and interviewing but there is also cost associated with the frame supplementation procedure for ABS. The ABS frame supplementation procedure discussed in this paper is called Check for Housing Units Missed, or CHUM, which was developed by RTI for use on in-person ABS studies (McMichael et al. 2008)

1.2 Address-Based Sampling Costs

ABS is an attractive alternative to in-person surveys because of its lower cost and efficiency. Purchasing address lists for area segments is less expensive than a listing operation. Address lists can be purchased and prepared closer to data collection while field enumeration must occur months in advance of screening and interviewing. The costliness of FE restricts the size of area segments: smaller segments are not as costly to list as large segments but small segments could lead to larger variance in estimates due to higher intra-cluster correlation. With ABS, the difference between the cost of smaller segments and larger segments is small because ABS does not require listers to visit each address on the ABS frame; the difference is only in the cost of additional addresses purchased. Thus, larger segments can be used with ABS which has the potential to reduce any intra-cluster correlation in the segments.

The drawbacks of ABS are that while nearly all households and group quarters have a mailing address, not all mailing addresses are locatable on the ground. An address must be locatable to be utilized for an in-person survey. Unlocatable addresses include PO Boxes which do not directly correspond to a housing unit that field staff can locate. Unlocatable addresses cannot be included on the ABS frame for in-person surveys which results in under-coverage of the target population. This under-coverage occurs more frequently in rural areas and areas with group quarters. Under-coverage can be mitigated with the use of a frame supplementation procedure; in this case, the CHUM.

Purchasing address lists is a cost that is unique to ABS. ABS also requires the production of maps; however, these maps can be quite different from listing maps. There are also costs associated with training field interviewers (FIs) on how to implement the CHUM during screening (used to roster the sampled household and select persons to be interviewed) and interviewing, the FIs implementing the CHUM, and any necessary field support during CHUM implementation.

Although there are cost savings associated with ABS, the coverage issues in rural segments and segments with group quarters are problematic. ABS can be used in these segments; however, in these cases the CHUM frame supplementation procedure would be

used to an extent that field enumeration may become a more cost-effective and error-minimizing approach.

1.3 National Survey on Drug Use and Health

The NSDUH is the nation's leading source of information on substance use behaviours and mental health and has been conducted by RTI under contract with the Substance Abuse and Mental Health Services Administration since 1988. Its target population is the civilian, non-institutionalized population aged 12 and older from both households and non-institutional group quarters. Examples of non-institutionalized group quarters include dormitories and convents. These types of living quarters will be important later in this paper when the hybrid frame is evaluated. Data is collected throughout the year in all 50 states and the District of Columbia. The NSDUH has approximately 700 field interviewers staffed in 7,200 local areas or segments completing 140,000 screenings and 67,500 interviews annually.

1.4 NSDUH Field Study

The objective of the NSDUH field study was to develop and test an ABS/FE hybrid frame that provides cost savings without sacrificing coverage. The evaluation of the frame required examining changes in coverage along with cost savings at several candidate ABS coverage thresholds (Iannacchione et al. 2010). A coverage threshold is the dividing line that distinguishes between segments that use FE or ABS under a hybrid frame. The hybrid frame was also evaluated for its effect on outcomes of interest (Morton et al. 2010). Theoretically, the hybrid frame offers the same coverage as an FE frame, but with cost savings. We evaluated the coverage differences between the hybrid frame and the field enumerated frame, assuming the FE frame has complete coverage of the NSDUH population.

The field study was implemented by sub-sampling 200 NSDUH segments from the 2009 quarter 1 sample. The DUs in the sample comprised of 3,878 screened and eligible[2] sampled DUs in a subsample of 200 NSDUH segments. Segments in Alaska and Hawaii were excluded from the field study sampling frame. To develop a hybrid frame of DUs, we attempted to match the addresses of eligible sampled DUs obtained from the NSDUH field enumeration process and updated during the screening to a list of mailing addresses purchased from a commercial vendor. We classified the segments by coverage threshold to evaluate how, theoretically, these segments would be allocated to FE or ABS under the hybrid frame.

The three candidate ABS coverage thresholds examined in the field study are 20 percent, 50 percent, and 80 percent. These arbitrary thresholds were chosen to understand and evaluate the variability of cost and coverage under different thresholds across the spectrum. At a 20 percent coverage threshold, all segments with less than 20 percent ABS coverage would be field enumerated and those with greater than 20 percent ABS coverage would utilize ABS supplemented with the CHUM procedure. Thus, at the 20 percent coverage threshold more segments fall under ABS than FE but at the 80 percent coverage threshold, more segments are enumerated and fewer utilize ABS. The ability to *accurately* predict BS coverage at the segment level is essential to utilizing the hybrid frame because failing to put a segment in the correct domain, ABS or FE, incurs costs

[2] An eligible DU for the NSDUH is either a housing unit (HU) for a single household or a non-institutional group quarters (GQ) where at least one civilian aged 12 years or older resides for the majority of a calendar quarter.

that may reduce the overall savings of the alternative frame and has coverage implications for the alternative frame.

2. Predicting Coverage of ABS Segments

As discussed previously, the NSDUH currently uses a field enumerated frame supplemented with HOI. A hybrid ABS/FE frame would achieve cost savings while retaining the desirable quality of nearly complete coverage. In a hybrid frame approach, ABS would be used in segments where ABS coverage is adequate and FE would be retained in segments were ABS coverage is deemed not sufficient. Research in to the viability of a hybrid frame was conducted by RTI in 2009. A critical part of this research was the development of a method to accurately predict whether a segment will have sufficient ABS coverage.

2.1 Sources of Inaccurate Predictions

The coverage prediction estimate used for the field study was the ratio of locatable mailing addresses to the number of households and group quarters in a segment.

$$\text{Predicted segment coverage} = \frac{\#\text{ Locatable Addresses}}{\#\text{Households} + \#\text{Group Quarters}}$$

Estimating ABS coverage by this formula brings in several potential sources of error. The numerator, number of locatable mailing addresses, is a count of addresses within a segment from the ABS mailing address list. The addresses on the list for that segment are subject to geocoding error. Addresses can be geocoded in to a segment they are not actually in or geocoded out of a segment in which they should be and that error can lead to over or under-coverage at the segment level (Shook-Sa et al. 2010). Another factor affecting the quality of mailing addresses in segments is the size of the segment. Research suggests that as segment size increases, geocoding accuracy improves as well (Morton et al. 2007).

The denominator of this estimated ABS coverage rate is also subject to error. For the 2009 study, RTI used updated segment-level census projections from Claritas as the denominator. As expected, projections further away from the 2000 census are more subject to error than those close to the census. These projections may or may not correctly account for the high rates of growth or decline that have occurred in some areas.

2.2 Implications of Inaccurate Predictions

The purpose of developing a hybrid frame is to achieve cost savings while retaining coverage. Incorrectly allocating segments to either FE or ABS can result in higher costs and lower coverage than would otherwise occur if a segment is correctly allocated to one of the two.

Allocating a segment to ABS when it should utilize FE is problematic in two ways. It can result in a loss of coverage because the field interviewer may have to use the CHUM for a substantial number of DUs which increases the likelihood of a mistake. This also increases costs because an FI using the CHUM for too many DUs would not be as efficient as a lister enumerating the same amount.

Allocating a segment to FE when it should utilize ABS also unnecessarily increases costs because the costs have been incurred for sending a lister to enumerate the whole segment when in reality the purchase of locatable mailing addresses along with the use of the CHUM procedure would have provided sufficient coverage for that segment.

For the NSDUH field study, we were able to compare the predicted coverage to the actual segment coverage for the 200 selected segments. We predicted coverage for each segment based on the formula discussed previously and then evaluated actual coverage based on the results of matching the 3,878 DUs (listed via FE) to the purchased ABS mailing address list for the segments. Field staff aided in the matching process by visiting DUs in the field that did not initially match to the ABS list.

There are four possible outcomes for each segment. For the purposes of this discussion, a 50 percent ABS coverage threshold is used although this same evaluation could be repeated for the 20 and 80 percent coverage thresholds. Actual segment coverage is defined as the number of matched DUs over the total number of DUs in a segment as listed via the FE procedure. This actual coverage calculation rate assumes that FE is complete with no error.

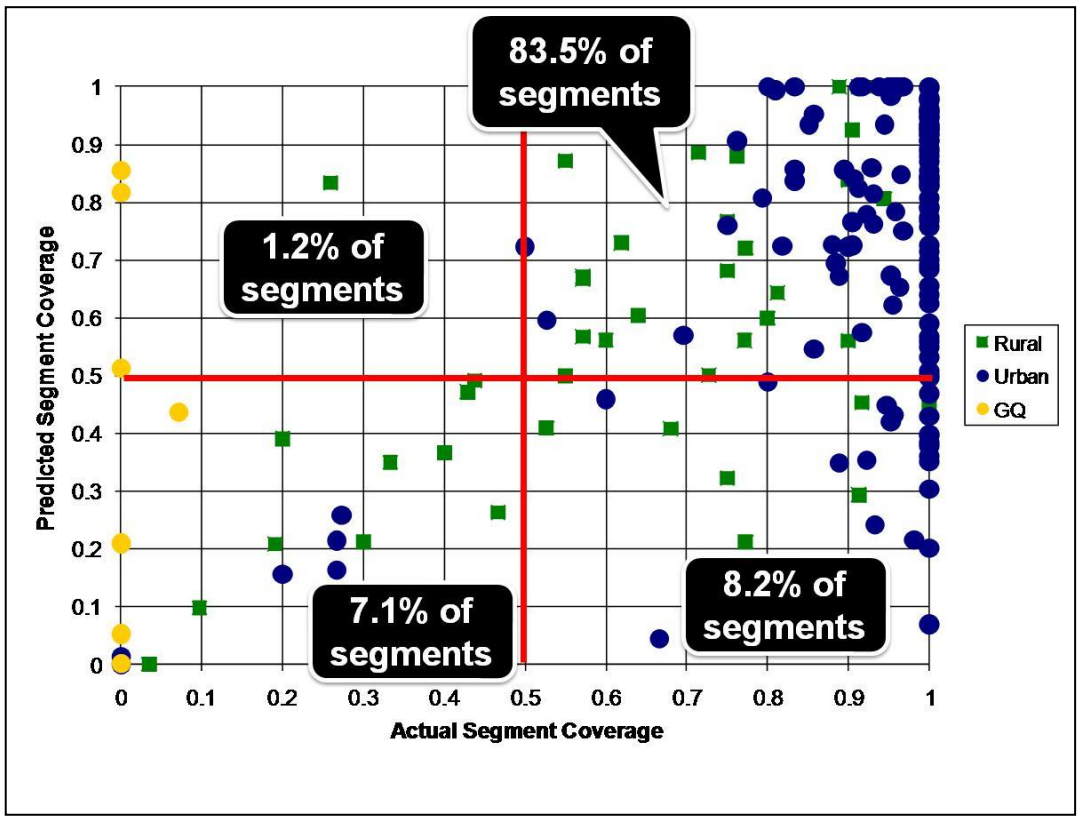

Figure 1: Weighted segment allocation by segment type and high percentage of group quarters units (GQU).

Segments in the top right and bottom left quadrants in Figure 1 were allocated correctly. Segments in the top right quadrant had ABS coverage that was predicted to be greater

than 50 percent and the actual coverage was also greater than 50 percent. Segments in the bottom left quadrant were predicted to have low ABS coverage: less than 50 percent. They were also found to have actual ABS coverage less than 50 percent. All of these segments were correctly allocated to the appropriate component, ABS or FE, of the hybrid frame.

Segments in the top left and bottom right quadrants are problematic. Segments in the top left quadrant were predicted to have high ABS coverage but in reality their actual ABS coverage is low. These segments would have incurred higher costs and potentially have under-coverage. Segments in the bottom right quadrant were predicted to have low ABS coverage but they actually have adequate ABS coverage. Segments in this quadrant would have unnecessarily increased the costs of the hybrid frame because they were enumerated when address lists could have been utilized.

2.3 Evaluation of Prediction Method

All 200 selected segments were evaluated for their predicted and actual ABS segment coverage. Eighty-five percent (weighted) of NSDUH segments had greater than 50 percent predicted coverage. Eighty-four percent of all NSDUH segments were correctly predicted to have adequate ABS coverage and had adequate ABS coverage. Only one percent of NSDUH segments were incorrectly allocated in to utilizing ABS, where their predicted ABS coverage was high but they were found to actually have low ABS coverage. Seven percent of NSDUH segments are also correctly allocated to FE but eight percent were incorrectly allocated to FE.

Figure 1 also shows the 200 selected segments and their classification based on predicted and actual coverage. Many more segments were found to have perfect ABS coverage (i.e., actual segment coverage rate=1) than were predicted to have perfect ABS coverage, although many still were allocated to the correct quadrant based on the predicted coverage rate. It is important to note that several of the segments that were incorrectly allocated to ABS when they should have been field enumerated have a high proportion of group quarters. Utilizing information about the presence of group quarters in segments will be critical in more accurately predicting segment coverage and allocation to ABS or FE. Also, the segments that were incorrectly allocated to FE (i.e., segments with low predicted coverage but high actual coverage) were a mix of both rural and urban segments. Misallocating to FE was more common than misallocating to ABS. Any improvement in the prediction method used to predict segment coverage of locatable addresses would need to identify the characteristics of these segments which would have resulted in correctly classifying them as using ABS.

3. Conclusions

Under a hybrid frame, correctly predicting whether segments should utilize ABS or FE is essential to retaining desired coverage properties (e.g., minimizing coverage bias) while achieving cost savings. Cost savings are achieved by appropriately allocating segments that have sufficient ABS coverage to ABS. Coverage is maintained by allocating segments where ABS coverage is low to FE. Allocating correctly, in both cases, reduces costs.

While the prediction method discussed in this paper provided favourable cost and coverage results, improved methods are sure to be developed as the ABS frame sources improve their coverage of locatable addresses, more accurate geocoding methods become available, and more recent census data comes online.

Acknowledgements

This project is funded by the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract no. 283-2004-00022 and Project no. 0209009. The views expressed in this paper do not necessarily reflect the official policies of the Department of Health and Human Services; nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

References

Iannacchione, V., K. Morton, J. McMichael, D. Cunningham, J. Cajka, and J. Chromy. 2007. Comparing the coverage of a household sampling frame based on mailing addresses to a frame based on field enumeration. In *JSM Proceedings,* Survey Research Methods Section, Alexandria, VA: American Statistical Association. 3323-3332.

Iannacchione, V., K. Morton, J. McMichael, B. Shook-Sa, J. Ridenhour, S. Stolzenberg, D. Bergeron, J. Chromy, and A. Hughes. 2010. The best of both worlds: a sampling frame based on address-based sampling and field enumeration. In *JSM Proceedings,* Survey Research Methods Section, Alexandria, VA: American Statistical Association.

Kish, L. (1965). *Survey Sampling* (p. 56). New York: John Wiley & Sons.

McMichael, J.P., J.L. Ridenhour, and B.E. Shook-Sa. 2008. A robust procedure to supplement the coverage of address-based sampling frames for household surveys. In *JSM Proceedings,* Survey Research Methods Section, Alexandria, VA: American Statistical Association. 4329-4335.

Morton, K. B., J.P. McMichael, J.C. Cajka, R.J. Curry, V.G. Iannacchione, and D.B. Cunningham. 2007. Linking mailing addresses to a household sampling frame based on census geography. In *JSM Proceedings,* Survey Research Methods Section, Alexandria, VA: American Statistical Association. 3971-3974.

Morton, K.B., J.P. McMichael, J.L. Ridenhour, and J. Bose. 2010. Address-based sampling and the National Survey of Drug Use and Health: evaluating the effects of coverage bias. In *JSM Proceedings,* Survey Research Methods Section, Alexandria, VA: American Statistical Association.

O'Muircheartaigh, C., E. English, and S. Eckman. 2007. Predicting the relative quality of alternative sampling frames. In *JSM Proceedings,* Survey Research Methods Section, Alexandria, VA: American Statistical Association. 3239-3248.

Shook-Sa, B.E., J.P. McMichael, J.L. Ridenhour, and V.G. Iannacchione. 2010. The implications of geocoding error on address-based sampling. In *JSM Proceedings,* Survey Research Methods Section, Alexandria, VA: American Statistical Association.