The Effects of Address Coverage Enhancement on Estimates from a Study using an ABS Frame

Michael Jones[1], J. Michael Brick[1], Andrea Piesse[1]
[1]Westat,1600 Research Blvd, Rockville, MD 20850

Abstract
For over a decade, address-based sampling (ABS) frames have often been used to draw samples for multi-stage area sample surveys in lieu of traditionally listed (or enumerated) address frames. However, it is well-known that ABS frames suffer from under-coverage due to, for example, households that receive mail via a PO Box rather than being delivered to the household's street address. Under-coverage of ABS frames has typically been more prominent in rural areas, but can also occur in urban areas where recent construction of households has taken place. Procedures have been developed to supplement ABS frames in order to address this under-coverage. In this paper we investigate a procedure called Address Coverage Enhancement (ACE) that supplements the ABS frame with addresses not found on the frame, and the resulting effects the addresses added to the sample through ACE have on estimates. Estimates are calculated with and without supplemental addresses.

Key Words: address coverage enhancement, address-based sampling, non-coverage bias, sampling frame coverage

1. Introduction

For much of the twentieth century, survey researchers often favored random digit dialing (RDD) frames to conduct telephone surveys over higher cost in-person surveys that relied on field-listed (enumerated) frames. However, the shift in telephone usage (Blumberg and Luke, 2017) and declining response rates to telephone surveys (Iannacchione, 2011; Brick and Williams, 2013) have necessitated alternative strategies. Also, limited budgets have impacted many surveys that relied on field-listed frames, pushing them toward alternatives. The advent of address-based sampling (ABS) frames in the 1990s offered a significantly less expensive option to survey researchers for both of these problems. Montaquila et al. (2013) show how the use of an ABS frame enabled a low-cost alternative to an RDD frame. This paper presents initial results of a broader investigation into the efficacy of ABS frame enhancement.

ABS frames generally provide high coverage nationally, although coverage in some areas remains poor (AAPOR, 2016). However, coverage challenges arise when an ABS frame is used to replace field listing, as described later. To supplement areas with poor ABS frame coverage for listing purposes, frame enhancement procedures have been developed. This paper presents results from a study about the effects of frame enhancement on estimates from an in-person study. We begin with a brief overview of ABS frames and Address Coverage Enhancement (ACE), the frame enhancement methodology developed and utilized by Westat. Section 3 discusses the Population Assessment of Tobacco and Health (PATH) Study. The approach we used for assessing the impact of ACE on the PATH Study is presented in Section 4, and results are presented in Section 5. We discuss conclusions and areas for further research in Section 6.

2. ABS Sampling Frames and Frame Enhancement

The core of an ABS frame consists of the addresses found on the United States Postal Service (USPS) Computerized Delivery Sequence file (CDS). Dohrmann et al. (2014) provide a comprehensive review of the CDS. They discuss how vendors qualify to hold a license in order to obtain the CDS, the processing vendors perform on the file prior to release, and the variables found on the file.

Overall, ABS frames provide excellent coverage of addresses across the Unites States, which makes them extremely effective for mail surveys. However, for studies that conduct in-person interviews, using these addresses in lieu of listing can create problems. Coverage rates can suffer due to non-locatable addresses such as those that receive mail only by PO Box. Also, at the time of sampling, new construction in high-growth areas might not be included on the CDS. Kali et al. (2014) point out some of the issues faced when using an ABS frame for the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC-III).

Frame enhancement procedures have been developed in order to improve coverage of ABS frames for in-person surveys. Three such enhancement procedures are reviewed in Harter and English (2018): Enhanced Listing, Check for Housing Units Missed (CHUM), and Address Coverage Enhancement (ACE). Developed at Westat, ACE is the enhancement procedure relevant to this paper.

The genesis of ACE can be traced to the frame enhancement methodology first proposed by Dohrmann et al. (2006) in which segment boundaries were created using contiguous census blocks or combinations of census blocks. Addresses geocoded within the segment boundaries defined the segment, knowing geocoding error could be present. Their approach amended the Waksberg approach to the half-open interval (HOI) method (Kish, 1965) by giving segments with expected poor frame coverage a greater chance of selection to undergo the enhancement procedure.

Dohrmann et al. (2012) revised and further developed the methodology by establishing rules for checking missed units against the frame and addressed geocoding error. An important distinction associated with ACE compared to other enhancement procedures is the recognition of *area* segments and *list* segments. In ACE, an area segment is a sampled geographic area (like a census block) that is sampled. In contrast, a list segment is all the ABS addresses that geocode to that area segment whether they are physically inside the segment or not. Figure 1 illustrates an example of an area segment outlined in blue and the addresses on the ABS frame that get geocoded to, or linked to, the area segment (shaded pink). The addresses shaded in the figure become the list segment.

Figure 1: Illustration of an area segment versus a list segment.

The steps to carry out the fully established ACE procedures are presented in Section 2 of Kali et al. (2014). Section 4.2 of Kalton et al. (2014) discusses treatment of probabilities of selection for segments to undergo ACE procedures and probabilities associated with address selection within ACE-selected segments. In general, ACE can be broadly described as follows:

1. Form area segments and link list segments to area segments
2. Assign probabilities of selection to the area segments and select a sample of those segments to undergo ACE procedures
3. Implement ACE procedures in the field
4. Sample from the addresses added to the ABS frame

Step 3, implementation of ACE in the field, involves loading the ABS addresses geocoded to a given list segment onto a laptop. Field staff trained to perform ACE then systematically canvass each ACE-selected area segment. They compare each housing unit they encounter within the boundaries of the area segment to the list of addresses on the preloaded list on the laptop. If the address for a housing unit is found on the preloaded list, they assign the address a status of “located.” If not found, they record the address into the system and the laptop application flags the address as “added in the field.” Addresses added in the field are reconciled at the home office to determine if they are already on the ABS frame but geocoded to another area segment. Any addresses not found on the ABS frame are added to it, and what results is an “enhanced ABS frame.”

2.1 Illustration of ACE

We next illustrate a simplified version of the ACE process. In this case the ACE procedure is very similar to enhanced listing. Notice that Figure 2 does not show the list segment shaded in pink that appears in Figure 1. Since our evaluation of the impact of ACE on estimates does not depend on geocoding error, we assume no geocoding error exists for the illustration. Under that assumption the list and area segments are the same. In Figure 2, housing units with addresses found on the unenhanced ABS frame that are geocoded into the segment are represented by the red circles. Those addresses are provided via laptop to the data collector assigned to conduct ACE for the segment. The data collector systematically canvasses the segment, comparing the addresses on the laptop to those seen on the ground, and adding the addresses not found on the laptop. The added addresses are represented in Figure 2 by the yellow circles.

Figure 2: Hypothetical segment with address status.

The complete list of addresses (those originally on the laptop and those added to it by the ACE procedure) is sent to the home office where the added addresses are checked for accuracy and against the larger unenhanced frame. The addresses not already on the frame and deemed accurate are added to the now enhanced frame. The addresses added to the frame become eligible for sampling. Figure 3 represents the segment following enhancement with sampled addresses denoted by and .

Figure 3: Hypothetical enhanced segment showing sampled addresses.

3. The PATH Study

We used the Population Assessment of Tobacco and Health (PATH) Study to assess the impact of frame enhancement, and in particular ACE, on study estimates. The PATH Study's primary goal is to collect information on tobacco-use patterns, risk perceptions, and attitudes toward current and newly emerging tobacco products. Therefore, the study produces many estimates related to tobacco use. However, a wide array of subjects are included, ranging from other substance use (e.g., alcohol, marijuana, and opioids) to self-reported physical and mental health, economic indicators, and technology use. The variation of subject matter provides the opportunity to make a reasonable assessment of the general impact that frame enhancement has on survey estimates.

The PATH Study is an in-person longitudinal study that completed its first wave of data collection in 2014 and has had subsequent waves of data collection each year following. In this paper, we examine select Wave 1 estimates for adults in the United States civilian noninstitutionalized population. The Wave 1 PATH Study sample was selected using a four-stage, stratified probability sample design. The first stage of the design involved the selection of 156 stratified primary sampling units (PSUs) consisting of counties or groups of contiguous counties. Selection of 6,049 second-stage units, or segments, from within the selected PSUs followed. See the PATH Study Restricted Use Files User Guide (United States Department of Health and Human Services, 2018) for additional details. An enhanced ABS frame was used in all but four counties to create the frame for address selection. The enhancement procedure, ACE, was performed on a subsample of 590 PATH Study segments resulting in over 21,000 eligible addresses added to the sampling frame. We call the addresses added to the frame by enhancement "ACE addresses," and refer to the addresses on the frame prior to enhancement as "ABS addresses." The address sampling frame for the segments in the counties that did not use an enhanced ABS frame used a field-listed frame. A total of 32,320 adults responded to Wave 1 of the study, including 936 (2.9 percent) from ACE addresses.

4. Approach

As explained by Brick (2013), one representation of bias partitions the population into respondent and nonrespondent strata so that nonresponse bias is a function of the nonresponse rates and characteristics of the units in the strata. The well-established result follows that the bias of an unadjusted estimator, $\hat{\bar{y}}_{un}$, is given by

$$bias(\hat{\bar{y}}_{un}) \approx NR * (\bar{Y}_r - \bar{Y}_m), \tag{1}$$

where *NR* is the proportion of units in the nonrespondent stratum, $\bar{Y}_r$ is the mean of the respondent stratum, and $\bar{Y}_m$ is the mean of the of the nonrespondent stratum. Given the strata defined by response, the problem with Expression (1) is that the mean of *Y* for nonrespondents is generally unknown.

When putting the expression in terms of coverage enhancement, or ACE, we can consider units from addresses added to the frame by ACE as an enhancement stratum, and units from ABS addresses as an unenhanced stratum, so that the approximate bias of an unenhanced estimator, $\hat{\bar{y}}_{UNH}$, can be expressed as

$$bias(\hat{\bar{y}}_{UNH}) \approx NC * (\bar{Y}_{ABS} - \bar{Y}_{ACE}), \tag{2}$$

where *NC* is the proportion of units in the enhancement stratum (i.e., not-covered units), $\bar{Y}_{ABS}$ is the mean of the unenhanced stratum, and $\bar{Y}_{ACE}$ is the mean of the enhancement stratum. The units in the enhancement stratum are those not covered by the ABS frame. Notice that all components of Expression (2) can be directly estimated post-data collection.

Table 1 presents the expected bias of an unenhanced estimate assuming varying values of *NC* and $\bar{Y}_{ABS} - \bar{Y}_{ACE}$. As shown, the bias of an unenhanced estimator is arguably negligible (i.e., half a percentage point or less), even when the difference between the unenhanced mean and enhanced mean is as high as 10 percent and the proportion of non-covered units is 0.05.

Table 1: Approximate Bias Given Size of Difference Between Means and Proportion of Non-Covered Units

			NC	
$(\bar{Y}_{ABS} - \bar{Y}_{ACE})$	0.03	0.05	0.10	0.20
			$bias(\hat{\bar{y}}_{UNH})$	
2%	0.06%	0.1%	0.2%	0.4%
10%	0.3%	0.5%	1.0%	2.0%
30%	0.9%	1.5%	3.0%	6.0%
50%	1.5%	2.5%	5.0%	10.0%

Next, we estimated the difference between $\bar{Y}_{ABS}$ and $\bar{Y}_{ACE}$ for various PATH Study estimates. Large differences reinforce the need for frame enhancement, while small differences bring into question its necessity. We then assessed the impact of coverage enhancement on PATH Study estimates by comparing estimates from the enhanced frame (i.e., estimates using ABS addresses and ACE addresses) to estimates assuming the study's address frame was not enhanced (i.e., estimates using only ABS addresses). We created estimates for an unenhanced frame by dropping ACE addresses from the sample and re-applying the PATH Study weighting procedures.

5. Results

This section presents preliminary results of our ongoing research. Table 2 shows differences between PATH Study estimates from ABS address respondents (denoted by $\hat{y}_{ABS}$) and estimates from ACE address respondents (denoted by $\hat{y}_{ACE}$), as well as ratios of those estimates ($\hat{y}_{ACE}/\hat{y}_{ABS}$).[1] Some ABS address estimates are very similar to their ACE address counterparts as shown by the small differences and ratios close to 1. For example, close to the same percentage of ABS respondents and ACE respondents indicated they visited a medical doctor in the last 12 months. Also, the estimate of ever use of heroin does not depend on whether a respondent was from the original ABS frame or added through frame enhancement. However, statistically significant differences exist between estimates for some variables. For example, the smokeless tobacco use rate for ACE respondents is almost 2.5 times that for ABS respondents. Additionally, daily internet use for ACE respondents is just over 15 percentage points lower than that for ABS respondents. Therefore, it appears that frame enhancement brings in some respondents who are different from ABS address respondents, suggesting some bias reduction is possible as a result of ACE.

Table 2: Differences and Ratios of Study Estimates by Address Type

Wave 1 Variable Description	$\hat{y}_{ACE} - \hat{y}_{ABS}$	$\hat{y}_{ACE}/\hat{y}_{ABS}$
Current Established Cigarette User	2.00	1.11
Current Established E-Cigarette User	-0.16	0.93
Current Established Smokeless Tobacco User	3.60*	2.41†
Ever Cigarette User	4.33	1.06
Ever E-Cigarette User	1.25	1.07
Ever Smokeless Tobacco User	6.96*	1.46†

[1] The estimates $\hat{y}_{ACE} - \hat{y}_{ABS}$ and $\hat{y}_{ACE}/\hat{y}_{ABS}$ are rounded. For example, when $\hat{y}_{ACE}/\hat{y}_{ABS} = 1$ you might expect $\hat{y}_{ACE} - \hat{y}_{ABS} = 0$ but these relationships are not exact due to the rounding.

Table 2: Differences and Ratios of Study Estimates by Address Type (continued)

Wave 1 Variable Description	$\hat{y}_{ACE} - \hat{y}_{ABS}$	$\hat{y}_{ACE}/\hat{y}_{ABS}$
Annual Household Income Less Than $50,000	8.54*	1.15†
Has Health Insurance	1.54	1.02
Had Visit with a Medical Doctor in Last 12 Months	0.06	1.00
Ever Told Have Cancer by Doctor	1.76	1.25
Self-perceived Mental Health At Least Good	-1.14	0.99
Ever Used Alcohol	-3.52	0.96
Ever Used Heroin	-0.16	0.97
Watches At Most 2 Hours of TV per Day	-5.67*	0.90†
Has Cell Phone for Personal Use	-3.63	0.96
Uses Internet At Least Once per Day	-15.04*	0.78†

* Indicates a statistically significant difference from 0 at the 0.05 alpha-level

† Indicates a statistically significant difference from 1 at the 0.05 alpha-level

Now we discuss direct evidence of bias that would result if the ABS frame were not enhanced. Under-coverage bias is estimated by the difference between the estimates based on the enhanced and unenhanced frames. Despite there being statistically significant differences between some ABS address estimates and ACE address estimates, the impact of frame enhancement on the PATH Study estimates examined appears to be minimal for most of the variables in Table 3.[2] The difference between the enhanced estimates (denoted by $\hat{y}_{ENH}$) and the unenhanced estimates (denoted by $\hat{y}_{UNH}$) is as small as zero for some cases (e.g., current established cigarette use and self-perceived mental health). For daily internet use, the variable with the largest difference between the ABS estimate and the ACE estimate (15 percentage points), the difference between the enhanced estimate and unenhanced estimate is 0.32 percentage points. This difference is statistically significant ($p < 0.0001$). Even though the ratio of daily internet use estimates ($\hat{y}_{ENH}/\hat{y}_{UNH}$) is statistically different from 1 at the 0.05 alpha-level, the value of the ratio to two significant digits (1.00) brings into question the practical significance of that difference. The differences between the enhanced and unenhanced estimates for current established smokeless tobacco use and ever smokeless tobacco use are also statistically significant at the 0.05 alpha-level. The ratio of the current established smokeless tobacco use estimates (0.96) exhibits the largest departure from 1 of any of the ratios examined ($p < 0.0001$). The generally small differences between the estimates are due to ACE addresses contributing only 2.9 percent (3.2 percent weighted) of Wave 1 adult respondents (see Table 1).

Table 3: Differences and Ratios of Study Estimates by Frame Type

Wave 1 Variable Description	$\hat{y}_{ENH} - \hat{y}_{UNH}$	$\hat{y}_{ENH}/\hat{y}_{UNH}$
Current Established Cigarette User	0.00	1.00
Current Established E-Cigarette User	0.02	1.01
Current Established Smokeless Tobacco User	-0.10*	0.96†
Ever Cigarette User	-0.01	1.00
Ever E-Cigarette User	-0.02	1.00
Ever Smokeless Tobacco User	-0.20*	0.99†

[2] The estimates $\hat{y}_{ENH} - \hat{y}_{UNH}$ and $\hat{y}_{ENH}/\hat{y}_{UNH}$ are rounded. For example, when $\hat{y}_{ENH}/\hat{y}_{UNH} = 1$ you might expect $\hat{y}_{ENH} - \hat{y}_{UNH} = 0$ but these relationships are not exact due to the rounding.

Table 3: Differences and Ratios of Study Estimates by Frame Type (continued)

Wave 1 Variable Description	$\hat{y}_{ENH} - \hat{y}_{UNH}$	$\hat{y}_{ENH}/\hat{y}_{UNH}$
Annual Household Income Less Than $50,000	-0.15	1.00
Has Health Insurance	-0.03	1.00
Had Visit with a Medical Doctor in Last 12 Months	0.02	1.00
Ever Told Have Cancer by Doctor	-0.02	1.00
Self-perceived Mental Health At Least Good	0.00	1.00
Ever Used Alcohol	0.13	1.00
Ever Used Heroin	0.02	1.00
Watches At Most 2 Hours of TV per Day	0.04	1.00
Has Cell Phone for Personal Use	0.05	1.00
Uses Internet At Least Once per Day	0.32*	1.00†

* Indicates a statistically significant difference from 0 at the 0.05 alpha-level

† Indicates a statistically significant difference from 1 at the 0.05 alpha-level

6. Discussion

Many in-person surveys have turned to ABS frames as a cheaper alternative to frames created via traditional listing. While ABS frames offer high national coverage overall in the United States, procedures have been developed to enhance ABS frames. Westat utilized ACE to enhance the PATH Study ABS frame. For some variables, statistically significant differences exist between estimates resulting from ABS addresses and estimates from addresses that were added to the frame through enhancement. However, the impact of frame enhancement on overall study estimates appears minimal. This result is due to the small amount that frame enhancement contributes to the overall sample.

Further investigation is required to determine the complete effects of frame enhancement on study estimates. Studies concentrated in highly urban areas might not benefit from frame enhancement enough to make the cost associated with the procedures worthwhile. However, if studies are interested in estimates in rural areas where ABS frame enhancement is expected to have a more significant impact, then frame enhancement might be necessary to produce unbiased estimates. The results in this paper were restricted to characteristics of adult respondents. Frame enhancement might have a more significant effect on a younger population, and that should be investigated as well.

Acknowledgements

This project is supported with Federal funds from the National Institute on Drug Abuse, National Institutes of Health, and the Center for Tobacco Products, Food and Drug Administration, Department of Health and Human Services, under contract to Westat (Contract No. HHSN271201100027C).

References

AAPOR Address-based Sampling Task Force (2016). "AAPOR Report: Address-based Sampling." Prepared for AAPOR Council by the Task Force on Address-based Sampling, operating under the auspices of the AAPOR Standards Committee. Deerfield, IL. (https://www.aapor.org/Education-Resources/Reports/Address-based-Sampling.aspx)

Blumberg, S.J., and J.V. Luke (2017). "Wireless substitution: Early release of estimates from the National Health Interview Survey, July – December 2016." Retrieved September 10, 2018. (https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201705.pdf)

Brick, J.M. (2013). "Unit Nonresponse and Weighting Adjustments: A Critical Review." *Journal of Official Statistics*, Vol. 29, No. 3, 329–353.

Brick, J.M., D. Williams (2013). "Explaining Rising Nonresponse Rates in Cross-Sectional Surveys." *The ANNALS of the American Academy of Political and Social Science*, Vol. 645, Issue 1, 36–59.

Dohrmann, S., D. Han, and L. Mohadjer (2006). "Residential Address Lists vs. Traditional Listing: Enumerating Households and Group Quarters." *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 2959–2964.

Dohrmann, S., G. Kalton, J. Montaquila, C. Good, and M. Berlin (2012). "Using Address Based Sampling Frames in Lieu of Traditional Listing: A New Approach." *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 3729–3741.

Dohrmann, S., T.D. Buskirk, A. Hyon, and J. Montaquila (2014). "Address Based Sampling Frames for Beginners." *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 1009–1018.

Harter, R., and N. English (2018). "Overview of Three Field Methods for Improving Coverage of Address-based Samples for In-person Interviews." *Journal Survey Statistics and Methodology*, 6, 360–375.

Iannacchione, V.G. (2011). "Research Synthesis: The Changing Role of Address-Based Sampling in Survey Research." *Public Opinion Quarterly*, 75, 556–575.

Kali, J., R. Sigman, W. Ren, and M. Jones (2014). "Experiences with the Use of Addressed Based Sampling in In-Person National Household Surveys." *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 3050–3059.

Kalton, G., J. Kali, and R. Sigman (2014). "Handling Frame Problems When Addressed Based Sampling is Used for In-person Household Surveys." *Journal of Survey Statistics and Methodology*, 2, 283–304.

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

Montaquila, J.M., J.M. Brick, D. Williams, K. Kim, and D. Han (2013). "A Study of Two-Phase Mail Survey Data Collection Methods." *Journal of Survey Statistics and Methodology*, Vol. 1, 66–87.

United States Department of Health and Human Services. National Institutes of Health. National Institute on Drug Abuse, and United States Department of Health and Human Services. Food and Drug Administration. Center for Tobacco Products. Population Assessment of Tobacco and Health (PATH) Study [United States] Restricted-Use Files, User Guide. ICPSR36231-v14 Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2018-03-15. http://doi.org/10.3886/ICPSR36231.userguide