

# Improving the Efficiency of Address-Based Frames with the USPS No-Stat File

Bonnie E. Shook-Sa\*

Keywords: usps no-stat file, address lists, hybrid frame designs, address-based sampling

DOI: [10.29115/SP-2014-0018](https://doi.org/10.29115/SP-2014-0018)

---

## Survey Practice

Vol. 7, Issue 4, 2014

---

Address-based sampling (ABS) frames are based on the Computerized Delivery Sequence (CDS) file, made available to qualified vendors by USPS. Because of low rural coverage for in-person surveys, researchers often resort to hybrid sampling frames that utilize ABS in areas where coverage is expected to be high and traditional field enumeration in areas where coverage is expected to be low. The more areas that are allocated to the ABS portion of the hybrid frame, the greater the cost savings. Since 2009, USPS has made available the No-Stat file, a supplement to the CDS file that contains approximately seven million predominately rural addresses not found on the CDS file. Although the overall coverage gains provided by the No-Stat file are modest, No-Stat addresses are clustered in relatively small geographic areas. In a hybrid frame design, clusters of No-Stat addresses could move some areas that otherwise would rely on field enumeration to the ABS portion of the frame, lowering field costs. This paper measures the efficiencies that are gained by including the No-Stat file for a specified hybrid frame design and coverage estimator. Efficiency gains vary widely across states, and are greater for higher coverage thresholds.

---

## INTRODUCTION

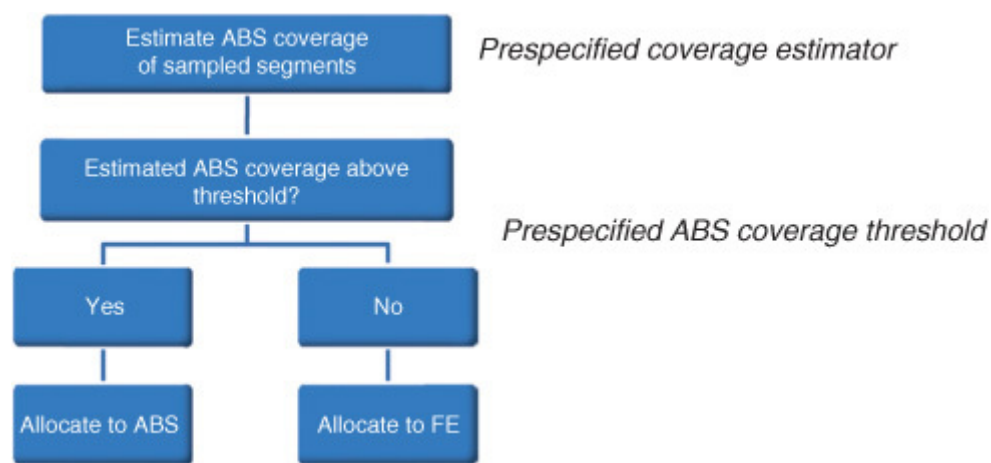
Address-based sampling (ABS) is commonly used either as an alternative to or in conjunction with traditional field enumeration (FE) methods for in-person surveys (Iannacchione 2011). FE frames are constructed by canvassing selected areas and enumerating potential housing units (HUs). ABS frames are derived from the U.S. Postal Service's (USPS) computerized delivery sequence (CDS) file, which is made available to select vendors through nonexclusive licensing agreements with USPS. ABS frames based on the CDS file provide both time and cost savings over traditional FE, because ABS frames do not require field staff to visit area segments in advance of sample selection.

Despite the time and cost savings of ABS designs, concerns remain about the household coverage of ABS frames for in-person surveys, particularly in rural areas (Dohrmann, Han, and Mohadjer 2007; Iannacchione et al. 2007; O'Muircheartaigh, English, and Eckman 2007). One approach to ameliorate low rural ABS coverage is to develop a hybrid sampling frame, where areas that are

---

expected to have adequate ABS coverage are allocated to the ABS portion of the frame, and areas with poor ABS coverage are allocated to FE (Lepkowski et al. 2010; Iannacchione et al. 2012).

Figure 1 depicts the implementation of a hybrid frame design. First, a sample of area segments is selected. Area segments are geographic areas that are typically formed based on Census geography (e.g., census block groups or collections of census blocks). When ABS is used as the primary frame source, segments can be larger than with traditional FE designs, because ABS does not require enumeration of HUs. With a hybrid frame approach, larger segments can be used, and those that are allocated to FE would be subdivided into smaller geographic areas where FE is feasible.



**Figure 1** Hybrid sampling frame design. ABS=address-based sampling, FE=Field enumeration.

The ABS coverage in each selected segment is estimated using a prespecified coverage estimation method. Various techniques exist for predicting ABS coverage. Montaquila et al (2011) propose utilizing a multiple regression model to predict areas where ABS coverage is expected to be sufficient. Iannacchione et al. (2012) evaluate a coverage model that estimates coverage by taking the ratio of ABS addresses in the segment to the estimated number of dwelling units.

A prespecified ABS coverage threshold is also established. Segments that exceed the ABS coverage threshold are allocated to the ABS portion of the frame, while segments where ABS coverage is not expected to achieve the threshold are allocated to FE.<sup>1</sup> The selection of an ABS coverage threshold is typically study-specific and depends on such factors as whether or not additional frame supplementation will occur in ABS segments (see Kalton et al. 2014 and

<sup>1</sup> Several variations of hybrid sampling frame designs exist. For example, an alternative enumeration method, such as enhanced listing, could be used in segments that do not meet the coverage threshold. With enhanced listing, field personnel update the ABS list prior to sample selection (Harter et al. 2010).

McMichael et al. 2008) and the budget of the study. The prespecified ABS coverage threshold is a minimum bound for allocating segments to ABS and is not an estimate of the expected coverage of the ABS portion of the frame.

Hybrid frame designs provide the cost savings of ABS where possible but retain FE where necessary to maintain adequate household coverage. In general, the more segments that are allocated to the ABS portion of the frame, the higher the cost savings. The highest cost savings are realized for lower ABS coverage thresholds, where more segments meet the prespecified ABS coverage threshold criteria and are allocated to ABS. However, there is a trade-off between cost and coverage, because unless frame supplementation procedures are utilized in ABS segments the expected coverage of the frame decreases as the ABS coverage threshold decreases.

The more complete the ABS frame, the more segments that are allocated to ABS and the greater the cost savings. Until recently, ABS frames for in-person surveys were limited to locatable addresses from the CDS file.<sup>2</sup> Starting in 2009, the USPS made available the No-Stat file, a supplemental file that contains approximately seven million locatable residential addresses. The following four types of locatable addresses are included on the No-Stat file:<sup>3</sup>

1. New Growth:<sup>4</sup> addresses of residences under construction that are not yet receiving mail
2. Rural Throwback: addresses for HUs on rural postal routes whose residents specify that their mail be delivered to a P.O. Box rather than to their residence
3. Internal Drop: locatable addresses with unit designators for a small proportion of the drop point addresses contained on the CDS file
4. Rural Vacant: addresses on rural postal routes that have been classified as vacant for 90 days or longer

There is very little overlap between the CDS and No-Stat files, so the two files can be combined into a single, No-Stat supplemented ABS frame.<sup>5</sup> Shook-Sa et al. (2013) found that approximately 21 percent of No-Stat locatable addresses are associated with occupied HUs (i.e., households). While this is much lower than the occupancy rate for HUs associated with CDS addresses (90 percent),

---

<sup>2</sup> Locatable mailing addresses include street numbers and names along with cities, states, and zip codes. Addresses that are not locatable cannot be linked to physical HUs on the ground and are typically excluded from the sampling frames for in-person surveys (e.g., P. O. Boxes, Rural/Highway Contract Routes, and Simplified Addresses).

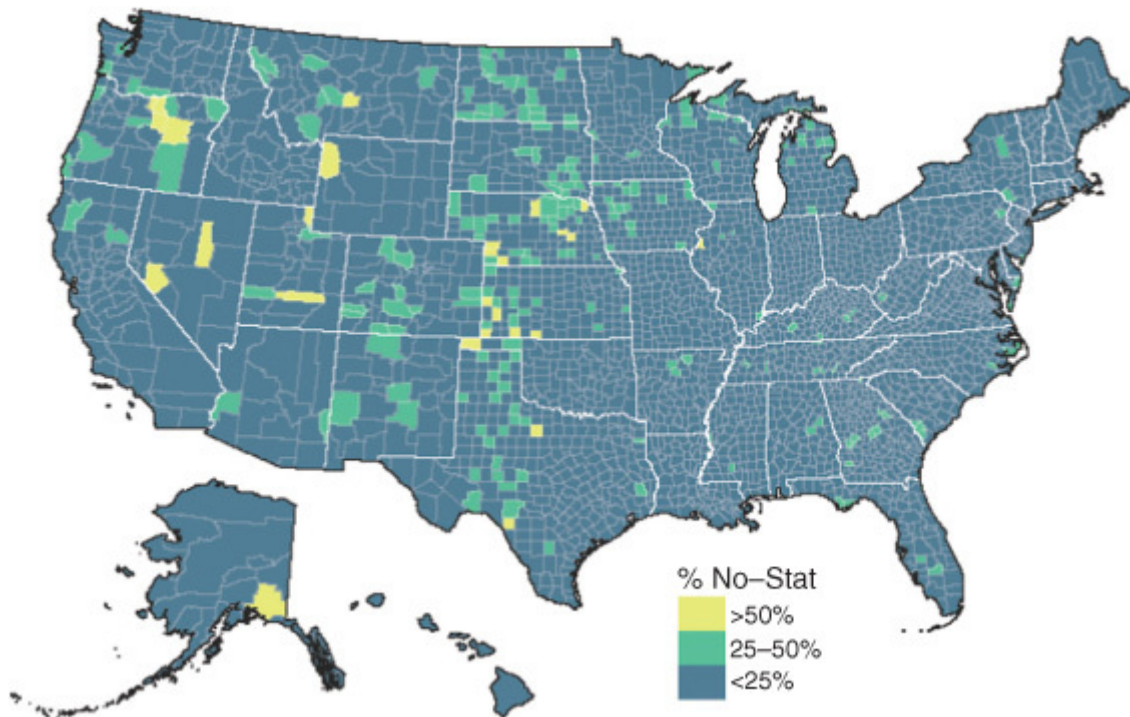
<sup>3</sup> No-Stat addresses can be classified into these four categories using the PO Box Throwback Indicator, the Delivery Point Type Code, and the CDS No-Stat New Growth Indicator as outlined in the CDS User Guide (United States Postal Service 2013).

<sup>4</sup> All addresses on the No-Stat file are classified based on their new growth status. For simplification, our classifications put all new growth addresses into a single category and then split out the non-new growth addresses into the three remaining categories.

<sup>5</sup> When combining the CDS and No-Stat files, No-Stat internal drop addresses should be excluded because they are already contained on the CDS file as drop points.

the No-Stat file contains the addresses for approximately 1.3 million households that are excluded from the CDS file. It provides an estimated 1.2 percent increase in coverage for households nationally and an estimated 3.8 percent increase in rural areas.

While the national coverage gains provided by the No-Stat file are modest, localized coverage gains could lead to significant improvements in the efficiency of hybrid frame designs. The map in Figure 2 displays the percentage of locatable addresses on the combined CDS/No-Stat file that come from the No-Stat file, at the county level. Because No-Stat addresses are clustered in primarily rural areas, the No-Stat file has the potential to move area segments that would otherwise rely on FE to the ABS portion of a hybrid frame. If this occurs in enough segments, the No-Stat file could provide sizeable cost savings in a hybrid design. This paper measures the national and state-level efficiencies that are gained by including the No-Stat file in a hybrid frame design.



**Figure 2** Percentage of locatable addresses from No-Stat file by county.

## METHODS

To evaluate the level of efficiencies resulting from the inclusion of the No-Stat file in a hybrid frame design, we compared two potential sampling frames: the CDS-only frame (the traditional ABS frame) and the combined CDS/No-Stat frame. We constructed a national list of 215,547 area segments, formed from census block groups or collapsed census block groups.<sup>6</sup> We obtained the March

<sup>6</sup> Less than one percent of census block groups had to be collapsed when forming area segments because they did not meet the minimum size

2013 CDS and No-Stat files that had been geocoded into census block groups and could therefore be linked to the area segments.<sup>7</sup> Within each area segment  $s$ , we calculated the number of locatable CDS addresses ( $L_{s,CDS}$ ) and the number of locatable addresses on the combined CDS/No-Stat frame ( $L_{s,CDS+}$ ).<sup>8</sup>

When allocating segments to ABS and FE, we are primarily concerned with the level of household coverage within sampled area segments. However, not every address on the CDS file corresponds with a household (e.g., the CDS contains addresses for vacant or demolished HUs). For this reason, we form our coverage model based on HU coverage and use this as a proxy for household coverage. We estimated the ABS coverage in each area segment for each frame (CDS-only and CDS/No-Stat) using the prespecified model defined below:

$$C_{s,f} = \frac{L_{s,f}}{HU_s}$$

Where  $C_{s,f}$  is the expected ABS coverage of segment  $s$  based on sampling frame  $f$ ,  $L_{s,f}$  is the number of locatable addresses on sampling frame  $f$  in segment  $s$  (as defined above), and  $HU_s$  is the estimated number of HUs in segment  $s$  (from the 2010 Census).

While the estimated ABS coverage for a given area segment and sampling frame is fixed, the allocation of that segment to ABS versus FE is dependent on the coverage threshold for a given study. As discussed previously, the choice of the ABS coverage threshold is often study specific. For this reason, we compared the allocation of segments to ABS and FE under a hybrid sampling design for various potential coverage thresholds. We assumed a national probability proportional to size (PPS) design where the size measure was the adult population from the 2010 Census. PPS sampling allows the probability of selecting each area segment to be proportional to its size measure. This ensures that persons have similar probabilities of selection, regardless of the size of the segments in which they reside.

We made national and state-level estimates for the differences in allocations between the two sampling frames by examining the expected distribution of segments across ABS and FE under this design.<sup>9</sup> The difference between the CDS-only and the CDS/No-Stat percentage of segments allocated to FE provides an estimate of the efficiencies gained by including the No-Stat file on the ABS frame. For a given ABS coverage threshold  $t$ , the percentage of segments moved from FE to the ABS portion of the hybrid frame by the

---

criteria of 100 HUs (at least one of which was occupied). HU estimates were derived from the 2010 Census.

<sup>7</sup> The CDS and No-Stat files were obtained from Compact Information Systems.

<sup>8</sup> Locatable addresses from the No-Stat file (excluding internal drops) were merged onto the CDS file. To avoid multiplicities on the combined frame, addresses appearing on both files were removed from the No-Stat file.

<sup>9</sup> Under the specified PPS design, the expected distribution is the weighted distribution of the list of area segments, where the weight is the size measure.

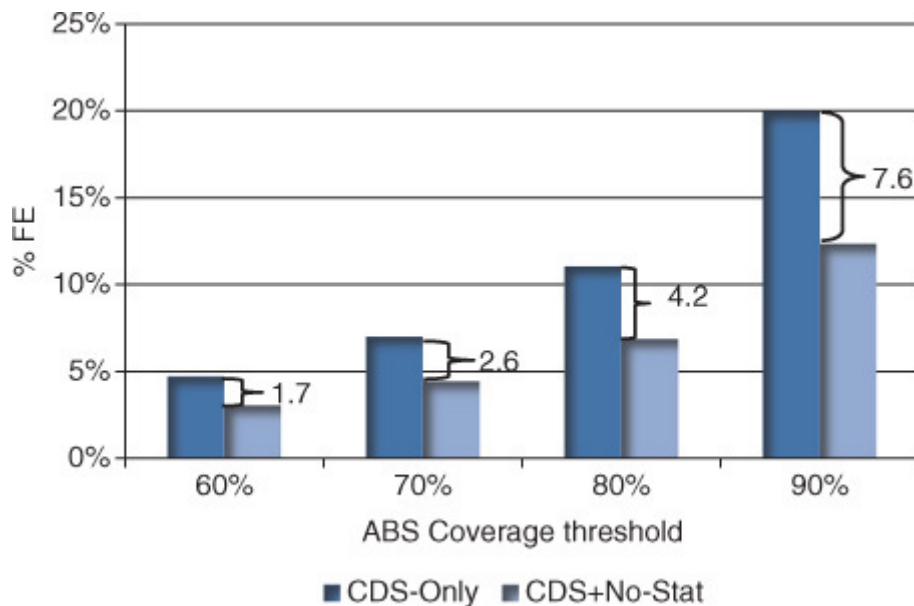
No-Stat file ( $P_{NS_t}$ ) is equal to the percentage of segments allocated to FE for the CDS-only frame ( $P_{CDS_t}$ ) minus the percentage of segments allocated to FE for the combined CDS/No-Stat frame ( $P_{CDS+_t}$ ). That is,

$$P_{NS_t} = P_{CDS_t} - P_{CDS+_t}$$

We estimate  $P_{NS_t}$  at the national and state levels for various potential ABS coverage thresholds that could be used in a hybrid frame design.

## RESULTS

Figure 3 compares the percentage of segments allocated to FE for the CDS-only frame ( $P_{CDS_t}$ ) to the CDS/No-Stat frame ( $P_{CDS+_t}$ ) for four different potential ABS coverage thresholds  $t$  at the national level. The differences between the bars represent the total percentage of segments that the No-Stat file would move from FE to ABS ( $P_{NS_t}$ ) for each potential ABS threshold. For all potential ABS coverage thresholds evaluated, the No-Stat file provided efficiency gains by boosting the expected coverage of segments enough that a sizeable number of segments moved from FE to ABS. Efficiency gains were larger for the higher coverage thresholds. For example, if a 90 percent coverage threshold were used to allocate segments between ABS and FE in the hybrid frame design, an additional 7.6 percent of segments could rely on ABS rather than FE when the No-Stat file is included, compared to a 1.7 percent gain if a 60 percent coverage threshold were used in this allocation.



**Figure 3** Percentage of segments allocated to FE by ABS coverage threshold.

Because No-Stat addresses are geographically clustered and would not be expected to achieve uniform coverage gains across the country, we also compared the efficiency gains resulting from No-Stat inclusion at the state



**Table 1** Percentage of segments No-Stat file moved to ABS portion of hybrid frame by state and potential ABS coverage threshold.<sup>1</sup>

State	ABS coverage threshold				State	ABS coverage threshold			
	60%	70%	80%	90%		60%	70%	80%	90%
IA	8.3	12.3	17.0	20.1	NH	3.2	5.0	5.3	8.1
SC	2.3	4.0	7.8	18.3	OK	3.4	4.1	6.1	7.7
AL	1.3	3.3	8.2	17.9	AZ	1.3	1.6	3.3	7.4
AR	4.7	7.7	12.4	17.8	NM	2.2	3.5	4.9	7.4
MS	2.2	3.9	6.9	17.7	IN	1.5	2.8	4.4	7.3
NC	2.5	4.6	8.5	16.0	PA	1.5	2.2	3.5	6.2
OR	4.7	6.7	9.8	15.8	NV	0.4	0.8	2.9	6.1
VT	8.6	13.4	14.5	15.8	LA	1.8	2.2	3.4	6.0
GA	1.3	2.7	5.5	13.9	OH	0.7	1.4	2.7	5.9
DE	1.4	3.4	5.8	12.9	IL	1.4	2.4	3.6	5.8
KY	1.8	3.5	7.0	12.4	TX	1.6	2.2	3.6	5.8
ME	5.7	9.6	10.4	12.0	MD	0.8	1.4	2.5	5.0
TN	0.6	1.4	3.9	11.8	CO	1.6	2.1	2.8	5.0
NE	8.1	10.3	11.9	11.5	NY	1.3	1.9	2.9	4.7
MT	4.1	4.6	7.9	10.0	UT	1.1	1.1	2.3	4.4
SD	7.6	7.2	9.6	9.6	RI	0.1	0.6	1.8	4.1
WA	2.7	3.1	5.1	9.5	WY	4.9	3.6	6.3	4.1
MI	2.6	3.8	5.5	9.3	CT	0.7	0.9	1.3	4.1
MN	3.7	5.1	7.1	9.2	MA	0.9	0.9	2.1	3.9
WI	2.7	4.4	6.4	9.1	HI	2.1	2.3	2.3	3.6
MO	1.2	2.4	4.6	8.8	ID	1.8	2.7	2.3	3.0
FL	0.8	1.7	3.4	8.7	NJ	0.6	0.8	1.3	2.8
VA	1.3	2.4	4.5	8.5	AK	2.9	4.5	4.1	2.3
WV	3.0	4.2	5.5	8.3	CA	0.4	0.6	0.9	1.9
ND	10.8	11.7	9.5	8.3	DC	0.1	0.1	0.0	0.7
KS	3.4	5.2	7.0	8.3					

<sup>1</sup>States are sorted by the 90 percent coverage threshold. The maximum value for each coverage threshold is bolded.

## DISCUSSION

For the specified ABS coverage estimation model, the No-Stat file improves the efficiency of hybrid frames for in-person surveys by moving segments that would otherwise rely on FE to the ABS portion of the frame. Under this model, efficiency gains are greater for higher coverage thresholds and vary widely across states. These results are consistent with an evaluation by Iannacchione et al. (2012), who found that the No-Stat file would result in between 2 and 10 percent of segments moving from FE to ABS with the inclusion of the No-Stat file at the national level, depending on the ABS coverage threshold specified.

The actual efficiencies gained from including the No-Stat file in a hybrid design are dependent on the coverage prediction model, ABS coverage threshold, and sample design for a given study so care should be taken when applying these results to alternative designs. In addition, these results are based on the

allocation of segments to FE and ABS under the specified coverage prediction model at the sample design stage, and the accuracy of the assignments has not been validated in the field. This coverage model assumes that CDS and No-Stat addresses correspond with HUs in a one-to-one manner. In a small field study, Dekker et al. (2014) found a sizable proportion of No-Stat addresses to be associated with vacant lots. Prior No-Stat findings indicate that approximately 21 percent of No-Stat addresses are associated with occupied households (Shook-Sa et al. 2013), but if a sizeable proportion of the remaining No-Stat addresses do not correspond with vacant HUs, then the specified coverage model could lead to inaccurate coverage estimates for some segments. Therefore, as with any coverage prediction model, following implementation of the hybrid frame allocation, the assignment of segments to FE and ABS should be validated with field data to ensure that the model leads to appropriate segment assignments and to assist in refining the coverage prediction model, as needed.

Despite these limitations, based on these and prior findings, the No-Stat file appears to be a useful source for supplementing the CDS file for in-person ABS designs. It provides modest household coverage gains, particularly in rural areas, and could lead to substantial cost savings in a hybrid frame design.

#### **ACKNOWLEDGEMENTS**

The author would like to acknowledge G. Lance Couzens for developing the maps in this paper as well as Vincent Iannacchione, Rachel Harter, Pat Wiley, and the editors of *Survey Practice* for their valuable contributions to this research.

## REFERENCES

- Dekker, K., N. English, and C. O’Muircheartaigh. 2014. “Examining the Threshold: Experiences in Evaluating the DSF When Listing May or May Not Be Necessary.” Presented at the Joint Statistical Meetings, Boston, MA.
- Dohrmann, S., D. Han, and L. Mohadjer. 2007. “Improving Coverage of Residential Address Lists in Multistage Area Samples.” In *Proceedings of the American Statistical Association, Survey Research Methods Section*.
- Harter, R., S. Eckman, N. English, and C. O’Muircheartaigh. 2010. “Applied Sampling for Large-Scale Multi-Stage Area Probability Designs.” In *Handbook of Survey Research*, edited by P. Marsden and J. Wright. Emerald, Bingley, UK.
- Iannacchione, V. 2011. “The Changing Role of Address-Based Sampling in Survey Research.” *Public Opinion Quarterly* 75: 556–75.
- Iannacchione, V., J. McMichael, B. Shook-Sa, and K. Morton. 2012. “A Proposed Hybrid Sampling Frame for the National Survey on Drug Use and Health.” Prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, RTI/0209009. <http://www.samhsa.gov/data/NSDUH/NSDUHMethodsRptHybrid2012.pdf>.
- 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 *Proceedings of the American Statistical Association, Survey Research Methods Section*.
- Kalton, G., J. Kali, and R. Sigman. 2014. “Handling Frame Problems When Address-Based Sampling Is Used for in-Person Household Surveys.” *Journal of Survey Statistics and Methodology* 2: 283–304.
- Lepkowski, J., W. Mosher, K. Davis, R. Groves, and J. Van Hoewyk. 2010. “The 2006-2010 National Survey of Family Growth: Sample Design and Analysis of a Continuous Survey.” *National Center for Health Statistics Vital and Health Statistics* 2 (150): 11–12.
- McMichael, J., J. Ridenhour, and B. Shook-Sa. 2008. “A Robust Procedure to Supplement the Coverage of Address-Based Sampling Frames for Household Surveys.” In *Proceedings of the American Statistical Association, Survey Research Methods Section*.
- Montaquila, J., V. Hsu, and J.M. Brick. 2011. “Using a ‘Match Rate’ Model to Predict Areas Where USPS-Based Address Lists May Be Used in Place of Traditional Listing.” *Public Opinion Quarterly* 75: 317–35.
- O’Muircheartaigh, C., E. English, and S. Eckman. 2007. “Predicting the Relative Quality of Alternative Sampling Frames.” In *Proceedings of the American Statistical Association, Survey Research Methods Section*.
- Shook-Sa, B., D. Currivan, J. McMichael, and V. Iannacchione. 2013. “Extending the Coverage of Address-Based Sampling Frames: Beyond the USPS Computerized Delivery Sequence File.” *Public Opinion Quarterly* 77: 994–1005.
- United States Postal Service. 2013. “CDS User Guide.” [http://ribbs.usps.gov/cds/documents/tech\\_guides/CDS\\_USER\\_GUIDE.PDF](http://ribbs.usps.gov/cds/documents/tech_guides/CDS_USER_GUIDE.PDF).