Using Address Based Sampling Frames in Lieu of Traditional Listing: A New Approach

Sylvia Dohrmann[1], Graham Kalton[1], Jill Montaquila[1], Cindy Good[1], and Martha Berlin[1]
[1]Westat, 1600 Research Boulevard, Rockville, MD 20850

Abstract:
Address lists originating from the Unites States Postal Service (USPS) have appeal as a possible alternative to the costly listing effort that has traditionally been used for multi-stage area household samples with in-person interviewing. A number of evaluations of the USPS lists have reached the consensus that the lists generally have excellent coverage in urban areas. However, these studies have consistently indicated poorer coverage in more rural areas, in areas with a large number of group quarters, and in fast-growing areas.

This paper presents an approach that applies an area linkage method developed specifically to address noncoverage in the USPS lists. The paper describes this approach, discusses alternative approaches considered, and reviews issues encountered during the development and field testing of this new approach. Besides serving as a means to provide sample coverage to residential addresses missing from the USPS lists, the approach also provides measures of the extent of noncoverage of these lists.

Key Words: ABS, area sample, USPS, quality control, coverage enhancement

1. Introduction

Historically, traditional listing has been used to create housing unit sampling frames for in-person household surveys. This listing operation requires that field staff canvass the sampled areas in-person months before data collection begins in order to list the addresses of all the housing units observed in those areas and to allow adequate time for sample selection. The operation can be both costly and time consuming. Furthermore, given the nature of this task, these lists may be incomplete and contain erroneous listings.

Over the past decade, researchers have been investigating commercially available address lists originating from the USPS's Address Management System as an alternative to traditional listing for creating housing unit sampling frames. The Address Management System contains the record of residential and commercial mailing addresses. The lists cannot be purchased directly from the USPS. Rather, a qualifying organization already in possession of a list of residential addresses purchases a license from the USPS. This license gives the organization the ability to confirm that its addresses are correct, obtain the delivery information for those addresses, and supplement its list with all other mailing addresses in the ZIP code. The organizations that purchase licenses have traditionally been vendors who deal with direct mail or other marketing agencies.

The USPS-based lists have some attractive features for sampling housing units for in-person surveys with multi-stage area sample designs. The lists are electronic, easily input into sample selection software, and are updated monthly or bi-monthly. Addresses can be sampled from the lists immediately before data collection, thus avoiding the cost and time needed to physically list the addresses prior to data collection. Sample designs using the postal service lists as housing unit sampling frames have come to be known as Address-Based Sampling (ABS).

Two complicating factors, however, impede the use of the postal lists as a complete replacement for the traditional methods. The primary factor is that the lists do not have full coverage of residential addresses, particularly in rural and certain other types of areas. The second factor is that errors occur in grouping the postal addresses into small geographic areas, in particular the secondary sampling units (SSUs or segments) used to cluster the sampled households within the sampled primary sampling units (PSUs) for efficient data collection.

Section 2 of this paper discusses the issue of noncoverage of the USPS lists when used as sampling frames for multi-stage in-person household surveys. Section 3 discusses the issue of grouping the addresses into segments. Section 4 then presents an approach that applies an area linkage method developed specifically to address these issues, and section 5 presents the results of a field test of this approach. A summary is provided in section 6.

2. Noncoverage

There are several reasons why USPS residential address lists may not provide complete coverage of the households in an area, especially in the context of area sampling with in-person data collection. Since the USPS Address Management System contains only residential addresses along a mail delivery route, addresses of housing units for which mail service is not available are excluded from the list. Residents living in these areas must collect their mail either at a general delivery facility or a Post Office (P.O.) box. Also, city-style addresses may not be provided for housing units along rural routes. All areas in the U.S. are expected to be eventually converted to city-style addresses for purposes of E-911 location[1]; in the interim, unconverted housing units are represented in the ABS frame by only their postal addresses (general delivery or P.O. box), so that the physical units are not locatable from the mailing address alone. If selected for an in-person survey, such housing units could not be located for interviewing, resulting in noncoverage of housing units that have only a P.O. box address or collect all their mail from a general delivery facility.

Coverage issues also exist for noninstitutional group quarters such as dormitories, assisted living facilities, halfway homes, and shelters. The presence of these units on the ABS frame depends on how the residents of the facility receive their mail. Some facilities, such as universities, operate their own “post offices,” and thus the USPS does not have information on the individual mailing addresses of the residents. Other facilities, such as assisted living facilities, halfway homes, and shelters may be operated by a business or charitable organization, and may not appear on an address list restricted to residential addresses (Dohrmann, Han and Mohadjer 2006).

[1] E-911 location refers to the ability of 911 emergency service vehicles to locate physical locations based on street addresses.

Placing the issue of group quarters aside, since not all household surveys include these units in their target population, the nature of the remaining coverage issues are such that the address lists have been found to have the least coverage in rural areas (Dohrmann, Han and Mohadjer 2006, Staab and Iannacchione, 2003, and O'Muircheartaigh, Eckman, and Weiss 2002). Montaquila et al. (2009) found that although coverage rates were generally higher in urban areas, there was variation in coverage rates at the segment level within PSUs, even in very urban PSUs.

Given these coverage limitations, most researchers have proposed that the ABS lists only be used in areas thought to have adequate coverage, relying on traditional listing in other areas (English, Dekker, and O'Muircheartaigh 2011; Montaquila, Hsu and Brick 2011; Iannacchione et al. 2010). This "all or nothing" approach has required first establishing a method of estimating coverage, and then developing a coverage threshold below which traditional methods would be used. The approach accepts coverage error to the extent that the lists do not have complete coverage of areas above the coverage threshold; additionally some areas may be subjected to traditional listing unnecessarily because they are incorrectly classified as falling below the coverage threshold.

Regardless of the extent to which the postal lists are used as housing unit frames, quality control procedures should be considered with ABS frames as with any other area sampling frame. These procedures should be used not only as a means to assess the quality of the sampling frame, but also to enhance the coverage of the frame. There are three levels for implementing quality control, or coverage enhancement, procedures with ABS: in none, all, or some of the areas in which the lists are to be used. Performing no quality control of the lists may be acceptable if all the areas surveyed are expected to have very high coverage, especially if the coverage estimates are reliable (see section 4.2). The most conservative approach is to perform quality control in *all* areas in which the lists are used; however, such an approach is resource-intensive and, as a result, might not be cost-effective.

The middle alternative is to perform coverage enhancement in *some* areas in which the lists are used. With this approach, a key step is the identification of the areas in which coverage enhancement is to be performed. In order to ensure that all addresses missing from the ABS frame are represented in the sample, this identification should be a probabilistic selection, similar to the methods used with traditionally listed samples, with all areas having a chance of being selected for the procedure. However, the coverage enhancement could be confined to areas that fall below a coverage threshold.

Two frame linkage procedures that are commonly used with traditionally listed samples are the half-open interval (HOI) procedure (Kish, 1965) and an approach developed by Waksberg in the early 1970's. When applied to give missed housing units the same chance of sample selection as listed units, the HOI procedure is performed on all sampled addresses across all the sampled segments while the Waksberg approach involves selecting a probability subsample of the segments and then conducting a thorough listing check in the subsample. Then all of the housing units found that were not on the original frame are added to the sample. Variants of both procedures have also been considered for use with ABS listings.

The HOI procedure does not immediately translate to the ABS framework since it relies on a pre-determined path of travel between a sampled unit and the next unit on the housing unit sampling frame. In the traditional listing context, this path of travel is determined by the field staff performing the listing procedure; with ABS, however, a definitive path of travel within the segment is absent unless specifically established by the field staff performing the procedure or by the home office. McMichael, Ridenhour, and Shook-Sa (2008) developed a modification of the HOI procedure for use with postal lists. However, they found that despite much training and home office support, interviewers captured only 48.5 percent of the potentially missing addresses from the sampling frame.

In an effort to present a more simple set of instructions to the field staff performing a type of HOI coverage enhancement procedure, McMichael et al. (2008) next introduced the Check for Housing Units Missed (CHUM) procedure. The simplification consisted of limiting the travel to within the city block of the selected housing unit; the procedure also included an additional check of blocks with no housing units on the ABS frame. This procedure was implemented in the 2008 American National Election Survey with some positive results. However, its use in areas with a low estimated coverage rate was found to be inefficient as the effort required to implement the procedure exceeded that of traditional listing (Iannacchione, 2011).

Dohrmann, Han and Mohadjer (2006) proposed a variant of the Waksberg approach for use with ABS. In this application, segment boundaries were created using U.S. Census Bureau geography (Census blocks or combinations of Census blocks) and the segment was defined as the set of addresses that were assigned via geocoding to fall within the geographic boundaries (see section 3), with the knowledge that not all addresses would be assigned to the correct are segment. An enhancement to the Waksberg approach gave segments with likely higher rates of noncoverage greater chances of selection for the procedure. Any units within the Census geographic boundaries of the area segment found not to be on the frame could be subsampled with probabilities that would result in an equal probability sample of all found units. Methods for checking missed units against the frame and other operational details were not established.

3. Geocoding

For an area sample survey, units need to be located “on the ground” so that they can be grouped into segments for efficient data collection. It is also desirable to have the ability to associate these segments with Census geographic boundaries so that the associated data (i.e., aggregate Census block-, block group-, and tract-level characteristics) can be used for stratification and in constructing the measure of size. However, the only “geographic” references on the USPS lists are the ZIP and ZIP+4 codes associated with the address. While the ideal definition of segments would be based on areas with precise geographic boundaries, the postal service defines these codes simply as collections of delivery points created for mail delivery efficiency. The units within the same ZIP code are usually clustered geographically, but not always, especially in rural areas.

Geocoding is the process of attaching geographic coordinates to a location. If an address has geographic coordinates attached to it, it can be matched to other geographic entities such as the Census geographic designations. The accuracy of geocoding is variable. With “rooftop geocoding,” latitude and longitude can be attached to an address at the exact position that the unit appears on a street; however, this form of geocoding is only available from some GIS companies and typically only in urban, stable areas. If the rooftop geographic coordinates are not available, the mid-point of the land parcel associated with the addresses may be available (known as “parcel geocoding”).

Most often, the geocoding process requires matching a particular address to a GIS database containing a mapping of the street network geographic coordinate space. In the GIS database, portions of each street, usually called "street segments," are defined in terms of street number ranges. When an address is geocoded, it is matched to the street segment with the street number range that contains the address of interest, rather than being matched to a specific address in the database. The position of the address is then determined by interpolating within the range along the street segment. With this "street-address geocoding," the specific geographic coordinates assigned to an address may not be at the precise location of the structure with that address, especially when the housing units are not evenly dispersed along a street. The structure may lie some distance down the street or even across the street from the location identified by the geographic coordinates.

When the street name or street number is not found in the database, two alternative methodologies to the street-address level geocoding process are available. The first assigns central geographic coordinates to the address based on the ZIP+4 associated with the address (e.g., 20850-1118); if the ZIP+4 is not recognized, the second alternative is to assign central geographic coordinates to the address based on the ZIP code associated with the address. The hierarchical assignment of geographic coordinates to an address, first attempting street-address geocoding, then these two alternatives, is known as street-level geocoding. Eckman and English (2012) estimate that 16.7 percent of the residential addresses on the USPS lists do not geocode at the street-address level, and thus may not be placed into the correct geographically defined segment for sampling. The extent of street-address level geocoding varies greatly depending on the area: 6.1% of metropolitan addresses and 45.1% of non-metropolitan addresses do not geocode at this level. While street-level geocoding may not place every address accurately, it does result in every address on an ABS frame being assigned to some segment, and thus given a chance of selection, when the segment is defined as the collection of addresses that geocode within a specified set of geographic boundaries.

4. A New ABS Approach

Our new approach, which is designed to allow for use of the USPS lists as ABS frames in nearly all areas, tackles both the coverage and geocoding issues while taking full advantage of the electronic nature of the USPS address lists (Kalton, 2010). The issue of coverage is addressed by adding a coverage enhancement procedure (CEP) similar to the enhanced Waksberg approach described in section 2; the issue of inaccurate geocoding is addressed with the use of the street-level geocoding methodology to define the segment and the use of a linked area segment for coverage enhancement. The formation of segments and their associated coverage enhancement areas is described in section 4.1. Section 4.2 describes the selection of segments for the CEP. Section 4.3 provides details on the implementation of the procedure.

4.1 Segment Formation, Coverage Enhancement Area Formation, and Linkage

The approach begins with segments created using Census geographic boundaries. A segment formed in this manner is an *area segment*. Next, addresses on the ABS frame are linked to area segments using the street-level geocoding method described in section 3. The street-level geocoding ensures that all addresses on the ABS frame can be associated with one, and only one, area segment. The collection of addresses that geocode to a particular area segment is termed a *list segment*. The list segments are the secondary sampling units.

As noted above, not all the addresses in the list segments will actually fall within their linked area segments. Many if not all of the addresses geocoded at the street-address level would be expected to fall within the area segment boundaries.[2] Some portion of the remainder will be geocoded at the ZIP+4 level. Generally, since only a small number of addresses has the same ZIP+4 (e.g., all those on the same side of a city block, or on a floor of an apartment building), geocoding at this level is unlikely to associate a housing unit with an area segment far from its actual location. If it is necessary to geocode at the ZIP code level, a classification that may contain thousands of addresses, the distance between the area segment and the actual location of the housing unit in question could be much larger. While an exact geographic match between the two types of segments would be helpful for data collection, the lack of an exact match presents an efficiency issue, not a coverage issue.

Figure 1 provides an illustration of the relationship between an area segment (defined by the blue boundary) and a list segment (grouped for illustration by the light red area) with no distinguishable boundary. Since the list segment is simply a collection of addresses with no definitive ordering or geographic reference, it is not useful for either the HOI or Waksberg approaches to coverage enhancement. However, since the area segment and the list segment have a one-to-one linkage, the area segment can be used as the frame of reference for the CEP. Thus, we will also refer to the area segment linked to a list segment as a "coverage enhancement area" when it is used for this purpose. As discussed next, there are alternative forms of coverage enhancement areas that could be considered.

Figure 1: Illustration of Corresponding Area and List Segments

[2] Some exceptions have been discovered in very urban area segments defined by 2010 Census geography.

Rather than using Census geography to define the coverage enhancement area, we also considered other approaches based on information provided on the USPS address lists. As noted above, the only identifiers on the original USPS file having general "geographic" relevance are ZIP and ZIP+4 codes. However, ZIP codes are not suitable for defining coverage enhancement areas because they may contain thousands of addresses and do not necessarily align with any geographic boundaries.

The use of the more detailed ZIP+4 designation would address the problem of the overly large ZIP codes for constructing coverage enhancement areas. However, since these smaller ZIP+4 categories contain very few addresses, forming efficient coverage enhancement areas based on the ZIP+4 designation would require combining multiple categories. The problem here is that the numbering system of the 4-digit suffix is not sequential so that addresses with ZIP+4 categories that are close in number are not necessarily close geographically (see the illustration in Figure 2). As a result, combining ZIP+4 categories to form efficient coverage enhancement areas would prove difficult. Furthermore, it would be an arduous, manual process to construct definitive geographic boundaries around the combined categories inside which the field staff could perform a CEP that would be used to identify housing units missing from the ABS frame (such as the one in Figure 2 indicated in by the).

Figure 2: Illustration of Possible ZIP+4 Assignments.

The USPS lists also identify the postal carrier route and the sequence in which the carrier travels along that route (see Figure 3). Since the routes comprise the addresses to which the postal carriers deliver mail within a day, the routes should not be too large to serve as coverage enhancement areas. Small routes could be combined to form more efficiently sized areas. However, as was the case with combinations of ZIP+4 categories, it would be a difficult, manual process to determine a well-defined geographic area in which to perform the coverage enhancement based on the carrier routes and walk sequences so that a unit not appearing on the ABS frame (indicated in Figure 3 by the) would have a chance of being included into the sample.

In light of these considerations, we decided to use Census geography both to define the area segments and to comprise the linked coverage enhancement areas. While ZIP codes, combinations of ZIP+4 codes or carrier routes could potentially be used to form list segments, some method of establishing a link with a geographic area in the PSU must be established. In the absence of any other geographic references on the USPS file, such a method is elusive. Additionally, using Census geography to define the area segments allows for the incorporation of Census data in the selection of segments for sampling and provides a straightforward approach for determining the linked coverage enhancement areas.

Figure 3: Illustration of One Possible Carrier Route and Walk Sequence.

4.2 Selecting Segments for the Coverage Enhancement Procedure

For a complete CEP, segments should be selected using a probabilistic approach that allows every segment a chance of being selected for the procedure. Our approach takes into account the fact that USPS lists have better coverage in some areas than in others by assigning the probability of area segment selection for the CEP based on the expected coverage of the list for the addresses in that area.

Let $P(i)$ be the probability of selecting segment *i* for the CEP. Let $P(j|i) = 1/k_i$ be the probability of selecting added address j for the survey if segment *i* is selected for the CEP. To give added addresses the same chance of selection as listed addresses, set $P(i) = k_i r_i$, where r_i is the within-segment sampling rate for segment *i*. If the number of units expected to be found through the procedure, $\widehat{M}_i$, is small, set $k_i = 1$, and bring all added addresses into the sample. If $\widehat{M}_i$ is large, k_i may be set to a value greater than 1 so that the number of actual addresses found that are added to the sample is more manageable.

One approach is to first determine a maximum number of addresses missing from the ABS frame that can be added to the sample in a segment, say m', based on the resources available in the field. If $\widehat{M}_i > m'$, set $k_i = \widehat{M}_i/m'$. If the estimate $\widehat{M}_i$ proves to be much smaller than the number of missed addresses actually found through the CEP, a subsampling fraction lower than $1/k_i$ may be applied to the added addresses, but this will require weighting up the subsampled addresses by an additional factor.

An important concern is how to determine the estimate $\widehat{M}_i$. One simple approach is to take the difference between the number of addresses in the list segment to the number of housing units in the associated area segment based on the most recent decennial census. The key issues to consider when using this approach is the age of the decennial census housing unit count, and whether large differences between the number of addresses in the list segment and the housing unit count is the a result of geocoding error or frame noncoverage. Other factors to consider are the number of P.O. box addresses in the list segment and the potential for large numbers of group quarters (if the target population includes people living in group quarters). Additional methods of estimating the segment coverage incorporate other characteristics of the area (see Montaquila, Hsu and Brick 2011 and McMichael et al. 2010 for more details).

In rural areas, the number of addresses in the list segment may be very small or even zero. If this is the case for a PSU or county within a multi-county PSU, it may be more efficient to resort to traditional listing in that area. This is similar to setting $P(i) = 1$ and $P(j|i)$ = r_i for all sampled segments in the area, without a need to distinguish originally listed addresses from added addresses.

The aim of the CEP described above was to give the sampled added addresses the same selection probabilities as addresses sampled from the list frame. In practice, it may be appropriate to select the added addresses with lower probabilities to acknowledge the costs associated with the application of the CEP. For example, for urban segments that can be predicted confidently to have no or very few unlisted addresses, the value of k_i could be set to a fraction rather than 1, thereby reducing the number of times the CEP is applied. The sampled added addresses found in those segments where the procedure is used will then need weights that take account of the use of a fractional k_i. For example, setting k_i=2/3 and including all added addresses in the sample would imply the need to increase the weights of the added addresses by a factor of 1.5.

4.3 Coverage Enhancement Procedure

This section describes how the CEP is implemented in practice. At the outset of the data collection period, all addresses sampled from the list segment for interviewing are transmitted to the assigned data collectors' computers. If a segment is also selected for the CEP, all addresses in the list segment (both sampled addresses and those not selected into the sample) are also transmitted to the data collector assigned to work that segment. The data collector performs the procedure before conducting interviews in the segment.

Using a specially developed software application, the data collectors access the preloaded list of all addresses in their list segment associated with the area segment chosen for the CEP. The data collectors are instructed to travel through their coverage enhancement area segment in a systematic manner. Their task is to determine for each housing unit they encounter on the ground within the boundaries of the area segment whether the address is on the preloaded list. If so, they assign the address a status of "located." If not, they add the address to the list (and the application flags the address as added). To reduce the potential for data entry errors, all the streets in the area segment are preloaded into the application, so the data collector may simply select the street name from a drop-down list rather than typing the name into the application. If the street name is not included in the preloaded list, the data collector adds the street manually. If components of an address are not discernible in the field, for example, if the house number is not visible, the data collector records identifying information associated with the unit and indicates the approximate location of the unit relative to the other nearby units.

For operational reasons, the data collectors are asked to assign a status to all addresses on their preloaded list of either located in the area, located outside the area, or not located (see Figure 4). After all addresses in the area are added or given a status, the data collector transmits the full list back to the home office. At this point, the data collector begins interviewing the sampled cases in the list segment using a separate software application on their computer.

Figure 4: Illustration of Address Statuses

At the home office, a detailed quality review of each data collector's work is conducted All added addresses and associated data collector comments are studied and any oddities, such as house/unit numbers appearing out of sequence, are investigated. Addresses added to the application with street names that were not preloaded are also reviewed. Additionally any addresses with unknown components are researched and reconciliation attempted using Internet resources including local government GIS websites.

After the home office review, the reconciled addresses added in the field (excluding those with unknown components) are checked against the frame held by the vendor to determine if the addresses are truly missing from the ABS frame, or if they are present in the ABS frame but geocoded into another list segment and thus given a chance of selection (see Figure 5). Depending on the probabilities associated with the CEP (as discussed in section 4.2) and the number of added addresses that were found to be missing from the list and those with unknown components, either all or a random subsample of these addresses are added to the sample and transmitted to the data collector for interviewing.

Figure 5: Illustration of Addresses in List and Area Segments after Frame Check

5. Field Test Results

A field test of the listing software application developed for the CEP was performed in 20 area segments. The scope of the test was limited, and therefore the results are not nationally representative. However, the segments were purposefully selected to test the application and the procedure in varying situations. The segments, which were defined in terms of 2000 Census geography, were selected from 10 states across the four regions of the U.S., with four of the segments from decidedly rural areas and seven segments dominated by multi-family structures (i.e., apartment buildings). The list segments associated with these area segments ranged in size from 59 to 195 addresses, with an average of 95.

Table 1 shows some results from the field test. The number of added addresses varied considerably by whether the segment was urban or rural. On average, fewer than two addresses were added in the urban segments, whereas the average added in the rural segments was over 18. When the added addresses were compared to the larger ABS frame, 50 percent of those in urban segments were found not to be present on the larger frame. In the rural segments, the corresponding figure was 73 percent. In an earlier field test in which the geocoding method was limited to the ZIP+4 and ZIP code levels, the average number of added addresses found elsewhere on the larger frame was 23 in urban segments and 78 in rural segments This underscores the value of accurate geocoding in helping to limit the resources required to implement the CEP.

On average only about five of the list segment addresses were located outside the area segment. This shows that for urban areas the hierarchical street-level geocoding process resulted in list segments that generally were in close alignment with their corresponding area segments. In the rural areas, the alignment was not as close with an average of 41 list segment addresses located outside the area segment. (Note that data collectors were not required to search for the exact location of all list segment addresses, but to confirm that they were not present in the boundaries of the area segment.)

For the urban segments, the net result was an excellent overall ABS frame coverage rate of 99 percent. In the rural segments, as expected, the coverage rate was lower, at 88 percent.

Table 1: Coverage Enhancement Field Test Results

	Number of segments	*Average list segment size*	*Average number of addresses added per segment*	*Added addresses not on ABS frame*		*Average number of list segment addresses not found in area segment*	*ABS frame coverage of segment*
				Average	*Percent of total added*		
Urban	16	96	1.5	0.8	50%	4.8	99%
Rural	4	90	18.3	13.3	73%	41.3	88%

6. Summary and Conclusions

This paper describes a methodology for using ABS list frames for household sampling that allows full ABS implementation and includes a coverage enhancement procedure that permits representation of all addresses while limiting the resource requirements. Since the coverage enhancement procedure (CEP) can be implemented immediately before data collection, using the same hardware as used for interviewing, data collectors can work initial sample cases while the addresses added through coverage enhancement are reviewed by the home office.

This methodology addresses both the coverage and geocoding issues previously cited as obstacles to using ABS for area surveys with in-person interviewing. The coverage issue is addressed with the inclusion of a coverage enhancement procedure applied more frequently in areas estimated to have poor coverage. Geocoding inaccuracy, while potentially complicating the CEP, does not increase noncoverage. While the geocoding may not place every address in the correct area segment as defined by Census geographic boundaries, every address on an ABS frame is assigned to a list segment that is defined as the collection of addresses that geocode to a specific area segment.

An effective one-to-one linkage of list and area segments defined in terms of Census geography gives researchers the ability to use Census data (or other data that can be linked to Census blocks, block groups, or tracts) in their sample design (for example, to oversample segments with high minority).

The hierarchical street-level geocoding method used with this approach, which first attempts to geocode an address at the street-address level before resorting to the ZIP+4-code and then ZIP-code level, proved very effective in allocating addresses correctly to area segments defined in terms of Census 2000 geography. However, geocoding addresses to area segments defined by Census 2010 geography appears to be more problematic; further research is needed on this issue.

Acknowledgement

The authors thank Michael Giangrande and Steve Williamson of Westat and David Malerek, Mansour Fahimi, Ashley Hyon of Marketing Systems Group (MSG) for their contribution to this work.

References

Dohrmann, S., Han, D., and Mohadjer, L. (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.

Eckman, S., and English, N. (2012). Creating Housing Unit Frames from Address Databases: Geocoding Precision and Net Coverage Rates. *Field Methods*. doi: 10.1177/1525822X12445141

English, N., Dekker, K., and O'Muircheartaigh, C. (2011). Choices of Frame Construction on the National Children's Study: Impacts on Address Quality and Survey Results. *Proceedings of the Survey Research Methods Section of the American Statistical Association*. 4250-4259.

Iannacchione, V. G. (2011). The Changing Role of Address-Based Sampling in Survey Research. *Public Opinion Quarterly*, *75*(3), 556-575. doi: 10.1093/poq/nfr017

Iannacchione, V., Morton, K., McMichael, J., Shook-Sa, B., Ridenhour, J., Stolzenberg, S., Bergeron, D., Chromy, J., and Hughes, A. (2010). The Best of Both Worlds: A Sampling Frame Based on Address-Based Sampling and Field Enumeration. *Proceedings of the Survey Research Methods Section of the American Statistical Association*.

Kalton, G. (2010). *Developing Linkage Rules to Reduce Noncoverage*. Presented at the AAPOR 65th Annual Conference.

Kish, L. (1965). *Survey sampling.* New York: J. Wiley.

McMichael, J.P., Ridenhour, J.L., Mitchell, S., Fahrney, K, and Stephenson, W. (2008). Evaluating the Use and Effectiveness of the Half-Open Interval Procedure for Sampling Frames Based on Mailing Address Lists in Urban Areas. *Proceedings of the Survey Research Methods Section of the American Statistical Association – 2008 AAPOR.* 4251-4257.

McMichael, J., Ridenhour, J., Shook-Sa, B., Morton, K., and Iannacchione, V. (2010). Predicting the Coverage of Address-Based Sampling Frames Prior to Sample Selection. *Proceedings of the Survey Research Methods Section of the American Statistical Association.* 4852-4859.

McMichael, J., Ridenhour, J., and Shook-Sa, B. (2008). A Robust Procedure to Supplement the Coverage of Address-Based Sampling Frames for Household Surveys. *Proceedings of the Survey Research Methods Section of the American Statistical Association- 2008 AAPOR.* 4329-4335.

Montaquila, J., Hsu, V., and Brick, M. (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(2):314–335.

Montaquila, J., Hsu, V., and Brick, M., English, N., and O'Muircheartaigh, C. (2009). ''A Comparative Evaluation of Traditional Listing vs. Address-Based Sampling Frames: Matching with Field Investigation of Discrepancies.'' *Proceedings of the Survey Research Methods Section of the American Statistical Association.* 4855–62.

O'Muircheartaigh, C., Eckman, S., and Weiss, C. (2002). Traditional and Enhanced Field Listing for Probability Sampling. *Proceedings of the Social Statistics Section of the American Statistical Association*, 2563-2567.

Staab, J.M., and Iannacchione, V.G. (2003). Evaluating the Use of Residential Mailing Addresses in a National Household Survey. *Proceedings of the Survey Research Methods Section of the American Statistical Association.* 4028-4033.