A Robust Procedure to Supplement the Coverage of Address-Based Sampling Frames for Household Surveys

Joseph P. McMichael[1], Jamie L. Ridenhour[1], and Bonnie E. Shook-Sa[1]
[1]RTI International, 3040 Cornwallis Road, Post Office Box 12194, Research Triangle Park, NC 27709

Abstract

Address-based sampling (ABS) is an alternative to field enumeration that uses a frame of mailing addresses from the delivery sequence file (DSF). In-person household surveys using ABS are more cost effective and are implemented faster than what is possible with field enumeration. The primary drawback of a frame based on mailing addresses is the under-coverage of households with unlocatable mailing addresses (e.g., simplified addresses). This drawback can be addressed with a frame supplementation procedure. The half-open interval (HOI), a frame supplementation procedure develop for and used on field enumerated frames, does not translate to ABS surveys because of its reliance on a predetermined path of travel for evaluating the interval between the sampled and expected next address. Although a frame based on the DSF has some proximal sorting from the carrier route delivery sequence, this information may not be a sufficient substitute to the path of travel that is essential to implementing HOI.

We have developed a new procedure that departs from the HOI and addresses many of the fundamental problems underlying HOI: instead of asking field staff to check an interval between the sampled address and the expected next address we instead ask the field staff to dynamically determine the next address on the ground. This robust, three component procedure called the Check for Housing Units Missed theoretically improves coverage on an ABS study to 100%. In this paper we discuss the implementation and testing of the procedure as well as future applications.

Key Words: Address-Based Sampling, ABS, Half-Open interval, HOI, frame supplementation

1. Address-Based Sampling and the Need for Frame Supplementation

Address-based sampling (ABS) is becoming commonly used as an alternative to traditional methods like field enumeration for household surveys. ABS is appealing because of its cost savings, timeliness, and geographic diversity (Iannacchione, *et al*. 2007). However, the coverage of ABS frames is not complete, and there is a need for both frame supplementation methods and evaluation of these methods.

RTI International conducted an evaluation of the coverage of an ABS frame against the coverage of a field enumerated frame. In this evaluation, field enumeration was considered to be the gold standard. The DSF was found to contain 82% of all housing units (DUs) enumerated. In this evaluation, a modified half-open interval (HOI) procedure based on the mail carriers' delivery sequence was implemented. The HOI added an additional 240 housing units to the frame; however, there were still over 600 households left uncovered by the ABS frame and this HOI method (Iannacchione, *et al*. 2007).

There are several sources of known undercoverage in ABS. Households that are new construction may not have been added to the delivery sequence file (DSF) at the time it was purchased. In addition, not every mailing address is locatable on the ground. P.O. Boxes in areas without home mail delivery and simplified Rural Routes, most common in rural areas, are not locatable by field staff. This source of undercoverage can lead to entire geographic areas (e.g., census blocks) that are not covered by the frame. Coverage on ABS frames also varies by urbanicity. Rural areas tend to have considerably lower coverage than urban areas (Dohrmann *et al*. 2006).

In traditional field enumeration, the frame supplementation method often used is the half-open interval (HOI) procedure. The HOI adds housing units to the frame by having field staff examine the interval between the selected

address and the next address on the frame. The HOI procedure relies on the fact the units on the frame are sorted in proximal order to each other.

With an ABS frame there is no proximal order created by a field enumerator. The carrier route delivery sequence can be used as a proximal order. The letter carrier usually proceeds down one side of the street and back up the other making the delivery sequence amenable to the HOI frame-linking procedure. However, an HOI cannot be constructed when the interval between the sampled DU and the next DU on the frame is ambiguous. Examples of ambiguous HOIs include clusters of mail boxes and some street intersections. Staab and Iannacchione, 2003, estimated that unambiguous HOIs could be constructed for at least 83.8 percent[1] of the addresses on an ABS frame for a national in-person household survey conducted in 2002.

In another in-person household study of six urban areas we have found that the HOI based on carrier route delivery sequence frequently breaks down due to the number of street crossing that cause ambiguous intervals (McMichael *et al*. 2008).

One way to adapt HOI to be used more effectively on an ABS study would be to create maps of all of the segments with a predetermined path of travel outlined on every one. Then, the FI would be able to simulate attempting to look for a missed household in an interval by following that path of travel. Developing maps for this approach could be very costly.

The approach we developed and outline in this paper is a procedure that does not rely on a priori knowledge of the interval between the sampled address and the next address. This new method allows field staff to identify addresses missing from the frame by following a path of travel determined by simple protocols and determining the next address on the ground. Unlike the HOI procedure described previously, this method does not require *predetermined* paths of travel.

2. Proposed Method - Check for Housing Units Missed (CHUM)

The Check for Housing Units Missed (CHUM) is a frame supplementation procedure that theoretically improves coverage on an ABS frame to 100%. The concept behind the CHUM is different from an HOI-type procedure because it does not rely on field staff to check a predetermined interval between the sampled address and a known next address. Instead, the CHUM requires the field staff to follow procedures much like those done in field enumeration to systematically identify DUs not on the ABS frame.

The CHUM methodology assumes that the areas or segments are based on census geography. While this method may be used for other geographic partitions, it has not been evaluated in this context. The CHUM method requires access to the entire address list sampling frame within each of the selected areas or segments. It also assumes segment maps are available to field staff.

The CHUM method uses three components to achieve complete coverage. Component One systematically identifies addresses (and by proxy DUs) that are missing from the ABS frame. This component is the most important and all other components are built off of it. Component Two identifies "missed areas" where Component One will have no effect because an area's (e.g., a city block) DUs are not on the ABS frame. Component Three identifies new streets that could interfere with the coverage properties of Component Two.

2.1 CHUM Component One – Identify the "Next Address"

Component One is the most important part of the CHUM procedure. It sets up a series of protocols that establishes a path of travel allowing field staff to identify a unique "next address" from the sampled address. A sampled address is most likely an address selected for study but could be an address selected for the sole purpose of frame supplementation. This unique "next address" is checked against the frame to determine whether or not it is missed. If it is not on the frame, it is added to the frame and given a positive probability of selection that is tied to the probability of selection of the sample address.

[1] Lower bound of a one-tailed 95 percent confidence interval.

The basic protocol for determining the next address on the ground, which is expected to be the most frequently used, instructs field staff to face the sampled address and travel to the left (or clockwise) around the block, without crossing a street to find the next address. If the next address is not on the frame, it is added. This step repeats itself until the next address identified is found on the frame (Figure 1, Example 1).

This basic protocol for identifying the next address needs to be modified for apartments. For the purpose of identifying a unique next address we propose sorting the apartment numbers within the building that contains the selected address in alphanumeric order. Field staff would use this sorted list to determine the next address. Figure 2 illustrates how to identify the next address in apartment-like situations.

Several practical supplements can be applied to CHUM Component One to decrease the amount of time between when the field staff begin looking for the next address in relation to the sampled address and when the field staff know that one of the next addresses is in fact on the ABS frame. To decrease this time the field staff could be supplied with potential next addresses for that sampled address. One way to supply field staff with possible next addresses is to use the carrier route delivery sequence information on the DSF. Another method of supplying potential next addresses is to determine a set of possible addresses that through geocoding are deemed likely to be close to the sampled address. Whether one or both of these methods of supplying field staff with likely next addresses is used, a supplied list of possible next addresses would mean that the field staff would only be asked to supply the full address if the next address that they encounter is not on their list of possible addresses. If the field staff encounters a string of addresses that are not on their list of possible next addresses the field staff could be asked to only give the next three addresses to minimize the amount of time the field staff spend in the field on this part of fieldwork. It would then be up to the sampling team in the home office to determine if one of those addresses matches an address on the ABS frame or if the field staff need to supply more next addresses

2.2 CHUM Component Two – Check for "Missed Areas"

Because Component One can restrict field staff to the block associated with the sampled address, there is the possibility of missed DUs on a block that has no coverage from the ABS frame (Figure 1, Example 2). Component Two identifies entire missed areas (likely census blocks) and their associated DUs that would not have a chance of being brought in to the sample via Component One. If an area is identified as a "missed area" its DUs are added to the frame. However, unlike Component One, the added DUs have probability of selection tied to the selection of the area or block. Any number of sampling schemes can be used to select areas or blocks for Component Two.

The field protocol for Component Two is similar to Component One except that there is not a selected address for field staff to begin identifying the next address. Instead, the area selected for evaluation has a start point or points placed on the segment map prior to field operations. Start points are placed such that the entire area can be canvassed using the protocols used for Component One, in effect acting as a surrogate for a selected address. From this start point field staff would then identify the next address as then did for Component One. It may take more than one start point to be able to completely canvas an area.

If during the implementation of Component Two field staff identify DUs not on the frame as well as DUs that are, then no DUs would be added to the frame using Component Two because these DUs already have a chance for inclusion in Component One (Figure 1, Example 3).

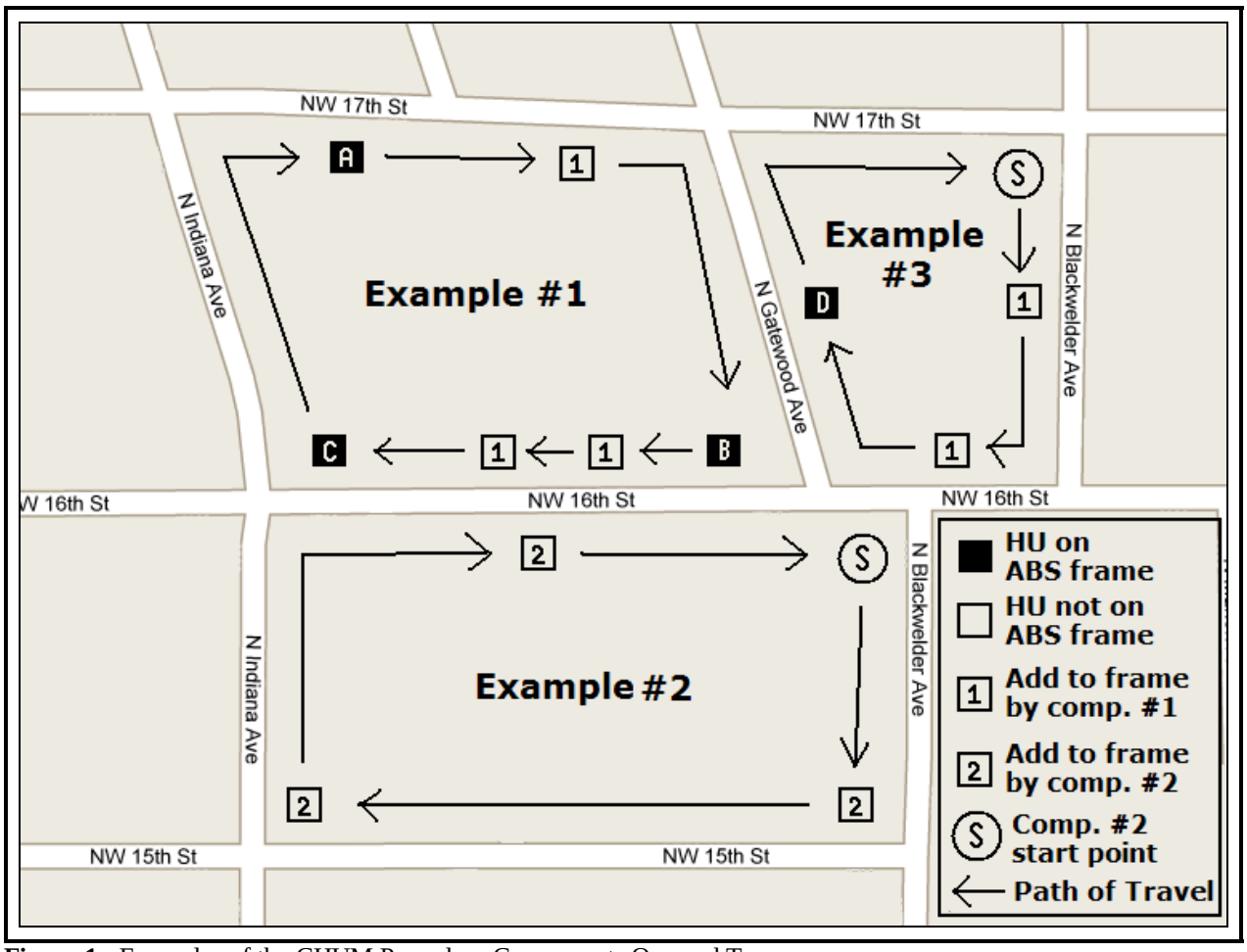

Figure 1: Examples of the CHUM Procedure Components One and Two.

Example 1 of Figure 1 illustrates an example of adding DUs to the frame by the CHUM Component One – Identify the "Next Address". Note that Component One creates a unique path of travel that not only allows field staff to identify the Next Address but also gives each DU not on the ABS frame a known probability of selection. If we assume DUs "A" and "B" are selected for Component One, one DU would be added to the frame through DU "A", and two DUs would be added through DU "B". No DUs would be added to the frame if DU "C" were selected.

Example 2 of Figure 1 illustrates where a start point could be placed and how DUs would be added to the frame if an area were randomly selected for the CHUM Component Two – Check for "Missed Areas". If we assume the city block with "Example #2" were randomly selected for the Check for Missed Areas, field staff would be provided a segment map with a start point that allows the entire block to be canvassed using the protocol established for Component One from the start point. In this case three DUs are added to the frame. Notice that the three DUs have no chance for inclusion solely using Component One.

Example 3 of Figure 1 illustrates what happens when an area is randomly selected for the CHUM Component Two – Check for "Missed Areas" and DUs both on and off the frame are identified. If we assume the city block with "Example #3" were randomly selected for the Check for Missed Areas, field staff would identify two DUs not on the frame before identifying one DU on the frame (DU "D"). Because DU "D" is on the frame the other two DUs identified as not on the frame would not be added to the frame thorough Component Two. This is because they already have a chance of being added through Component One if DU "D" were selected for Component One.

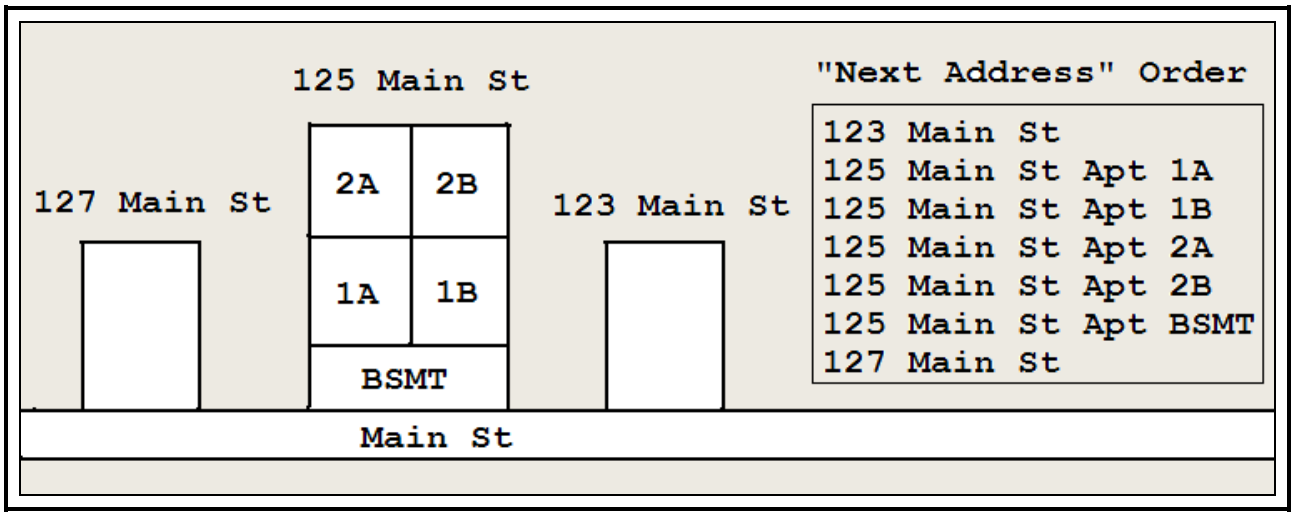

Figure 2: Examples of CHUM Procedure Involving Apartments

The CHUM procedures involving apartments are similar to those outlined previously; however, because the ordering of apartments cannot always be determined in a geographic space we suggest that an alphanumeric sort of apartments is the most effective way to determine a consistent ordering.

Example 1 – If sampled address is 125 Main St., Apt 1B then the next address would be 125 Main St., Apt 2A . If this address is on the frame, the field staff would stop listing addresses. Otherwise, 125 Main St., Apt 2A would be added to the frame and the field staff would continue listing addresses in the order of the "Next Address Order" until reaching an address that is on the ABS frame.

Example 2 – If sampled address is 123 Main St the next address would be 125 Main St., Apt 1A. If this address is on the frame, the field staff would stop listing addresses. Otherwise, 125 Main Street, Apt 2A would be added to the frame and the field staff would continue listing addresses in the order of the "Next Address Order" until reaching an address that is on the ABS frame.

Note: Apartments within 125 Main St. are sorted in alphanumeric order in the "Next Address Order" list.

2.3 CHUM Component Three – Check for "New Streets"

The third component of the CHUM accounts for a situation where DUs in the segment could be missed by Components One and Two. When field staff complete Component Two of the CHUM they are using a map which may not have accurate information about an area, particularly if that area is experiencing new growth and development. Because field staff could encounter new streets that are not displayed on the segment map used to designate the starting points for Component Two, and thus affecting its efficacy, the Component Three of the CHUM procedure asks field staff to verify that all of the streets in the selected sub-segment are shown on the map. If a new street is identified and it interferes with the effectiveness of Component Two, additional start points may need to be added (Figure 3). Note that this situation is expected to be very rare.

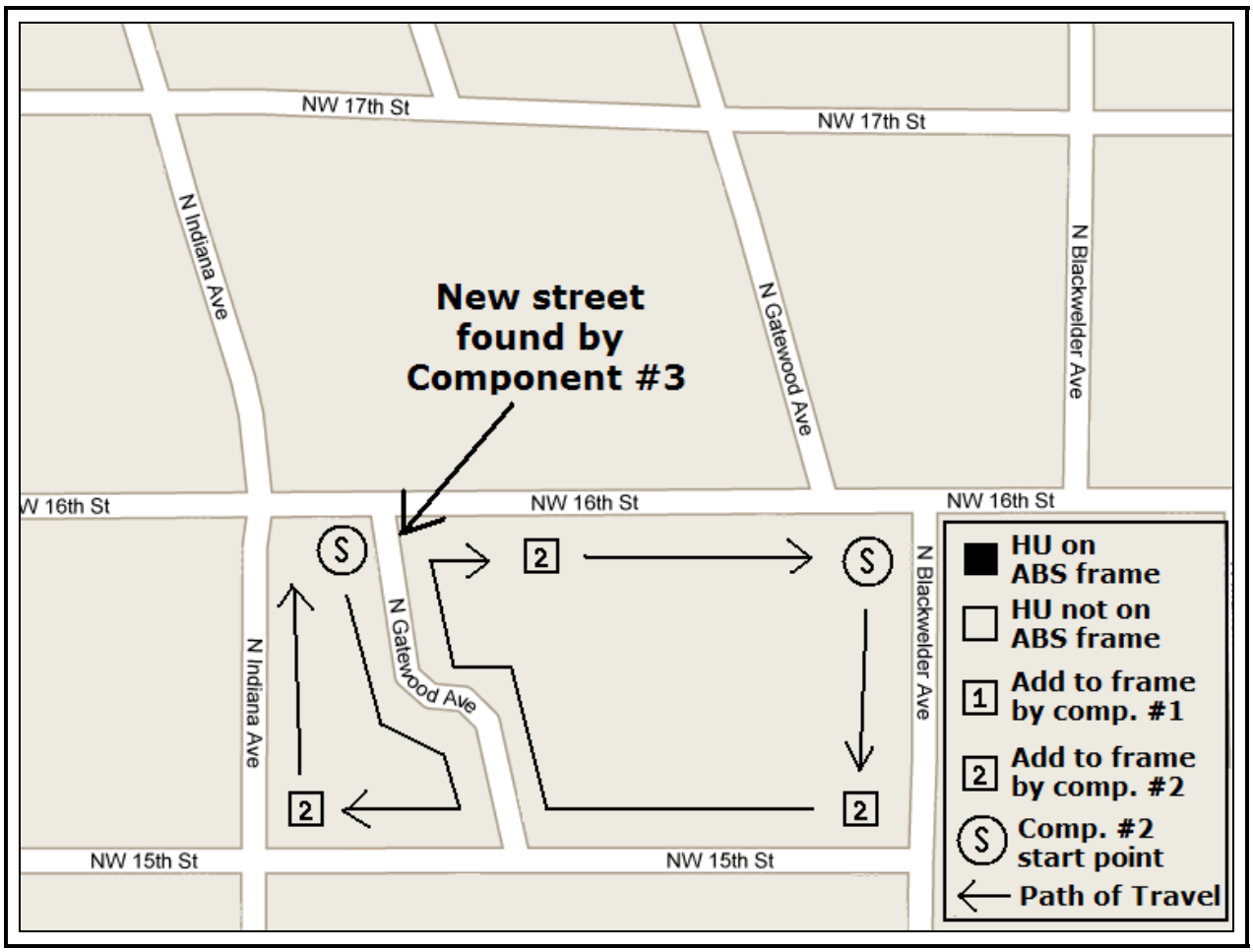

Figure 3: Example of the CHUM Procedure, Component Three. This example illustrates the implementation of CHUM Component Three – Check for "Missed Streets". Assume we implement the Check for Missed Streets on the same block used in Figure 1, Example 2. Field staff would be provided a segment map and would now be instructed to identify new streets not on the segment map. If after evaluating a new street it is determined that it will interfere with the coverage of Component Two, then another start point would need to be added to be able to completely canvas the selected area with Component Two. Note that this is an extremely rare situation.

3. Evaluation of Check for Housing Units Missed

To evaluate the efficacy of the CHUM procedure, we used data already collected by RTI International for the evaluation of the coverage rates of an ABS frame against a field-enumerated frame (Iannacchione, et al. 2007). This study evaluated 6,408 DUs in 50 geographic segments in North Carolina. Out of those 6,408 DUs, 837 (13%) were not on the DSF. We evaluated these 837 addresses using field enumeration maps, GPS coordinates captured during field operations, and the DSF frame for this study to see if these missed DUs could have been picked up by either Component One or Two of the CHUM procedure. Component Three could not be evaluated due to the time delay between when that data was collected and this evaluation. The field enumeration maps provided a location of these missed DUs on a map along with information about the neighboring DUs. We knew from the DSF frame for this study whether or not the neighboring DUs were on the frame.

We determined that all 837 missed DUs could have been picked up by the combination of Components One and Two of the CHUM procedure. Component One picked up 662 of these missed DUs (79%) while Component Two picked up the remaining 175 DUs (21%). These findings assume that the CHUM methodology would be correctly implemented in the field. One caveat to these findings is that some of the segments in this data were rural and had very low coverage

from the DSF. These segments resulted in a higher number of addresses being found via Component Two of the CHUM than would probably otherwise happen.

4. Concluding Statements

The Check for Housing Units Missed (CHUM) procedure is a robust, comprehensive three component method for supplementing coverage on an address-based sampling study. It theoretically provides 100% coverage. The evaluation of this method used data from a study with relatively small geographic segments and because this method has not yet been field tested we are aware that there may be new challenges with implementing this method in a study with larger geographic segments. Also, the CHUM procedure assumes that there is a complete ABS frame for each geographic area of interest in a study. If a study were only taking a simple random sample of addresses from the national DSF then a supplementation method like the CHUM could not be used.

This method, as with all frame supplementation methods, relies heavily on field staff to implement it correctly. The key to making the method successful is thorough training for field staff. This reliance on field staff for correct implementation could imply that our theoretical coverage would be unlikely in a study due to any errors introduced by field staff. We think that the procedures that comprise the CHUM will make it easier for field staff to implement than other HOI-type procedures that have previously been applied to ABS studies. RTI will be using and further evaluating the CHUM procedure on two upcoming national household surveys.

Acknowledgements

The authors would like to thank RTI International for providing the resources to conduct this research. The authors would like to thank Vincent Iannacchione for his ongoing support and guidance. They would also like to thank Katherine Morton for suggesting the memorable yet somewhat unfortunate acronym for this procedure.

References

Dohrmann, Sylvia, Daifeng Han, and Leyla Mohadjer. 2006. Residential Address Lists vs. Traditional Listing: Enumerating Households and Group Quarters. *Proceedings of the American Statistical Association, Section on Survey Research Methods.*

Dohrmann, Sylvia, Daifeng Han, and Leyla Mohadjer. 2007. Improving coverage of residential address lists in multistage area samples. *Proceedings of the American Statistical Association, Section on Survey Research Methods.*

Iannacchione, Vincent, Katherine Morton, Joseph McMichael, David Cunningham, James Cajka, and James Chromy. 2007. Comparing the Coverage of a Household Sampling Frame Based on Mailing Addresses to a Frame Based on Field Enumeration. *Proceedings of the American Statistical Association, Section on Survey Research Methods.*

McMichael, Joseph, Jamie L. Ridenhour, Susan Mitchell, Kristine Fahrney, and Wanda Stephenson. 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 American Statistical Association, Section on Survey Research Methods.*

Staab, Jennifer, and Vincent Iannacchione. 2003. Evaluating the Use of Residential Mailing Addresses in a National Household Survey. *Proceedings of the American Statistical Association, Section on Survey Research Methods*