RTI Press

Methods Report

February 2016

The CHUM: A Frame Supplementation Procedure for Address-Based Sampling

Bonnie E. Shook-Sa, Rachel M. Harter, Joseph P. McMichael, Jamie L. Ridenhour, and Jill A. Dever

RTI INTERNATIONAL

RTI Press publication MR-0034-1602

Suggested Citation

Shook-Sa, B. E., Harter, R., McMichael, J., Ridenhour, J., and Dever, J. (2016). *The CHUM: A Frame Supplementation Procedure for Address-Based Sampling*. RTI Press Publication No. MR-0034-1602. Research Triangle Park, NC: RTI Press. http://dx.doi.org/10.3768/rtipress.2016.mr.0034.1602

This publication is part of the RTI Press Methods Reports series.

RTI International
3040 East Cornwallis Road
PO Box 12194
Research Triangle Park, NC
27709-2194 USA

Tel: +1.919.541.6000
E-mail: rtipress@rti.org
Web site: www.rti.org

Contents

About the Authors

Bonnie E. Shook-Sa, MAS, is a research statistician at RTI International. Her research focuses on sampling frame development and evaluation, sample design and optimization, and the analysis of complex data for household and establishment surveys.

Rachel M. Harter, PhD, is a senior research statistician and program director at RTI. Areas of interest include household and establishment surveys, area probability survey designs, address-based sampling, imputation, and small area estimation.

Joseph McMichael, BS, is a research statistician in RTI's Division for Statistical and Data Sciences.

Jamie Ridenhour, MStat, is a research statistician at RTI. Her research interests are sample design, weighting, and methodological challenges associated with address-based sampling and dual-frame random-digit-dial surveys.

Jill A. Dever, PhD, is a senior research statistician at RTI. Her current research interests are variance estimation with calibrated analysis weights for complex survey designs and statistical issues related to samples drawn without a defined probabilistic structure.

Acknowledgments
The authors would like to acknowledge the contributions of Vince Iannacchione and Katie Morton with RTI International for their valuable contributions to this research.

Abstract

RTI developed the check for housing units missed (CHUM) methodology to compensate for housing unit undercoverage of address-based sampling (ABS) frames for in-person, area probability surveys. The CHUM systematically identifies housing units missing from the ABS frame, giving each housing unit a chance of selection with known probability. The CHUM poses several advantages over alternative supplementation approaches. Because only a subset of housing units within selected areas must be evaluated, the CHUM is less costly than supplementation techniques that require the verification of all addresses within selected areas. Because it is conducted after housing units are selected instead of the frame-building stage, the CHUM provides timelier frame updates. This paper presents details for designing ABS studies that incorporate the CHUM, appropriately incorporating missed units into area probability samples, and training field personnel to implement the CHUM. It also compares the CHUM with other frame supplementation approaches and discusses the advantages and limitations of each approach.

Introduction

Area probability sampling has long been used for in-person surveys of the US housing unit population. With area probability sampling, the goal is to sample housing units and persons within those housing units from selected geographical clusters within a multistage probability design. Because of the availability of demographic information and clearly defined boundaries, geographic areas as early stage sampling units are often defined and sampled based on census delineations (e.g., counties, census tracts, census block groups, census blocks).

At the design stage, prior to housing unit selection, geographies often referred to as *area segments* are selected. Area segments are relatively small geographic units that ensure proper population coverage but also provide enough clustering to minimize the number of field personnel needed for screening and interviewing. For some sample designs, clusters also allow for the enumeration of housing units within segments. Following the selection of area segments, housing units and persons within housing units are subsampled for inclusion in the study. To appropriately account for the complex designs of area probability samples, survey researchers calculate survey weights to scale the sample up to the population. The base weights for selected persons are the inverse of their probabilities of selection, taking into account all stages of sampling. Valliant and colleagues (2013) provide a detailed description of area probability sampling and the calculation of survey weights.

Traditionally, field enumeration techniques have been used within area segments to obtain housing unit frames. Because of rising data collection costs for field-enumerated surveys, many in-person household surveys in the United States are now moving to address-based sampling (ABS) frames. ABS frames are extracts from vendor address files that incorporate mailing addresses from the US Postal Service's Computerized Delivery Sequence (CDS) file (Iannacchione, 2011). Since 2009, a supplemental file also maintained by the US Postal Service called the No-Stat file has been available; it has the potential to increase the coverage of in-person ABS studies while reducing costs associated with field supplementation procedures by bringing the ABS frame closer to a complete frame of housing units in the United States (Shook-Sa et al., 2013; Shook-Sa, 2014). The No-Stat file contains address types that are not included on the CDS file, and thus the CDS file in combination with the No-Stat file contains all US postal delivery points. Although there is not always a one-to-one correspondence between residential mailing addresses and housing units, mailing addresses serve as reasonable proxies for the physical locations of housing units.[1]

Although the CDS and No-Stat files provide the mailing addresses for nearly all housing units in the United States, in-person ABS studies suffer from undercoverage of the housing unit population, primarily due to two factors. First, certain types of postal delivery points (e.g., Post Office boxes, rural route boxes) are not suitable for in-person studies because they cannot be physically located based on their mailing addresses and are thus excluded from ABS frames prior to sample selection. ABS frames for in-person surveys have been shown to suffer from undercoverage in rural areas where a higher proportion of mailing addresses are not locatable (Dohrmann et al., 2007; Iannacchione et al., 2007; O'Muircheartaigh et al., 2007).

The other key factor that can lead to ABS undercoverage is *geocoding error*. The process of assigning mailing addresses to geographic segments, called *geocoding*, can lead to undercoverage[2] in selected areas where geocoding is less accurate and addresses are assigned to the wrong area segments (Morton et al., 2007; Eckman & English, 2012). In addition to these primary sources of undercoverage, changes in housing units between when the frame was constructed and when the sample is fielded can lead to housing unit undercoverage and sampling

1 When the target population includes residents of group quarters as well as residents of housing units, the broader term *dwelling units* is appropriate. However, this paper uses the term *housing units*, assuming that residents of group quarters are excluded from the target population.

2 Geocoding error does not cause overall frame undercoverage, but geocoding error can cause both undercoverage and overcoverage for selected geographic segments. Overcoverage leads to loss of efficiency by including addresses not eligible for the selected segments. The more serious problem, however, is undercoverage, which is why frame supplementation methods have been developed.

RTI Press Publication No. MR-0034-1602. Research Triangle Park, NC: RTI Press. http://dx.doi.org/10.3768/rtipress.2016.mr.0034.1602

inefficiencies (e.g., newly constructed or demolished housing units).

Because housing unit undercoverage for in-person surveys tends to be geographically clustered, ABS studies that do not supplement the frame can be susceptible to biased estimates, particularly for studies where the target population is underrepresented on the frame (e.g., rural populations). *Supplementing* the ABS frame can improve housing unit coverage. Frame supplementation refers to methods that are conducted in sampled area segments that aim to add missed units to the sampling frame for improved coverage. The *check for housing units missed* (*CHUM*) methodology corrects for geocoding error and gives every housing unit missing from the frame a chance of selection into the study with a defined probability while providing survey practitioners with flexibility in terms of the timing and scope of implementation (McMichael, Ridenhour, & Shook-Sa, 2008).

In this report, we expand upon the CHUM, discuss design considerations and details for appropriately incorporating missed units into area probability samples, and outline considerations for training field personnel to implement the CHUM. We also compare the CHUM with other ABS frame supplementation procedures, outlining the advantages and disadvantages of each approach.

Types of Supplementation Procedures

Although some surveys are based on ABS or field enumeration alone, many national, in-person ABS designs (e.g., National Survey of Family Growth, Residential Energy Consumption Survey, General Social Survey) now use a mixture of ABS and field enumeration methods, sometimes referred to as *hybrid sampling frames*, depending on the expected coverage provided by the address frame for the selected geographies. The sampling frame for a particular study depends on many factors, including the budget of the study and the target population. Whether a sampling frame is ABS only or includes a field enumeration component, frame supplementation procedures can be used to obtain more complete coverage of the US housing unit population.

Traditionally, the *half-open interval (HOI) method* has been used in field enumeration studies to provide coverage for missed housing units (Kish, 1965). With HOI, field personnel search from a sampled housing unit up to but not including the next housing unit on the ordered frame, and any new housing units found in that interval are also selected into the sample. By linking the previously missed housing unit to the one listed just before it on the ordered list, the probabilities of selection for missed housing units are set to be the same as the originally sampled unit. This method requires a well-specified listing sequence so that field personnel can follow the same path as the enumerator whose information was used to create the field enumerated frame. The problem with the HOI method for ABS is that the address lists are in mail delivery sequence order, which does not lend itself to well-defined half-open intervals because of the tendency of postal lists to cross streets and jump from block to block. For this reason, HOI is not typically considered to be a viable frame supplementation technique for ABS (McMichael, Ridenhour, Mitchell, et al., 2008).

Enhanced listing, also referred to as dependent listing, combines ABS with field enumeration by using the address list for the segment and asking field personnel to update the frame by adding, deleting, and correcting addresses from the ABS list based on the housing units observed in the segment. Unlike the HOI, enhanced listing is conducted prior to selecting housing units as part of frame-building (English et al., 2013). This type of enhanced listing is beneficial when area segments will be used for multiple studies, as the supplementation procedure does not need to be conducted independently for each study, provided that studies are conducted close enough together in time to avoid undercoverage and inefficiencies resulting from newly added or removed housing units. However, because all addresses within supplemented segments must be verified, enhanced listing can be costly. As a result, enhanced listing may require practitioners to limit the size of area segments, which has direct implications on the precision of the study estimates (Valliant et al., 2013). In addition, because supplementation is completed prior to any sample selection, frame updates are less timely than methods conducted after the initial sample has been selected.

With the *address coverage enhancement (ACE)* procedure, field personnel canvass a subset of area segments either prior to or during data collection and identify missed housing units. The ACE methodology mitigates the problem of geocoding error by including all sampled addresses that geocode into sampled area segments in the sample, rather than requiring them to be identified through the listing process (Dohrman et al., 2012; Kalton et al., 2014). Although canvassing area segments in urban areas can be quick, it can be more time and cost intensive in large rural areas (Kalton et al., 2014). Because addresses that incorrectly geocode into area segments are considered eligible for the study, the ACE procedure requires all missed units identified in the field to be matched against the entire ABS frame rather than a comparison with only those ABS addresses that geocoded into the area segment.

In contrast to other field supplementation methods, the CHUM provides flexibility in terms of the timing and scope of implementation for ABS studies. Unlike enhanced listing, the CHUM is performed after sample selection and usually during data collection, making frame updates timelier. Additionally, the CHUM requires fewer addresses to be verified in the field compared to enhanced listing or the ACE procedure. With the CHUM, field personnel search randomly chosen housing units for any missed housing units, and then search from the selected housing units to the next housing units on the frame, following a prescribed order. Field personnel also search a subset of selected blocks to ensure that housing units in blocks without addresses on the frame have a chance of selection. Further details about the sampling and weighting implications of the CHUM are discussed in the following section, followed by associated operational issues. A more detailed comparison of the CHUM with the other supplementation techniques is included in the Benefits and Limitations section.

The CHUM Methodology

Like the HOI, the CHUM has rules for assigning probabilities of selection for all housing units missing from the ABS frame and thus (theoretically) results in 100 percent coverage of housing units within selected areas. The CHUM was implemented within the 2008 American National Election Study, where it was estimated that the procedure boosted coverage of the ABS frame by 5.1 percent (McMichael et al., 2009). CHUM coverage rates vary based on the target population of the study, whether the frame is ABS-only or a hybrid of ABS and field enumeration, and the quality of CHUM implementation.

For an ABS sample, the spatial layout of housing units within selected area segments is typically not known at the sample selection stage. Instead of defining a prespecified end point, as with the HOI, the CHUM has a defined start point and allows field personnel to determine the end point of the geographic interval in the field. Probabilities of selection for housing units missing from the ABS frame are assigned based on one of two CHUM procedures—the *check for missed units (CHUM1)* or the *check for missed blocks (CHUM2)*—described below.

Check for Missed Units (CHUM1)

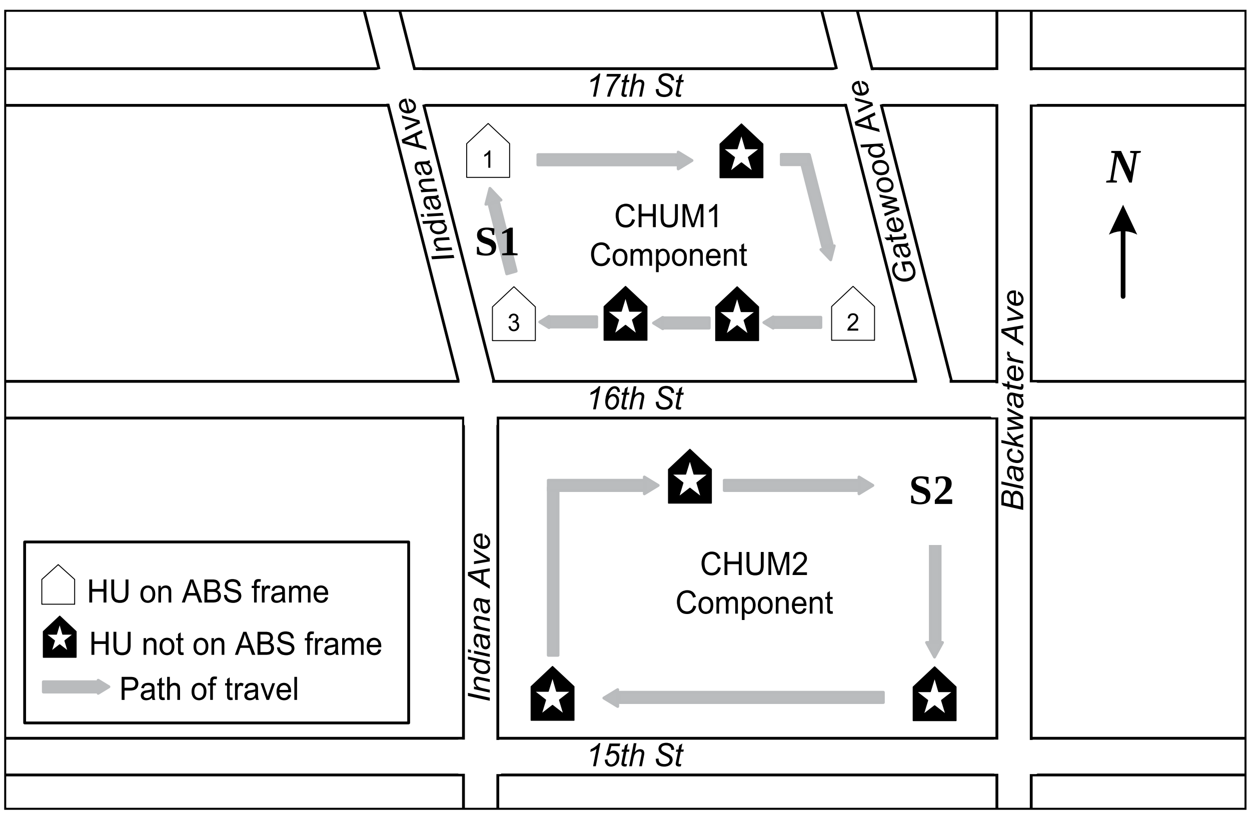
CHUM1 is designed to identify housing units missing from the ABS frame that are within *geographic intervals* linked to selected housing units in the ABS sample. A geographic interval consists of a selected housing unit and the geographical area from the selected housing unit up to the next housing unit that is listed on the ABS frame, based on the CHUM path of travel.[3] The start points for the CHUM1 procedure are the housing units associated with sampled addresses from the ABS frame. The CHUM has a preestablished path of travel that allows field personnel to identify the "next housing unit" from the starting point, where the path of travel is a clockwise direction around the block. Field personnel, often field interviewers, are instructed not to cross streets in order to ensure that any missed housing unit can be identified from a single address on the frame.

[3] CHUM geographic intervals typically consist of all housing units within a census block.

http://dx.doi.org/10.3768/rtipress.2016.mr.0034.1602

After field personnel determine the address of the next housing unit, they check whether the address is included on the ABS frame. If the address is on the ABS frame, then the CHUM1 procedure is complete—there are no missed housing units in the interval following the start point. However, if the address of the next housing unit is not contained on the ABS frame, then the address is recorded and the interviewer continues to navigate the geographic interval until a housing unit with its address on the ABS frame is located (the end point). If the block contains no other housing units on the frame, the path of travel will eventually return to the selected housing unit, which in this case constitutes the end as well as the beginning of the interval. The CHUM1 procedure is complete in the block only when the end point has been reached; any newly identified missed housing units in the geographic interval between the start and end points are assigned a probability of selection and may be randomly chosen for the study (see the section Sampling Issues on p. 5 for further details). Figure 1 illustrates the CHUM 1 procedure for an example geographic area. (It also illustrates the CHUM2 procedure described next.)

Figure 1. CHUM1 (check for missed units) and CHUM2 (check for missed blocks) procedures



HU = housing unit; ABS = address-based sampling

CHUM1: In this example, if housing unit 1 were selected for the study, then the field interviewer would travel clockwise around the block until reaching housing unit 2, picking up the starred housing unit missing from the address-based sampling (ABS) frame. Alternatively, if housing unit 3 were selected, no missed units would be detected before reaching the endpoint (housing unit 1).

CHUM2: If the block at the top of the figure were sampled for the CHUM2, the field interviewer would locate the start point, S1, and travel clockwise until reaching housing unit 1, which is contained on the ABS list. The CHUM2 procedure would then terminate and no dwellings would be added to the frame. However, if the block at the bottom of the figure were sampled for the CHUM2, the field interviewer would locate the start point, S2, and circumnavigate the entire block without finding the addresses of any housing units on the ABS frame. This is a missed block, and all three missed housing units on this block would be sampled for inclusion in the study.

Check for Missed Blocks (CHUM2)

If only the CHUM1 procedure is implemented, then housing units in geographic intervals within the selected areas with no ABS coverage do not have a chance of selection. The CHUM2 procedure ameliorates this problem. During the sample selection stage, one or more census blocks within each area segment chosen for the study are randomly selected for the CHUM2. To minimize design effects associated with the base weights, CHUM2 blocks can be selected such that the weights of missed housing units are similar to the weights of sampled addresses from the ABS frame. An alternative approach is to select CHUM2 blocks based on probability proportional to size sampling, giving higher probabilities of selection to blocks expected to contain large numbers of housing units missing from the frame, as approximated by a coverage prediction method. For example, ABS coverage can be predicated by taking the ratio of ABS frame counts to census counts or using a coverage predication model (Iannacchione et al., 2012; Montaquila et al., 2011).

For each segment, project staff select CHUM2 start point(s) to ensure complete coverage of the census block by field personnel based on the path of travel. Multiple CHUM2 start points are sometimes required to allow complete coverage of the selected CHUM2 block, depending on the geographic characteristics of the census block (e.g., if a street bisects the census block, two start points are needed to allow coverage of both geographic intervals within the census block). The start points are typically selected in a clearly defined location (e.g., the intersection of two streets) so that field personnel can easily determine where to begin implementing the CHUM2 procedure. This differs from the CHUM1 procedure that uses a sampled housing unit as the start point.

Field personnel must first canvass selected CHUM2 blocks to ensure the accuracy of the maps used to select CHUM2 areas. If field personnel identify discrepancies between the map and the structure of roads on the ground, they notify project staff for further instruction. After verifying the accuracy of CHUM2 maps and locating the prespecified start point, the protocol for implementing the CHUM2 procedure is the same as the protocol for CHUM1. From the start point, field personnel follow the specified path of travel to locate the "next housing unit" and check whether or not the address of the next housing unit is on the ABS frame. If it is, the procedure is complete. Otherwise, field personnel continue listing and checking housing units until they either locate a housing unit with an address on the ABS frame or return to the start point. If the CHUM2 geographic interval contains any housing units associated with ABS addresses, it is an area covered by the frame and no housing units are added to the frame. In this case, any missed housing units in the geographic interval are covered by the CHUM1 procedure. However, if no housing units in the CHUM2 interval are included on the ABS frame, then this area is not covered by the frame or CHUM1, and all housing units in the CHUM2 interval are assigned a probability of selection and are eligible for inclusion in the study. Figure 1 provides an illustration of the CHUM2 procedure.

Sampling Issues for the CHUM Methodology

The base probability of selection for a missed housing unit $m1i$ identified with the CHUM1 procedure (π_{m1i}) is the unconditional probability of selection for the sampled housing unit $l1i$ to which it is linked, taking into account all stages of sampling (π_{l1i}). However, it is not always practical to sample all missed housing units identified in the CHUM1 procedure, particularly when large numbers of missed units are identified. Subsampling of missed housing units can control the overall sample size in the segment. The base probability of selection must be adjusted for any subsampling of missed housing units. The adjusted probability of selection for a missed housing unit is

$$\pi_{m1i} = \pi_{l1i} \times \frac{n_{s1}}{n_{a1}},$$

where n_{s1} represents the number of subsampled missed housing units out of the n_{a1} missed housing units identified through CHUM1 from the interval beginning with housing unit i. For example, if the starting housing unit had an unconditional probability of selection of 0.05, and if 7 of the 20 missed housing units identified in the CHUM1 procedure were sampled for inclusion in the study and were linked to this unit, the subsampled missed housing units would have a base probability of selection of 0.05 × 7 / 20 = 0.0175.

With the CHUM2 procedure, the (unconditional) probability of selection for a missed housing unit $m2i$ (π_{m2i}) in a selected block with no housing units listed on the frame is the product of the segment selection probability (π_s) and the conditional probability of selecting the block for CHUM2, given the segment (π_{b2}). As with the CHUM1, the probability of selection for missed housing units identified with CHUM2 must be adjusted for any subsampling. That is,

$$\pi_{m2i} = \pi_s \times \pi_{b2} \times \frac{n_{s2}}{n_{a2}},$$

where n_{s2} represents the number of subsampled missed housing units out of the n_{a2} missed housing units identified in this interval with CHUM2. For example, consider a segment with a selection probability of .01. If CHUM2 blocks were selected based on a simple random sample of blocks in the segment, and if 2 of the 30 census blocks in the segment were selected, then the conditional block-level probability of selection would be 2 / 30 = 0.067. The unconditional probability of selection for the CHUM2 block is 0.01 × 0.067 = 0.00067. If all missed housing units identified in this CHUM2 block were sampled for inclusion, the base probability of selection for these missed units would be 0.00067.

After determining the unconditional probability of selection for each sampled housing unit i (π_i), including both housing units sampled from the ABS frame and the CHUM, the base weight is calculated as the inverse of the probability of selection (i.e., $1/\pi_i$). As with any survey, adjustments for ineligibility, nonresponse, and coverage (i.e., calibration) are applied to the base weights to limit biasing effects associated with each issue to form the analysis weights, regardless of whether the housing units were included on the ABS frame or were added with the CHUM.

As previously mentioned, missed housing units identified with the CHUM procedures may be subsampled as a cost- and time-saving measure. However, differential subsampling can increase the variation in base weights and potentially lower the precision of the estimates (e.g., Kish, 1965; Valliant et al., 2013). Thus, statisticians must consider both factors in their decision to subsample, and if so, at what rate. Large numbers of missed housing units can lead to challenges in controlling the sample size and subsequently, if significant subsampling is implemented, statistical inefficiencies will result from differential weights. Alternative approaches, such as a hybrid sampling frame where field enumeration is used in areas where ABS coverage is deemed inadequate, can mitigate these statistical inefficiencies.

CHUM Operational Issues

Although the CHUM theoretically provides complete coverage for housing units missing from the ABS sampling frame, it is dependent on field personnel correctly implementing the methodology. Several operational aspects of the CHUM procedures can be tailored to the resources and goals of a particular study.

Tailoring Where and When to Implement CHUM

The CHUM provides survey researchers with flexibility in designing ABS studies, as it does not need to be completed for all sampled addresses or area segments in the study. If a study does not have the resources available to implement the CHUM everywhere, or if the target population resides in areas where address undercoverage is less of a concern (e.g., urban areas), then the CHUM1 procedure can be completed for a subset of sampled addresses. If resources are limited, the CHUM can be focused on areas where ABS undercoverage is expected. Similarly, depending on the resources available, the expected ABS coverage, and the sample design, any number of CHUM2 blocks can be selected for inclusion. As described earlier, subsampling for cost savings can adversely affect precision, so the sample design should balance cost savings associated with limited implementation of the CHUM with precision goals.

In addition to tailoring where the CHUM is conducted, researchers can tailor when the CHUM is implemented. The CHUM can be implemented at any point after the sample of addresses has been selected. Field personnel can make a trip to sampled segments to complete the CHUM prior to screening and interviewing, or they can implement it during data collection. Implementing the CHUM prior to data collection allows time to develop the final sample of addresses before the start of screening and interviewing and to determine appropriate subsampling rates. However, additional resources are required to visit segments prior to data collection, and frame updates are less timely.

Developing CHUM Materials

Regardless of the sampling frame within a selected segment (ABS, field enumeration, or hybrid), area probability surveys require maps to help field interviewers navigate to and between selected addresses and to determine the segment boundaries. When the CHUM is incorporated in the design, field interviewers are generally also provided with lists of sampled units for CHUM1 and the location of CHUM2 start points. To implement the CHUM, field interviewers also need a method for searching the ABS frame within the sample segment to identify housing units whose addresses are already on the frame. This methodology can be as simple as a printout of addresses in that segment or as complex as an application that allows interviewers to search for addresses electronically. Either approach can provide prepopulated lists of candidate next addresses, but if such lists are provided, care should be taken to avoid confirmation bias by field personnel—that is, survey researchers should follow quality control measures to ensure that the field personnel are not treating the list as correct and complete and therefore assuming no updates are needed (Eckman and Kreuter, 2011).

Quality Checks for CHUM Implementation

Prior research has shown that field personnel do not always complete field work correctly, whether implementing the HOI, an enhanced listing procedure, or the CHUM (Eckman and Kreuter, 2011; Eckman and O'Muircheartaigh, 2011; Iannacchione et al., 2012). One method for monitoring the quality of CHUM fieldwork is to create situations where field personnel should find at least one missed housing unit; that is, remove addresses from the address lists that are likely to be the next housing unit for their CHUM start points. Survey researchers have manipulated field personnel's address lists to evaluate the effectiveness of other supplementation techniques as well, including enhanced listing (Eckman and Kreuter, 2011). Even though this seeding method is not perfect, it allows sampling personnel to measure how well the CHUM is being implemented. Seeding does have cost implications, as it puts an additional burden both on field personnel and sampling personnel to implement and verify unnecessary CHUM intervals, but the data quality benefits and validation that the CHUM is being implemented correctly typically outweigh these costs.

The seeding technique has been used to evaluate the CHUM for a few studies. In the 2008 American National Election Study, field staff identified 72.2% of the 231 CHUM1 and 100% of the 117 CHUM2 seeded intervals (McMichael et al., 2013). In the 2010 National Children's Study in Los Angeles County (NCS-LA), field staff correctly implemented 94.3% of the 53 seeded CHUM1 intervals. The improved CHUM1 implementation rate for the NCS-LA in comparison with the American National Election Study is attributed to enhanced training on the procedure and the urban population.

More sophisticated monitoring techniques are possible when field personnel have global positioning system (GPS) capabilities in the field, but these techniques have not yet been evaluated. With GPS, researchers can monitor the location of field personnel to ensure they are following the correct path of travel; if not, they can identify and correct mistakes quickly enough that the CHUM procedure can be repeated when necessary. Although GPS accuracy rates are not perfect, GPS capabilities could be useful for gauging field personnel performance. Field tests are needed to fully assess the feasibility and costs associated with GPS monitoring of field personnel.

RTI Press Publication No. MR-0034-1602. Research Triangle Park, NC: RTI Press. http://dx.doi.org/10.3768/rtipress.2016.mr.0034.1602

Benefits and Limitations of Using ABS with CHUM

As previously discussed, there are other options for reducing undercoverage in ABS studies, including enhanced listing and the ACE procedure. The timeline, segment size, and impact on variances are all considerations in choosing the best frame supplementation approach for a particular study.

If used, enhanced listing, another frame correction method, must be completed prior to sample selection, whereas CHUM1 is necessarily completed after selecting the sample, typically at the start of screening and interviewing. The ACE procedure can be completed either prior to or following selection of the sample, but typically is done after selection in conjunction with data collection. There are advantages and disadvantages associated with the timing of each approach. Implementing frame supplementation following sample selection is beneficial for studies with aggressive data collection schedules. In addition, supplementing the frame during data collection reduces the number of field personnel visits to the segment, which can result in cost savings. Supplementing the frame during data collection also provides more timely updates compared to methods that supplement prior to sample selection, providing coverage for housing units constructed between the sample selection and data collection stages. However, supplementing the ABS frame during data collection puts additional burden on field personnel during the interviewing stage.

The CHUM requires less matching of addresses to the ABS list than the other procedures do. Both enhanced listing and the ACE procedure require field personnel to compare all segment addresses found in the field to addresses on the ABS list. This process can be ambiguous, particularly in the case of "fuzzy matches" (e.g., the ABS frame has Apartment 1–3 even though the units are actually labeled A–C). The ACE procedure further requires all found housing units—not just addresses associated with sampled segments—to be matched to the entire ABS frame. In contrast, the CHUM requires addresses only in designated intervals to be compared to the ABS list for the sample of selected addresses and CHUM2 blocks within the same segment.

Enhanced listing and the ACE procedure require field personnel to canvass entire area segments to supplement the ABS frame; in contrast, the CHUM is implemented only from randomly selected starting points to the next address on the frame. Because the CHUM is therefore less time consuming for field personnel, it may be the better choice for studies with large segment sizes. For example, some ABS studies use census block groups as the area segments, which average around 500 households. Checking all addresses associated with census block group–sized segments would be quite time-consuming and potentially resource-prohibitive.

The selection of a frame supplementation procedure also affects the variances of resulting estimates. As previously mentioned, the CHUM is more feasible in large area segments than procedures that require entire area segments to be canvassed. For the same target sample size, larger segment sizes pose advantages from a statistical efficiency perspective. Area cluster designs are subject to reduced efficiency due to similarities in respondents within area clusters, and these cluster effects are often larger for smaller area segments (Valliant et al., 2013).

Nevertheless, enhanced listing or the ACE (implemented prior to sample selection) may be better options in studies where a large number of addresses are likely to be added to the frame. Including them ahead of time allows the statistician more control over the probabilities of selection associated with the added addresses, reducing inefficiencies from differential weights. In areas where ABS frame coverage is very low, the work associated with CHUM approaches that of the other methods.

Based on a predicted amount of coverage from the ABS frame, a study can utilize the benefits of both ABS and field enumeration. Where address coverage is expected to be sufficient, ABS supplemented with the CHUM can be implemented and field enumeration can be retained in low coverage areas. This combination mitigates the concern of areas with large numbers of missing addresses (Iannacchione et al., 2012).

Figure 2. Comparison of address-based sampling (ABS) supplementation procedures

Check for Housing Units Missed (CHUM)

- **Summary:** Field personnel search for missed housing units in selected geographic intervals within selected segments.
- **Timeliness:** Frame updates are made during data collection (timely updates); however, this puts additional burden on field personnel.
- **Cost Considerations:** Because updates are made only in selected intervals, this procedure can be less costly to implement than other procedures; however, it can become inefficient when segment coverage is very low. Furthermore, if many missed housing units are found, subsampling may be desirable to keep data collection costs low.
- **Precision Implications:** Probabilities of selection for added units are associated with randomly selected start points, so procedure requires careful control of subsampling rates. Subsampling rates have a direct impact on the precision of estimates. Because this procedure does not require verifying all addresses in area segments, it allows for larger segment sizes, which can improve the precision of resulting estimates.

Enhanced Listing

- **Summary:** Field personnel update the ABS list in sampled area segments (add, remove, and correct addresses) before selecting the housing unit sample.
- **Timeliness:** Frame updates are made before sample selection (less timely), which can lead to undercoverage of newly constructed units.
- **Cost Considerations:** Efficient when frame is used for multiple studies; can be costliest of the three methods when the frame is developed for a single study.
- **Precision Implications:** Because supplementation occurs before the selection of housing units, this procedure allows more control over sampling rates and the resulting precision of estimates; however, because of the costliness of updating all addresses in sampled segments, segment sizes are sometimes kept small. For fixed sample sizes this can lead to lower statistical precision.

Address Coverage Enhancement (ACE)

- **Summary:** Segments are defined by geographical boundaries and include all housing units that geocode into them. Field personnel canvass a subset of sampled area segments and identify housing units not on the ABS frame. All addresses that geocode into area segments are included on the frame.
- **Timeliness:** Frame updates are made either before sample selection or during data collection. When implemented prior to sample selection, updates are less timely and can lead to undercoverage of newly constructed units. When implemented during data collection, this procedure puts additional burden on field personnel.
- **Cost Considerations:** Entire segments are searched, but the search is limited to a subsample of segments. Fewer corrections need to be made than with the other methods because corrections are not made for geocoding error. Addresses must be checked against the entire ABS frame, not just the frame for the sampled segment.
- **Precision Implications:** When supplementation occurs before housing units are selected, this procedure allows more control over sampling rates and the resulting precision of estimates; however, because entire segments are searched, lower-budget studies will need to either limit segment sizes or sample fewer segments for the ACE. Both options affect statistical precision.

Figure 2 compares the timeliness, cost considerations, and precision implications for the three most commonly used ABS supplementation procedures: the CHUM, enhanced listing, and ACE.

Summary

ABS is increasingly the preferred approach for in-person area probability sampling studies because of the potential cost savings associated with the ABS methodology. Although the combined Computerized Delivery Sequence and No-Stat frames provide mailing addresses for nearly all housing units in the United States, they do not provide full housing unit location coverage sufficient for in-person surveys. Several approaches have been developed to supplement the ABS frame where needed, including using field enumeration sampling frames where using ABS frames are deficient or using ABS with one or more supplementation methods: enhanced listing, the ACE procedure, or the CHUM.

Enhanced listing and the ACE procedure (when implemented prior to sample selection) provide more statistical control for sampling addresses identified in the field, as researchers can select the sample of addresses from a complete frame that has already been supplemented. This eliminates the need for subsampling that often occurs with the CHUM. However, the CHUM provides some operational advantages over enhanced listing. The CHUM occurs at the start of screening and interviewing rather than at the frame-building stage, so address updates are timelier. Furthermore, the CHUM can avoid multiple trips to the segment, which saves resources. The CHUM is implemented from only a sample of addresses in each segment, which makes it less time consuming than methods that require field personnel to canvass entire segments looking for missed addresses. These time savings also allow the CHUM to be implemented in larger geographic segments compared to methods that canvass the entire segments, which has the likely benefit of reducing intraclass correlations. Because of the advantages associated with the CHUM, it should be considered for in-person ABS studies.

Regardless of which supplementation technique is implemented, field personnel must be appropriately trained, monitored, and supported while performing field work. The logistics of these procedures can be tailored to the needs and resources of the study while maintaining high coverage and ensuring data quality.

References

Dohrmann, S., Han, D., & Mohadjer, L. (2007). *Improving coverage of residential address lists in multistage area samples*. In *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 3219–3226. Retrieved January 5, 2016, from http://www.amstat.org/sections/SRMS/Proceedings/y2007/Files/JSM2007-000570.pdf

Dohrmann, S., Kalton, G., Montaquila, J., Good, C., & Berlin, M. (2012). *Using address based sampling frames in lieu of traditional listing: A new approach*. In *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 3729–3741. Retrieved January 5, 2016, from http://www.amstat.org/sections/SRMS/Proceedings/y2012/Files/304436_73106.pdf

Eckman, S., & English, N. (2012). Creating housing unit frames from address databases: Geocoding precision and net coverage rates. *Field Methods, 24*(4), 399–408. http://dx.doi.org/10.1177/1525822x12445141

Eckman, S., & Kreuter, F. (2011). Confirmation bias in housing unit listing. *Public Opinion Quarterly, 75*(1), 139–150. http://dx.doi.org/10.1093/poq/nfq066

Eckman, S., & O'Muircheartaigh, C. (2011). Performance of the half-open interval missed housing unit procedure. *Survey Research Methods, 5*(3), 125–131. http://dx.doi.org/10.18148/srm/2011.v5i3.4737

English, E., Dekker, K., & O'Muircheartaigh, C. (2013). *Correcting coverage deficiencies in address-based frames: The use of enhanced listing*. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Washington, DC.

Iannacchione, V. G. (2011). The changing role of address-based sampling in survey research. *Public Opinion Quarterly, 75*(3), 556–575. http://dx.doi.org/10.1093/poq/nfr017

Iannacchione, V., Morton, K., McMichael, J., Cunningham, D., Cajka, J., & Chromy, J. (2007). Comparing the coverage of a household sampling frame based on mailing addresses to a frame based on field enumeration. In *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 3323–3332. Retrieved January 5, 2016, from http://www.amstat.org/sections/SRMS/Proceedings/y2007/Files/JSM2007-000675.pdf

Iannacchione, V. G., McMichael, J. P., Shook-Sa, B. E., & Morton, K. B. (2012). *A proposed hybrid sampling frame for the National Survey on Drug Use and Health, final report*. Retrieved December 5, 2016, from Substance Abuse and Mental Health Services Administration website: http://www.samhsa.gov/data/sites/default/files/NSDUHMethodsRptHybrid2012/NSDUHMethodsRptHybrid2012.pdf

Kalton, G., Kali, J., & Sigman, R. (2014). Handling frame problems when address-based sampling is used for in-person household surveys. *Journal of Survey Statistics and Methodology, 2*(3), 283–304. http://dx.doi.org/10.1093/jssam/smu013

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

McMichael, J., Ridenhour, J., & Shook-Sa, B. (2008). *A robust procedure to supplement the coverage of address-based sampling frames for household surveys*. In *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 4329–4335. Retrieved January 5, 2016, from http://www.amstat.org/Sections/Srms/Proceedings/y2008/Files/ridenhour.pdf

McMichael, J., Ridenhour, J., & Shook-Sa, B. (2009). *Evaluation of address based sampling (ABS) frame supplementation methods for in-person household surveys*. Hollywood, FL: American Association for Public Opinion Research.

McMichael, J., Ridenhour, J., Mitchell, S., Fahrney, K., & 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*. In *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 4251–4257. Retrieved January 5, 2016, from http://www.amstat.org/sections/srms/Proceedings/y2008/Files/mcmichael.pdf

McMichael, J., Shook-Sa, B., Ridenhour, J., & Harter, R. (2013). *CHUM: A frame supplementation procedure for address-based sampling*. Paper presented at the Federal Committee on Statistical Methodology Research Conference, Washington, DC.

Montaquila, J. M., Hsu, V., & Brick, J. 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), 317–335. http://dx.doi.org/10.1093/poq/nfr008

Morton, K., J. McMichael, J. Cajka, R. Curry, V. Iannacchione, & Cunningham, D. (2007). *Linking mailing addresses to a household sampling frame based on census geography*. In *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 3971–3974. Retrieved January 5, 2016, from https://www.amstat.org/sections/srms/Proceedings/y2007/Files/JSM2007-000774.pdf

O'Muircheartaigh, C., English, E., & Eckman, S. (2007). *Predicting the relative quality of alternative sampling frames*. In *JSM Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 3239–3248. Retrieved January 5, 2016, from https://www.amstat.org/sections/srms/Proceedings/y2007/Files/JSM2007-000575.pdf

Shook-Sa, B. E. (2014). Improving the efficiency of address-based sampling frames with the USPS No-Stat File. *Survey Practice, 7*(4), 1–10. Retrieved January 5, 2016, from http://www.surveypractice.org/index.php/SurveyPractice/article/view/264/html_8

Shook-Sa, B. E., Currivan, D. B., McMichael, J. P., & Iannacchione, V. G. (2013). Extending the coverage of address-based sampling frames beyond the USPS computerized delivery sequence file. *Public Opinion Quarterly, 77*(4), 994–1005. http://dx.doi.org/10.1093/poq/nft041

Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York: Springer.

www.rti.org/rtipress **RTI Press publication RR-0034-1602**