

Evaluating the Use and Effectiveness of the Half-Open Interval Procedure for Sampling Frames Based on Mailing Address Lists in Urban Areas

Joseph P. McMichael¹, Jamie L. Ridenhour¹, Susan Mitchell², Kristine Fahrney¹, and Wanda Stephenson¹

¹RTI International, 3040 Cornwallis Road, Post Office Box 12194, Research Triangle Park, NC 27709

²RTI International, 701 13th Street, N.W., Suite 750, Washington, DC 20005

Abstract

The use of sampling frames constructed from mailing address lists is growing as researchers seek a cost-efficient alternative to counting and listing for sample frame development. The potential for coverage bias using mailing lists is known to exist because the lists are not always complete or up to date. Researchers have developed methods to supplement this undercoverage. The Half-Open Interval (HOI) frame-linking procedure (Kish, 1965) is a method that is often used to help reduce the undercoverage associated with household sampling frames. In a field survey, the HOI procedure adds dwelling units (DUs) to an existing frame by instructing interviewers to search for new units in the interval between the selected DU and the next DU on the frame. New DUs that are discovered are automatically included in the sample. The current literature is lacking information about how to effectively train interviewers to implement HOI procedures and how successful the interviewers are in identifying missed addresses.

This paper discusses the protocol used to implement the HOI procedure in the Study of Community Family Life (SCFL), an in-person survey of about 13,000 households in six low-income urban areas, sponsored by the Administration for Children and Families, U.S. Department of Health and Human Services. We answer the following research questions about HOI and its implementation: How were interviewers trained to identify missed addresses? How often were they able to construct an interval? What obstacles did they encounter? Finally, based on an experimental design, we report on the interviewer error rate in identifying missed addresses. Findings will add to the literature on the effectiveness of reducing coverage bias by using HOI procedures in field surveys.

Key Words: Mailing Address List, Address Based Sampling, ABS, Half-Open interval, HOI

1. Study Background

Data for this research come from the Study of Community Family Life (SCFL), part of the Evaluation of the Community Healthy Marriage Initiative. This study was conducted by RTI International and sponsored by the Administration for Children and Families, U.S. Department of Health and Human Services. The objective of the SCFL was to evaluate the community impacts of intervention programs designed to improve marital stability and satisfaction, and family and child well-being among low-income families. The SCFL sampled 13,134 households from 40 zip codes in six urban areas. Zip codes were grouped to form contiguous areas in each of the urban areas. .

Face-to-face computerized interviews were conducted with a total of 4,023 respondents in Dallas, TX; Milwaukee, WI; St. Louis, MO; Cleveland, OH; Ft. Worth, TX; and Kansas City, MO. Baseline data collection began in September 2007 and was completed in March 2008.

This study used address-based sampling (ABS) rather than field enumeration (FE) to develop the sampling frame. ABS is gaining strength as an alternative to traditional field enumeration for studies that do not have either the time or resources to develop a field enumerated frame prior to sampling. In 2006, RTI International conducted an evaluation of ABS frames against traditional field enumeration in North Carolina and found that the overall coverage rate, assuming field enumeration as a gold standard, of ABS to be roughly 82% (Iannacchione *et al.* 2007). Coverage with ABS is not perfect; however, the coverage in urban areas, the focus of the SCFL, is comparable to FE household frames (Dohrmann *et al.* 2006). Sampling frames created with ABS need frame supplementation just as those created through field enumeration do.

ABS studies build their frames from a list of postal addresses, called the Delivery Sequence File (DSF). For the SCFL, a two-stage stratified random sample of addresses was selected from the DSF, purchased from a commercial vendor. Given the known coverage issues with the DSF (O’Muircheartaigh *et al* 2007), and in spite of the fact that this study was conducted in urban areas, we expected that frame supplementation would be key to improving the coverage of the frame. As we describe in this paper, adapting frame supplementation techniques used with field enumeration to an ABS frame can be problematic.

1.1 The Half-Open Interval

In traditional field enumeration, the frame supplementation method often used is the half-open interval (HOI) procedure. HOI is a method where field staff systematically identify addresses not on the frame by examining the interval between the sampled 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. This proximal ordering is natural on the FE frame because it is created by field staff who determine the path of travel for the field interviewer (FI) to follow. For example, if the sampled address was 123 Main Street, the field interviewer would know by using the FE frame that the next address should be 125 Main Street.

If the next address were not the address specified in the path of travel, the FI would report this possible missed housing unit to the home office. The home office would research if this possible missed housing unit is in fact missing. If the possible missed housing unit is indeed missing from the frame, the missed unit would be added to the sample to improve coverage. Housing units may be missing from the frame because of errors at the frame development stage, or because new housing units were built in the time between frame development and data collection.

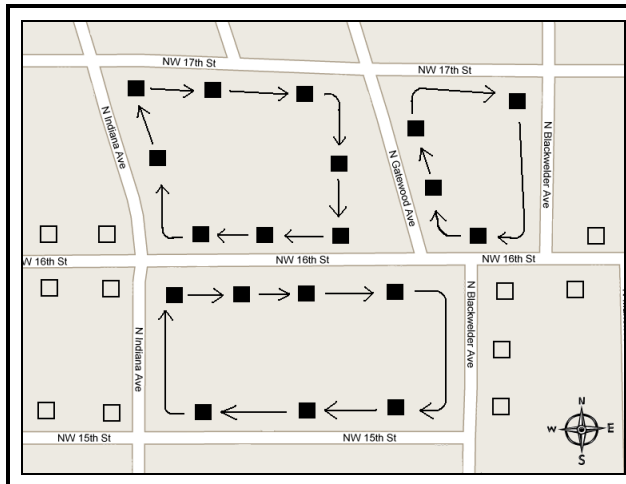


Figure 1a: Path of Travel created by a field enumerated (FE) frame. Arrows not only show the path of travel but indicate the HOI that would be searched for missed housing units.

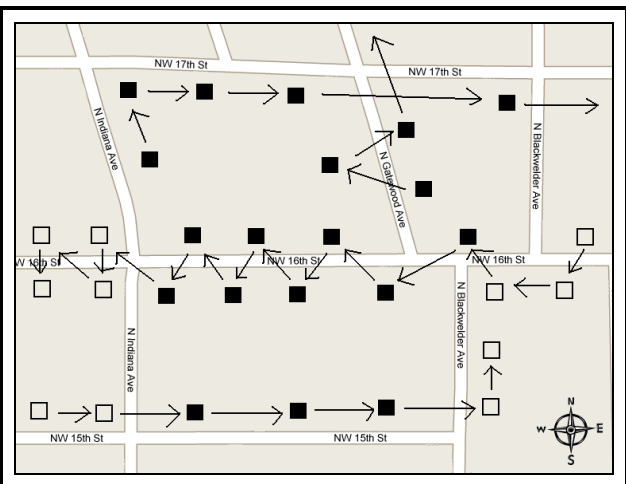


Figure 1b: Path of Travel created by the mail carriers delivery sequence which can be found on the Delivery Sequence File (DSF). Arrows not only show the path of travel but indicate the HOI that would be searched for missed housing units. Note that many HOIs are ambiguous due to street crossings.

Figure 1a displays the path of travel that field staff would create while listing housing units for the sampling frame. Each of the listed houses on the map would also be on a list. In a situation where it is difficult to determine what the address of a house is the field enumerator could provide a house description instead of or in supplement to the address of the house. This additional information provided by the field enumeration process is not available in the list of postal addresses that create the frame in ABS. When the field interviewer returns to a sampled address in this area, the FI would follow this predetermined path of travel to check the interval between the sampled address and the expected next address. Also, if the FI were struggling to find the sampled address there might be additional information, such as a house description, available from field enumeration about the sampled address.

Figure 1b illustrates the equivalent sampling frame as Figure 1a but lacks a clear predetermined path of travel that would inform the FI as to which direction to travel in order to close the interval between the sampled address and the expected next address. The closest approximation to this path of travel is the carrier route delivery sequence information and as illustrated it may not establish a path of travel suitable for survey research.

1.2 Challenges with Applying HOI to Address-Based Sampling

ABS frames originating from the DSF are not organized in a way that is optimal for researchers conducting household surveys; rather, they are organized in a fashion that benefits the mail carrier. The closest information to a proximal sort available on the DSF is the letter carrier's delivery sequence. This information organizes the addresses on the DSF into carrier routes and an order in which the letter carrier would deliver the mail within that route. As shown in Figure 2, when the sampled address and expected next address are not located next to each other it can be difficult to determine what the appropriate path of travel is between the two. The delivery sequence on a letter carrier's route accommodates one-way streets and the multiple routes in which the mail carrier works. Again, this structure is not necessarily ideal for researchers interested in creating a predetermined path of travel from a list of addresses.

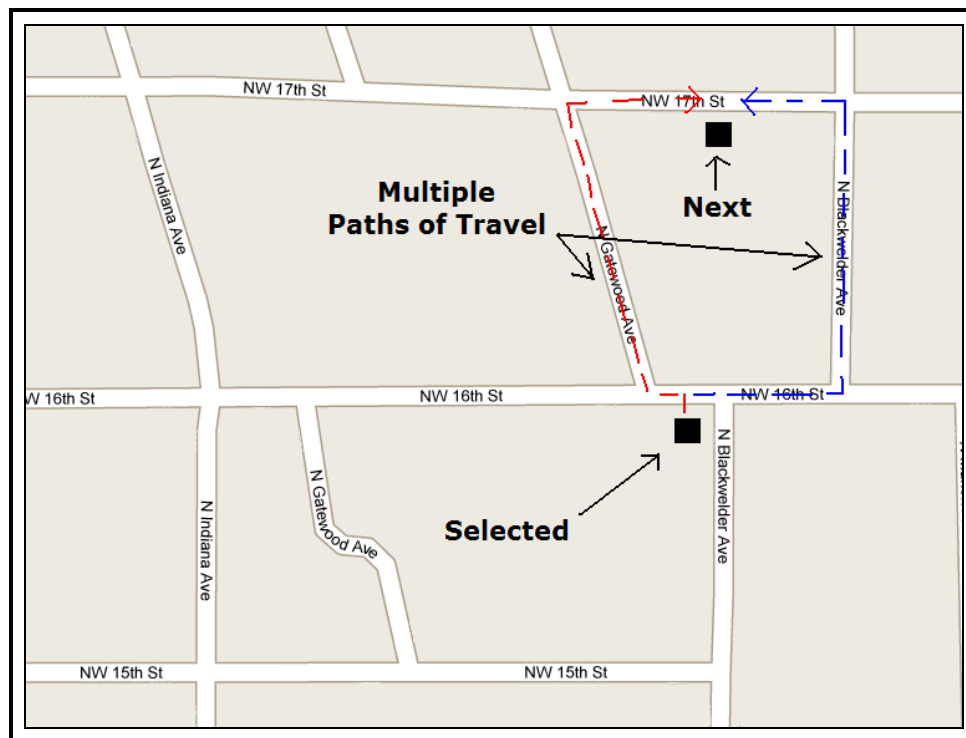


Figure 2: Many paths of travel can exist between the sampled and expected next addresses on an ABS frame.

Determining a path of travel can also be difficult in areas with cluster boxes (e.g. apartments and condominiums). An FI might observe one order of households by the arrangement of the mailboxes in an apartment complex, but this order may not resemble the actual ordering of apartments in apartment buildings. Although determining a path of travel is one of the primary issues with ABS frames, there are other challenges: a small proportion of addresses are not locatable on the ground. FIs cannot locate rural route addresses or post office boxes.

Given all of these challenges the success of a study, from a sampling perspective, rests heavily on the abilities of the FIs to correctly handle any problem they may encounter in the field. One of the ways to best insure FI success is to have home office support that can provide the FIs with real-time direction on how to handle problems and information on when a possible missed DU is truly a missed DU. The home office may not have the opportunity to help FIs address these questions if the FIs are reluctant to implement the HOI procedures because they are unsure about how to correctly implement the fieldwork or because they are rewarded for completing interviews.

2. Field Implementation

2.1 Training Field Interviewers

On the SCFL, field interviewers received extensive training on the HOI procedure. The second day of the five-day, in-person training program focused specifically on procedures for validating the whether DUs were on the ABS frame. The presentation-style training detailed an explanation of the procedures and different scenarios the FIs might encounter during their fieldwork. Specifically, training consisted of an explanation of how the housing units were selected from the DSF, what constituted a missed unit, where to look for missed units (by checking the geographic space and in/on the selected unit's property), how to answer the questions in the screener that helped determine if a unit was actually missed, and the challenges they might encounter when looking for missed units.

Based on prior experience, we anticipated four major challenges the FIs would encounter in the field: the addresses not listed in numerical or alphabetical order, inconsistent numbering of addresses (most likely in a cluster-box type situation), addresses found across the street or in some other location not immediately local to the selected address, and selected addresses and expected next addresses (ENAs) separated by a street. These scenarios were introduced via a lecture using PowerPoint graphics and reiterated through the use of 10 practice exercises. At the end of that day of training, a homework packet was distributed which consisted of four additional exercises. A certification exam was administered the following day.

Because mailing address coverage is fairly high in urban settings, we did not accurately anticipate how often the FIs would encounter problems in constructing an HOI in the field. Further, as fieldwork progressed it became apparent that some FIs simply did not understand how to implement an HOI in the field, despite our training. We did not have good metrics for evaluating HOI progress during data collection which prevented us from knowing if a particular FI was struggling or simply not attempting the HOI if the FI did not inform us of the issue.

The problem the FIs encountered far more than any other was when travel between the selected address and the expected next address required them to cross a street. As shown previously in Figure 2, there can be many ways of traveling between the two addresses in this type of situation. This is problematic because without a pre-specified path of travel or a training protocol that specifically outlines how to handle this situation the FIs would not determine the path consistently thus leading to the possibility that some missed addresses would have multiple opportunities to be picked up for the study and some missed addresses would have no opportunity at all to be found. Not only are there multiple ways to travel between the two addresses but every street crossing is opening up the interval between the two addresses in a way that makes it difficult, or impossible, to close. Shortly after data collection began, the sampling team devised a job aid with an alternate set of instructions for dealing with common problems. This job aid instructed FIs how to proceed when specific problems arose and significantly reduced the volume of cases sent to the home office for resolution.

2.2 The Field Protocol: Accommodating New Challenges

The central idea in the protocol for checking for a missed address is that the FIs would examine the geographic space between the selected address and the expected next address. If the FI encountered a DU in that space before reaching the ENA this possible missed address was reported to the sampling team at the home office for further evaluation. FIs were also instructed to check for additional DUs in or on the selected DU's property. If a selected address had another DU, for example a basement apartment or a converted pool house, a series of questions in the screener helped to determine if the DU was missing from the frame. Interviewers were instructed to report all potential missed addresses to the home office, where the sampling team checked those addresses against the frame to determine if the address was in fact missed. Many times, these possible missed addresses were not missed; they were on the frame but sorted by the carrier route delivery sequence which does not order the addresses in a geographically ordered way.

As data collection progressed, several situations that FIs encountered time and time again in the field necessitated some revisions to the protocol they had learned in training. To accommodate situations where the expected next address was either across the street (e.g. the sampled address was odd and the expected next was even on the same street) or on an entirely different street altogether, the field protocol was modified as explained in the next section.

2.2.1 Field Protocol Revisions

If the ENA was found across the street (or more than one street away) from the sampled address, instead of crossing the street and opening up the interval between the sampled address and the next address the FIs were instructed to instead determine the actual next address on the ground by enumerating the next addresses from the sampled address. This issue of street crossings was critical to address because crossing a street would open up the interval between the sampled address and the next address in a way that would make it difficult to close, and subsequently difficult to determine which DUs fell inside the interval and which did not.

Thus, we developed a revised protocol only for situations that required the FI to cross streets in the interval between the sampled and next addresses. Ultimately this shift meant that instead of determining the expected next address from the frame the FIs assisted the sampling team in establishing a new next address in the field.

To implement this new protocol, the FI was instructed to face the sampled DU and travel in a clockwise direction (moving left) and staying on the same block (i.e. not crossing a street). If the FI traveled all the way around the block without finding any other addresses and arrived back at the sampled address the task of determining the next address on the ground was complete. The FIs were instructed to list up to three houses in completing this task for any sampled address in this type of situation. The FI then sent an email to the sampling team at the home office with this listing information. While the sampling team investigated whether these listed DUs were on the frame the FI was free to screen the sampled DU. In rare situations where none of the three addresses the FI listed were on the frame the FI was asked to continue listing until an address that was on the frame was found in the field.

2.2.2 Apartments with Cluster Boxes

Another situation that was consistently challenging for the FIs was cluster boxes. When FIs visited an apartment or other dwelling unit that had a cluster box, they checked for missed addresses by looking at the mail boxes. First, they looked for addresses that fell numerically or alphabetically between the sampled unit and the ENA. For example, if the sampled address was Apartment 3 and the ENA was Apartment 4 an address that would be considered to fall in between the two would be Apartment 3A. This situation is also an example of another type of situation the FIs were trained to evaluate: check to see if the units were numbered in an inconsistent manner. Inconsistent numbering is typically something that is a combination of letters and numbers that is not used consistently throughout the range of units. To continue the previously given example, if a building had Apartment 3 and Apartment 3A but the following units were Apartment 4, Apartment 5, and Apartment 5 then this would be a case of inconsistent numbering. If the FIs noted missed units or an inconsistency, they took notes about the numbering scheme and counted the units.

Numbering that differed between the frame and field (e.g. four apartments on the frame are A, B, C, D but in the field they are 1, 2, 3, 4) was another situation in which the FIs would seek assistance from the sampling team. Although we did not change the protocol for handling cluster boxes during data collection, it was the other main source of confusion for the FIs and communication between the FIs and the home office.

3. Results

In spite of all of these challenges, the FIs were able to check for missed addresses in all but 17 cases. This excludes controlled access situations where FIs were unable to access the sampled address (e.g., the sampled address was in a gated community). Often when FIs were not able to construct an HOI it was because they could not determine the house number.

The frame supplementation procedure for this study was a blend between the HOI we set out to do at the beginning of the study and the new protocol we developed to address the common problems we were seeing with implementing the HOI on an ABS frame. After the new protocol was developed most of the communications with FIs were about reporting potential missed addresses rather than attempting to resolve situations in which the FI did not know how to proceed. As mentioned, these addresses were almost always found on the sampling frame. The frame supplementation procedure did result in 59 housing units being added to the 13,134 on the sample frame, improving coverage in a small way.

To test how reliable the interviewers were in reporting potential missed addresses, we embedded an experiment in the data collection. For 263 cases, we gave incorrect expected next addresses to the field interviewers. Typically, this

meant we skipped over the next address and gave the following address to the FI as the expected next address. For example, if 123 Main Street was the selected address and 125 Main Street was the next address on the frame, we would tell the FI that 127 Main Street was the expected next address. We expected the FI to tell us that 125 Main Street was a potential missed address. Interviewers were informed about this experiment during training as a means of evaluating their performance. In 125 out of 263 cases (48.5%) the field interviewer successfully reported the possible missed address. This proportion was somewhat lower than we expected given that the missed addresses were easy to discern and interviewers were aware that their compliance was being monitored; however, given the change in protocol during data collection it is not surprising that if checking the interval between the sampled address and the next expected address required the FI to cross a street then they would abandon the old protocol in favor of traveling clockwise around the block in order to establish a new expected next address.

4. Lessons Learned

The most important lesson that we took away from this study is that field interviewers need a simple, basic method to apply in the field that is customized to ABS studies. Too often, the expected next address is not a sensible next address on the ground. The delivery sequence used by mail carriers is somewhat idiosyncratic and not constructed for research purposes. Therefore, the HOI procedure does not translate well from field enumeration to ABS.

Also, real-time monitoring of field staff is essential to the successful implementation of frame supplementation procedures. Too often, FIs either did not recognize they were implementing the HOI procedure incorrectly or did not want to ask for help. If we had recognized compliance problems sooner, we would have provided the job aids and retraining sooner, and targeted help to specific interviewers. We recommend adding a monitoring capability to frame supplementing procedures, such as “seeding” cases with incorrect information as we did on an experimental basis, to identify interviewers with low rates of compliance early on.

Communication with field staff was a persistent problem. Although by the end of the data collection period, the RTI sampling team had fielded over 800 calls and emails from field staff, interviewers often grew frustrated if their calls were not answered immediately. Moreover, there were problems relaying information because the situations were often difficult to describe without drawing pictures.

Timely resolution of HOI problems is important to minimize interviewer travel costs. Because of the large volume of calls and emails we received, we were not always able to respond the same day, particularly during the evening or on weekends. Often interviews were finalized with the sampled address before the HOI interval was formed. Once the interview was finalized, interviewers no longer had information about the case in their possession, and we were limited in what we could do about the HOI problem. When missed units were added to the frame, it often necessitated a return trip to the neighborhood for a single case.

Frame supplementation is essential to ensuring good coverage of a household survey. As stated previously, there are known coverage issues associated with ABS. The HOI frame supplementation method typically used with field enumeration breaks down frequently when implemented in ABS. New methods are needed that are more applicable to the types of situations encountered in an ABS study. These procedures need to be easily taught and understood by FIs to ensure missed housing units have a chance of being included in the study. In response to this finding, RTI has developed a frame supplementation procedure called the Check for Housing Units Missed (CHUM) that is tailored to the nuances of ABS studies and is intended to be much easier for the FIs to implement (McMichael *et al.* 2008). We expect the CHUM will ameliorate many of the problems encountered on this study applying frame supplementation to an ABS study.

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*.

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*.

Kish, Leslie. 1965. *Survey Sampling*. New York: John Wiley & Sons.

O'Muircheartaigh, Colm, Edward English, Stephanie Eckman. 2007. Predicting the relative quality of alternative sampling frames. *Proceedings of the American Statistical Association, Section on Survey Research Methods*.