

## Comparing the Coverage of a Household Sampling Frame Based on Mailing Addresses to a Frame Based on Field Enumeration

Vincent Iannacchione, Katherine Morton, Joseph McMichael,  
David Cunningham, James Cajka, and James Chromy

RTI International

**KEY WORDS:** Locatable mailing addresses, Delivery Sequence File (DSF), Global Positioning System (GPS)

### Abstract

Cost savings, timeliness, and geographic diversity are primary advantages of using mailing addresses instead of field enumeration as a sampling frame for household surveys. The question is whether the advantages of mailing addresses are accompanied by a decrease in the coverage of the household population. Our research is based on a probability sample of 50 segments which were assembled from Census Blocks in North Carolina. Within the geographic confines of each segment, we constructed two frames: one based on locatable residential mailing addresses and the other based on field enumeration. We used Global Positioning System technology to match the housing units (HUs) from each frame, without presuming that either approach is the “gold standard.” Overall, we found that field enumeration included approximately 98 percent of the HUs compared to 82 percent coverage for mailing addresses. When restricted to occupied HUs however, the coverage increased to approximately 99 percent and 95 percent respectively. Equal coverage was found in the 59 percent of occupied HUs located in urban areas. In rural areas however, mailing addresses were found to have significantly lower coverage than field enumeration. Locatable mailing addresses were nonexistent for the 0.4 percent of HUs located in areas without home delivery of mail. We estimate that field enumeration combined with the half-open interval frame supplementation methodology would yield virtually complete coverage of occupied and unoccupied HUs. An analogous methodology based on the letter carrier’s delivery sequence would increase the coverage of locatable mailing addresses by at least 3.4 percentage points.

### 1. Introduction

Field enumeration often is assumed to be the “gold standard” for frame construction for in-person household (HH) surveys. A typical field enumeration process begins with a partitioning of the geographical area associated with a survey population into small areas known as *segments*. A sample of segments is then drawn and field staff are dispatched to enumerate all potential dwelling

units in each selected segment. Several months may be needed to fully canvass a national sample of segments to develop a complete area sampling frame. Without an alternative methodology for frame development, many studies have neither the resources nor the time to implement an in-person mode of data collection that typically results in relatively high response rates (Hox and de Leeuw, 1994).

Mailing addresses are offered to the public by the U.S. Postal Service (USPS) through a nonexclusive license agreement with qualified private companies. The addresses are based on the Delivery Sequence File (DSF), which is a computerized file that contains all delivery point addresses serviced by the USPS, with the exception of general delivery. The primary advantages of using mailing addresses instead of field enumeration as a sampling frame for household surveys are listed below.

**Cost Savings.** Just as mass mailers use address lists as a cost-effective way of saturating their client base, mailing addresses can be used to develop a sampling frame that costs a fraction of field enumeration. The low cost associated with mailing addresses means that fewer resources are needed for frame development, enabling more resources for training interviewers and improving response rates.

**Timeliness.** A sampling frame based on mailing addresses can be developed in a matter of weeks instead of the months usually required for field enumeration. This time savings enables improved coverage of areas with high growth. It also gives survey planners time to retarget the sample in reaction to findings obtained from a pilot survey.

**Geographic Diversity.** Area designs based on field enumeration often are constrained by costs to relatively small segments based on Census Blocks. These small segments can reduce efficiency by introducing intracluster correlation into survey estimates. In contrast, segments based on mailing addresses can comprise entire Census Block Groups or even Census Tracts. These larger

segments can produce a geographically diverse sample of households which can improve statistical efficiency.

Although mailing addresses offer a ‘faster and cheaper’ method of frame construction than field enumeration, the coverage properties of mailing addresses are largely unknown. This leads to the following research question:

*In what situations are the coverage of households using residential mailing addresses equal, better, or worse than field enumeration?*

We attempted to answer this question by using our experience with mailing addresses and field enumeration to design and evaluate a field study that we conducted in North Carolina in May and June 2006. This paper summarizes our methods and findings.

## 2. Residential Mailing Addresses

Although the USPS is prohibited from providing mailers with a complete list of residential mailing addresses, it is authorized to provide services and products to qualified mailing list compilers that enable them to validate the accuracy of their residential mailing addresses down to the physical delivery point. In addition, the USPS can assist qualified list compilers in obtaining accurate delivery address information, identifying erroneous addresses, and updating delivery sequence address information by carrier route.

With over 126 million residential mailing addresses, one such qualified compiler of mailing lists is ADVO, Inc. (American List Counsel 2007). ADVO claims that its residential mailing address database reaches virtually all households in the United States and provides the most complete consumer mailing list available. The ADVO database is updated monthly through direct feeds from the USPS.

The basic data elements of a standardized mailing address include street/box number, city, state, nine-digit ZIP, carrier route, and delivery sequence number. In addition, the USPS maintains a vacant address indicator flag and a seasonal address file with monthly indicators for seasonal delivery. A vacant address is an address that has been unoccupied for 90 days or more. Seasonal and vacant flags are assigned at the discretion of the local post office and letter carrier. Additional information such as name of resident, address coordinates, Census geography (e.g., Tract or Block), and telephone number are not provided by the USPS but may be purchased (with varying levels of accuracy and completeness) from commercial vendors.

While it is not unreasonable to assume that virtually every household in the United States has a mailing address, not

all mailing addresses are suitable for household surveys because interviewers must be able to locate a mailing address “on the ground.” A 2004 study by the U.S. Census Bureau (Corlett, et al. 2006) estimated that 2.68 million HHs or 2.4 percent of the 109.9 million HHs in the United States (2004 American Community Survey) were located in Census Blocks with no city-style (i.e., locatable) mailing addresses. Although the prevalence of unlocatable mailing addresses is diminishing because emergency 911 (E-911) programs are encouraging local governments to convert rural addresses to city-style street addresses, this estimate can be considered a lower bound of the under-coverage associated with a sampling frame based on mailing addresses.

The primary types of mailing addresses that are not locatable for in-person surveys are simplified rural addresses<sup>1</sup> and residential P.O. boxes. Although residential P.O. boxes account for the vast majority of unlocatable household addresses, most can be excluded without loss of coverage because mail also is delivered to the resident’s street address. There are two situations where P.O. boxes present a coverage problem.

***P.O. Boxes in Areas with Home Mail Delivery.*** In this situation, a resident chooses to have mail delivered to a P.O. box instead of to the household. Typically, these households are scattered along postal carrier routes. As we describe later, we have a methodology for including households with P.O. boxes that are on postal carrier routes during the field interviewing process. The second situation is much more problematic.

***P.O. Boxes in Areas without Home Mail Delivery.*** Households that are not on a carrier route do not receive home mail delivery. Instead, the USPS provides P.O. boxes free of charge to residents of households that are located in these areas. Because these areas do not have any locatable mailing addresses, they are difficult to include during the field interviewing process.

In 2004, we used the following commercially available data products offered by the USPS to identify and locate areas without home delivery of mail. These products are updated on a monthly basis.

1. The Delivery Statistics File provides address counts by carrier route for all routes associated with a ZIP Code. We used this product to identify the delivery types (i.e., city, rural, or box) of all addresses associated with 27,306 post

<sup>1</sup> A simplified rural address does not have a street address. Mail delivery is based on the resident’s name, city, state, and ZIP Code. Typically, simplified rural addresses are assigned to all households on a rural carrier route.

offices nationwide. From these, we identified 4,678 post offices that serviced 1.3 million active P.O. boxes. These post offices were designated as serving areas without home delivery of mail.

2. The Address List Management System contains the identification number and address of all postal facilities in the United States. We used this product to associate an address<sup>2</sup> and a set of coordinates with each post office that serviced residential P.O. boxes.

Among the 4,678 post offices without home delivery, the number of active P.O. boxes per post office varied significantly ranging from zero to 10,022 boxes. In September 2004, we called the 30 largest post offices (in number of P.O. boxes) to verify that they did not offer home delivery and to ask about the population they serve. All 30 indicated no home delivery and all provide P.O. boxes at no charge to residents of their local communities. Two of the post offices indicated that they serve prisons while six serve Indian reservations. One of the post offices serves residents of Catalina Island, California, and another serves residents of Grand Canyon National Park. In all, the 30 post offices accounted for almost 98,000 active P.O. boxes.

**Simplified Rural Addresses.** Households on rural carrier routes with simplified rural addresses have mail delivered to their homes but do not have city-style (street) addresses. As a result, the mailing addresses are not locatable for survey purposes. In 2006, Marketing Systems Group (MSG) estimated that about 2.8 percent of the HHs in the United States were assigned to rural carrier routes with simplified rural addresses.

Simplified rural addresses are assigned to an entire rural carrier route which means that entire local areas do not have locatable mailing addresses. Like areas without home mail delivery, HHs in these areas are problematic for coverage purposes. Unlike areas without home mail delivery however, locatable addresses can be obtained for many HHs with a simplified address.

MSG has compiled a listing of *augmented addresses* for simplified rural addresses from commercially available residential databases. MSG estimates that augmented addresses are available for approximately 80 percent of simplified addresses nationwide. Because augmented addresses are commercially available in electronic form, we include them in our evaluation of coverage.

<sup>2</sup> About two thirds of the Post Office addresses were P.O. boxes. For these, we used the geographic centroid of the town or city associated with the ZIP Code of the post office.

**Prior Experience with Mailing Addresses.** RTI has gained insight into the completeness and usability of mailing addresses by implementing two major household surveys that used mailing addresses as a sampling frame. In 2000, we used residential mailing lists, instead of a list derived from field enumeration, to develop a sampling frame for a probability-based in-person survey of 15,000 households in Dallas County, Texas (Iannacchione, et al. 2003). We estimated that the expense of purchasing and processing the entire list of 800,000 residential mailing addresses for Dallas County was less than one tenth of the cost of field enumeration.

In 2002, we used residential mailing lists as a sampling frame for the EQ-5D national household survey that RTI conducted for the University of Arizona (Staab and Iannacchione, 2003). To comply with the ambitious schedule for the survey, we developed a sampling frame in a matter of weeks instead of the months required for onsite enumeration. In addition, RTI field interviewers were able to locate all but 44 of the 12,000 active mailing addresses selected from 120 ZIP Codes from across the country. Among located addresses, 771 were vacant and 186 were nonresidential. The remaining 10,999 addresses (92 percent) were occupied HUs.

### 3. Field Enumeration

Field enumeration entails dispatching listers to completely canvas all potential dwelling units (DUs) in a selected segment. A DU is a generic term used to include both HUs and group quarters<sup>3</sup>. During the enumeration process, if listers have any doubts about whether to list a structure, they are told to list it. HUs include a group of rooms or a single room occupied or intended for occupancy as separate living quarters by nine or fewer unrelated persons. An occupied HU meets the requirement of having separate living quarters if the occupants do not live and eat with any other family group or persons residing in the structure, and there is direct access to the unit from either the outside or through a common hall. Examples of HUs include:

- single family houses;
- flats, or apartments;
- vacant houses or apartments that are intended for occupancy;
- residential units under construction;
- mobile homes or trailers; and
- seasonal dwellings such as summer homes or resort cottages.

<sup>3</sup> A group quarters structure is any single structure in which 10 or more unrelated persons reside and do not live and eat separately from each other.

Because listers do not determine the occupancy of DUs when listing, many listed DUs are later deemed ineligible during data collection. These ineligible DUs include vacant DUs, seasonal and vacation homes, unoccupied new construction, vacant mobile home or trailer location in a trailer lot or mobile home park, and units which are used solely for business purposes. In addition, listers may miss potential DUs because of carelessness or may list a DU that appears to be a single family home but actually contains multiple HUs. In other situations, DUs may be missed because they are difficult to view from roads, or may be contained within locked buildings or gated communities which cannot be accessed.

**Prior Experience with Field Enumeration.** RTI uses field enumeration to develop national area sampling frames for the National Survey of Drug Use and Health (NSDUH) (Morton, et al. 2007) which is fielded annually for the Substance Abuse and Mental Health Services Administration. Since 1999, the NSDUH has been based on an independent, multistage area probability sample of 67,500 respondents drawn from 7,200 area segments from the 50 states and the District of Columbia.

A national validity study of field enumeration conducted for the NSDUH (Cunningham, et al. 2006) found a greater percentage of enumeration errors in rural areas (3.4%) than in urban areas (1.5%). Because North Carolina is more rural than the nation as whole, we expected the coverage of the field enumeration to be slightly lower than the national estimate of 95.1 percent coverage that was found for the NSDUH.

#### 4. Research Design

The objective of the field study is to compare the coverage of a sampling frame based on field enumeration to a frame based on locatable mailing addresses. A field study with national scope was not considered because of the cost of field enumeration. Instead, the study population was confined to HUs in North Carolina. Although group quarters often are included in area sampling frames based on field enumeration, they are not included in the survey population because their inclusion would entail the purchase and extensive screening of business mailing lists (Dohrmann, et al. 2006).

As **Table 1** shows, North Carolina is a good choice for the field study because it is more rural and has more vacant HUs than the country as a whole. In addition, HUs in North Carolina are more likely to have unlocatable mailing addresses as those in the United States. As a result, the findings from the field study are likely to be conservative with respect to the coverage of mailing addresses.

**Sampling Approach.** Initially, we planned to base the sampling design for the field study on clusters of HUs defined by postal carrier routes. Our rationale for using postal geography (in the form of carrier routes) instead of Census geography (in the form of Census Blocks) was to avoid errors in the geocoding process that is needed to crosswalk a mailing address into a Census Block. Instead, the vendor simply queries the database for all addresses falling into a given postal carrier route. The carrier route attribute is not derived by geocoding or any other potentially inaccurate process.

Before finalizing the research design, we implemented a pilot study on an urban and a rural carrier route in the Triangle region of North Carolina to determine the feasibility of performing field enumeration of an area defined by a carrier route. We found that carrier routes often are composed of disjoint streets, which complicates the field enumeration process. In addition, we purchased postal carrier route boundaries from a commercial vendor (TeleAtlas, NA) and found that the boundaries are updated much less frequently than lists of commercial mailing addresses. We concluded that the dynamic nature of postal carrier routes, especially in areas of high growth, makes it difficult to define areas that can be synchronized with a commercial mailing list.

Based on these findings, we developed a sampling design based on Census geography which typically is used for field enumeration. Census Blocks are contiguous with boundaries based on discernable landmarks. In addition, Census geography covers the entire land mass of North Carolina which enabled us to include HUs in areas with unlocatable mailing addresses in the evaluation.

The target population for the field study comprises all HUs (vacant or occupied) in the state of North Carolina in May and June, 2006. The construction of the sampling frame began with a file containing one record for each of the 232,363 Census Blocks in North Carolina. Census Blocks were then partitioned into three strata:

1. areas without home mail delivery,
2. areas with simplified rural addresses, and
3. other areas.

Using data provided by MSG, we constructed a list of incorporated place<sup>4</sup> names that consisted entirely of P.O. boxes and a second list of place names that consisted entirely of simplified rural addresses.

---

<sup>4</sup> An incorporated place is a type of governmental unit incorporated under state law as a city, city and borough, municipality, town, borough, or village that has legally prescribed limits, powers, and functions.

**Table 1. North Carolina Versus the United States**

		North Carolina		United States	
		#	%	#	%
<b>Housing Units<sup>1</sup></b>					
	<b>Occupied</b>	3,454,068	85.8%	111,617,402	88.4%
	<b>Vacant</b>	<u>572,490</u>	<u>14.2%</u>	<u>14,694,421</u>	<u>11.6%</u>
		4,026,558	100.0%	126,311,823	100.0%
<b>Urban</b>					
	<b>Occupied</b>	2,072,508	87.6%	86,710,850	89.9%
	<b>Vacant</b>	<u>293,067</u>	<u>12.4%</u>	<u>9,743,981</u>	<u>10.1%</u>
		2,365,575	100.0%	96,454,831	100.0%
<b>Rural</b>					
	<b>Occupied</b>	1,381,560	83.2%	24,906,552	83.4%
	<b>Vacant</b>	<u>279,423</u>	<u>16.8%</u>	<u>4,950,440</u>	<u>16.6%</u>
		1,660,983	100.0%	29,856,992	100.0%
<b>Mailing Addresses<sup>2</sup></b>					
	<b>Locatable<sup>3</sup></b>	3,163,459	85.8%	108,186,809	90.6%
	<b>Unlocatable<sup>4</sup></b>	<u>522,439</u>	<u>14.2%</u>	<u>11,226,965</u>	<u>9.4%</u>
		3,685,898	100.0%	119,413,774	100.0%

<sup>1</sup>Source: 2006 American Community Survey.

<sup>2</sup>Source: Marketing Systems Group 2006.

<sup>3</sup>Locatable mailing addresses are city-style street addresses. Augmented addresses are not included.

<sup>4</sup>Unlocatable mailing addresses include P.O. boxes and simplified rural addresses.

Then we used the lists to classify all Census Blocks associated with each place name as either in an area without home mail delivery or in an area with simplified addresses. Census Blocks that were not associated with either list were classified as “other.” Note that all areas without place names (including areas with only P.O. boxes or simplified addresses) were classified as “other.” Of the 232,363 blocks in North Carolina, 1,140 were P.O. box only, 4,980 were simplified address, and 226,243 were assigned to the “other” stratum.

Because Census Tracts served as the primary sampling units (PSUs), Census Blocks were aggregated to the Census Tract level within strata. PSUs were required to contain a minimum of 30 HUs<sup>5</sup>. PSUs not meeting this minimum were collapsed with a nearby PSU by first sorting the PSUs by latitude and longitude and then combining as necessary within strata. The final sampling frame consisted of 1,668 PSUs, each of which contained between 31 and 9,423 HUs.

We selected the sample in two stages. First, PSUs were selected within strata. Prior to selecting PSUs, the sampling frame was implicitly stratified (i.e., sorted) by a rural or urban indicator<sup>6</sup>, region (Coastal, Piedmont, and

Mountain), and percent growth<sup>7</sup>. Five PSUs were selected with probability proportional to size from the “no home mail delivery” stratum and from the “simplified rural address” stratum; 40 PSUs were selected from the “other” stratum.

At the second stage, segments were formed within the selected PSUs by combining nearby Census Blocks until each segment contained a minimum of 30 HUs. One segment was selected with probability proportional to the number of HUs in each of the selected PSUs. The final sample segments consisted of 22 rural and 28 urban segments and were distributed as follows: 17 segments in the Coastal region, 19 in the Piedmont region, and 14 in the Mountains. The segments contained between 31 and 777 HUs, with the median equal to 64 HUs.

We assigned a design weight to a segment as the inverse of its probability of selection within strata. Because we used a two-stage design, the design weight equals the inverse of the probability of selecting the PSU (or Census Tract) multiplied by the inverse of the probability of selecting the segment within the sampled PSU.

**Field Implementation.** The sample of 50 segments was loaded into a Geographic Information System (GIS) as spatial layers of Census Blocks and carrier route boundaries. To account for geocoding inaccuracies (see, Bichler and Balchak 2007), all the Census Blocks

<sup>5</sup> HU counts were obtained from the 2000 decennial Census adjusted to 2007 population projections obtained from Claritas.

<sup>6</sup> In order for a PSU to be considered “rural” all of the blocks within the PSU had to be rural. Otherwise, it was considered “urban.”

<sup>7</sup> Percent growth is defined as the ratio of the 2007 Claritas housing unit projections to the 2000 Census counts.

surrounding the 50 segments were selected as well as all carrier routes that touched any of the surrounding blocks. This list of carrier routes was then sent to MSG to extract the set of active (i.e., non-vacant and non-seasonal) and locatable mailing addresses that would blanket the segments. After we received the list of mailing addresses, we sent them to TeleAtlas, NA for geocoding.

The field protocol specified that the same team visit each segment twice. The first visit was the field enumeration visit in which DUs were recorded. The second visit was the mailer visit in which every address on the mailing list was recorded. The visits were done in this order to prevent possible contamination of the field enumeration with mailing addresses.

The field enumeration team and the mailing list team used the same Global Positioning System (GPS) device<sup>8</sup> and protocol for recording the coordinates and addresses of the DUs. The intention was to capture a coordinate that would represent the DU and also be repeatable for the next visit.

After the field work was completed, we created two sets of points for each segment: the field enumeration points and the mailer points. Each point contained coordinates and a street address. A two-step matching process was then performed. First, we compared the house number, street name, and apartment number. Identical mailer entries and field enumeration entries were considered matches. The remaining entries were passed through an interactive GIS application that enabled an analyst to quickly flag matching pairs of addresses.

The entries that remained unmatched were present on one frame but not the other. Two survey methodologists revisited the segments to physically examine each unmatched entry and to search for HUs not found on either frame. The methodologists resolved transcription errors, verified that the unmatched entries corresponded to actual HUs, and searched for HUs that were not found on either frame. The union of the two frames combined with the follow-up examination enabled us to assume complete coverage of the HUs in the selected segments.

**Estimating Occupancy Status.** Occupancy status has obvious implications for evaluating coverage of the household population. (Households often are referred to as occupied HUs.) However, determining the occupancy status of every HU in the sample of segments was beyond the scope of the study. Instead, we used the following

<sup>8</sup> The team selected the Garmin GPSMAP 76 GPS receiver as the GPS device for this project. The reasons for selecting this model were its low cost, its large storage capacity, its ease of use, and its accuracy (within 10–20 feet of a true location).

process to estimate the number of occupied HUs in each segment.

First, we applied the 2006 Census projected occupancy rate to the total number of HUs in each segment. Then we compared the projected number of occupied HUs to the number of active and locatable mailing addresses (all of which were assumed to be occupied), and then assigned the larger of the two numbers to the segment. The rationale for this process is the assumption that the number of occupied HUs obtained from the projected Census occupancy rates should equal or exceed the number of active and locatable mailing addresses.

After we assigned the number of occupied HUs to each segment, we post-stratified the design weights so that the sample estimate of the total number of occupied HUs equals the 2006 American Community Survey estimates in urban and rural areas of North Carolina.

## 5. Results

A total of 6,408 DUs were enumerated in the 50 selected segments. Of these, 6,317 DUs were listed by the field enumeration team and determined to be HUs (i.e., there were no businesses or group quarters among the listed units). Of those listed, 837 did not have an active and locatable mailing or an augmented address.

A total of 4,530 active and locatable mailing addresses and 1,033 augmented addresses were associated with the enumerated HUs. Of these, 83 were not included in the field enumeration. The 8 HUs not found on either frame were identified during the follow-up activity that was done after the frames were constructed.

As shown in **Table 2**, the estimated coverage of HUs with active and locatable mailing addresses was 82.1 percent compared to 98.1 percent for HUs found during the field enumeration. If the actual coverage rates were applied to the national distribution (i.e., more urban, fewer simplified addresses) then the coverage rates for HUs would be slightly higher for mailing addresses and about the same for field enumeration.

The 98.1 percent state wide coverage rate for the field enumeration was higher than the expected rate of 95 percent. We speculate that the use of experienced listers and the small size of the segments used in the field study contributed to this result. The minimum segment size for the field study was 30 DUs and the mean was 130 DUs. By comparison, the NSDUH requires a minimum of 150 DUs per segment in urban areas and 100 DUs in rural areas. The 2006 NSDUH averaged 243 DUs per segment. As a result, the small segment sizes are not representative of what listers typically work.

**Table 2. Comparison of the Coverage of All Housing Units in North Carolina:  
Field Enumeration versus Locatable and Active Mailing Addresses**

Type of Area	<u>Number of HUs<sup>1</sup></u>		<u>FE Coverage Rate<sup>2</sup></u>		<u>MA Coverage Rate<sup>3</sup></u>		<u>Coverage Rate Difference</u>	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
<b>No Home Mail Delivery<sup>4</sup></b>	16,946	7,354	96.1%	2.7%	3.2%	3.2%	92.8%	3.1%
<b>Simplified Rural Addresses<sup>5</sup></b>	511,070	143,175	97.6%	1.5%	70.6%	9.5%	27.0%	9.0%
<b>Other Rural</b>	1,172,988	213,851	98.8%	0.4%	77.5%	5.7%	21.4%	5.7%
<b>Other Urban</b>	2,325,554	261,755	97.9%	0.7%	87.5%	4.2%	10.4%	4.2%
<b>Total</b>	<b>4,026,558</b>	<b>307,761</b>	<b>98.1%</b>	<b>0.4%</b>	<b>82.1%</b>	<b>3.0%</b>	<b>16.1%</b>	<b>2.9%</b>

<sup>1</sup>The estimates of total HUs in urban and rural areas were post-stratified to equal the 2006 American Community Survey.

<sup>2</sup>All dwelling units found during the field enumeration were HUs.

<sup>3</sup>Estimates of active and locatable mailing addresses.

<sup>4</sup>Housing Units in areas with no home mail delivery are assigned a P.O. box. A few street addresses were found in Census Blocks that overlapped the catchment of the post office.

<sup>5</sup>Augmented addresses were classified as active and locatable mailing addresses in these areas.

The 82.1 percent state wide coverage rate of all HUs with mailing addresses was significantly lower than the 95 percent rate we expected prior to the field study. We attribute much of the shortfall to the use of active mailing addresses which are assumed to be associated with occupied HUs. In fact, the screening results of the NSDUH field enumeration indicate that unoccupied and seasonal HUs account for approximately 15 percent of HUs. Unoccupied HUs are especially prevalent in areas with low population density and in areas with seasonal housing.

The estimated coverage rates for *occupied* HUs are shown in **Table 3**. The 99.6 percent coverage of occupied HUs with field enumeration reflects nearly complete coverage of the household population. The 95.7 percent coverage of occupied HUs with active and locatable mailing addresses is close to what we expected going into the field study. Both sampling frames yielded nearly complete coverage of occupied HUs located in urban areas.

**Table 3. Comparison of the Coverage of Occupied Housing Units in North Carolina:  
Field Enumeration versus Locatable and Active Mailing Addresses**

Type of Area	<u>Number of OHUs<sup>1</sup></u>		<u>FE Coverage Rate<sup>2</sup></u>		<u>MA Coverage Rate<sup>3</sup></u>		<u>Difference in Coverage Rates</u>	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
<b>No Home Mail Delivery<sup>4</sup></b>	7,354	3,380	100.0%	-	7.5%	6.1%	92.5%	6.1%
<b>Simplified Rural Addresses<sup>5</sup></b>	420,321	137,721	99.8%	0.2%	85.8%	8.0%	14.0%	7.9%
<b>Other Rural</b>	977,133	206,593	99.8%	0.2%	93.0%	3.8%	6.8%	3.8%
<b>Other Urban</b>	2,049,260	237,754	99.5%	0.3%	99.3%	0.6%	0.2%	0.7%
<b>Total</b>	<b>3,454,068</b>	<b>282,676</b>	<b>99.6%</b>	<b>0.2%</b>	<b>95.7%</b>	<b>3.0%</b>	<b>4.0%</b>	<b>1.5%</b>

<sup>1</sup>The estimates of occupied HUs in urban and rural areas were post-stratified to equal the 2006 American Community Survey.

<sup>2</sup>All dwelling units found during the field enumeration were HUs.

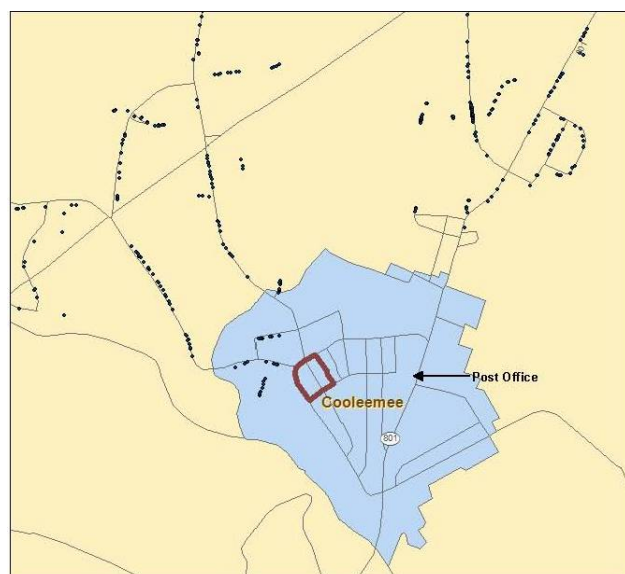
<sup>3</sup>All active and locatable mailing addresses were assumed to be occupied.

<sup>4</sup>HUs in areas with no home mail delivery are assigned a P.O. box. A few street addresses were found in Census Blocks that overlapped the catchment of the post office.

<sup>5</sup>Augmented addresses were classified as active and locatable mailing addresses in these areas.

**Coverage in Areas with No Home Mail Delivery.** By definition, these areas do not have locatable mailing addresses. We did however find 13 locatable mailing addresses in one of the five segments with no home mail delivery. We attribute this to a minor overlap between the catchment of the post office and that of the Census Blocks. The field enumeration in these areas accounted for 96 percent of the HUs and was only slightly lower than the state wide coverage rate of 98 percent.

We did a follow-up of the segment that was selected in Cooleemee, North Carolina (shown in **Figure 1**) to confirm with the Post Office that the 34 enumerated HUs in the selected segment and the HUs in the surrounding Census Blocks did not have mail delivered to their homes.



**Figure 1. An Area without Home Mail Delivery.** The shaded area shows the Census Blocks that comprise the town limits of Cooleemee, North Carolina that was designated as an area without home mail delivery. The selected segment is highlighted. The dots represent locatable mailing addresses that are outside the catchment of the post office. The Census Blocks in the northwest corner of the town border the catchment and contain locatable and unlocatable mailing addresses.

**Coverage in Areas with Simplified Rural Addresses.** In addition to the five selected segments that were known to contain simplified mailing addresses, we identified six other segments that only contained simplified addresses after the sample was selected. The Census Blocks that comprised these segments were not designated as areas with simplified rural addresses on the sampling frame because they either did not have a place name or their place name did not match to the list of place names provided by MSG.

Augmented addresses accounted for an estimated 70.6 percent of all HUs and 85.8 percent of occupied HUs located in areas with simplified rural addresses. Augmented addresses typically are associated with occupied HUs so the results are similar to the 80 percent coverage rate that we expected prior to the field survey.

**Coverage in Other Rural Areas.** HUs in other rural areas accounted for approximately 29 percent of all HUs in North Carolina in 2006. Field enumeration yielded nearly coverage (98.8 percent) in these areas compared to only 77.5 percent coverage for locatable mailing addresses. Among occupied HUs however, the coverage of mailing addresses was 93.0 percent. We attribute much of the difference in coverage to relatively high vacancy rates in these areas.

**Coverage in Other Urban Areas.** HUs in other urban areas accounted for approximately 59 percent of all HUs in North Carolina in 2006. As expected, the coverage of both sampling frames was high in urban areas. Among all HUs, the 97.9 percent coverage rate for field enumeration was significantly higher than the 87.5 percent rate for locatable mailing addresses. Among occupied HUs however, both frames yielded nearly identical coverage rates of 99.5 percent and 99.3 percent respectively.

## 6. Supplementing Coverage

The Half-Open Interval (HOI) frame-linking procedure (Kish, 1965, p. 56) often is used to help reduce the undercoverage associated with field enumeration. The HOI procedure adds HUs to an existing frame by searching for new units in the interval between the selected HU and the next HU on the frame. New HUs that are discovered in the interval during field interviewing are automatically included in the sample.

To be effective, the HOI procedure requires that the HUs on the frame be sorted in geographically proximal order. With field enumeration, this is achieved through a contiguous enumeration process which allows for consecutive HUs to be adjacent whenever possible. Our examination of the 91 HUs that were missed by the field enumeration indicated that all would be included with the HOI procedure.

A variation of the HOI procedure based on the delivery sequence<sup>9</sup> of the postal carrier route can be applied to a frame based on mailing addresses. The delivery sequence is the path a letter carrier follows to deliver the mail. The

<sup>9</sup> The carrier route and the delivery sequence number of a mailing address are included in the basic data elements of a standardized mailing address.



HOI procedure can reduce the incompleteness of the mailing addresses regardless of whether it is caused by new construction, maturity of the list, or the exclusion of unlocatable addresses from the sampling frame.

The delivery sequence usually proceeds down one side of the street and back up the other, which makes it amenable to the HOI frame-linking procedure. HOIs cannot be constructed when the interval between the sampled HU and the next HU on the frame is ambiguous. Examples of ambiguous HOIs include clusters of mail boxes and some street intersections. Because delivery sequence numbers are not available for augmented addresses, we assigned HOIs according to the street numbering sequence.

The sampling frame based on active locatable mailing and augmented addresses excluded 845 HUs that were found during the field enumeration or follow-up phases of the field study. HOIs based on the delivery sequence would have added 240 of these HUs to the frame. At the state level, we estimate an increase in coverage of at least 3.4 percentage points<sup>10</sup>. At the national level, an increase of at least two percentage points was found for the 2002 EQ-5D survey (Staab and Iannacchione 2003).

Locatable mailing addresses (and hence delivery sequence numbers) are nonexistent for the 0.4 percent of HUs in North Carolina that are located in areas without home delivery of mail. As a result, the HOI procedure cannot be implemented. Alternative methods such as field enumeration or property tax records (Kalsbeek, et al. 2005) should be used to construct sampling frames in these areas.

## 7. Conclusions and Recommendations

Our field study compared the coverage of two sampling frames of HUs in North Carolina: one based on locatable residential mailing addresses and the other based on field enumeration. We found that field enumeration produced significantly higher coverage of occupied and vacant HUs than mailing addresses. When restricted to occupied HUs however, comparable coverage was found for the 59 percent of occupied HUs in urban areas. In rural areas, mailing addresses were found to have significantly lower coverage than field enumeration. However, the use of augmented addresses in areas with simplified rural addresses produced noticeable gains in coverage.

We estimate that field enumeration combined with the HOI procedure would yield virtually complete coverage of occupied and vacant HUs. An analogous HOI methodology based on the letter carrier's delivery

sequence would increase the coverage of locatable mailing addresses by about two percentage points.

The following limitations of the research design are noted.

- The occupancy status of HUs in the sample of segments could not be determined because contact with residents was not authorized. As a result, direct estimates of the coverage of the household population could not be made. Instead, we made direct estimates of the coverage of all HUs (occupied and vacant) and indirect estimates of occupied HUs.
- There is not a one-to-one correspondence between a residential mailing address and an HU. A single locatable mailing address may be associated with multiple HUs and vice versa.
- The evaluation of coverage does not include households in areas with restricted access (e.g., gated communities) because our protocol required that every address be physically located.
- Cost considerations required that the segments for the field study have fewer DUs than most household surveys. As a result, the small segment sizes are not representative of what listers typically work. The small segment sizes also precluded realistic cost comparisons between field enumeration and mailing addresses.

In spite of these limitations, the field study provides evidence that a sampling frame comprised of locatable mailing addresses and augmented addresses and supplemented with a frame-linking procedure like the HOI will yield reasonable coverage of the household population.

In this era of declining response rates, the benefits of in-person surveys should be reconsidered. The high cost of field enumeration has inhibited this mode of data collection for all but the largest surveys. For small and moderate sized household surveys, a sampling frame based on mailing lists can help make in-person interviewing a viable mode of data collection. (The surveys cited in this paper could not have been done if field enumeration was a requirement.) For large scale studies, a holistic approach is needed when considering which frame to use. Mailing lists may have slightly less coverage than field enumeration but their use could enable the transfer of resources from frame development to activities like additional training for refusal conversion, an examination of total survey error, or to fund a nonresponse follow-up study.

<sup>10</sup> Lower bound of a 95% confidence interval.

### Acknowledgements

This research was funded by RTI International as part of an internal research and development project. The authors thank Dale Kulp and David Malarek of the Marketing Systems Group, Tim Kennel of the U.S. Census Bureau, and RTI staff members Ross Curry, Jamie Ridenhour, and Amanda Lewis-Evans for their help with this research.

### References

- American List Counsel. 2007. ADVO advantage residential database (accessed October, 2007). <http://search.alcdata.com/market?page=research/datacard&id=71768>
- Bichler, Gisela, and Stephanie Balchak. 2007. Address matching bias: Ignorance is not bliss. *Policing: An International Journal of Police Strategies & Management*, 30(1):32-60.
- Corlett, Marjorie, Timothy Kennel, Tina Flannigan, Xijian Liu, and Cliff Loudermilk. 2006. Frame assessment for current household surveys analytic report: comparing the area frame listings to the MAF. Demographic Statistical Methods Division, U.S. Census Bureau.
- Cunningham, David, Susan Hunter, Laura Justin, Katie Morton, and Stephanie Stolzenberg. 2006. 2005 National Survey on Drug Use and Health: Count and List Validity Study Overall Summary Report. Prepared for the Substance Abuse and Mental Health Services Administration, Contract No.283-2004-00022. Research Triangle Park, NC: RTI.
- 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*.
- Hox, Joop and Edith de Leeuw. 1994. A Comparison of Nonresponse in Mail, Telephone, and Face-to-Face Surveys. *Quality and Quantity*, 28:329-344.
- Iannacchione, Vincent, Jennifer Staab, and David Redden. 2003. Evaluating the use of residential mailing lists in a metropolitan household survey. *Public Opinion Quarterly*, 67(2):202-210.
- Kalsbeek, William, Sarah Kavanaugh, and Jingjing Wu. 2005. Using GIS-based property tax records as an alternative to traditional household listing in area samples. *Proceedings of the American Statistical Association, Section on Survey Research Methods*.
- Kish, Leslie. 1965. *Survey Sampling*. New York: John Wiley & Sons.
- Morton, Katherine, James Chromy, Susan Hunter, and Pielan Martin. 2007. Sample design report. In *2006 National Survey on Drug Use and Health: Methodological resource book* (Section 2, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Phase I, Deliverable No. 8, RTI/0209009.230.004). Research Triangle Park, NC: RTI International. [Available as a PDF at <http://www.oas.samhsa.gov/nsduh/methods.cfm#2k6>]
- 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*.