EVALUATING THE USE OF RESIDENTIAL MAILING ADDRESSES IN A NATIONAL HOUSEHOLD SURVEY

Jennifer M. Staab and Vincent G. Iannacchione

RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709

Key Words: Sampling Frame, Residential Mailing Address Lists, Household Surveys, Coverage, Half-Open Interval, E-911 Address Conversion

Introduction

On-site enumeration is generally regarded as the gold standard for sampling frame construction for in-person household surveys. However, on-site enumeration is expensive and time consuming because of the amount of time needed to completely enumerate selected areas. A more expedient and cost-effective alternative method of household enumeration is the use of residential mailing lists.

In 2000, we used residential mailing lists, instead of a list derived from on-site enumeration, to develop a sampling frame for a probability-based in-person survey of 15,000 households in Dallas County, Texas (Iannacchione, Staab, and Redden 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 on-site enumeration. In addition, we found that the mailing lists, when supplemented with missed households identified by a frame-linking procedure provided reasonably complete coverage of the metropolitan household population.

Although these results demonstrated the utility and completeness of using residential mailing lists to develop a sampling frame, our evaluation in 2000 was limited to one large metropolitan area. In this paper, we evaluate the completeness of mailing lists for a national household survey by comparing counts of residential mailing addresses to Census projections of households. We also demonstrate the utility of mailing lists by describing their use in a national household survey that we conducted in 2002.

Comparing Counts of Residential Mailing Addresses to Counts of Households

It is not unreasonable to assume that virtually all households in the United States have a mailing address. While the US Postal Service (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.

Methods. With over 121 million residential mailing addresses, one such qualified compiler of mailing lists is ADVO, Inc. (2003). ADVO claims that their residential mailing address database reaches virtually all households in the United States and provides the most complete consumer mailing list available (American List Counsel, 2003). The ADVO database is updated monthly through direct feeds from the USPS. ADVO uses a proprietary Advanced Address Monitoring System to edit the database.

While the ubiquity of mailing addresses can be demonstrated, not all mailing addresses are suitable for household surveys. To be usable for survey purposes, addresses must be specific enough that interviewers are able to locate them "on the ground." The primary types of mailing addresses that are not locatable are Rural Route Boxes and P.O. Boxes.

Rural Route Boxes. Rural route and highway box numbers (e.g. RR 1 Box 100 or HC 2 Box 200) are not locatable because they cannot be used to determine the physical location of a household. However, the prevalence of rural route box numbers has diminished in recent years because of the emergency 911 (E-911) program which has encouraged local governments to convert rural route addresses to city-style addresses. The Locatable Address Conversion System (LACS) is a USPS file that contains addresses converted from rural routes, highway route, and box numbers to city-style addresses. Most of the converted addresses on the LACS have been changed as a result of the E-911 program.

During the past two years, ADVO has improved their database by converting many rural route boxes to city-style addresses (e.g. 1001 HWY 82). In 2002, ADVO had acquired city-style addresses for at least 85 percent of their rural route delivery addresses (American List Counsel, 2003). Ultimately, the conversion will depend on how quickly local governments comply with the E-911 address conversion of rural route box numbers to city-style addresses.

Post Office Boxes. Although PO Boxes are not locatable for survey purposes, they do not always represent a coverage problem because many households maintain a PO Box in addition to their street address. In these situations, PO Boxes can be excluded from a sampling frame without loss of coverage.

PO Boxes can present a coverage problem when a household only receives mail through a PO Box. However, if such a household is located on a carrier route, it can be covered by the use of a frame-linking procedure that we describe later in the paper. Households receiving mail exclusively through a PO Box that are not on a carrier route (i.e., no home mail delivery) present more difficult coverage problems.

The USPS provides PO Boxes free of charge to residents of households that are located in areas where home delivery of mail is not provided. Because we were unable to obtain counts of the number and location of households where the USPS does not provide home delivery of residential mail, we estimated these counts by identifying local areas comprised exclusively of PO Boxes.

Merging the files. We obtained counts of residential mailing addresses from ADVO for over 29,000 local areas across the country and then compared them to corresponding Census household projections generated by Claritas, Inc. The goal of the comparison was to estimate the completeness of the mailing addresses at the local level, and to estimate the proportion of households that have locatable residential mailing addresses. This analysis also helped us to determine whether certain types of local areas (e.g., rural areas) are more prone to unlocatable mailing addresses than other areas.

The comparison required us to merge two distinct sources of information:

- Counts of mailing addresses for 37,952 residential Zip Codes (Effective date: October, 2002); and,
- Projections of households based on Census 2000 for 30,003 Zip Code Tabulation Areas (ZCTAs) (Effective date: June, 2002).

Before the merge, we collapsed Zip Codes with the same 'City' portion of the mailing address into local areas. Similarly, we collapsed ZCTAs into local areas using the ZCTA geographic label. In spite of the fact that ZCTAs were based on 2001 Zip Code definitions, only 0.07% of all residential addresses and 0.03% of all projected households failed to merge.

We used local areas as the unit of analysis to help alleviate some of the artificial noise that would be introduced by comparing counts at the Zip Code level. The dynamic nature and possible small size of Zip Codes can lead to spurious results. We acknowledge that this does not eliminate all inconsistencies in the comparisons because the boundaries of local areas also are subject to change.

Another decision we faced was whether to compare the number of mailing addresses to the corresponding number of housing units or to the number of households (i.e., occupied housing units). While neither is entirely consistent with mailing addresses, we decided to use households because most vacant households, seasonal addresses, and other out-of-date addresses are removed from residential mailing lists as a normal part of list hygiene.

As we show in **Table 1**, the national total of 108.8 million residential mailing addresses (excluding PO Boxes) compares favorably with the estimated 107.7 million households nationwide. Even if the comparison is restricted to locatable mailing addresses, nearly all of the 70.7 million households in areas with 10,000 or more households have locatable addresses. This is not the case for the remaining 37 million of households in areas with less than 10,000 households. In these predominantly rural areas, the coverage of the mailing addresses primarily depends on whether or not there is home delivery of mail, and on the penetration of E-911 address conversion.

We estimate that 1.4 million households (1.3 percent of households nationwide) are located in areas without home delivery of mail. Taken together with the 4.2 million RR Boxes, these unlocatable addresses represent potential sources of non-coverage. In the next section, we describe how we dealt with the coverage issues of unlocatable addresses in implementing a national household survey.

Table 1. Distribution of Mailing Addresses and Households by Size of Local Area[1]

	Size of Local Area (in HHs)				Total
	< 10,000		10,000 - 22,000	> 22,000	
	No Home Delivery	Home Delivery			
Number of Local Areas	5,490	21,947	1,238	775	29,450
Locatable Mailing Addresses					
City Routes	0	13,318,676	13,524,396	48,162,006	75,005,078
Rural Routes	0	17,344,585	5,232,834	6,945,906	29,523,325
	0	30,663,261	18,757,230	55,107,912	104,528,403
Unlocatable Mailing Addresses					
RR Boxes	0	3,597,822	373,575	263,723	4,235,120
PO Boxes	1,499,641	7,198,162	1,209,458	2,936,206	12,843,467
Est. Number of Households	1,421,451	35,549,431	18,125,769	52,612,784	107,709,435
% of HHs with Locatable Adds.	0.0%	86.3%	103.5%	104.7%	97.0%

[1] A local area comprises all Zip Codes with the same 'City' portion of the mailing address.

Using Residential Mailing Lists as a Sampling Frame for a National Survey

We used residential mailing lists as a sampling frame for the *2002 U.S. Valuation of the EuroQol EQ-5D Health State Survey* that RTI conducted for the University of Arizona with funding from a grant provided by the Agency for Health Care Research and Quality (AHRQ). The target population for the EQ-5D Survey comprised the estimated 212 million civilian, non-institutionalized English- and Spanish-speaking adults, aged 18 and older, who resided in the United States (50 states and D.C.) at the time of data collection (June through October, 2002).

The analytical objectives of the EQ-5D Survey required that we over-sample eligible Hispanics and Non-Hispanic Blacks. Therefore, we began the sample allocation by classifying each of the 27,724 Zip Code Tabulation Areas (ZCTAs) (U.S. Census 2000) in the 50 states and D.C. on the basis of minority concentration. ZCTAs are approximate geographic representations of USPS Zip Code service areas that cover the U.S. The use of ZCTAs enabled us to use Census demographic data to allocate the sample without having to cross-walk mailing addresses to Census Tracts or Blocks.

To achieve the desired study objectives with an initial sample size of 12,000 households, we developed the following four-stage probability-based sampling design.

First Stage of Selection: 3-Digit ZCTAs. We selected a first-stage sample of 60 3-digit ZCTAs from a sampling frame of 883 3-digit ZCTAs that were formed by collapsing 5-digit ZCTAs to their first 3 digits. We assigned size measures to the 3-digit ZCTAs that were proportional to the weighted number of minority and non-minority households in the 3-digit ZCTA. (Counts of minorities received larger weights than counts of non-minorities.) We selected the first-stage sample with a PPS (probabilities proportional to size) systematic selection algorithm. Prior to selection, we sorted the frame by the 3-digit ZCTA identifier to distribute the sample as evenly as possible across the country. The 60 selected 3-digit ZCTAs accounted for 17,737,921 (16.8%) of the households enumerated in Census 2000.

Second Stage of Selection: 5-Digit Zip Codes. Prior to selecting the second-stage sample, we obtained counts of residential mailing addresses (as of April, 2002) for each of the 2,387 5-digit Zip Codes associated with the 60 selected 3-digit ZCTAs[1]. We assigned a size measure to each 5-digit Zip Code in the same way that we assigned size measures to 5-digit ZCTAs except that up-to-date counts of mailing addresses (obtained from ADVO) were used instead

[1] We did not use 5-digit ZCTAs as second-stage sampling units because the size and number of ZCTAs were based on Zip Codes as they existed two years prior when the Census 2000 was conducted.

of counts of households from the Census 2000. For the 2,237 5-digit Zip Codes that matched to a 5-digit ZCTA, we used the ZCTA demographic information to assign them to a minority concentration category. We used the minority concentration of the corresponding 3-digit ZCTA to assign the remaining 18 5-digit Zip Codes to a minority category.

We excluded from the second-stage sampling frame 84,841 PO Boxes that were associated with Zip Codes without home mail delivery (using our definition). As we described earlier, these exclusions represented approximately 1.3 percent of the national household population.

We calculated a size measure for each 5-digit Zip Code that was proportional to the number of addresses that the 5-digit Zip Code accounted for in its assigned minority concentration category. Counts of residential PO Boxes in the Zip Code were excluded from the size measures. We collapsed a number of small Zip Codes with adjacent Zip Codes so that each sampling unit could support the minimum number of addresses specified by the sample design.

We again used a PPS systematic selection algorithm to select two 5-digit Zip Codes from each of the 60 selected 3-digit ZCTAs. Prior to selection, we sorted the Zip Codes by their numeric identifiers to spread the sample throughout each 3-digit ZCTA.

Third Stage of Selection: Addresses. The 120 selected Zip Codes contained 1,293,368 city-route addresses and only 175,092 rural-route addresses. Like many population-based surveys, the Zip Codes selected for the EQ-5D Survey over-represented large urban areas. The over-sampling of Zip Codes with concentrations of minority households further contributed to the predominance of city-route addresses.

The low cost associated with purchasing mailing addresses enabled us to obtain the entire list of residential addresses in each of the 120 selected Zip Codes. Obtaining all of the mailing addresses enabled us to increase the efficiency of the design by dispersing the sample throughout each Zip Code instead of restricting it to a number of small geographic clusters (e.g., Census Blocks) required for on-site enumeration. Having the entire set of mailing addresses in each Zip Code also facilitated the creation of custom maps of neighborhoods surrounding each selected address.

We purchased the city-route addresses from ADVO and the rural-route addresses from Donnelley Marketing Services and Americom. We purchased rural addresses from these vendors because of ADVO's use of simplified rural-route addresses. Using address lists from multiple vendors enabled us to obtain city-style addresses for all of the rural routes in our selected Zip Codes.

Our field interviewers were able to locate all but 44 of the 12,000 addresses selected for the EQ-5D Study. Among located addresses, 771 were vacant and 186 were non-residential. The remaining 10,999 addresses appeared to be occupied households. Overall, we were able to determine the survey eligibility status for 76.1 percent of the addresses in the sample.

Fourth Stage of Selection: Persons. A total of 7,943 households were screened and eligible for the survey. We selected one eligible adult for interview from each successfully screened Hispanic or Non-Hispanic Black household. We randomly selected one adult for interview in about 43 percent of households with other race/ethnicities. We generated address-specific randomized selection tables for use by the field interviewer in the selection of individuals.

We selected a total of 5,237 persons for the survey and obtained completed interviews from 4,048 participants for a 77.3 percent interview completion rate among addresses that were successfully screened. The overall response rate of 59.4 percent (using the AAPOR RR3 formulation) reflects our ability to determine the survey eligibility of the address sample and to obtain the participation of persons selected for the survey interview. The use of Census demographic data enabled us to come close to the target number of completed interviews for Hispanics (1,255 completes compared to a target of 1,200) and for Non-Hispanic Blacks (1,132 completes compared to a target of 1,200).

Expanding the Coverage of the Sampling Frame

We used the Half-Open Interval (HOI) frame-linking procedure (Kish, 1965) to help reduce under-coverage caused by the incompleteness of the mailing lists regardless of whether it was caused by new construction, maturity of the list or the exclusion of unlocatable addresses from the sampling frame. To be effective, the HOI procedure requires that the addresses on the sampling frame be sorted in geographical proximal order. With traditional on-site enumeration, this is achieved through a contiguous

enumeration process which allows for consecutive housing units to be adjacent whenever possible. We achieved a reasonable approximation of geographic proximity with addresses on the residential mailing lists by sorting the addresses by the delivery walk sequence of the postal carrier route.

The delivery walk sequence of the postal carrier route is simply the path the postal carrier follows to deliver the mail. Generally, the delivery walk sequence proceeds down one side of the street and back up the other, which makes it amenable to the HOI frame-linkage procedure. The only time that a HOI cannot be constructed is when the interval between the sampled housing unit and the next one on the list cannot be uniquely determined. This occurs when the delivery walk sequence crosses the street or moves to a new street, as illustrated by the circled housing units on **Figure 1**.

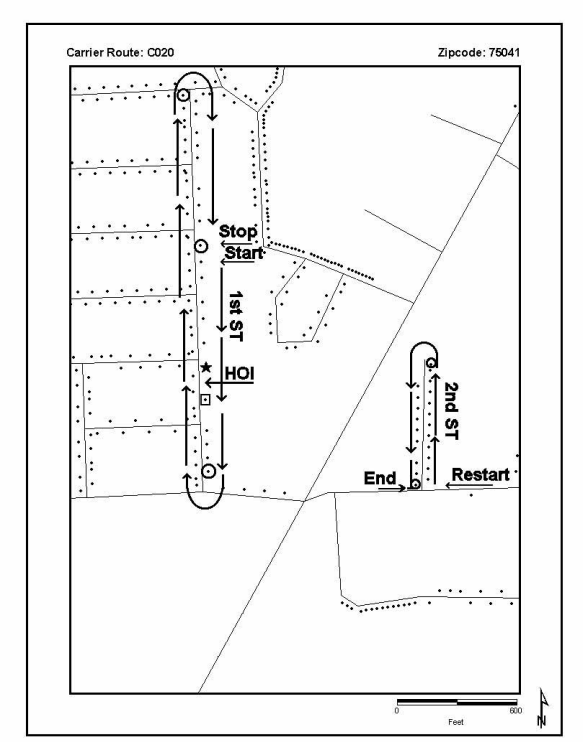

Figure 1. Example of a Delivery Sequence for a Postal Carrier Route. *In this example, the carrier route comprises addresses on two streets. The delivery sequence begins in the middle of 1st Street and then proceeds in a clockwise direction along each side of the street. It then restarts on 2nd Street proceeding in a counter-clockwise direction. The HOI arrow points to the half-open interval between the selected address (starred) and the next address in the delivery sequence (boxed). The five circles identify addresses where half-open intervals cannot be uniquely determined.*

We selected a sub-sample of 3,999 addresses for HOI examination. Our field interviewers were able to construct and locate HOIs for 3,305 of the sub-sampled addresses. Extrapolating from the sub-sample to the national level, we estimate that HOIs could be constructed and located for at least 83.8 percent (lower bound of 95% CI) of the usable mailing addresses on the sampling frame.

Using logistic regression models, we found that addresses were significantly less likely to have an HOI associated with them if they were located in a Zip Code that was:

- Outside a Metropolitan Statistical Area (MSA);
- Predominantly white;
- In the Mid-West region; or,
- Experiencing a high rate of growth.

The use of the HOI procedure resulted in the addition of 76 'new' addresses to the sample, i.e., addresses not on the mailing lists. Nationally, we estimate the prevalence of encountering a missed address was at most 1.75 percent (Upper bound of 95% CI). We used a log-linear regression model to determine that missed addresses were significantly less likely to be encountered in the Mid-West Region. Other factors such as MSA status, percent growth, and region did not significantly affect the prevalence of new addresses.

These findings suggest that future surveys use the HOI frame-linkage procedure to target selected addresses in the Mid-West, areas outside of MSAs, predominantly white areas, and areas of high growth areas because they appear to be areas where the mailing lists are most vulnerable to under-coverage.

Households with a Street Address and a PO Box

We included a question in the EQ-5D screener that asked whether or not the occupants of a household also received mail through a residential PO Box. This question was an attempt to estimate how many of the households maintained PO Boxes in addition to receiving mail at their street address. Unfortunately, the large variation in the number of PO Boxes in the selected Zip Codes combined with the relatively few Zip Codes in our sample caused the associated standard error of the estimate to be unacceptably large. We also suspect that the results were further perturbed by residents who maintained PO Boxes in a different Zip Code than their residence.

Conclusions and Recommendations

Our evaluation of the use of residential mailing lists as a sampling frame for a national household survey led to following findings.

- Residential mailing addresses are commercially available for virtually all households nationwide. Approximately 97 percent of all households have locatable mailing addresses.
- The prevalence of locatable mailing addresses diminishes with population density, and is non-existent in areas where home delivery of mail is not provided.
- Except in areas without home mail delivery, the exclusion of PO Boxes from a sampling frame does not lead to appreciable under-coverage, especially if the HOI procedure is use to supplement the mailing lists.
- The half-open interval frame-linking procedure can supplement the coverage of mailing lists (in areas with locatable mailing addresses) by as much as 1.75 percent.

These findings lead us to conclude that there are important caveats associated with the use of residential mailing addresses as a national sampling frame. The most important is the systematic non-coverage of the estimated 1.4 million households (1.3 percent of households nationwide) that are located in areas without home delivery of mail. Because households in these areas do not have locatable addresses, alternative enumeration strategies such as property tax records, satellite imagery, or standard on-site enumeration are needed in these areas if these households are to be included in the coverage of the sampling frame.

Households with RR Boxes present a more nebulous and dynamic coverage issue. For the EQ-5D Survey, we were able to obtain city-style addresses for all selected rural routes by purchasing addresses from multiple list vendors. In addition, households with RR Boxes that are located on rural delivery routes that contain city-style addresses can be linked to the frame with the HOI procedure. However, households located on rural routes comprised exclusively of RR Boxes are problematic because the HOI procedure cannot be implemented on a carrier route that does not have any locatable addresses.

The prevalence of RR Boxes (currently about 15 percent of all rural addresses) is expected to decrease as E-911 address conversion continues to be implemented by local governments. In fact, we found evidence of the dynamic nature of the conversion process when we accessed a Zip Code listing of counts of addresses converted by the Locatable Address Conversion System (LACS) (USPS 2003) and found just over one million more locatable addresses in 3,300 Zip Codes than were identified on the ADVO database. Although some LACS conversions result from renaming or renumbering existing city-style addresses, the vast majority of conversions involve changing RR Boxes to city-style addresses.

Except for households in areas without home delivery of mail, our use of residential mailing lists for the EQ-5D Survey resulted in reasonably complete coverage of the national household population. Although we did not happen to select any rural routes comprised exclusively of RR Boxes, we acknowledge that they represent another source of non-coverage that cannot be remedied with the HOI procedure. In spite of this, we believe that these results demonstrate the utility and relative completeness of using residential mailing lists to develop a sampling frame for a national in-person household survey.

References

ADVO. (2003) "Resident List." Retrieved May 10, 2003 from http://www.advo.com/html/solutions/extended/res_list.htm

American List Counsel. (2003) "ADVO Advantage Residential Database." Retrieved May 10, 2003 from http://search.alcdata.com/market?page=research/datacard&id=71768

Iannacchione, V.G., J.M. Staab, and D.T. Redden (2003). "Evaluating the Use of Residential Mailing Lists in a Metropolitan Household Survey." *Public Opinion Quarterly*. Vol 67. No.2 pp 202-210.

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

US Postal Service (2003). "Delivery Point Validation Product and Second Generation Delivery Sequence File – Licensed Service." v 4 Retrieved May 10, 2003 from http://www.ribbs.usps.gov/files/DSF2/CERTCHNG.DOC

US Postal Service (2003). "Locatable Address Conversion System (LACS)" Retrieved November 18, 2003 from http://www.usps.com/ncsc/addressservices