

The Presence and Characteristics of Households at Addresses Obtained by Traditional Field Listing and from USPS-Based Lists

Sarah P. Shore¹, Jill M. Montaquila¹, and Valerie Hsu¹

¹Westat, 1600 Research Blvd., Rockville MD, 20850

Abstract

This paper builds upon an earlier evaluation of the United States Postal Service (USPS)-based residential address frames, conducted in the Vanguard Study locations of the National Children's Study. The earlier research compared a USPS-based address frame with an address frame created by traditional field listing. This paper extends the earlier research using data obtained in connection with a household enumeration of all addresses obtained by the traditional field listing. The household enumeration data were used to establish whether addresses were occupied by households, an important consideration since missed unoccupied residential addresses do not contribute to household noncoverage. Analyses are reported that compare the presence of households in addresses that were on both the traditional field listing and the USPS-based list with that of addresses that were only on the traditional field listing. During the course of the household enumeration, coverage enhancement procedures were applied to identify units missed by the traditional field listing efforts, thus supplementing the list to give more complete coverage. To examine the characteristics of households that are more likely not to be covered by the USPS-based list, the demographic characteristics of households that were on both the supplemented address list and the USPS-based list are compared with those that were only on the supplemented list.

Key Words: Address-based sampling, National Children's Study, Computerized Delivery Sequence File, disposition codes

1. Introduction

For many years traditional field listing has been the standard approach for generating frames for household area probability sample surveys. However, the recent availability of address lists based on the United States Postal Service (USPS) Computerized Delivery Sequence (CDS) file has prompted evaluation studies to help determine whether USPS-based lists can replace the resource-intensive traditional field listing of addresses. Several studies comparing the coverage of household frames based on USPS address lists to addresses obtained by traditional field listing have yielded promising results, at least for urban areas (O'Muircheartaigh et al. 2006; Iannacchione et al. 2007; English et al. 2009; Montaquila et al. 2009). Additionally, models have been developed to predict when USPS-based lists are a viable alternative to traditional field listing (O'Muircheartaigh, English, and Eckman 2007; Hsu, Montaquila, and Brick 2010).

This paper builds upon an earlier evaluation of the USPS-based residential address frame, conducted in the Vanguard Study locations of the National Children's Study (Montaquila et al. 2009). The earlier research compared a USPS-based address frame with an address

frame created by traditional field listing (prior to supplementation based on coverage enhancement procedures). This paper extends the earlier research using data obtained in connection with a household enumeration of all addresses obtained by the traditional field listing. The household enumeration data were used to establish whether addresses were occupied by households, an important consideration since missed unoccupied residential addresses do not contribute to household noncoverage. During the course of the household enumeration, coverage enhancement procedures were applied to identify units missed by the traditional field listing efforts, thus supplementing the list to give more complete coverage. To examine the characteristics of households that are more likely not to be covered by the USPS-based list, the demographic characteristics of households that were on both the supplemented address list and the USPS-based list are compared with those that were only on the supplemented list. The next section contains an overview of previous research involving comparisons of USPS-based address frames with address frames developed using traditional field listing. Section 3 describes the approach taken in our current research. The results of the current research are given in Section 4, and Section 5 contains a discussion and conclusions.

2. Background and Previous Research

In Section 2.1, we begin with summary of traditional address listing approaches and a review of previous evaluations that have compared USPS-based lists to traditional field listing. Section 2.2 gives a brief overview of the Vanguard Study of the National Children's Study (NCS), the study used in our research, and describes the matching research conducted in 2009 upon which our research is built.

2.1 Primary Address Listing Approaches and Summary of Previous Research

For decades, traditional field listing has been the usual frame generation approach for household area probability sample surveys. With this approach, field staff known as "listers" record all residential addresses in defined geographies in a systematic fashion (Kish 1965). Coverage enhancement procedures, also conducted by field staff, are often applied to identify addresses that the listers missed during the initial listing. The two main types of coverage enhancement procedures are the missed structure procedure, where entire missed structures are added to the frame, and the missed dwelling unit (DU) procedure, where separate units within a structure (e.g., a basement or garage apartment) are added to the frame.

Recently, several studies have been undertaken to evaluate the coverage of USPS address-based sampling frames by comparing them to frames constructed using traditional field listing. O'Muircheartaigh et al. (2006) used field verification to arrive at a "best" address frame for a nationally representative set of well-defined geographical areas (referred to as segments). They compared geocoded USPS-based lists and traditional listings to this "best" frame. The authors concluded that the USPS-based list was, in most situations, superior to traditional field listing.

Iannacchione et al. (2007) compared the coverage of the USPS-based lists to traditional field listings for a probability sample of 50 segments in North Carolina. When restricted to occupied DUs, comparable coverage was found in urban areas. In rural areas, the USPS-based list was found to have significantly lower coverage than traditional field listing. However, the use of "augmented addresses" (additional, non-USPS-based

addresses provided by the vendor to augment the frame in rural areas with simplified addresses) produced substantial gains in coverage.

English et al. (2009) compared the coverage of two versions of the USPS-based address lists (Valassis and Customer Identification Services (CIS)) to traditional field listings in 17 segments in Waukesha, WI, a suburban area. After matching the traditionally listed addresses with the two versions of the USPS-based lists, they returned to the housing units that were missed by one or more sources to collect additional data. Their multiple list evaluation found that in these areas in Waukesha, traditional field listing produced results that gave the highest coverage of their “best” frame (which was constructed by combining the traditional listing and USPS-based frames, with field verification and supplementation).

2.2 Earlier Matching Research Based on the National Children’s Study

Our research uses data from the Vanguard Study of the NCS, a study designed to examine the health and development of 100,000 children across the United States, following them from before birth until age 21. The sample design for the NCS is a multi-stage area probability household sample, with 110 primary sampling units (PSUs) selected for the Study (Montaquila, Brick, and Curtin 2010). PSUs are single counties or groups of adjacent counties, with roughly 10,000 addresses sampled per PSU. The samples within each PSU are generally clustered in 10 to 15 segments (i.e., geographically contiguous census blocks) that vary in size between about 500-1,200 households.

For the NCS Vanguard Study, a pilot study conducted in seven of the NCS PSUs, traditional field listing was conducted in the sampled segments in 2008. In 2009-2010, household enumeration was attempted at each address in each sampled segment. Women of child bearing age (ages 18-49) are currently being contacted periodically, and if they give birth during the enrollment period, their children are eligible for the Vanguard Study. The seven PSUs are:

- Duplin County, NC;
- Brookings County, SD; Yellow Medicine, Pipestone, and Lincoln Counties, MN (BYPL);
- Montgomery County, PA;
- Orange County, CA;
- Queens County, NY;
- Salt Lake County, UT; and
- Waukesha County, WI.

Our research uses listing and enumeration data from these seven PSUs. Even though the Vanguard Study is limited to seven purposively chosen PSUs, they provide a range of geographic diversity, with representation from all four census regions. Also, Duplin and BYPL are rural PSUs while the other five are urban PSUs.

For the seven PSUs in the Vanguard Study, previous research (Montaquila et al. 2009) involved matching traditionally listed addresses in the sampled segments to USPS-based addresses that geocoded to blocks within the sampled segments. First, lists of residential addresses in the ZIP codes associated with the sampled segments were obtained from a vendor. Addresses that geocoded to census blocks within the sampled segments were

retained; all non-geocodable addresses and addresses that geocoded to a block outside the segments were dropped. Non-city-style addresses (e.g., Post Office boxes and rural route addresses) were treated as non-geocodable.

Matching of the traditionally listed addresses with addresses on the USPS-based address list was performed using an automated matching program followed by manual matching. The retained USPS-based and traditionally listed addresses were merged by street number, street name, street suffix, pre- and post-direction, unit designator, unit number, and state abbreviation. City and ZIP code were not used in the matching because listers might not have been able to acquire accurate information about these two address variables while in the field. All unmatched addresses after the automated matching were investigated manually to resolve minor discrepancies, such as differences in spelling or typos (e.g., “Weatherby Rd.” vs. “Wetherby Rd.”), differences in street type (e.g., “Oak St.” vs. “Oak Ln.”), and “No number” addresses (e.g., matching a “no number” address listed between 123 Main St. and 127 Main St. with a “125 Main St.” listing on the USPS-based list).

Lastly, the match rates, the percentage of traditionally listed addresses that were on both the traditional list and the USPS-based list, were calculated and examined. Results indicated that match rates were higher in urban PSUs than in rural PSUs, and lower in high-population-growth areas. However, within a PSU, there was considerable variation in match rates at the segment level. Thus, it is possible (and in fact anticipated) that in many cases, the USPS-based list would be adequate for some segments within a PSU but would be considerably inferior to traditional field listing for other segments within the same PSU.

3. Current Methodology

In this paper, we extend the earlier research by examining NCS Vanguard Study household enumeration data for all addresses obtained using the traditional field listing approach. The current research includes the missed units found using the coverage enhancement procedures (both the missed structure procedure and the missed DU procedure). Using the earlier matching results and additional data, our research objective was to compare addresses on USPS-based list to addresses obtained through traditional field listing or identified during missed unit procedures. The USPS-based listings were contemporaneous with the traditional listings, but the data collection (including the missed unit procedures) lagged these by nearly a year. While the earlier research focused on the presence of addresses on the sampling frame, the current research takes a further step to determine whether the addresses were occupied by households. An omitted address has coverage implications only if it is occupied by a household. We then use respondents’ data to examine the characteristics of households and persons according to whether or not their addresses were on the USPS-based frame.

Figure 1 identifies the various address groups of interest and shows their relative sizes in the NCS evaluation. It should be noted that the relative sizes of the groups are partly a reflection of the particular PSUs in our study, as the figures were not weighted to be nationally representative. For our research, we identified addresses that were obtained by traditional listing (T), on the USPS-based list (U), or picked up during missed unit procedures (M), along with various intersections of these sets. The missed unit set (M) includes addresses obtained from both types of missed unit procedures (missed structure and missed dwelling unit), since these were not distinguishable in the enumeration data.

Since the Vanguard Study data collection was based on the traditional field listing approach, enumeration data are not available for the USPS-only addresses (5.3 percent of the total number of addresses under consideration).

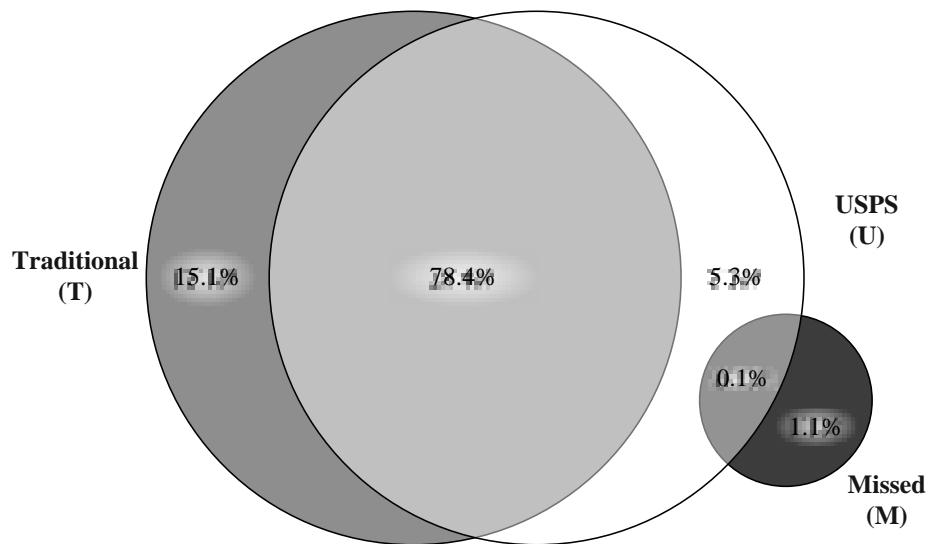


Figure 1: Address groups in the NCS Vanguard Study address list evaluation

Note: Percentages indicate the sizes of the groups in the evaluation; these are unweighted.

The purpose of our research was to help address the question of when USPS-based sampling frames can adequately replace the traditional field listing approach. For our analyses, we first examined the eligibility and response status of all addresses identified through the approach of traditional field listing plus missed unit procedures ($T \cup M$). We then compared the eligibility of addresses (i.e., the presence of households at addresses) that were both traditionally listed and on the USPS-based list ($T \cap U$) to those that were traditionally listed but not on the USPS-based list ($T \cap U^c$). Ideally, we would have liked to have compared the addresses obtained through the traditional listing approach ($T \cup M$) to those obtained through an alternative approach of USPS addresses that geocoded to blocks within sampled segments together with its own missed unit procedure (i.e., $U \cup M$); however, because the traditional listing approach was used for the NCS Vanguard Study, information was not available for addresses other than those obtained through the traditional listing approach. Lastly, we used respondent data to compare characteristics of households between addresses on the USPS-based list and addresses not on the USPS-based list, both with and without missed units.

4. Results

This section gives the results of our evaluation. Section 4.1 examines address eligibility. Section 4.2 discusses characteristics of responding households.

4.1 Results on the Presence of Households at Addresses

We begin with a characterization of the dispositions of the addresses obtained through the traditional listing approach, including those found through the coverage enhancement procedures. Here, eligibility refers simply to whether a particular address is a household;

it does not consider characteristics of the household members. The results in Table 1 show that, although the eligibility rate was much lower for missed units than for traditionally listed addresses (just over 80 percent, compared to 93 percent) and the response rate for missed units was lower than for traditionally listed addresses (67 percent versus 84 percent), the missed units have little effect on the overall eligibility and response rates attained through the traditional listing approach, due to the fact that they comprise a very small proportion of all addresses identified through that approach.

Table 1: Eligibility and Response Status of Addresses Obtained through the Traditional Listing Approach

<i>Eligibility Status</i>	<i>Address Subset</i>		
	T	M	T ∪ M
Household respondent	78.1%	54.3%	77.8%
Household nonrespondent	15.2	26.2	15.4
Not a household	6.6	19.5	6.8
Total	100.0%	100.0%	100.0%
Number of addresses	81,120	1,096	82,216

NOTE: Due to rounding, details may not sum to 100 percent.

Next, we compared the eligibility of addresses that were both traditionally listed and on the USPS-based list to those that were traditionally listed but not on the USPS-based list. Missed units are not considered here, because this discussion is meant to characterize only traditionally listed addresses according to whether or not they appear on the USPS-based list. However, when missed units were included, very similar results were obtained. Among the traditionally listed addresses, 84 percent were on the USPS-based list, and the remaining 16 percent were only on the traditional listing. These figures pertain to the particular PSUs in our study, and are not weighted to be nationally representative.

Table 2 shows that compared to addresses that were traditionally listed but not on the USPS-based list, those on both traditional and USPS-based lists have a higher eligibility rate (96 percent vs. 78 percent). This is evidence of the fact that listers were trained to be conservative and record every address that appeared to be residential, even those that were ambiguous. The primary difference is that only 2 percent of traditionally listed addresses on the USPS-based list were found to be vacant, compared to 10 percent of traditionally listed addresses not on the USPS-based list. Eligible addresses on both traditional and USPS-based lists also have a higher overall response rate (86 percent, compared to 71 percent for addresses only on traditional lists). These higher eligibility and response rate findings were consistent across the seven PSUs.

Although the traditional-only subset was rather large in our study (about 15 percent of all addresses, as shown in Figure 1), the fact that about 20 percent of these are not households means that this set is indicative of undercoverage of only about 12 percent of households in our study (and the undercoverage is predominantly restricted to rural PSUs). It is likely that the undercoverage rate is actually even lower, since some of the 12 percent of households that appeared to be in the traditional-only subset may also be contained in the USPS-only subset due to the inability to match the addresses.

Table 2: Eligibility of Addresses on Both Traditional and USPS-based Lists and for Addresses only on Traditional Lists

<i>Eligibility Status</i>	<i>Address Subset</i>	
	<i>Both traditionally listed and on USPS-based lists</i> ($T \cap U$)	<i>Only traditionally listed</i> ($T \cap U^c$)
Eligible respondent	82.4%	55.8%
Eligible nonrespondent	13.9	22.4
Not a household	3.7	21.8
Total	100.0%	100.0%
Number of addresses	68,036	13,084

Table 3 breaks down ineligibility and response rate results from Table 2 into four sub-categories: segments in urban PSUs that have a high match rate; segments in urban PSUs that have a low match rate; segments in rural PSUs that have a high match rate; and segments in rural PSUs that have a low match rate. Segments with a match rate of 80 percent or greater were classified as “high”, and segments with a match rate below 80 percent were classified as “low”. The results demonstrate that the pattern of higher eligibility and response rates for addresses on both the traditional and USPS-based lists are consistent regardless of urbanicity of the PSU and match rate.

Table 3: Ineligibility and Response Rates of Addresses on Both Traditional and USPS-based Lists and of Addresses Only on Traditional Lists, by Characteristics of the PSU and Segment

	<i>Address Subset</i>	
	<i>Both traditionally listed and on USPS-based lists</i> ($T \cap U$)	<i>Only traditionally listed</i> ($T \cap U^c$)
Segments with high match rate in urban PSUs		
Not a household	2.8%	23.4%
Response rate	88.4	79.5
Segments with low match rate in urban PSUs		
Not a household	2.3	22.2
Response rate	80.0	68.1
Segments with high match rate in rural PSUs		
Not a household	5.4	38.0
Response rate	82.9	82.4
Segments with low match rate in rural PSUs		
Not a household	7.2	20.6
Response rate	75.8	69.1

4.2 Analysis of Characteristics of Responding Households

Using respondent data, we next compared characteristics of households on the USPS-based list to those not on the USPS-based list. Only 5 percent of respondents in urban PSUs were only on traditional listings, and 28 percent of respondents only on traditional

listings were in urban PSUs. Since the majority of addresses not on the USPS-based list were from the two rural PSUs, we restricted this analysis of demographic characteristics to the two rural PSUs.

Within the two rural PSUs, traditionally listed and missed addresses on the USPS-based list comprised about 70 percent of respondents, while traditionally listed and missed addresses not on the USPS-based list comprised about 30 percent of respondents. The substantial percentage of addresses not on the USPS-based list was due in part to the large number of cases where listers were unable to obtain complete street addresses. Rural areas are also more likely to have households that get their mail at P.O. boxes only.

The characteristics considered include the type of dwelling unit and characteristics of the responding household member. We included whether or not the household had adult females ages 18-49, since this is the subpopulation the NCS seeks to enroll. Results are given in Table 4.

Table 4: Characteristics of Addresses on Both Traditional and USPS-based Lists vs. Addresses Only on Traditional Lists, for Addresses in the Two Rural PSUs

<i>Household Characteristic</i>	<i>Address Subset</i>		<i>p-value</i>
	<i>Both traditionally listed and on USPS-based lists</i> ($T \cap U$)	<i>Only traditionally listed</i> ($T \cap U^c$)	
Single-family home	91.9%	97.3%	0.017*
More than one adult in household	70.9	70.5	0.790
Household has adult males	80.7	82.0	0.281
Household has adult females	83.6	84.2	0.687
Household has adult females 18-49	43.2	40.4	0.082
Interview conducted in English	98.5	96.3	0.007*
Respondent Hispanic	4.3	8.3	0.013*
Respondent Black	9.0	10.1	0.699
Number of responding households in the two rural PSUs	12,092	5,239	

*indicates significance at the 0.05 level

Comparing addresses on both traditional and USPS-based lists to addresses only on traditional lists, the differences between these two groups were significant for three characteristics: type of dwelling unit, interview conducted in English, and Hispanic respondent. These significant differences remained the same after adding in the missed units to the two groups analyzed.

Since the results only apply to the two rural PSUs in the NCS Vanguard Study (and may, in fact, be due at least in part to differences in characteristics of segments in the rural PSUs for which the USPS-based list provides good coverage and those for which the USPS-based list provides poorer coverage), we caution against focusing on the particular findings. Instead, we suggest that the results are simply an indication that coverage bias could result if addresses missing from a USPS-based address frame are not captured using coverage enhancement procedures.

5. Discussion and Conclusions

By using household enumeration dispositions and data, our research extends beyond the address matching that previous studies have used to compare USPS-based lists to traditional field listing. In our study, addresses that were both traditionally listed and on the USPS-based list comprised 84 percent of all addresses that were traditionally listed. Among the addresses that were only on traditional lists, only 78 percent were eligible households, compared to an eligibility rate of 96 percent for addresses on both lists. Eligible addresses on both lists also had a higher overall response rate (86 percent, compared to 71 percent for addresses only on traditional lists). These higher eligibility and response rate findings were consistent at the PSU-level and also at the segment-level within groups of segments defined by match rate and PSU urbanicity. Overall, our research supports the case for the replacement of traditional field listing with USPS-based address lists when generating frames for household area probability sample surveys, with a few important considerations for doing so.

In our study, the majority of the traditionally listed addresses not on the USPS-based list were from the two rural PSUs. Respondent data within these rural PSUs showed that compared to traditionally listed addresses not on the USPS-based list, traditionally listed addresses also on the USPS-based list have some differing demographic characteristics. This is an indication that if USPS-based lists are to replace traditional field listing of addresses, coverage enhancement procedures need to be implemented. Missed dwelling units are an issue of approximately equal burden no matter which method is used and can be identified during data collection using existing approaches. However, the proportion of and procedures required for missed structures could vary between these two methods. With traditional field listing, since all addresses in an area are listed in a systematic, pre-specified sequence, missed structure procedures such as the half-open interval are easily used. With the USPS-based approach, there is a need to consider alternative approaches for identifying missed structures, unless the entire household frame in a segment can be sequenced in a manner similar to that available from traditional field listings.

Our findings are consistent with previous research in that USPS-based lists provide generally better coverage in urban areas compared to rural areas, but we were able to delve further into what types of addresses and households could be missed by failing to include traditionally listed addresses not on USPS-based lists. Continued evaluation of the USPS-based lists is warranted, as enhancements such as increases in the proportion of city-style addresses and improvements to geocoding databases occur over time.

References

- English, N., O’Muirheartaigh, C., and Eckman, S. 2009. Coverage Rates and Coverage Bias in Housing Unit Frames. *Proceedings of the Survey Research Methods Section, American Statistical Association*.
- Hsu, V., Montaquila, J. M., Brick, J. M., 2010. Using a Match Rate Model to Predict Areas Where USPS-Based Address Lists May Be Used in Place of Traditional Listing. *Proceedings of the Survey Research Methods Section, American Statistical Association*.
- Iannacchione, V., Morton, K., McMichael, J., Cunningham, D., Cajka, J. and Chromy, J. 2007. Comparing the Coverage of a Household Sampling Frame Based on Mailing

- Addresses to a Frame Based on Field Enumeration. *Proceedings of the Survey Research Methods Section, American Statistical Association.*
- Kish, Leslie. 1965. *Survey Sampling*. New York: John Wiley and Sons, Inc.
- Montaquila, J. M., Brick, J. M., Curtin, L. R. 2010. Statistical and practical issues in the design of a national probability sample of births for the Vanguard Study of the National Children's Study. *Statistics in Medicine*, 29(13), 1368 - 1376.
- Montaquila, J.M., Hsu, V., Brick, J.M., English, N., and O'Muircheartaigh, C. 2009. A Comparative Evaluation of Traditional Listing vs. Address-Based Sampling Frames: Matching with Field Investigation of Discrepancies. *Proceedings of the Survey Research Methods Section, American Statistical Association.*
- O'Muircheartaigh, C., English, N., Eckman, S., Upchurch, H., Garcia, E., and Lepkowski, J. 2006. Validating a Sampling Revolution: Benchmarking Address Lists against Traditional Listing. *Proceedings of the Survey Research Methods Section, American Statistical Association.*
- O'Muircheartaigh, C., English, N., Eckman, S. 2007. Predicting the Relative Quality of Alternative Sampling Frames. *Proceedings of the Survey Research Methods Section, American Statistical Association.*