Validating a Sampling Revolution: Benchmarking Address Lists against Traditional Listing

Colm O'Muircheartaigh[1], Ned English[1] , Stephanie Eckman[1], Heidi Upchurch[1] , Erika Garcia[1], and James Lepkowski[2]
National Opinion Research Center, 55. E Monroe St, Chicago, IL, 60603 [1]
University of Michigan, Ann Arbor, MI [2]

Abstract

The paper presents empirical findings from a comparison of two approaches for national area probability sampling. The National Opinion Research Center (NORC) at the University of Chicago and the Institute for Social Research (ISR) at the University of Michigan are collaborating to compare two national area-probability sampling frames for household surveys: (i) the frame produced by traditional listing, using survey field staff, and (ii) the list of postal addresses compiled by the United States Postal Service (USPS). We conducted this comparison in an ongoing survey operation which combines the current wave of the Health and Retirement Survey (HRS) with the first wave of the National Social Life, Health, and Aging Project (NSHAP). Since 2000, survey samplers have been exploring the potential of the United States Postal Service (USPS) address lists to serve as a sampling frame for probability samples from the general population. Though the early work has demonstrated the strengths of the USPS address lists, there has not been a comparison of the two methods on a national scale. We report the relative coverage properties of the two frames, as well as predictors of the coverage and performance of the USPS frame. The research provides insight into the coverage and cost/benefit trade-offs that researchers can expect from traditionally listed frames and USPS address databases. The results in this paper are not weighted to represent the population of the US as a whole.

Keywords: USPS List Frames, GIS, Area Probability Sampling

1. Introduction

While it has been generally assumed that field listing is the gold standard for generating area-probability samples in terms of coverage and accuracy, until now there has been no national evaluation of the process. NORC has, since 2001, been carrying out an examination of the alternative approach of using the United States Postal Service (USPS) list as a basis for frame construction for area probability surveys. In 2004, NORC and ISR embarked on a national benchmark comparison, whose goal was to provide a quantitative analysis describing the benefits and drawbacks of traditional listing (the "gold standard") vs. the USPS list (O'Muircheartaigh et al. 2005).

In our earlier report on this research, we have compared a traditionally-listed list frame to a USPS-based frame in the same set of areas (O'Muircheartaigh et al. 2005). When discrepancies arose between the two frames, however, it was not possible to determine the source of the error, or which frame, was more accurate.

In order to address this question, we conducted additional field work in a set of areas that covered the major dimensions of variation identified earlier. This "best" frame provides a basis for determining the performance of the two approaches relative to this best frame, and to distinguish the circumstances under which each approach may be preferable.

2. Background and Problem

The research was undertaken as a methodological supplement to the National Social Life and Health in Aging Project (NSHAP) using field listing and screening for the Health and retirement Survey (HRS), both NIH/NIA projects. [1]

ISR at the University of Michigan conducted a listing in 2004 of 549 national segments for the HRS survey. NORC and ISR collaborated on the below research as a methodological supplement to evaluate the overall quality of the ISR listing. NORC then used the results of ISR screening as the sampling frame for the NSHAP survey.

The two components of the comparison are as follows. ISR provided the whole frame of listings (and not just the sampled households) for this project. NORC licenses the national USPS delivery point database from ADVO, which contains a record for every

[1] This research was supported by National Institute on Aging (NIA) grant R01-AG021487-02S1; L Waite, Principal Investigator.

mailing address in the US; NORC identified, by geocoding, the relevant parts of the this database for the comparison.

In 2005 NORC compared the addresses from the traditional ISR listing (to be referred to as 'T') to those on the USPS list (to be referred to as 'U') for the 549 segments in the HRS/NSHAP sample. The results were somewhat disappointing in that the overall match rates were lower that expected (O'Muircheartaigh et. al 2005). There were strong indications that each of the approaches was unsuccessful in particular situations; the failure of traditional listing was particularly significant, given its dominance in national studies.

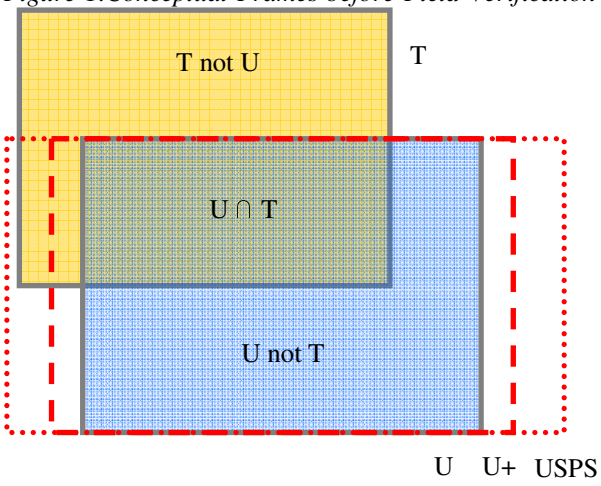
Consequently, we decided to undertake validation on a sample of the 549 segments. We selected 100 segments for field verification/testing of the two lists in order to identify the levels of agreement and then to explore factors related to quality of each frame. We conducted the field validation during the summer of 2005 as a supplement to the NSHAP field work.

3. Methodology

Two datasets, the 'T' or traditional and 'U' or USPS lists, were the basis of this study. The first, the 'T' list, was generated from the HRS listing, a field listing of a national probability sample of 549 segments. This listing was conducted by field staff using TIGER-based maps; each address was identified in the dataset as belonging to the sampling unit and segment in which it was listed.

The second dataset used in this study, the 'U' list, was derived from the USPS list that NORC licenses. NORC's USPS list contains all delivery points in the United States, but without any *a-priori* segment information. In order to produce a frame for comparison, we therefore needed to utilize GIS technology to geocode each USPS address and associate them with Census boundaries. Geographic processing permitted us to define delivery points that geocoded inside the segment boundaries as the 'U' frame. Because we have learned that geocoding can be inexact due to irregularities in street block ranges, we defined a 'U+' frame as those delivery points that geocoded within a 300' buffer of each segment, intended to represent the width of a large city street. Finally, we define as the 'USPS' frame all delivery points in the US, whether they geocoded inside an HRS segment or not. Figure 1 is a graphical representation of the intersection of T, U, U+, and the USPS. Important intersections are labeled.

Figure 1.Conceptual Frames before Field Verification



Cost and time considerations precluded carrying out fieldwork in all 549 segments. We stratified the set of segments according to a set of criteria that had potential as predictors of frame quality. Using a geographic dataset describing the structural characteristics of the 549 segments, we classified each segment as being urban, suburban, or rural. From our examination of the T/U match pattern, we categorized each segment as being either 'interesting' e.g. having characteristics that would be valuable to explore in the field, as opposed to 'other' e.g. not having notable characteristics. Finally, using the results of the previous research (O'Muircheartaigh et al. 2005) we classified each segment in these larger categories as follows: having high match rates in both T and U; having moderate match rates in both T and U; a U frame that was much smaller than the T frame; a T frame that was much smaller than the U frame; low matches presumably due to geocoding error; low matches due to unknown issues.

Following stratification, we selected a national sample of 100 segments, as mapped in Figure 2 and summarized in Table 1. Table 1 shows the selected segments in each cell, with the frame count in parentheses.

We then created a hybrid or comprehensive list within each segment, containing all lines from each source frame (T union U). This list was known as the 'raw' or 'R' frame because it was simply the sum of all addresses and therefore also contained any errors from either the USPS or the traditionally-listed frame. The 'raw' frame was thus expected to contain some addresses that were not actually in the segments, but would also fail to contain 'real' addresses that were missing from both frames.

Our field staff was provided with the R frame (which did not contain information on the source of each listing). They checked the status of each address on the list and each address was either confirmed as being in the segment or non-existent, demolished, or outside the segment; for each address we also determined whether or not it was equivalent to another address on the list. Furthermore, each address was classified as appearing 'newly constructed' or 'old'. In addition, listers were instructed to add addresses determined to be missing from the R frame; this is known as 'enhanced' listing' (O'Muircheartaigh et. al 2003). The purpose of this field verification was to create an optimal description of reality, and so generate what we call the 'Best' or 'B' frame.

Field verification allowed us to identify all housing units actually in each segment, as opposed to those erroneously added or missed by the original traditional or 'T' listers. We were also able to identify those addresses accidentally included due to geocoding error on the U frame. We argue that the B frame constitutes a new gold standard for list comparison, as it contains the best of both traditional and USPS lists.

Figure 3 illustrates the new partitioning into now have the following representations of the world: traditional or 'T' frame; the 'U' frame; the 'U Plus frame; and finally the 'B' Frame. Figure 3 also shows the complete intersections of these four frames. The question is then how do the T and U frames themselves relate to reality, or the 'B' frame.

Figure 3. Conceptual Frames after Field Verification

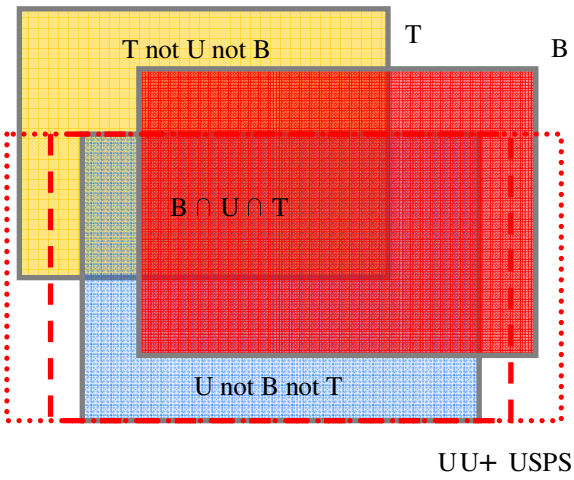


Figure 2 Counties by Selected Segment Count

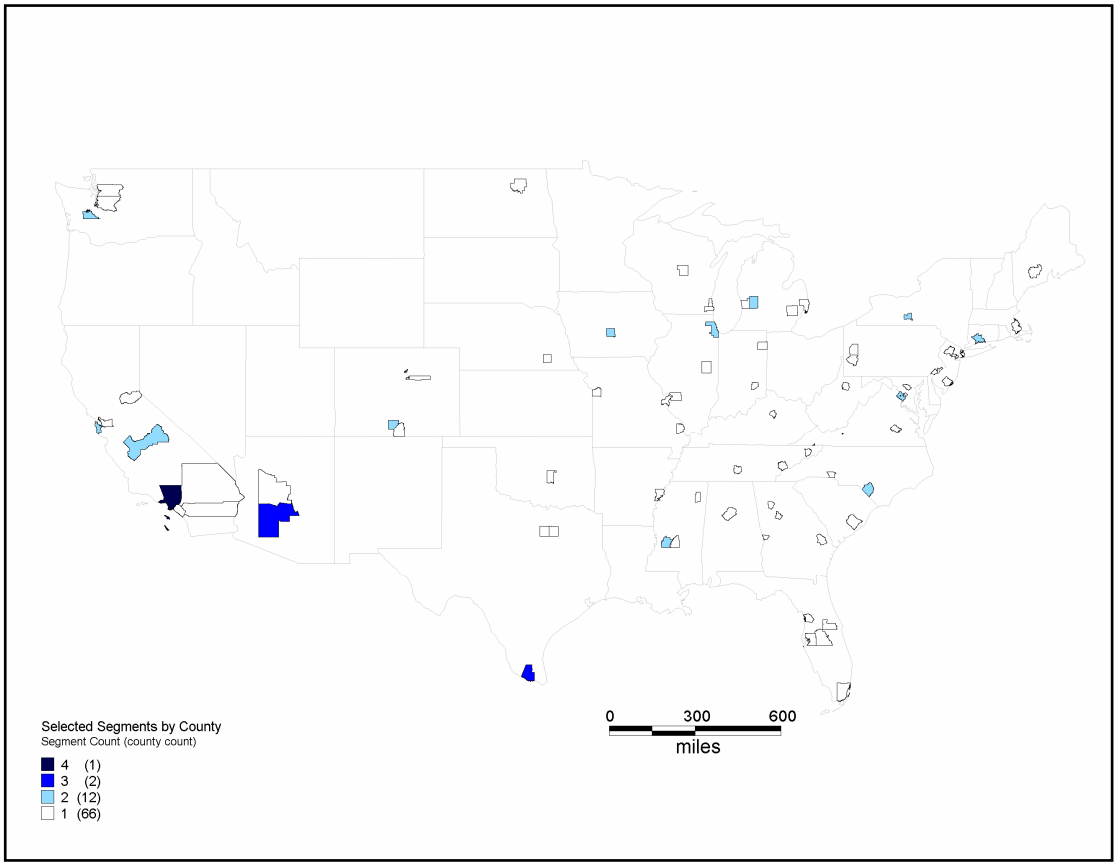


Table 1- 100 Selected Segments with Frame Counts by Cell

	Urban			**Suburban**			**Rural**		
	Interesting Cases	Other	*Total Urban*	Interesting Cases	Other	*Total Suburban*	Interesting Cases	Other	*Total Rural*
High Match	2 (4)	11 (83)	*13 (87)*	2 (3)	11 (63)	*13 (66)*	0 (0)	1 (2)	*1 (2)*
Moderate Match	5 (16)	5 (34)	*10 (50)*	2 (12)	6 (38)	*8 (50)*	0 (1)	1 (4)	*1 (5)*
U lower than T	4 (11)	1 (5)	*5 (16)*	3 (25)	1 (12)	*4 (37)*	4 (12)	0 (9)	*4 (21)*
T lower than U	5 (20)	1 (12)	*6 (32)*	2 (13)	3 (8)	*5 (21)*	1 (2)	0 (0)	*1 (2)*
Low match: Unknown	3 (5)	1(8)	*4 (13)*	7 (14)	1 (13)	*8 (27)*	1 (3)	0 (4)	*1 (7)*
Low match: Geocoding	1 (3)	4 (18)	*5 (21)*	6 (12)	1 (30)	*7 (42)*	0 (5)	4 (18)	*4 (23)*
Total	20 (59)	23 (160)	***43 (219)***	22 (79)	23(164)	***45 (243)***	6 (23)	6 (37)	***12 (60)***

4. Results and Discussion

Table 2 shows the overall intersections among B, T, U, U+, and the entire USPS for the 100 selected segments. As demonstrated, overall 73% of B is in T, 77% of B is in U, 83% of B is in U+, and 84% of B is in the entire USPS. If one excludes four segments with particular structural issues that diminished their match rates (such as trailer housing or the absence of USPS addresses entirely) then the U matches improve by a few points.

Table 2. Overall Intersections of B with U and T

% of B

Frame	*All segments*	*Without excluded segments*
T	73%	72%
U	77%	80%
U+	83%	86%
USPS	84%	87%

It is important to note that while the HRS listing and USPS frame are contemporaneous, there was a time lag between the HRS listing in 2004 and the field verification in 2006. Therefore, there was the possibility of change in any segment through demolition and/or construction. Our data base did, however, permit the inclusion of addresses during field verification that were not on the T or U lists, and to describe such additions as appearing 'new' (e.g. since 2004, implying new construction) or 'not new' (not likely since 2004, and thus a dataset deficiency). For some purposes of comparison we removed 'new' additions from B to produce a frame called B'. B' is believed to be closer in time to the T list than B. If one compares B' to B, one generally sees 2% to 4% improvements in rates, with the 'truth' likely between B and B'. Table 3 summarizes the overall intersections as in Table 2, but for B' rather than B.

Table 3. Overall Intersections of B' with U and T

% of B'

Frame	*All segments*	*Without excluded segments*
T	77%	76%
U	80%	83%
U+	86%	89%
USPS	88%	91%

A major underlying question is how do different categories of segments perform with regard to matching to B. Table 4 shows the results for 20 rural segments, defined as our pre-field 'rural' categorization plus any others enumerated by the US Census Bureau as such. As shown, T is considerably more effective than U in rural areas, with 79% of B appearing in B and only 51% of B appearing in U.

Table 4. Intersections of B and B' with U and T in 20 Rural Segments

Frame	% B	% B'
T	79%	87%
U	51%	56%
U+	56%	61%
USPS	58%	64%

The converse is true in the 80 non-rural segments, shown in Table 5. U is considerably more effective than T in non-rural areas, with 82% of B in U in comparison with 72% of B in T. Geocoding error plays a role in these physically

compact urban and suburban segments (often on the scale of one or two blocks), as demonstrated by the gain in coverage when we consider U+ rather than U.

Table 5. Intersections of B and B' with U and T in 80 non-Rural Segments

Frame	% B	%B'
T	72%	74%
U	82%	85%
U+	87%	91%
USPS	89%	92%

Beyond the descriptors of urbanicity, it is also useful to examine some functional groups of segments derived from other factors that could classify segments *a-priori*. For example, one could consider the 18 segments where U is considerably smaller than the Census count, defined as less than 70%. As shown in Table 6, for these segments 81% of the B addresses are in T, 48% are in U, 53% are in U Plus, and 56% are on the entire USPS. If one determines that the U list in an area is below a certain tolerance with respect to the Census count, then clearly traditional listing is worth considering.

Table 6. Intersections of B' with U and T in 18 Segments where U << Census

Frame	% B	% B'
T	81%	91%
U	48%	54%
U+	53%	60%
USPS	56%	63%

The opposite effect is present in the 12 segments where U is considerably larger than the Census count, as shown in Table 7. These segments would be expected to have experienced rapid growth since Census 2000, and saw 70% of B in T, 81% of B in U, 90% of B in U+, and 92% of B in USPS. U is more effective than T in these segments.

Table 7. Intersections of B' with U and T in 12 Segments where U >> Census

Frame	% B	%B'
T	70%	73%
U	81%	85%
U+	90%	94%
USPS	92%	96%

One can also consider the 69 most 'promising segments' for U, as summarized in Table 8. These segments exclude any rural segments or those where the U count is much less than the Census count. For the most promising segments, 71% of B is in T, 86% of B is in U, 91% of B is in U+, and 93% of B is in USPS.

Table 8. Intersections of B' with U and T in 69 'Most Promising' Segments

Frame	% B	%B'
T	71%	72%
U	86%	88%
U+	91%	93%
USPS	93%	95%

Lastly, there are also the 31 least-promising segments, as summarized in Table 9. These segments are either classified as rural or are situations where the U frame is considerably smaller than the Census count. In the least-promising segments 78 % of B is in T, while only 54% of B is in U. Adding a buffer zone does not help appreciably, as the percentage of B in U+ is 59% and in the entire USPS is 62%.

Table 9. Intersections of B' with U and T in 31 'Least Promising' Segments

Frame	% B	%B'
T	78 %	87 %
U	54 %	60 %
U+	59 %	66 %
USPS	62 %	69 %

Field verification removed addresses from the 'raw' or R frame that were not actually in a segment, which we can call 'over-coverage'. Over-coverage comprises wasted T lines and wasted U lines, e.g. superfluous cases from either list. Table 10 shows the extent of over-coverage overall for each frame. Clearly, there were more false inclusions on the U frame as a result of geocoding or other dataset errors, but there were also some on the T list.

Table 10. Overall Extent of Overcoverage

	% of B	
Frame	*All segments*	*Without excluded segments*
U-U∩B	10%	11%
T-T∩B	5%	6%

Summary

Overall, U is more effective than T with 77% of B in U and 73% of B in T; if we relax the geocoding criteria, the coverage of USPS rises to 84% overall. In some areas, however, T is indeed better than U. Our research shows that potentially problematic segments can be determined *a-priori*.

We found areas of poor U performance to fall into three categories. Firstly, areas with complex geographies, especially those with irregular street patterns (old cities, those dissected with water) were more subject to geocoding error. Secondly, if the U count was considerably smaller than the Census count, there would be no chance for a satisfactory intersection with the B frame. Lastly, areas with high population growth rates, and thus out-of-date listing, tended to be characterized by better intersections with T than U. Nonetheless, taken as a whole, U is superior to T.

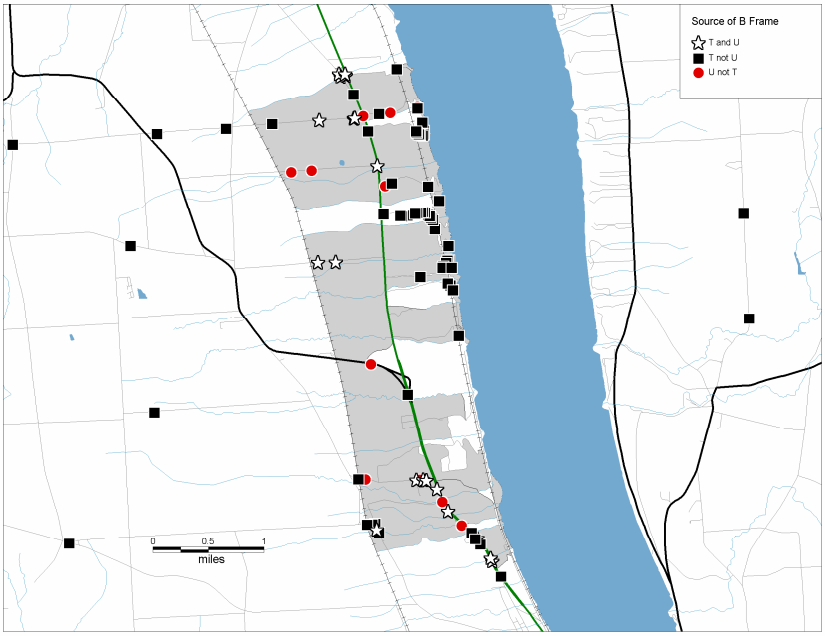
It is useful to consider examples of segments where U was more effective than T and vice-versa. Figure 3 shows a suburban segment where 99% of B was in U, but only 46% of B was in T. This segment could have been flagged as a potential problem *a-priori* with a U count that was considerably higher than the T count. One can see that the traditional listers omitted some streets that were included by the U frame.

Figure 4 shows an example of a rural segment where 91% of B is in T, but only 22% of B is in U. This segment has a substantial number of PO BOX addresses, which are non-geocodable and thus do not contribute to the U frame. Also, it is clear that this segment has numerous blocks dissected by streams. One could conclude *a-priori* that this segment was a candidate for traditional listing.

Geocoding itself is an issue for samples in areas that are geographically compact, such as segments of only a few blocks. Overall 7%, of B was mis-geocoded, as shown by the difference between B in U and B in USPS. The problem was exacerbated for the HRS sample design by the very small size of the HRS segments. Therefore, if location within small areas is a critical design element, one should be cautious about depending solely on geocoded lists. Mis-geocoding could also affect the precision of stratification in any survey, with members of a particular cell being located in an incorrect area. If sampling units themselves are larger Census geographies, however, then the level of error will be reduced even if some selections will be in unexpected locations. Postal list-based frames, such as those constructed from ZCTAs, would theoretically avoid geocoding issues altogether but would not have the same availability of Census data for stratification.

Figure 3 Example Segment where U is More Effective than T

Figure 4 Example Segment where T is More Effective than U



It is clear from our research that there is no single best frame for all circumstances. One can argue, however, that in most non-rural areas the U frame is at least as effective as the T frame with regard to coverage. Fundamentally, the most suitable situations to employ the U list would be in urban areas with city-style addresses. The U list is not recommended in rural areas, as can be determined *a-priori* by the prevalence of PO-BOXes or U counts considerably smaller than Census counts. Even in very rural areas with few city-style addresses, however, 51% of B is in U. We contend that even in such situations, where U alone would be inadequate, it should be considered a basis for field listing.

While we feel our results demonstrate the promise of USPS-derived lists, the situation is changing dynamically as a greater share of US addresses is becoming city-style and thus geocodable within a small area. One would therefore expect the performance of the U list to improve over time.

It is also important to keep in mind that our results are preliminary and are not weighted to represent the US population as a whole. Even though weighting will alter some of the coverage values presented, we feel that weighting the results will not affect the narrative. We will report the weighted results in a later paper.

5. Conclusions

In conclusion, our results demonstrate that the U list is a better representation of reality than the T list in most situations. These results also indicate the situations that would specifically benefit from the T list, e.g. rural areas or those with a small ratio of U to Census. Going forward, we intend to develop a more detailed model of quality prediction, based on the classification of each segment into functional groups.

Acknowledgements

The authors would like to thank the NORC NSHAP team, especially Stephen Smith, Project Director, and Sara Zuckerbraun, Associate Project Director.

References

O'Muircheartaigh, Colm, Stephanie Eckman, Ned English, James Lepkowski, and Steven Heeringa. *Comparison of Traditional Listings and USPS Address Database as a Frame for National Area Probability Samples.* Presented at American Association for Public Opinion Research Conference, May 2005, Miami Beach, Fl.

O'Muircheartaigh, Colm, Stephanie Eckman and Charlene Weiss. 2003. "Traditional and Enhanced Field Listing for Probability Sampling." 2002 Proceedings of the Section on Survey Research Methods, American Statistical Association.