Is Housing Unit Undercoverage Random?

Timothy L. Kennel[1,2], Xijian Liu[1,2]
[1]US Census Bureau, 4600 Silver Hill Road, Washington, DC 20233
[2]This paper is released to inform interested parties of research and to encourage discussion. Any views expressed on methodological or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

Abstract
The coverage of many Address Based Sampling Frames has been well documented; however, much less attention has been paid to the mechanisms generating coverage errors. A clearer understanding of the mechanisms generating undercoverage can help us more effectively target areas needing improvement or listing. To investigate correlates of Housing Unit undercoverage on Address Based Sampling Frames, we used preliminary results of the US Census Bureau's Address canvassing operation to assess the coverage of an Address Based Sampling Frame. We then explored the relationship between various block characteristics to the Housing Unit undercoverage rate. Our results provide insight into the distribution of omissions on an Address Based Sampling Frame.

Key Words: Address Based Sampling, Coverage, Delivery Sequence File, Frame Creation, Master Address File, Listing

1. Introduction

The advent of national address lists derived from the United States Postal Service's Delivery Sequence File is revolutionizing survey frame creation activities and data collection methods. New questions and research opportunities accompany the shift toward national address lists. Early research focused on national coverage rates and quickly progressed to coverage measurement by geography and sub-domains. New research has included estimates of bias and discussions of various operational issues. Numerous studies have all found strong evidence of differential coverage between urban and rural areas, spurring many survey research centers to supplement their address list with field listings in rural areas.

Yet, some rural areas have excellent coverage, while others do not. The drive to partition the nation into areas needing coverage improvement and those that do not has sparked much research on targeting. Operationally, targeting has the potential to reduce costs and improve coverage, since only sample areas needing coverage improvement are subject to the expensive listing operation. In this paper, we use a large national canvassing of the United States to determine which blocks should be targeted for Coverage Improvement.

2. Background

2.1 The Delivery Sequence File

To facilitate the sorting and delivery of mail, the Postal Service has created a national file of addresses, called the Delivery Sequence File (DSF). Prior to 1994, Section 412 of United States Code 39 stated, "Except as specifically provided by law, no officer or employee of the Postal Service shall make available to the public by any means or for any purpose any mailing or other list of names or addresses (past or present) of postal patrons or other persons." In 1994, the United States Congress amended the title to require the US Postal Service to provide addresses and address-related information to the US Census Bureau. As a result, the US Postal Service sends a copy of the DSF to the Census Bureau twice a year. Additionally, a limited number of restricted licenses are available for establishments to purchase access to the DSF. Numerous survey research centers have purchased frames and samples from data brokers who have access to the DSF through a license with the US Postal Service.

The DSF provides access to nearly all postal addresses in the United States. In some areas, local post offices sort their own mail, and thus addresses in such areas are not included on the DSF. Furthermore, some Native American lands have their own postal system and are not included in the DSF. Some data vendors enhance the DSF with supplementary address lists from sources such as credit card transactions, white pages, home sales data, local property tax assessments, and other administrative records. These additional sources may provide some additional coverage of housing units that do not appear on the DSF. They may also provide complete city style addresses for housing units that have incomplete or rural style addresses on the DSF. Further enhancements can be made by attaching household characteristics and person level data from other sources or models.

In 2000, RTI created a sampling frame using addresses purchased from ADVO and Donnelley Marketing Services (Iannacchione, 2003). Although their survey was limited to a large metropolitain area, RTI noted that coverage in rural areas of the country may be deficient because "home delivery of mail is less prevalent in rural areas." They concluded that "a combination of mailing lists in urban areas and on-site enumeration in rural areas is desirable for household surveys that are national in scope." Since then, the US Census Bureau, National Opinion Research Center, Westat, the Survey Research Center at the University of Michigan, and other survey research centers reached similar conclusions based on evaluations of address lists.

Considerable research has been conducted on the characteristics of addresses that are completely omitted from the DSF. Some housing units do not receive any mail. For example, some basement apartments, in-law suits, and attic apartments share an address with the main housing unit. Other clandestine housing units, such as guest houses, garages, and pool houses, may be rented to illegal immigrants, fugitives, or others wishing to be "off the grid." Areas without strong governmental resources, such as colonias, may also not be fully represented on the DSF. Sometimes mail for multi-unit apartment buildings is dropped off in a drop box and sorted by a building employee. If the DSF only contains one record for the basic street address, then all of the units within the building will be undercovered by the DSF (Montaquila, 2009). Some very rural post offices still sort their own mail without the assistance of the DSF. Addresses in these simplified ZIP codes are excluded from the DSF, but can sometimes be captured by supplemental frames (Iannacchione, 2007).

Non-city style addresses pose special problems for face-to-face surveys. Non-city style addresses are addresses that lack a house number, street name, or both. For example, rural routes and post office box addresses are considered non-city style addresses. Dozens of survey methodologists have noted that non-city style addresses cannot be efficiently located for face-to-face surveys. Thus, for face-to-face surveys non-city style addresses are usually included in undercoverage estimates.

Even if the DSF contains an address, there may be a variety of reasons why the address may be erroneously excluded from the sampling frame. If the data vendor subsets the DSF to residential units, then addresses that may be classified as non-residential may not be included in the frame. For example, group quarters are often classified as non-residential on the DSF and may be erroneously excluded (Dohrmann, 2006). Also, households receiving mail at a business address may also be excluded from the frame (Dohrmann, 2007). If only certain blocks or tracts are purchased, ungeocoded and incorrectly geocoded addresses may be excluded from the sampling frame (Dohrmann, 2007). If Exclude from Delivery Statistics (EDS) addresses are also excluded, then some of the new construction will also be erroneously excluded (Martin, 2009).

2.1 The Master Address File

The Master Address File (MAF) is an address file maintained by the Census Bureau. It contains almost all city style addresses on the DSF. It is further enhanced every 10 years with decennial Census addresses. Several small scale operations also update the MAF on an ongoing basis, including the Demographic Area Address Listing (DAAL) operation.

In 2009, the address canvassing operation listed nearly all blocks in the US in preparation of the 2010 decennial census. The results of the address canvassing operation updated the MAF. In addition to updating addresses, the address canvassing operation also captured GPS coordinates for over 127 million housing units. After the address canvassing operation, only 100,000 of the 135 million addresses on the MAF lacked both a complete address and Global Positioning System (GPS) coordinates. Thus, over 99% of the addresses on the MAF should be locatable by either a complete address or a GPS coordinate.

2.1 Coverage Improvement

Since the DSF may be omitting some units and the addresses for some housing units may be erroneously excluded during frame creation, most data vendors seek to compensate for these deficiencies by adding records from other sources. In urban areas, the half-open interval or some variation of it is often used to extract additional units at the time of interview. However, this technique is prone to quite a few errors (Eckman, 2011). Listing is often done in areas with rural addresses to get precise coordinates of the housing units.

For surveys opting to do some coverage improvement, a key research question is how best to partition the frame into areas needing improvement and area that do not need improvement. In the past, the current household demographic surveys conducted by the Census Bureau focused on improving coverage in blocks where 5% or more of the addresses were non-city style and also blocks that were not covered by a building permit office. The Type of Enumeration (TEA) area, percent of non-city style addresses, and the urbanicity of the area have been used in the private sector to partition segments into those needing listing from those not.

Montaquila, Hsu, and Brick (2011) used a model to predict the percent of housing units on the ground that could not be matched to a vendor file at the block group level. Their covariates included measures of the age of housing units, the mobility of individuals, the urbanicity of the block, the percent of seasonal housing units, the percent of occupied units, the percent of single units, the percent of transit users, and the ratio of USPS addresses to Census counts.

The Census Bureau has been planning to move from a four framed approach to sampling a dynamic address file derived from the DSF for its Title 13 current household surveys (Liu, 2008). As a part of deciding to transition to the new frame, the Census Bureau investigated the potential coverage bias that would be incurred by transitioning to the new frame (Liu, 2009). Liu estimated coverage bias of key estimates at the national and state level resulting from using the DSF based frame. He found some evidence of bias for state level estimates, the smallest geographic area of estimation for the current household surveys. Using Liu's estimates of bias, along with coverage estimates and other state level characteristics, we decided that 14 states needed coverage improvement in order to mitigate the risk of state level coverage biases. The 14 states are: West Virginia, New Mexico, Maine, Alaska, Montana, Wyoming, Mississippi, Oklahoma, Arkansas, Hawaii, New Hampshire, Kentucky, Alabama, and Vermont. At the national level, we noted that estimates from the MAF-based frame did not differ from estimates from the current four frame approach. Thus surveys only making national estimates do not need coverage improvements.

Our goal in the 14 states was to further define a set of blocks which would capture most of the omissions. That is, we only wanted to list the blocks that had the most omissions. By listing the blocks with the most omissions rather than the blocks with the highest omission rates, we would better be able to minimize the state level omission rate. This is one key difference between this research and the research described by Montaquila, Hsu, and Brick (2011). They used the match rate to determine which blocks to enhance. Listing the block groups with the lowest match rates will not necessarily increase the state or national coverage rates because block groups vary quite a bit in size. To efficiently increase state level coverage rates, the block groups with the most omissions should be improved.

Determining the blocks with the most omissions can be a challenge. On the one hand, if the number of omissions are roughly uniformly distributed among the blocks, we would not be able to effectively target the blocks with the most omissions. On the other hand, if omissions are clustered in certain areas or correlated with known block characteristics then we might be able to effectively target the blocks with the most omissions.

In the next section, we discuss the methods and data we used to define a frame of blocks needing coverage improvement.

3. Methods

3.1 Methods

We used several techniques to partition blocks into those that would be eligible for coverage improvement from those blocks that would not be eligible for coverage improvement. Using results from the address canvassing operation, we investigated which blocks had the most omissions. The set of blocks with the most omissions formed the gold standard for targeting. We compared models to this gold standard, seeking to list the fewest blocks and housing units, while capturing the most omissions. We compared four different techniques to target blocks: one based on listing the largest blocks, one based on listing the blocks with the fewest matches to city style addresses on the DSF, one based on listing the blocks with the lowest match rates to city style addresses on the DSF, and one based on fitting a Poisson regression model.

In each targeting process, we sort the list of blocks in the 14 states by a variety of attributes. Then we divide the sorted list of blocks into 100 groups, each containing about 7,607 blocks. In the case where multiple blocks had the same attribute, we further sorted the blocks by the number of addresses on the MAF prior to address canvassing. For each percentile, we counted the total number of adds found during the address canvassing operation.

3.2 Data

In an effort to create a comprehensive address list to mailout Census 2010 forms, most of the blocks in the US were listed in the spring and summer of 2009. Using a handheld electronic device, field representatives captured the GPS coordinates for the front door of over 127 million of the 135 million addresses on the MAF. Field representatives also updated address information on the MAF for the housing units on the ground. When appropriate, they added new units that were missing from the MAF and deleted units that were on the MAF, but not on the ground. The results of the address canvassing operation forms the basis for the results and analysis that follow.

Some of the units added from the address canvassing operation may be vacant units. Others may have converted to nonresidential units or have been demolished in the time between the address canvassing operation and Census 2010. Although most blocks were canvassed, not all blocks were canvassed, especially some in remote Maine and Alaska. For these reasons and others, not all adds from the address canvassing operation turned out to be valid Census 2010 housing units. Furthermore, some units captured in Census 2010 were not included in the address canvassing operation. Thus, the address canvassing operation data does not completely represent MAF undercoverage. For these, and other, reasons, the estimated rate of units added to the MAF during address canvassing is expected to be an overestimate. Furthermore, because the address canvassing operation was subject to nonsampling errors, it does not correspond exactly to what one would expect to be on the ground.

We subset our analysis to the 14 states needing coverage improvement. In those 14 states, there were 817,723 housing units on the MAF before the address canvassing operation that were not geocoded. That is, we knew what county the address was in, but could not determine which census tract and tabulation block the address was in. These ungeocoded records were removed from all analysis. Since the address canvassing operation captured geocodes for all units, all omissions are geocoded. Furthermore, the ungeocoded records are not included in the set of omissions. During the address canvassing operation, a total of 1,302,023 omissions were found in those 14 states.

Overall, the 14 states contained 13,232,786 geocoded addresses in 1,295,164 blocks on the MAF before the address canvassing operation. Of those, 1,295,164 blocks, 534,483 were zero blocks, meaning they did not contain any valid housing units. The 534,483 zero blocks contained 60,138 omissions. The number of omissions found per block is extremely low for this category, indicating that listing zero blocks is an extremely operationally inefficient way to capture omissions. Since 41% of the blocks were zero blocks and only contained 4.6% of all omissions, they were removed from further analysis. Zero blocks include water blocks, blocks in national and local forests, and other blocks without any housing units.

After removing the ungeocoded records and zero blocks, the final universe contained 760,681 blocks in the 14 states. This set of blocks contained 1,241,885 adds found during address canvassing and 13,232,786 housing units prior to the address canvassing operation.

4. Results

4.1 Best Possible Targeting Analysis

In the 14 coverage improvement states, the address canvassing operation added 1,241,885 housing units. We call these MAF omissions, although they have not been thoroughly verified as good housing units. Overall, before the address canvassing operation, the MAF had 13,232,786 addresses in those 14 states. Using the address canvassing data, we counted the number of omissions found in each block. Then we sorted the 760,681 blocks in the 14 states by the total number of omissions in each block. Then we partitioned the blocks into 100 groups of about 7,607 blocks based on the total number of omissions in each block. Table 1 shows the results of this analysis. As we see, the 7,607 blocks with the most omissions contained 372,446 of the 1,241,885 omissions. All 1,241,885 omissions are found in just 35% of the blocks.

Table 1: Best Possible Targeting

Percentile	Blocks	Housing Units	Adds	Adds per Block	Adds per HU	HUs per Block
1	7,607	797,227	372,446	49	0.47	105
2	15,214	1,249,046	509,506	33	0.41	82
5	38,035	2,265,379	735,598	19	0.32	60
10	76,069	3,376,956	930,457	12	0.28	44
20	152,137	4,983,432	1,120,634	7	0.22	33
30	228,205	6,622,706	1,203,175	5	0.18	29
50	380,341	11,132,497	1,241,885	3	0.11	29
100	760,681	13,232,786	1,241,885	2	0.09	17

If we were able to perfectly target the blocks with the most omissions, we would be able to capture all of the omissions in the 14 states by only listing 35% of the blocks.

Table 1 shows the number of HUs per block prior to listing. It clearly shows that the blocks with the most omissions tend to have more housing units. This strong correlation between block size and omissions led us to explore more carefully the relationship between block size and omissions.

4.2 Block Size Analysis

Given the relationship between block size and number of omissions in Table 1, we produced a similar table by sorting blocks based on the number of housing units on the MAF prior to address canvassing. The first row of Table 2 shows the 7,607 largest blocks and the number of omissions in those blocks.

Table 2: Distribution by block size

Percentile	Blocks	Housing Units	Adds	Adds per Block	Adds per HU	HUs per Block
1	7,607	2,101,568	128,549	17	0.06	276
2	15,214	3,062,021	205,637	14	0.07	201
5	38,035	4,824,761	369,700	10	0.08	127
10	76,069	6,592,951	531,648	7	0.08	87
20	152,137	8,762,570	714,688	5	0.08	58
30	228,205	10,176,422	839,869	4	0.08	45
50	380,341	11,909,489	1,010,165	3	0.08	31
100	760,681	13,232,786	1,241,885	2	0.09	17

As we see, larger blocks tend to have more omissions. Targeting based on block size makes some sense. As we see, 839,869 of the 1,241,885 or nearly 68% of the omissions are in the largest 30% of blocks. Of course, nearly 77% of all housing units are in those blocks. Thus, even though most of the omissions are concentrated in a small number of blocks, most housing units are also concentrated in those blocks. If the cost of listing were determined by the number of blocks listed, there could be great reductions in cost by targeting the largest blocks. On the other hand, if the cost of listing is more determined by the total number of housing units in the block, then there are few gains to be made by targeting the larger blocks.

In comparing Table 1 to Table 2 we see that targeting blocks based on block size is not as effective at capturing omissions as the optimal targeting. Certainly there is some room for improvement beyond targeting the largest blocks. Nevertheless, it should be clear from both tables that larger blocks tend to have more omissions than smaller blocks.

If all housing units have the same fixed probability of being omitted from the MAF, then we would expect the number of omissions per HU to be relatively constant in each percentile. Indeed, we see rough confirmation of this hypothesis in Table 2. The omission rate is fairly stable, hovering between 7 and 9 percent. Given this stability, it may be a challenge to target blocks. If listers made relatively random errors in constructing the MAF and new construction is randomly missed as well, we would expect a constant omission rate per housing unit. Given difficulties with finding well fit models to target omissions and the relatively constant number of omissions per housing unit in Table 2, we find much support for the hypothesis that all housing unit have the same probability of being omitted.

4.3 Poisson Regression Analysis

We fit a large Poisson regression model to estimate the total number of omissions in each block using the number of housing units in the block prior to the address canvassing operation as an offset. We used 97 covariates in our model. We used tract level data from the 2010 tract level planning database, block characteristics from ACS 5 year data, other block level characteristics related to the housing unit growth over time, the style of mail delivery in the ZIP code, the state, whether the block was covered by a building permit office, and an urban/rural indicator among other block level counts. We then used that model to predict the number of omissions in the every block, including the blocks that were not selected for sample. Sorting the blocks by the predicted number of omissions and ranking them allowed us to make tables similar to Table 1 and Table 2.

Table 3: Modeled Distribution of Adds

Percentile	Blocks	Housing Units	Adds	Adds per Block	Adds per HU	HUs per Block
1	7,607	884,485	222,089	29	0.25	116
2	15,214	1,515,530	317,489	21	0.21	100
5	38,035	2,972,292	487,234	13	0.16	78
10	76,069	4,606,851	664,175	8	0.14	61
20	152,137	6,642,050	816,051	5	0.12	44
30	228,205	8,002,700	909,224	4	0.11	35
50	380,341	9,990,669	1,005,982	3	0.10	26
100	760,681	13,232,786	1,241,885	2	0.09	17

Table 3 shows the distribution of omissions based on sorting the blocks by the expected number of omissions. The model is rather complicated, but seems to perform better than simply targeting the largest blocks. However, there is much room for improvement.

4.4 DSF Non Match Rate

The DSF is the primary source of new city style addresses on the MAF. Areas with few city style addresses and areas with many omissions from the DSF are likely to have higher undercoverage rates on the MAF, because new non-city style addresses are not entering the MAF on a regular basis. To explore if our hypothesis that MAF undercoverage would increase as the rate of DSF matches decreased, we calculated the percent of valid housing units on the MAF that matched to city-style DSF addresses.

Table 4: Distribution of Expected Adds When Sorting By DSF Non Match Rate

Percentile	Blocks	Housing Units	Adds	Adds per Block	Adds per HU	HUs per Block
1	7,607	349,832	126,409	17	0.36	46
2	15,214	487,612	170,285	11	0.35	32
5	38,035	712,427	240,197	6	0.34	19
10	76,069	892,094	296,497	4	0.33	12
20	152,137	1,018,436	335,624	2	0.33	7
30	228,205	2,192,833	638,572	3	0.29	10
50	380,341	4,897,621	962,850	3	0.20	13
100	760,681	13,232,786	1,241,885	2	0.09	17

Table 4 shows our results based on targeting the blocks with the highest DSF non match rates. The most efficient blocks to list are those with the most omissions per block, which are the blocks with DSF match rates between 0 and 50 percent. These blocks also have among the greatest adds per HU rates supporting the theory that listing these blocks is also efficient in terms of the cost of listing housing units. 167,988 of the 760,681 blocks didn't have any matches to the DSF. These blocks contained about 27% of all omissions and 8% of all housing units.

The coverage improvement universe defined by percent of DSF matches has considerable appeal because it is simple and has a theoretical justification. Montaquila (2011) describes a similar method that appears to work for improving the coverage of a different DSF-based frame. Though, there may be better methods to target coverage improvement blocks.

4.5 DSF Non Match Count

The primary goal of coverage improvement is to mitigate state level coverage biases. One way to reduce the coverage bias is to reduce the state level undercoverage rate through coverage improvements. The most efficient way to reduce the state level undercoverage rates is to target the blocks with the most omissions. With this in mind, we explored the impact of targeting the blocks with the most MAF addresses that didn't match to the DSF prior to address canvassing.

Table 5: Distribution of Expected Adds When Sorting By DSF Non Matches

Percentile	Blocks	Housing Units	Adds	Adds per Block	Adds per HU	HUs per Block
1	7,607	989,898	225,495	30	0.23	130
2	15,214	1,502,686	325,406	21	0.22	99
5	38,035	2,577,472	502,711	13	0.20	68
10	76,069	3,792,176	670,451	9	0.18	50
20	152,137	5,495,672	856,442	6	0.16	36
30	228,205	6,817,930	968,709	4	0.14	30
50	380,341	9,149,520	1,101,293	3	0.12	24
100	760,681	13,232,786	1,241,885	2	0.09	17

Table 5 shows the results of this analysis. Like other methods, this method also tends to target the blocks with more housing units, but it is also quite efficient at capturing adds.

4.6 Summary

The previous tables pooled all of the blocks and omissions together across all 14 states. In order to reduce state level biases, we would like to independently target blocks within each state. Effective targeting should be done at the state level, since our goal is to mitigate the risk of state-level coverage bias.

To compare the different targeting processes, we first arbitrarily set a threshold of needing to reduce the omission rate to less than 4% for each state. We then looked at how many blocks and housing units would be screened into a coverage improvement universe in order to capture enough omissions to reduce the omission rate to 4%.

Table 6: Total Number of Blocks Needing Improvement to Reduce the Omission Rate to 4%

State	Blocks					
	Total	Optimal	Block Size	Poisson	DSF %	DSF Count
Alabama	175,220	1,995	9,851	8,504	29,410	8,560
Alaska	21,874	938	3,974	2,861	7,154	2,525
Arkansas	141,178	1,745	8,161	5,257	22,186	5,575
Hawaii	18,990	445	2,139	1,233	3,105	1,208
Kentucky	122,141	1,807	7,557	4,945	17,481	4,127
Maine	56,893	3,730	10,696	8,472	27,471	7,094
Mississippi	136,150	2,851	12,143	9,855	16,025	10,403
Montana	99,018	1,921	11,134	5,115	17,632	5,878
New Hampshire	34,728	579	2,773	1,661	4,672	1,276
New Mexico	137,055	3,373	16,976	11,524	21,163	8,158
Oklahoma	176,064	3,142	21,450	8,009	38,611	7,865
Vermont	24,824	1,934	4,570	3,608	6,599	3,172
West Virginia	81,788	5,887	15,345	17,232	25,852	9,844
Wyoming	67,264	931	7,047	3,273	7,873	3,043

Table 6 shows the total number of blocks in each state and the total number of blocks that would be needed under each of the targeting methods in order to reduce the state level omission rate to 4%. For example, West Virginia has 81,788 blocks with housing units. If we were able to know exactly what blocks had the most omissions, we could reduce the omission rate to 4%, by just listing 5,887 blocks. However, given that we do not know which blocks have the most omissions, the next best we can do is to list the 9,844 blocks with the most DSF non matchs.

In general, listing the blocks with the most DSF nonmatches will get us enough omissions to reduce the state level omission rates to 4% by listing the fewest blocks. If listing costs were completely determined by the total number of blocks, the DSF count method would be the best of the four methods presented.

Of course, the cost of listing is also determined by the total number of housing units in the listed blocks, not just the total number of blocks. Table 7 shows the total number of housing units that would need to be listed to reduce the state omission rate to 4%.

Table 7: Total Number of Housing Units Needing Improvement to Reduce the Omission Rate to 4%

State	Housing Units					
	Total	Optimal	Block Size	Poisson	DSF %	DSF Count
Alabama	2,340,284	198,753	986,557	590,546	354,149	501,775
Alaska	330,852	55,916	225,873	121,229	96,723	121,455
Arkansas	1,415,610	110,562	623,408	326,969	175,399	283,564
Hawaii	544,256	70,873	372,078	129,788	59,969	174,305
Kentucky	2,094,673	185,114	926,630	414,472	237,227	339,172
Maine	794,307	187,564	521,945	344,695	279,721	329,522
Mississippi	1,415,969	163,312	713,770	473,892	201,152	393,382
Montana	510,119	68,844	334,230	162,940	105,620	145,762
New Hampshire	637,624	53,107	289,256	128,746	65,084	108,141
New Mexico	955,706	162,593	673,629	400,229	165,010	271,364
Oklahoma	1,749,847	150,700	1,017,016	349,443	216,941	303,837
Vermont	365,507	102,162	237,316	168,530	106,000	156,402
West Virginia	1,049,852	308,491	679,896	537,898	338,726	447,681
Wyoming	270,065	29,784	186,385	87,565	49,483	66,639

The DSF count method results in the fewest number of blocks requiring to be listed; however, the DSF count method also tends to target blocks that are larger than the DSF non match rate procedure. In terms of the total number of housing units needing to be listed, we see that the targeting the blocks with the greatest DSF non match rate tends to yield the most omissions, in general.

Tables 6 and 7 confirm that the MAF is highly reliant on the DSF for address updates. In areas where the DSF is deficient or where there is a high degree of non-city style addresses, the MAF also tends to be deficient. Targeting blocks with high DSF non match rates or with a great number of DSF non matches are two relatively efficient methods to determine where to list. For survey research centers without access to the MAF, housing unit counts based on population projections, census 2010 results, the American Community Survey, or other sources may be compared to the DSF counts instead of comparing the count of MAF addresses to the DSF.

5. Limitations

The data used in this study were computed using data from the address canvassing operation. Although not subject to sampling error, is it subject to nonsampling errors. Certainly listers can make errors, which may bias some of the estimates. Since the final Census 2010 status of housing units has undergone much more scrutiny, a similar analysis using Census 2000 data would have fewer nonsampling errors.

To be an effective means for targeting, the undercoverage trends in the past 10 years need to be consistent with the future trends. The add rates presented in this document represent about eight years of new growth that was not capture by the DSF. Add rates in the few years following the 2010 census should be much lower than those presented in this paper.

6. Conclusion

All housing unit undercoverage on the MAF in the 14 states with undercoverage problems is clustered in only 35% of all blocks. These blocks tend to be the largest blocks in terms of the total number of housing units in the block, suggesting that all housing units may have a constant probability of being omitted. However, a closer analysis reveals that the ability of the MAF to capture new addresses from the DSF at the block level is a major correlate of MAF undercoverage. Advanced models can be found to improve targeting. Targeting blocks with the highest DSF non match rate or with the most DSF non matches are both effective at reducing the state level omission rates. However, no method considered so far comes close to the optimal targeting method. More research is needed to foster a theory for undercoverage on enhanced address based frames.

Acknowledgements

The authors would like to thank Alan Peterson for reviewing this paper.

References

Dohrmann, Sylvia. Diafeng Han, and Leyla Mohajer. 2006. "Residential Address Lists vs. Traditional Listing: Enumerating Households and Group Quarters." In *JSM Proceedings, Survey Research Methods Section*, 2959-2964. Alexandria, VA: American Statistical Association.

Dohrmann, Sylvia. Diafeng Han, and Leyla Mohajer. 2007. "Improving Coverage of Residential Address Lists in Multistage Area Samples." In *JSM Proceedings, Survey Research Methods Section*, 3219-3226. Alexandria, VA: American Statistical Association.

Eckman, Stephanie. Colm O'Muircheartaigh. 2011 "Performance of the Half-Open Interval Missed Housing Unit Procedure." Unpublished Manuscript.

Iannacchione, Vincent, Katherine Morton, Joseph McMichael, David Cunningham, James Cajka, and James Chromy. 2007. "Comparing the Coverage of a Household Sampling Frame Based on Mailing Addresses to a Frame Based on Field Enumeration." In *JSM Proceedings, Survey Research Methods Section*, 3323-3332. Alexandria, VA: American Statistical Association.

Iannacchione, Vincent G., Jenifer M. Staab, and David T. Redden. 2003. "Evaluating the Use of Residential Mailing Addresses in a Metropolitan Household Survey." *Public Opinion Quarterly* 67: 202-10.

Liu, Xijian. 2008. "Using A MAF-Based Frame for Demographic Household Surveys." In *JSM Proceedings, Survey Research Methods Section*, 2864-2871. Alexandria, VA: American Statistical Association.

Liu, Xijian. 2009. "Impact of MAF-Based Frame Coverage on Survey Estimates." In *JSM Proceedings, Survey Research Methods Section*, 1213-1222. Alexandria, VA: American Statistical Association.

Martin, Joel M. Clifford L. Loudermilk. 2009. "Assessing the Filter Rules for Extracting Addresses From the Master Address File to Construct a Housing Unit Frame for Current Demographic Surveys." In *JSM Proceedings, Survey Research Methods Section*, 1359-1366. Alexandria, VA: American Statistical Association.

Montaquila, Jill M., Valerie Hsu, J. Michael Brick, Ned English, and Colm O'Muircheartaigh. 2009. "A Comparative Evaluation of Traditional Listing vs. Address-Based Sampling Frames: Matching with Field Investigation of Discrepancies." In *JSM Proceedings, Survey Research Methods Section*, 4855-62. Alexandria, VA: American Statistical Association.

Montaquila, Jill M., Valerie Hsu and J. Michael Brick. 2011. "Using A 'Match Rate' Model to Predict Areas Where USPS-Based Address Lists May Be Used In Place Of Traditional Listing." *Public Opinion Quarterly*, 75: 317-335.