FRAME IMPROVEMENT FOR THE DEMOGRAPHIC SURVEYS SAMPLE REDESIGN

Xijian Liu and Timothy Kennel[1]
U.S. Census Bureau
4600 Silver Hill Road, Washington, DC 20233

Abstract
This paper gives a summary and interpretations of the results of a series of research conducted by the U.S. Census Bureau to support the development of a frame improvement system for demographic household surveys in the 2010 Demographic Surveys Sample Redesign. Coverage evaluations concluded that a household survey frame developed using the Master Address File as the sole source can provide a frame for producing comparable quality U.S. estimates and most state estimates. Coverage improvement will be needed to mitigate the risk of coverage bias in several states. The U.S. Census Bureau conducted subsequent research to identify the states that needing coverage improvements and the characteristics of the blocks that coverage improvements will be targeting. This paper will also describe the coverage improvement frame methodology for the 2010 Sample Redesign.

Keywords: Coverage Improvement, Master Address File, Address Base Sampling

1. Background

The U. S. Census Bureau uses a common frame system to cover housing units and certain group quarters in the U.S. to support several major demographic household surveys. These surveys include the Current Population Survey (CPS), the Survey of Income and Program Participation (SIPP), the Consumer Expenditure Survey (CE), the American Housing Survey (AHS), the National Crime and Victimization Survey (NCVS), and the State Child Health Insurance Program (SCHIP).

The current frame system consists a unit frame, a group quarters (GQ) frame, a permit frame, and an area frame. In most areas of the U.S., housing units and group quarters existed at the time of Census 2000 are represented in the unit frame and the GQ frame; new housing units added after Census 2000 are represented in the permit frame. The sources of addresses for these three frames are the address lists from Census 2000 and building permits. In areas that do not have a high percent of city-style addresses or no building permits coverage, we use the area frame. This final frame is developed using ongoing field listings. The current area frame covers approximately 12% of housing units in the U.S. The field listing operations for the area frame and the permit frame are the most expensive of the four described. The permit frame sampling system is very complex.

In the next Demographic Surveys Sample Redesign (DSSR) after 2010, the U.S. Census Bureau will use the continually updated Master Address File (MAF) as the primary source of addresses for developing the Title-13 household survey frames. The MAF is

---

U.S. Census Bureau's inventory of addresses for all known living quarters in the U. S. Decennial census operations provided major updates for the MAF. Every six months, the U.S. Census Bureau uses the Delivery Sequence File (DSF) from the United States Postal Service to update the MAF. There are various smaller scale operations that update the MAF as well. The American Community Survey (ACS) uses the MAF as the source to construct its sample frame. The most obvious change from the four-frame system to the frame system used in the 2010 DSSR is that the DSF will replace the building permits as the primary source of post-census new construction. It is also desirable to eliminate or reduce the costly area listing.

This paper will give a summary of the research conducted by the U.S. Census Bureau to support the development of the frame improvement system and will briefly describe the coverage improvement frame used in the 2010 DSSR.

In the remaining sections of this paper, we will not discuss group quarter. The results will only focus on housing units.

2. Why Is Coverage Improvement Needed for the MAF-Based Frame?

From 2003 to 2008, the U.S. Census Bureau conducted a series of evaluations to compare the coverage quality of a MAF-based frame and the current four frames. A major assumption used by these evaluations was that the decennial censuses and the DSF were the main sources of updates to MAF. In other words, the MAF-based frame for these evaluations excluded the improvements from listing operations of the current survey area frame and the Community Address Update System, a coverage improvement operation for the ACS.

The readers can find more detailed results from these evaluations in Kennel and Corlett (2005), Flanagan and Loudermilk (2006), Loudermilk and Li (2009), Liu (2008), and Liu (2009). This section will provide a brief summary of these results.

Table 2.1 is from Liu (2008). It provides a summary of the coverage comparison of the two frame systems at the U.S. level using a time reference of 2007. It shows the coverage differences of the two frame system at the U.S. level were small.

Table 2.1. Coverage Comparison of the MAF-Based Frame and the Current Four Frames (2007)

	On MAF-based Frame, Not on Current Four Frame	On Current Four Frame Not on MAF-based Frame
Unit frame sub-universe	2.5%	0
Permit frame sub-universe	≥0.5%	1.1%
Manufactured homes	0.2-0.3%	0
Area Frame sub-universe	0	1.6%

The 1.6% on the bottom row shows the contribution of the current area frame listing to the entire U.S. housing unit frame.

Liu (2009) further investigated how the coverage differences impact CPS Civilian Labor Force Participation Rate and Unemployment Rate estimates for the U.S. and for states. The results showed some concerns of non-trivial coverage bias only for a few states that have a large proportion of housing units in rural areas. In Arkansas, Mississippi,

Montana, New Mexico, and West Virginia, 4-5% of the CPS samples in Nov 2005, March 2006, and July 2006 were added from area frame listing. These states had the highest risk of coverage bias due to rural MAF undercoverage. For example, excluding the units added from area frame listing can cause nearly nine percent relative difference to the West Virginia unemployment rates.

In recent years, more addresses in the U.S. are converted to city-style addresses. Census 2010 Address Canvassing collected GPS coordinates for most addresses in the U.S. Kennel and Martin (2010) found that there were just under 2% of the valid addresses on the MAF were incomplete in 2010, down from about 5% in 2000 and that almost all MAF addresses in 2010 were locatable, i.e., were complete addresses, had GPS coordinates, or were geocoded addresses with location descriptions. Hence, collecting information to help locating sample units is no longer a major benefit of coverage improvement listing operations.

In summary, we concluded that coverage improvement to the MAF-based frame will be needed for surveys that produce estimates in some states to mitigate the risk of coverage bias to state estimates due to MAF undercoverage in rural area.

3. Where Is Coverage Improvement Needed?

The next question is which states will need coverage improvement and which blocks in those states the coverage improvement listing should target.

It is unrealistic to find a scientific rule that can be used to determine which states will need improvement and which states will not. We use a conservative approach to determine the states needing coverage improvement is conservative.

As mentioned earlier, Liu (2009) had shown coverage concerns in West Virginia, New Mexico, Montana, Mississippi, and Arkansas. These five states will be included on the list of states needing coverage improvement.

Kennel and Martin (2010) recommended additional states needing coverage improvements. They investigated the units added from Census 2010 Address canvassing operation. Between Census 2000 and Census 2010, the main source of update to the MAF was the DSF. Assume the 2010 Census Address Canvassing operation can find the housing units missed by DSF updates and add them back to the MAF. The percent of Census Address Canvassing adds can serve as an indicator of the quality of the DSF updates. Coverage improvements will be needed in areas where the DSF updates have weakness. Table 3.1 is from Kennel and Martin (2010) that shows the proxy omission rates, (or the Address Canvassing add rates). The proxy omission rate is used to help determine the priority order of states for improvement needs. It is not an accurate measure of MAF undercoverage, e.g., a nineteen percent undercoverage rate for West Virginia is apparently too high.

Based on the results from Kennel and Martin (2010), we include nine additional states: Alabama, Alaska, Hawaii, Kentucky, Maine, New Hampshire, Oklahoma, Vermont, and Wyoming, as needing coverage improvement.

Table 3.1. Omission Rate and Percent of Omissions in the Improvement Universe Blocks

State	Omission Rate
West Virginia	19.0%
Alaska	14.8%
Vermont	14.2%
New Mexico	12.4%
Maine	12.3%
Wyoming	10.9%
Montana	10.7%
Hawaii	9.3%
Mississippi	7.9%
Oklahoma	7.7%
Kentucky	7.6%
Alabama	7.4%
New Hampshire	7.4%
Arkansas	7.3%

Source: Kennel and Martin (2010)

Next, we turn to the characteristics of blocks that the coverage improvement listing will be targeting. We will call these blocks Coverage Improvement Blocks. Kennel and Martin (2010) compared several alternatives and recommended the use of blocks that had less than 75% of their addresses matched to the DSF as Coverage Improvement Blocks in the 14 states selected. Table 3.2 is from Kennel and Martin (2010) and shows the Coverage Improvement Blocks are quite effective in capturing units missed by the DSF. Using West Virginia as an example, the Coverage Improvement Blocks contain about half of the housing units and ninety-three percent of the proxy omissions in this state.

Table 3.2. Effectiveness of the Coverage Improvement Blocks

State	% of Omissions in CI Blocks	% of HUs In CI Blocks
West Virginia	93	51
Alaska	93	48
Vermont	87	47
New Mexico	88	33
Maine	84	41
Wyoming	89	33
Montana	87	35
Hawaii	80	25
Mississippi	68	23
Oklahoma	85	26
Kentucky	77	23
Alabama	66	20
New Hampshire	82	28
Arkansas	73	26

Source: Kennel and Martin (2010)

Finally, the Coverage Improvement Blocks is in 14 states and contains only about 3% of the housing units in the U.S. The current area frame is in all 50 states and the District of Columbia and contains about 12% of the housing units in the U.S.

Table 3.3. Size of the Improvement Universe in Selected States

State	% of HUs In CI Blocks	% of HUs In 2000 Area Frame
West Virginia	51	55
Alaska	48	43
Vermont	47	53
New Mexico	33	35
Maine	41	46
Wyoming	33	34
Montana	35	47
Hawaii	25	16
Mississippi	23	40
Oklahoma	26	31
Kentucky	23	46
Alabama	20	37
New Hampshire	28	22
Arkansas	26	46
U.S.	3	12

Source: Kennel and Martin (2010)

4. The 2010 Coverage Improvement Frame

In the 2000 Sample Design, the entire universe is divided into two sub-universes: the unit frame blocks and the area frame blocks. In the unit frame blocks, housing unit samples are selected from the Census 2000 address list and building permits. In the area frame, sample of blocks are sent out for listing. Then housing units samples are selected from the address list in the blocks listed. For the 2010 Redesign, one option is to use this traditional approach and divide the housing unit universe into two sub-universes according whether a block is a Coverage Improvement Block. The unit frame contains housing units not in Coverage Improvement Blocks. The coverage improvement frame will contain all housing units in Coverage Improvement Blocks. Housing unit samples are selected annually from the MAF extracts for this unit frame sub-universe. This is the blue color portion of the following illustration chart. From the coverage improvement frame, survey samples can be selected using a method similar to the 2000 area frame. This is the gray color portion of the illustration chart below. We determine the sample of blocks that need to be listed, list the blocks, and select hosing unit samples from the address lists after the blocks are listed. Because the listing is a dependent listing starting from the address list existed on the MAF, the address list coming back from listing has addresses existed on the MAF and MAF omissions, i.e., addresses previously not on the MAF but were added during the listing. In other words, coverage improvement frame of this traditional approach includes MAF addresses and MAF omissions. The unit frame and the coverage improvement frame are approximately 97% and 3% of the entire universe for this approach.

Illustration for the Traditional Method:

In Non-Improvement Universe Block On MAF	
In Improvement Universe Block On MAF	In Improvement Universe Block MAF Omission

The second approach divides the housing unit universe into two sub-universes according to whether the addresses are on the MAF. The coverage improvement frame contains only the MAF omissions. This is the gray color portion of the illustration chart below. We will select a sample from Coverage Improvement Blocks and list them. After listing, we will identify the MAF omissions from the listing results and select coverage improvement samples from the MAF omissions. The coverage improvement samples will be a small portion of the entire U.S. samples. Assuming the coverage improvement universe blocks contain an average of 10% omissions before 2020, the coverage improvement frame will contain (or represent) approximately 0.3% of the entire housing unit frame in the U.S. The housing units on the MAF are on the MAF frame, which is the blue color portion of the illustration chart. The main housing unit samples (or MAF samples) will be selected from the MAF extracts annually after removing the omissions.

Illustration for the New Method.

In Non-Improvement Universe Block On MAF	
In Improvement Universe Block On MAF	In Improvement Universe Block MAF Omission

There are many advantages of using the new method for coverage improvement frame. If a survey chooses not to use coverage improvements it will only have one sampling system for selecting MAF samples. Only surveys that use coverage improvement will have a second sampling system for selecting coverage improvement samples. However, the traditional method will require all surveys to have two sampling systems, one for the non-Coverage Improvement Blocks and the other for the Coverage Improvement Blocks, regardless of whether the survey use coverage improvements. The new method also allows for more flexible methods of selecting samples from omissions. It also has the flexibility of turning off the coverage improvement frame at any time, e.g., after Census 2020 if we determine that the MAF coverage will be good enough and the coverage improvement listing will not provide much benefit. There will also be operational simplicity with the new method.

The new method has some disadvantages too. It will result in less clustering sample units. So it will be slightly less efficient in this aspect. It will have more sample units that are isolated. The listing identifies omissions, but it also improves the quality of the MAF addresses. In the new method, the MAF addresses in the Coverage Improvement Blocks may or may not contain into the MAF samples. Therefore, the surveys may not reap the full benefits of the listing.

We have made a decision to use the new method because of the many advantages.

6. Summary

The research conducted by the Census Bureau showed that when using the MAF-based frame in the 2010 Demographic Surveys Sample Redesign frame coverage improvement is needed to mitigate the risk of coverage bias to certain state estimates due to MAF undercoverage in rural areas. The significance of this finding is that it allows the

coverage improvement listing to focus on only fourteen states, resulting in approximately 75% reduction in area listing from last sample design.

The Census Bureau is currently developing the detailed methods for a more flexible sampling system to select coverage improvement samples.

7. References

Flanagan, T. and Loudermilk, C., How quickly do permit addresses appear on the Delivery Sequence Files? Proceedings of the Join Statistical Meeting, 2006

Kennel, T. and Corlett, M., Comparing the US Census Bureau's Master Address File to the Survey of Income and Program Participation area listings. Proceedings of the Join Statistical Meeting, 2005

Kennel, T. and Liu, X., Creating a Coverage Improvement Universe for an Address Based Sampling Frame, American Association for Public Opinion Research 66th Annual Conference, 2011

Kennel, T. and Martin, J, Summary Report for Evaluating MAF Content Quality After the 2010 Decennial Address Canvassing, Internal Census Bureau Document, 2010

Kennel, T., Is Housing Unit Coverage Random? Proceedings of the Join Statistical Meeting, 2011

Liu, X., Using a MAF-Based Frame for Demographic Household Surveys, Proceedings of the Join Statistical Meeting, 2008

Liu, X., Impact of MAF-Based Frame Coverage on Survey Estimates, Proceedings of the Join Statistical Meeting, 2009

Loudermilk, C. and Li, M., A National Evaluation of Coverage for a Sampling Frame Based on the Master Address File (MAF), Proceedings of the Join Statistical Meeting, 2009

U. S. Census Bureau, Design and Methodology, Current Population Survey Technical Paper 66, 2006