A National Evaluation of Coverage for a Sampling Frame Based on the Master Address File (MAF)

Clifford L. Loudermilk ***** , Mei Li
U.S. Census Bureau, Washington, DC 20233

Abstract

A primary goal of the 2010 Demographic Surveys Sample Redesign is to consider switching to using a housing unit sampling frame based on the Master Address File (MAF) for current demographic household surveys if the frame provides an acceptable level of coverage. To measure coverage for a MAF-based frame, we selected a nationally representative sample of census blocks and sent them for field listing. By comparing the addresses collected during listing to the MAF addresses, we produce estimates of coverage at the national and regional levels. We also analyze selected subclasses, such as addresses in rural blocks, to identify strengths and weaknesses in the coverage provided by a MAF-based sampling frame.

Keywords: Master Address File; coverage; sampling frame

1. Introduction

For the past few decades, the Census Bureau has used a multiple-frame sampling approach to select sample for the current demographic household surveys[1] following each Decennial Census. The multiple frames include a unit frame (addresses from the census), an area frame (addresses collected via field listing operations), a permit frame (new construction addresses collected from building permit offices), and a group quarters (GQ) frame. Given the considerable listing costs and operational complexities of the area and permit frames, an alternative frame approach is being considered for the 2010 Demographic Surveys Sample Redesign.

An obvious alternative already exists within the Census Bureau: the Master Address File (MAF)[2]. The MAF, created and maintained by the Census Bureau, is a national inventory of addresses for living quarters in the United States. It contains the addresses from the most recent decennial census and is continually updated with addresses from the Delivery Sequence File (DSF) from the United States Postal Service (USPS) and other post-census field operations. The American Community Survey (ACS) already uses the MAF as the sole source for its sampling frame.

So, a major research question for the 2010 Sample Redesign is:

---

[1] *The current demographic household surveys: American Housing Survey, Consumer Expenditure Survey, Current Population Survey, National Crime Victimization Survey, National Health Interview Survey (NHIS uses only an area and a permit frame), Survey of Income and Program Participation, and State Children's Health Insurance Program Survey.*
[2] *The term Master Address File (MAF), while still in use at the Census Bureau, has largely been superseded by the term MAF/TIGER Database (MTdb) to reflect the integration of the MAF with the TIGER database of geospatial features. We will use MAF in this paper to refer to the address list component of the MTdb.*

> ➢ Can one sampling frame based on the continually updated MAF replace the current multiple frames for the current surveys while maintaining or improving the present levels of coverage and content quality?

To answer this question, the Census Bureau designed a series of MAF-related evaluations, collectively known as the Frame Assessment for Current Household Surveys (FACHS) program. The National Evaluation (FACHS-NE) is the first of these FACHS studies to attempt to provide national estimates of coverage for a MAF-based sampling frame. This paper will present and discuss the major results of the National Evaluation.

2. Methodology

The major goal of the National Evaluation research was to derive gross estimates of undercoverage and overcoverage at the national level for a MAF-based housing unit (HU) sampling frame for current surveys. To accomplish this, we selected a nationally representative sample of census blocks, the National Evaluation Sample (NES), to be visited and listed by experienced Census field representatives (FRs). Once we had the "ground truth" for these sample blocks, we compared it to the HUs on the MAF to measure undercoverage (HUs found by the FRs but not included in the MAF-based frame) and overcoverage (HUs in the MAF-based frame determined to be invalid or nonexistent by the FRs). In this section, we will discuss the NES sample design in greater detail and will describe our methods for deriving our estimates from the NES data.

2.1 *NES Sample Design*

The NES actually consists of two independent national samples, a 2007 sample and a 2008 sample. The 2007 sample blocks were listed in the field in two waves of equal size, the first ending in January 2007 and the second in August 2007. The 2008 NES was selected independently of the 2007 NES and was one-half its size; listing for these blocks was completed by February 2008. The 2008 sample was intended as a supplemental sample to improve the reliability of the 2007-based estimates. When we use the term NES in the rest of this report, we refer to the full sample formed by combining the 2007 and 2008 national samples.

The universe from which the NES sample was selected was the set of combined blocks defined for the entire nation for the current surveys for the 2000 Redesign. Combined blocks consist of one or more census tabulation blocks joined together to give zero blocks (those containing no HUs) a chance of selection and to decrease variance by reducing the number of very small blocks.

The only combined blocks excluded from the NES universe were those previously listed for the current surveys 2000-based area frame sample and those in the 2008 Census Dress Rehearsal sites. These blocks were excluded to avoid interference with ongoing field operations. Another reason for excluding the area frame sample blocks was to avoid crediting the MAF with enhancements from a listing operation that may not exist in the post-2010 environment.

The NES consisted of a unit frame sample and an area frame sample. The unit frame sample was stratified by census region, combined block size, and the rate of "growth" (defined as the proportion of valid HUs in a block that came from post-census DSF adds).

The area frame sample was stratified by building permit-issuing status, combined block size, and the degree of address conversion activity in the block. The combined block characteristics were based upon data from the January 2006 MAF.

Samples were selected with probability proportional to size (i.e., number of HUs in combined blocks) in all but the small block strata. Samples in small block strata (combined blocks with 1-5 HUs) were selected with equal selection probability.

We selected 5,277 combined blocks for the NES, which translates to 8,560 tabulation blocks. There are sample blocks in 1,773 (56%) of the 3,141 counties in the nation. The blocks were listed in 2007 and 2008 by FRs trained in the Demographic Area Address Listing (DAAL) program. These DAAL FRs conducted these NES listings using the same listing instrument and procedures as for the current surveys area frame listings.

2.2 Deriving Coverage Estimates

Using the results of the NES listings, we derived estimates of gross undercoverage and overcoverage at the national and regional levels. Also, we produced gross undercoverage estimates for selected subclasses, such as new construction HUs, mobile homes, urban vs. rural HUs, HUs in unit frame blocks vs. area frame blocks, etc. In the discussion that follows, note that the denominators for our gross undercoverage and gross overcoverage estimates are different; the implication is that it is not possible to produce a net coverage estimate by combining the gross coverage measures.

Undercoverage

To derive gross undercoverage estimates for the nation and for each census region, we first produced measures for these two components of undercoverage.

➢ Omissions are those HUs that were added in an NES block during listing that were not on the MAF prior to listing.

➢ Erroneous exclusions are those HUs validated in an NES block during listing that were already on the MAF, but did not meet the filtering criteria for inclusion in a MAF-based sampling frame.

The estimate for each component is the percentage of all added/validated HUs in the sample blocks that were omissions or erroneous exclusions, respectively. The combination of these two measures produces the gross undercoverage rate. We did not consider an HU to be undercoverage if its block on the MAF was different (or missing) than the block assigned by the DAAL FR, as long as the MAF had it in the same county. We took this approach since the current surveys sampling is county-based, not block-based.

We feel it is important to produce separate estimates for the two undercoverage components, since they represent different problems. Omissions represent certain coverage loss – the units simply are not on the MAF. The erroneous exclusions, however, are on the MAF but classified in a way that causes them to be filtered out of the MAF-based sampling frame. Presumably, some portion of these filter exclusions could be reclaimed by improving the filter rules.

To be classified as an omission, an HU had to be added by the FR in a sample block and never matched to any address already on the MAF in that same county by the GEO address matching. Initially, though, we found that some addresses were being classified as omissions even though there was an exact address match in the same block on the MAF. To counter this, we did additional address matching within NES blocks to try to identify such unflagged duplicates. We also did a match of addresses within 3-digit ZIP codes for a limited set of counties. Our matching procedure involved a limited standardization of the addresses and identified only exact matches on the address fields.

We classified a field-validated address as an erroneous exclusion if it was on the MAF extracts before listing but invalid for ACS (that is, rejected by the ACS MAF filter). We considered the ACS MAF filter to be the best available proxy for a current surveys MAF filter when determining which HUs from the MAF extracts before listing would have been included in our MAF-based sampling frame.

Overcoverage

Overcoverage was defined as those HUs in the sample blocks on the MAF extracts before listing that would have been included in a MAF-based sampling frame for current surveys, but were determined by the FRs to be nonexistent, nonresidential, duplicates, GQs, or otherwise out-of-scope for an HU frame. Again, we used the ACS MAF filter to determine which MAF addresses would have been included in our frame. The estimate we produced is the percentage of all HUs from sample blocks on the MAF extracts before listing that would have been included in our frame (based on the ACS filter) that were determined by the FRs not to be valid HUs.

There are two important limitations to our overcoverage estimates. The first is that the DAAL FRs were listing specific blocks and could not canvass all addresses in the county. So FRs may have designated some HUs on their dependent lists as "nonexistent" when, in fact, the HU existed in some other part of the county. A second limitation is that we could not include ungeocoded MAF addresses in our analysis, since our design was block-based and we could not determine which of these addresses were in our sample blocks. Neither of these limitations applies to the undercoverage estimates.

3. Results

We will first present the major findings of the National Evaluation, which are the gross undercoverage and gross overcoverage estimates for the MAF-based frame at the national and regional levels. Then we will present and discuss undercoverage estimates for selected subnational geographic levels, as well as for various subclasses of HUs (e.g., new construction HUs, mobile homes) of interest. For those subclasses with undercoverage issues, we will discuss possible reasons for the undercoverage and, when possible, offer comparisons to the multi-frame approach now used by the current surveys.

3.1 Undercoverage at the National and Regional Level

For each undercoverage type, estimates are provided separately for omissions versus erroneous exclusions, since they are fundamentally different. The MAF omissions represent certain coverage loss -- the units are not on the MAF. The filter exclusions, however, are on the MAF but are classified in a way that causes them to be rejected by

the filter. Presumably, then, some portion of this coverage loss could be reclaimed by improving the filtering rules.

It is important to note that the coverage estimates presented in this report are specific to the time period in which our measurements were made. That is, we provide estimates for a housing unit (HU) sampling frame based on an updated MAF as it exists 6-7 years after Census 2000. Had we done our study immediately after the census, our coverage rates would likely have been considerably better than what we present here. Conversely, if we had done our field listings at a later time than we did, MAF-based frame coverage would probably look a little worse. The MAF constantly evolves, so the coverage provided by a MAF-based sampling frame also changes over time.

Table 1 shows total gross undercoverage estimates for the entire nation and for each census region[3].

Table 1:
Undercoverage Rates for a MAF-Based Sampling Frame:
U.S. and Census Regions

		% of HUs in Sample[4]	Total Undercoverage (SE)	Omissions (SE)	Erroneous Exclusions (SE)
U.S.			6.39 (0.28)	4.67 (0.27)	1.73 (0.11)
REGION	**Northeast**	18.4	6.06 (0.59)	4.03 (0.53)	2.03 (0.29)
	Midwest	22.2	4.71 (0.41)	3.41 (0.37)	1.30 (0.11)
	South	38.2	8.47 (0.56)	6.34 (0.54)	2.13 (0.23)
	West	21.2	4.69 (0.48)	3.51 (0.46)	1.18 (0.14)

We estimated a national undercoverage rate of 6.39%, with more of the undercoverage attributed to MAF omissions (HUs completely missing from the MAF) than to erroneous exclusions. Erroneous exclusions are those HUs validated by listing and physically on the MAF, but excluded from the MAF-based frame by the filtering rules. There are many reasons for excluding a MAF address in filtering, such as: the address was flagged as nonresidential; the address was deleted by the census and never appeared on the DSF; the address was on an earlier DSF version but later dropped off; etc. Research into the filtering rules may allow us to recover some of the valid HUs lost through filtering, but the current FACHS research does not suggest significant coverage gains are likely (Martin, 2008).

Consistent with expectations, the South region was found to have a significantly[5] greater undercoverage rate than any of the other regions. There are a number of possible reasons for this:

---

[3] *Census regions are geographic groupings of states for the presentation of decennial census data. The four Census Regions are the **Northeast** (CT, MA, NH, NJ, NY, PA, RI, VT, ME); **Midwest** (IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, WI); **South** (AL, AR, DE, DC, FL, GA, KY, LA, MD, MS, NC, OK, SC, TN, TX, VA, WV); and **West** (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA, WY).*

[4] *This column, which appears in tables throughout this report, represents the weighted percentage of HUs in our sample for each subclass.*

[5] *Any difference noted as "significant" or "statistically significant" in this report is the result of hypothesis testing at the 10% level of significance using the t distribution.*

- The South has had a higher HU growth rate since Census 2000 than any other region (see Table 2). New construction HUs typically present a greater coverage challenge than old construction HUs.
- The MAF is presumed to have weaker coverage in rural areas and in current area frame blocks; the South has a considerably greater share of its HUs in area frame blocks (versus unit frame blocks) than any other region (Table 2). We'll explore rural undercoverage on the MAF and possible reasons for it later in this paper.
- Mobile homes, a common coverage concern, make up a higher proportion of HUs in the South than in the other regions (Table 2). We will provide estimates for mobile home undercoverage in the MAF-based frame in a later section.

Table 2:
Growth Rates Since 2000 and
Area Frame and Mobile Home HU Percentages
by Census Region

		HU Growth Rate Since 2000[6]	% of Census HUs in Area Frame Blocks	Mobile Homes as % of All HUs (Census 2000)
U.S.		9.0%	12.3%	7.6%
REGION	**Northeast**	3.8%	6.9%	3.0%
	Midwest	7.2%	12.7%	5.4%
	South	11.7%	18.4%	11.6%
	West	10.9%	6.5%	7.1%

For a rough comparison of coverage in the MAF-based frame versus the current multi-frame approach, we obtained the HU coverage ratios from the Census Bureau's Housing and Vacancy Survey (June 2008) for each region. The HU coverage ratio is the ratio of the estimated number of HUs using the first stage weight to the independent Census Bureau HU estimates for June 2008. The estimates using the first stage weight (prior to applying population and housing unit controls) represent the number of HUs captured by the current survey frames. Table 3 shows the HU coverage ratios by census region, with the MAF-based coverage rates (that is, one minus the undercoverage rates in Table 1) included for comparison.

Table 3:
Comparison of HU Coverage Ratios
to the MAF-Based Coverage Rate (by Region)

		Estimated HU Coverage Ratio	MAF-Based Coverage Rate (1 – Undercoverage Rate for MAF-Based Frame)
U.S.		92.6%	93.6%
REGION	**Northeast**	94.4%	93.9%
	Midwest	94.0%	95.3%
	South	90.5%	91.5%
	West	93.5%	95.3%

The comparisons by region in Table 3 suggest similar coverage rates for the current frames and the MAF-based frame. For both, the greatest coverage weaknesses are in the South. The factors discussed above -- a disproportionately high share of rural HUs, new

---

[6] *The growth rate has been calculated as the change in HUs since Census 2000 compared to the Census Bureau (Population Division) official estimate of housing units as of June 2006.*

construction HUs, and mobile homes in the South -- are probable contributors to the undercoverage problem in the South in both frames, though the relative contribution of each may differ by frame.

3.2 Overcoverage at the National and Regional Level

To estimate the gross overcoverage of the MAF-based frame, we calculated the proportion of erroneous exclusions in the NES sample blocks. Erroneous exclusions are MAF units that would have met the criteria for inclusion in a MAF-based sampling frame in the NES blocks that were deleted as nonexistent, identified as a duplicate, or identified as "not an HU" (nonresidential, GQ, boarded up, etc.) by the lister.

Table 4 shows overcoverage estimates for the entire nation and for each census region.

Table 4: Overcoverage Rates for a MAF-Based Sampling Frame: U.S. and Census Region			
		% of HUs in Sample	**Total Overcoverage (SE)**
U.S.			10.23 (0.30)
REGION	**Northeast**	19.0	8.13 (0.52)
	Midwest	22.6	8.71 (0.56)
	South	37.1	13.44 (0.57)
	West	21.3	8.14 (0.63)

As with undercoverage, the South region has significantly worse overcoverage than any of the other regions. The South has a larger proportion of rural HUs than other regions, so the result in Table 4 may provide some support to the commonly accepted assumption that duplication on the MAF is worse in rural areas than in urban areas.

3.3 Undercoverage: Urban vs. Rural

Table 5 shows MAF-based frame undercoverage estimates for urban blocks vs. rural blocks (using Census 2000 urban/rural classifications).

Table 5: Undercoverage Rates for a MAF-Based Sampling Frame: Urban vs. Rural				
	% of HUs in Sample	**Total Undercoverage (SE)**	**Omissions (SE)**	**Erroneous Exclusions (SE)**
U.S.		6.39 (0.28)	4.67 (0.27)	1.73 (0.11)
Urban blocks	74.9	4.60 (0.30)	3.13 (0.26)	1.47 (0.14)
Rural blocks	25.1	11.73 (0.59)	9.22 (0.62)	2.51 (0.15)

The significantly higher rate of undercoverage in rural blocks is in line with expectations, for several reasons:

- There were more census misses in rural areas.
- Addresses in rural blocks are more likely to be non-city-style (e.g., lacking either a house number or street name). Non-city-style addresses from the DSF are not used to update the MAF, so coverage of new construction addresses suffers in rural areas.

- The DSF is a file of mail delivery points, so may miss addresses in rural jurisdictions where there is no home delivery of mail.
- The ACS filter excludes DSF adds in certain rural areas where there is a fear of duplication (this affects only the erroneous exclusion rate).

3.4 Undercoverage: Unit Frame Blocks vs. Area Frame Blocks

We derived separate coverage estimates for blocks that were screened into the unit frame for the 2000 redesign of current surveys versus those blocks that were screened into the area frame. Unit frame blocks are more urban than area frame blocks and have fewer incomplete or non-city-style addresses, on average. In unit frame blocks, coverage is provided by the census address list; post-census (new construction) addresses in these blocks are covered by the permit frame. In area frame blocks, addresses are captured through field listings of the blocks (with supplementation from the permit frame in permit-issuing areas).

Table 6 shows MAF-based frame undercoverage estimates for unit frame blocks vs. area frame blocks.

	% of HUs in Sample	Total Undercoverage (SE)	Omissions (SE)	Erroneous Exclusions (SE)
Table 6: Undercoverage Rates for a MAF-Based Sampling Frame: Unit Frame vs. Area Frame Blocks				
U.S.		6.39 (0.28)	4.67 (0.27)	1.73 (0.11)
Unit Frame Blocks	88.7	5.50 (0.31)	3.98 (0.29)	1.52 (0.12)
Area Frame Blocks	11.3	13.38 (0.67)	10.03 (0.60)	3.35 (0.28)

We found that undercoverage is significantly greater in area frame blocks than unit frame blocks. This was expected; the current surveys expend considerable effort and cost on area frame listings to improve coverage in area frame blocks. Table 6 shows, in effect, the value added by the listings in area frame blocks. More detailed information about MAF coverage in area frame blocks was provided by an earlier FACHS evaluation, the Area Frame Study (Corlett, 2006; Kennel, 2007).

Note that the Table 6 estimates are for <u>all</u> HUs in the blocks that were screened into the unit and area frames. So, each estimate includes those new construction addresses which, in the current multi-frame approach, would be part of the permit frame. The estimate for unit frame blocks, then, is not a measure just of the coverage provided by the static census address list from 2000, but of the census address list plus any post-census adds on the MAF.

This suggests an important question for the current surveys: Can the MAF-based frame, with its continual updates from the DSF (and other sources), provide better coverage for old construction addresses in unit frame blocks than the current unit frame? After all, the unit frame for current surveys is a static list of census addresses, never updated. Wouldn't the DSF provide us with the chance of capturing some census misses that are permanently missed by the unit frame?

To answer this question, we tried to identify those HUs validated by the FRs in the NES unit frame blocks that represented old construction (i.e., addresses that existed at the time of the last census). For each of these old construction HUs, then, we identified its source. That is, we determined if the HU was (1) on the MAF as a census address, (2) on the MAF as a post-census add, or (3) not on the MAF at all (i.e., added by the FRs). The estimated percentage for each source is shown in Table 7:

Table 7: Sources of Old Construction HUs (Except Mobile Homes) in Unit Frame NES Blocks	
Source	**% of Total (S.E.)**
Census Addresses on MAF	93.0% (0.3%)
Post-Census Adds on MAF	3.4% (0.2%)
Not on MAF (Added by FACHS-NE)	3.6% (0.3%)

We estimate that 3.4% of the old construction HUs in the unit frame blocks came from post-census adds to the MAF, primarily from the DSF. These HUs would have been missed by the current unit frame because they were not in the census. This points to one certain advantage that a MAF-based frame would have over the current unit frame: any old construction HUs missed by the census would have a chance to be picked up as DSF updates (or other types of post-census updates) to the MAF.

3.5 Undercoverage: Single-Unit vs. Multi-Unit Addresses

One of our goals was to explore the differences in undercoverage between single units and units within multi-unit addresses. The MAF, however, is a collection of addresses at the unit level, with no explicit grouping of those units that are in the same multi-unit address. For our analysis, we used a variable assigned by ACS (Number of Units at Basic Street Address) to distinguish between single-unit and multi-unit addresses. This variable is based on an ACS algorithm that attempts to combine MAF extract records that share the same basic street address.

Table 8 shows MAF-based frame undercoverage estimates for single units and units within multi-unit addresses.

Table 8: Undercoverage Rates for a MAF-Based Sampling Frame: Single Units vs. Multi-Unit Addresses				
	% of HUs in Sample	**Total Undercoverage (SE)**	**Omissions (SE)**	**Erroneous Exclusions (SE)**
U.S.		6.39 (0.28)	4.67 (0.27)	1.73 (0.11)
Single HUs	75.4	5.42 (0.27)	4.12 (0.27)	1.30 (0.06)
HUs in Multi-Unit Addresses	24.6	9.37 (0.74)	6.33 (0.64)	3.05 (0.38)

We found that undercoverage is significantly higher for units in multi-unit addresses than single units. But how much of that undercoverage is due to the MAF missing the entire multi-unit address rather than just missing some units within a multi-unit that is on the MAF?

The difference is important to the current surveys, since time-of-interview field procedures for the 2010 redesign could include the complete relisting of certain multi-

unit addresses. (Currently, all multi-units in the permit frame are listed prior to interview, versus a small subset of those in the unit frame.) As long as the multi-unit address itself is on the MAF, then, any MAF-missed units within the address could be recovered by these field procedures. Such listings are costly, however, so the current surveys would need to balance the potential coverage gain against the expense.

We explored this question of undercoverage within multi-unit addresses by defining undercoverage at the level of the basic street address. That is, the MAF was given credit for all HUs in a multi-unit address if the basic street address (each unique combination of the ZIP, house number, and street name fields, excluding unit identifiers) was represented on the MAF. We found that this modified definition dropped the overall undercoverage from 6.39% (0.28%) to 5.07% (0.26%), while the undercoverage for units in multi-unit addresses was approximately halved, from 9.37% (0.74%) to 4.02% (0.61%).

3.6 Undercoverage: New Construction

A major concern for the current surveys in moving to a MAF-based sampling frame is the quality of coverage for new construction housing units -- i.e., those units built after the most recent census. Does the MAF capture new construction as completely and as promptly as the current permit frame?

For the current surveys, "new construction" is defined as those housing unit addresses built after the most recent census, while "old construction" refers to those units built before the census. In order to produce accurate undercoverage measures for new construction units, then, we would have to determine when (before or after the census) each of the housing units found in the NES blocks was built. This is a task that is not possible in a conventional DAAL block listing without incurring great expense. Therefore, we produced undercoverage estimates for new construction using this proxy:

  ➢ **New construction** was defined as the set of housing units (excluding mobile homes) that were determined by the lister <u>by observation</u> to have been built after April, 2000.

Certainly, errors will be made by field listers in trying to definitively place units on one side or the other of the census by observation alone. But it should be true that the units identified by the listers as "after census" units will be, for the most part, the "newest" HUs in the NES blocks. So, if we produce undercoverage measures for this set of "newest" HUs, then we should gain valuable knowledge about the effectiveness of the DSF in capturing new construction.

Table 9 shows the undercoverage rates for the MAF-based frame for new construction and old construction.

Table 9:
Undercoverage Rates for a MAF-Based Sampling Frame:
New Construction vs. Old Construction

	% of HUs in Sample[7]	Total Undercoverage (SE)	Omissions (SE)	Erroneous Exclusions (SE)
U.S.		6.39 (0.28)	4.67 (0.27)	1.73 (0.11)
New Construction	8.6	21.71 (1.44)	18.94(1.44)	2.78 (0.45)
Old Construction	89.1	3.77 (0.22)	2.38 (0.18)	1.40 (0.11)

As expected, undercoverage is a significantly greater problem for new construction than old construction. The major source of post-census MAF updates is the DSF, so any evaluation of new construction coverage on the MAF is, to a large extent, a referendum on the quality of the DSF as a source of new units in the post-census period. In section 3.3, we already discussed how the DSF (and the DSF update process) may factor into rural undercoverage:

- Non-city-style addresses from the DSF are not used to update the MAF.
- The DSF may have coverage deficiencies in areas without home mail delivery.
- The ACS filter excludes DSF adds in certain rural areas where there is a fear of duplication.

All these factors are likely contributors to new construction coverage, as well, given the essential role of the DSF in supplying post-census updates to the MAF. But another FACHS study, the New Construction Study 2008 (Flanagan et al, 2008), provided an important additional explanation for new construction undercoverage in the MAF-based frame:

- New construction addresses appear on the DSF 5-6 months later, on average, than in the current permit frame.

The study also found that 95.6% (0.5%) of the occupied permit frame HUs from November 2005 CPS/SCHIP interviews were found on the January 2007 MAF extracts. Together, these results suggest that many of the MAF omissions reflected in the Table 9 estimates are temporary in nature. That is, many of the missing new construction HUs are missing because of DSF lag -- the DSF does not have the addresses yet, but eventually will capture them.

The New Construction Study 2008 also included an analyses of the undercoverage created in the current permit frame because of the "start dates" for permit sampling. This permit frame undercoverage offsets much of the undercoverage found in the MAF-based frame. The study concluded that, on balance, the MAF-based frame is a reasonable replacement in coverage terms for the current permit frame.

---

[7] *The new construction and old construction percentages do not add to 100% because mobile homes were excluded. Table 10 provides estimates for mobile homes.*

3.7 Undercoverage: Mobile Homes

Another area of interest for current surveys is coverage of mobile homes, particularly those placed since the census. The current permit frame is thought to provide poor coverage of mobile homes (relevant only in unit frame blocks) since building permits are not typically issued for the manufacture of mobile homes. We were able to produce separate undercoverage estimates (Table 10) for those units identified as mobile homes by the DAAL listers.

Table 10:
Undercoverage Rates for a MAF-Based Sampling Frame:
Mobile Homes vs. Other HUs

	% of HUs in Sample	Total Undercoverage (SE)	Omissions (SE)	Erroneous Exclusions (SE)
U.S.		6.39 (0.28)	4.67 (0.27)	1.73 (0.11)
Mobile Homes	5.5	18.88 (1.35)	15.22 (1.31)	3.66 (0.38)
All Other HUs	94.5	5.66 (0.29)	4.05 (0.27)	1.61 (0.11)

Just as for new construction, our data shows that mobile homes are underrepresented on the MAF compared to other types of HUs. We do not have an undercoverage estimate for mobile homes in the current frames, but we believe it likely that the MAF-based frame provides better coverage. Since both frames will have those mobile homes captured by the census, the primary difference would be for post-census placements of mobile homes. The DSF, as an inventory of mailing addresses, is expected to be a considerably better source of mobile home placements than the current permit frame.

4. Limitations

The National Evaluation results are subject to these limitations:

- The DAAL operation is not specifically designed for coverage evaluation. We took advantage of the systems, procedures, and personnel already in place for the DAAL program for reasons of practicality, budget, and operational convenience.

- The accuracy of the results is limited to the quality of the DAAL listings.

- DAAL FRs have a limited ability to search for HUs outside the blocks they are listing, so they may not be able to determine whether a HU on the dependent list for the block is nonexistent or actually exists somewhere else in the county. As a result, we may have misclassified some geocoding error as overcoverage.

- Ungeocoded MAF addresses (that is, addresses that could not be assigned to a block) were not included in our overcoverage analysis. Our evaluation was block-based, so we could not include addresses which we could not allocate to blocks. This affected only the overcoverage estimates; the undercoverage estimates were not affected.

- It is not possible to derive a "net" coverage measure from the gross undercoverage measure and the gross overcoverage measure presented by this report. That is because the base (denominator) for the two measures is different in that the denominator for gross undercoverage includes ungeocoded units and units geocoded to a non-NES block while the denominator for gross overcoverage excludes these units.

- Since the analysis is based upon comparisons of the MAF before and after the updates from the NES block listings have been applied, the efficacy of our address matching programs is an important limiting factor. If, for example, a unit added in block listing already exists on the MAF as an ungeocoded DSF unit, then the MAF will be unfairly charged with undercoverage if we fail to identify the address match. Conversely, if we match a field add with an existing MAF unit when in fact they represent separate units, then the MAF is credited with correct coverage of the added unit instead of undercoverage.

5. Conclusion

Our national estimate of gross undercoverage in the MAF-based frame is 6.4%. Gross overcoverage for the nation was estimated at 10.2%. These estimates from the National Evaluation are coverage measures of the MAF-based frame as it existed in mid-2007. We presume that MAF coverage is best immediately after each decennial census and gradually deteriorates until the next census, so any MAF coverage estimates are very time-sensitive.

Our regional estimates show significantly higher gross undercoverage and overcoverage rates for the South than other census regions. MAF coverage rates by region are quite similar to the HU coverage ratios from the Housing and Vacancy Survey (June 2008).

Our undercoverage measures show that maintaining a high quality national address list between decennial censuses is very challenging. In particular, we found relatively high undercoverage rates in area frame blocks (13.4%), for new construction (21.7%), and for mobile homes (18.9%). We recommend that the Census Bureau continue to invest in MAF improvements and MAF improvement research programs that concentrate on these areas of concern. An example is the plan by Geography Division to research ways to make use of the rural (non-city-style) updates from the DSF, which are now excluded from the MAF because of duplication concerns.

We conclude that a MAF-based frame provides better coverage of old construction than the current unit frame. The current unit frame is a static inventory of census addresses, while a MAF-based frame would include the same census addresses plus updates from the DSF (and other sources) that would capture some portion of the census misses. We found evidence of this in unit frame blocks, where 3.4% of old construction HUs were not in the census but were added to the MAF by the DSF and other sources. We expect that our unit and permit frames would fail to pick up these HUs.

We recommend that the current surveys continue to measure the quality of the demographic household survey frames and the proposed MAF-based frame, and to develop strategies to improve these survey frames when needed.

6. References

Flanagan, Tina, Clifford Loudermilk, and Xijian Liu, Final Analytical Report for the Frame Assessment for Current Household Surveys – New Construction Study for 2008, U. S. Census Bureau Internal Memorandum, 2008.

Kennel, Timothy, and Marjorie Corlett, "Comparing the U.S. Census Bureau's Master Address File (MAF) to the Survey of Income and Program Participation Area Listings", Proceedings of the Joint Statistical Meeting, 2005.

Kennel, Timothy, "A Coverage Profile of Area Frame Blocks on the United States Census Bureau's Master Address File", Proceedings of the Joint Statistical Meeting, 2006.

Li, Mei, Clifford Loudermilk, and Xijian Liu, Final Analytical Report for the Frame Assessment for Current Household Surveys – National Evaluation, U.S. Census Bureau Internal Memorandum, 2008.

Loudermilk, Clifford and Timothy Kennel, "Deciphering the DSF: Which Addresses from the Delivery Sequence File Should Be Included in the Sampling Frames for Demographic Surveys?", Proceedings of the Joint Statistical Meeting, 2005.

Martin, Joel, and Clifford Loudermilk, "Assessing the Filter Rules For Extracting Addresses from the Master Address File To Construct a Housing Unit Frame for Current Demographic Surveys", Proceedings of the Joint Statistical Meeting, 2008.