Evaluation of the Census Bureau's Master Address File Using National Health Interview Survey Address Listings

Kathy Ott, Randy Parmer, Barbara Reilly, Cliff Loudermilk, Yolanda McMillan, Tom Coughlin

U.S. Census Bureau
Washington, DC 20233

This paper describes a research project done by the Census Bureau to evaluate the coverage of the Master Address File (MAF). Basically, the MAF is intended to eventually be a list of all addresses in the United States. It will be used for the mailout of Census forms for Census 2000 and for the new American Community Survey, as well as a sampling frame for the Census Bureau's current surveys. In order to better understand the accuracy and completeness of the MAF, the Demographic Statistical Methods Division of the Census Bureau formed an Evaluation Group to study the coverage of the MAF.

To do the evaluation, we compared the addresses contained in the MAF to recent area segment listings produced for the National Health Interview Survey (NHIS). The comparison was done within Census blocks, small geographic areas created and used by the Census Bureau to collect and tabulate data.

The MAF is created using a number of steps. The initial MAF building process begins with the input of city-style addresses from two basic address sources-the Address Control File and the Delivery Sequence File.

The Address Control File (ACF) is a nationwide inventory of all living quarters enumerated during the 1990 Decennial Census of Population and Housing. Although somewhat out-of-date, the ACF is an important address source because it contains housing units that may not be mail delivery points.

The second source of addresses is the United States Postal Service (USPS) Delivery Sequence File (DSF). The DSF is a nationwide address file of all residential and commercial units that receive mail delivered by the USPS. This file provides information about existing addresses, new addresses, and demolished addresses and is constantly updated by the USPS letter carriers.

The addresses from the DSF and ACF are merged, unduplicated, and "geocoded" or mapped to a certain geographic area for tabulation using the TIGER system. In the future, noncity style addresses, group quarters, and commercial addresses will be added and geocoded. A few times a year, the MAF will is updated using new versions of the DSF. For this evaluation, only the city style addresses were on the MAF.

Up to the time of our evaluation, MAFs had been created for only a select group of counties. They were the MAFs created for the Census Bureau's American Community Survey 1996-1997 test sites. The counties where we had MAFs to use for the evaluation:

- Brevard County FL,
- Douglas County NE,
- Franklin County OH (including all of Columbus City),
- Houston TX (Harris Co. and Fort Bend Co.),
- Multnomah County OR (including all of Portland City), and
- Rockland County NY.

All six test areas are metropolitan in character and all sites contain county wide address systems. In general, then, mail delivery in each of these areas is accomplished using city-style addresses. All MAFs for these counties were updated with the April 1996 DSF. We estimate that this DSF reflects addresses that existed as of early 1996.

As an independent source of addresses, we used keyed listings done for the NHIS, a survey that does field listings of all sample areas. The NHIS listings used in our study were listed between the fall of 1994 and the summer of 1996. Thus, much of the NHIS listing is older than the MAF listing, and a small portion may be newer than the MAF listings. We attempted to identify mismatches caused from this timing discrepancy in our field verification operation.

In order to evaluate the coverage of the MAF, we verified all records that were on the NHIS file that were not on the MAF, and all records on the MAF that were not on the NHIS keyed file for specific Census blocks. We used computer programs, a clerical operation, a field verification operation in which field representatives verified the existence or nonexistence of each address, and resolution to match the addresses between the two files. In this way, every address on either file had some sort of outcome.

Once we got the results back from the field verification operation, we evaluated the MAF for undercoverage, overcoverage, and geocoding errors. The definitions we used follow.

- Undercoverage was measured by determining the cases that were legitimately on the NHIS listings, but were not included on the MAF. The MAF somehow missed these units.
- Overcoverage was measured by determining the cases that were on the MAF, but shouldn't have been. These would be MAF units that were not

on the NHIS listings or found during field verification, and should therefore not be included.

- Geocoding errors were measured for the MAF only. These were records that had an incorrect tract or block code on the MAF. Units that were on both MAF and NHIS were not checked for geocoding errors. Therefore, this number could be under representing the true percentage of geocoding errors.

Results

We calculated weights for each block or sub-block (a smaller land area) used in the evaluation. The 123 blocks we used in this evaluation were not a simple random sample because we used the NHIS listings for any block in the MAF counties. In fact, we know we have a disproportionate number of large blocks. Therefore, we needed to apply weights to represent all of the blocks that were not selected in the 7 counties. The weights reflect the inverse of the probability that the block was selected for NHIS sample.

The weight for blocks was the NHIS sampling interval divided by the (combined block measure of size *4). The weight for sub-blocks was the NHIS sampling interval divided by the number of units listed for NHIS. (See the end of this article for the definitions of the terms used in the weighting).

Standard errors were also calculated for the housing unit and basic address rates at the county level. We computed variances by two methods, first using pq/n * the 90 Housing Unit Coverage Study (an evaluation done after the 90 census) design effects. Second, by treating the sample as a simple random sample of blocks and computing simple variances between weighted block level error rates.

We compared the two methods. They were very close for overcoverage and geocoding errors, but not for undercoverage. Generally, the second method yielded lower standard errors, except in Harris where it was quite a bit higher since the errors in Harris were clustered.

We decided the second method was better, since it gives only a slight overestimate of the true variance because we aren't reflecting the variance gain due to the NHIS Within-Primary Sampling Unit sort and stratification. The first method could be quite inaccurate since it is dependent upon the error causing mechanisms which aren't consistent between the MAF and ACF.

Table 1 shows the overall coverage of the MAF, including the number of housing units (HUs) and basic street addresses (BSAs) that were used.

Table 1: Overall Coverage of the MAF in the 1996-1997 ACS Test Sites

	Under-coverage	Over-coverage	Geocode Errors
# HUs	12,915	3,893	3,893
# Nonmatch HUs	206	170	31
Weighted HU %	1.42%	4.04%	1.91%
# BSAs	4,854	2,834	2,834
# Nonmatch BSAs	43	82	31
Weighted BSA %	1.25%	3.89%	1.93%

Below, I will show two tables for each specific measurement we calculated. One table is for housing units, and the other for basic street addresses. There are breaks in the table numbers because they match the table numbers from the full paper. The full paper contains additional tables for each measurement showing size of structure breakdowns by county.

MAF Undercoverage. Undercoverage was measured by dividing the Number of NHIS Nonmatches by the (Number of NHIS HUs or BSAs - the Number of NHIS invalid HUs or BSAs). In other words, any unit that was listed on the NHIS listing, but was found to be nonexistent, commercial, or otherwise invalid, was taken out of this evaluation. The nonmatches, then, were the cases that were not on the MAF, but were legitimately on the NHIS listings and verified through the field follow-up operation.

In most cases, the tables show only those numbers that were significant (with an '*'). We list Harris County in all of the tables. This is part of the Houston, TX test site, which we divided into Fort Bend and Harris for the tabulation of data.

We also pooled together data for the Non Harris counties, shown as "Non Harris" in the tables because of the small numbers in many counties.

Table 2: MAF HU Undercoverage by County

County Name	Weighted %	Std Error
Fort Bend, TX	2.11%	0.96%*
Non Harris	0.48%	0.19%*

Harris, TX	2.40%	0.89%*
All Counties	1.42%	0.49%

As shown in the above table, the HU undercoverage weighted rate was 1.42%. The HU undercoverage rate in Harris (2.4%) is significantly higher than in the pooled Non Harris areas (0.5%). (The p-value is .036.) Although we don't know the cause of this, we have a hunch that it could be because of the fast growth in this county or because there seems to be an unusually high level of mixing of commercial and non-commercial units in this county.

Also, Harris, Fort Bend, and the pooled Non-Harris areas were significantly different from zero. None of the other areas were significant.

Table 3: MAF BSA Undercoverage by County

County Name	Weighted %	Std Error
Fort Bend, TX	1.13%	0.85%*
Non Harris	0.14%	0.12%*
Harris, TX	2.42%	0.88%*
All Counties	1.25%	0.48%

There were similar results for BSAs, shown in table 3. The overall BSA weighted rate was 1.25%. The same counties were significantly different from zero, and again Harris was significantly higher than the pooled Non-Harris counties.

These undercoverage rates are fairly low, but are statistically different from zero. In considering the use of MAF as a survey frame, the BSA undercoverage rate is more important than the HU undercoverage because sampling and field procedures can correct for units that are not included, as long as the BSA was included.

The census, however, relies on the coverage of housing units to mail out census forms and encourage response. Low undercoverage, especially in city areas where additional listing operations are not often used, is critical to a complete census.

MAF Overcoverage. Overcoverage was calculated as the Number of Invalid MAF HUs or BSAs divided by the Total Number of MAF HUs or BSAs. If an address was on the MAF, but not on NHIS and not found during field follow-up, it was included in the counts since these units were assumed to not exist. These numbers give us an idea of the extent to which the MAF contained units that are not legitimate, and should therefore not be included.

Table 6: MAF HU Overcoverage by County

County Name	Weighted %	Std Error
Douglas, NE	3.79%	1.74%*
Franklin, OH	2.18%	1.18%*
Fort Bend, TX	2.73%	1.05%*
Non Harris	3.14%	1.16%*
Harris, TX	4.94%	1.46%*
All Counties	4.04%	0.93%

As shown in Table 6, the HU overcoverage weighted rate was 4.04%. All of the areas in the table were significantly different from zero. However, none of the areas were significantly different from one another as Harris and Non Harris were for undercoverage.

Table 7: MAF BSA Overcoverage by County

County Name	Weighted %	Std Error
Douglas, NE	4.14%	1.82%*
Franklin, OH	1.41%	0.65%*

Fort Bend, TX	2.82%	1.08%*
Non Harris	2.99%	1.13%*
Harris, TX	4.81%	1.36%*
All Counties	3.89%	0.89%

As shown above, the BSA overcoverage weighted rate was 3.89%. The same areas are significantly different from zero for BSAs as for HUs.

The overcoverage rates, while higher than the undercoverage rates, seem fairly promising. In considering MAF for use as a survey sampling frame or for the census, overcoverage of housing units or basic addresses is not as damaging as undercoverage. Survey procedures can account for overcoverage, so it is correctable. In a census environment, housing unit overcoverage could lead to more costly operations (such as enumeration, nonresponse follow-up, duplicate mailings, corrections to MAF, or multiple visits to same house), but does not contribute to the loss of household or person information.

We have not looked yet at the source of the errors. We will be looking at whether the "bad" addresses are coming from the Delivery Sequence File or the Address Control File to help us improve the overcoverage rate.

MAF Geocoding Errors. The third and final measurement we calculated was MAF geocoding errors. The only HUs or BSAs that were included in the follow-up operation were in blocks where there were address nonmatches from MAF. Blocks were not sent out if they only had geocoding discrepancies. Therefore, these totals underrepresent the actual amount of geocoding error probably in the MAF. Field Representatives did, however, verify all geocoding in the blocks that were sent out for follow-up.

We calculated MAF geocoding errors as the Number of Geocoding Errors divided by (the total Number of MAF BSAs or HUs - the Number of MAF Invalid BSAs or HUs). The geocoding error rate gives us an idea of how accurately the MAF places an address in the correct geographic area.

Table 10: MAF HU Geocoding Errors by County

County Name	Weighted %	Std Error
Non Harris	2.64%	1.83%
Harris, TX	1.16%	0.65%*
All counties	1.91%	0.97%

The overall HU weighted geocoding error rate was 1.91%. Only Harris (1.16%) was significant, and it was not significantly different from any other areas. The pooled Non Harris areas were not significant, but are shown for information.

Table 11: MAF BSA Geocoding Errors by County

County Name	Weighted %	Std Error
Non Harris	2.66%	1.83%
Harris, TX	1.19%	0.68%*
All Counties	1.93%	0.97%

The BSA weighted rate was 1.93%. Again, only Harris is significant, and it is not significantly different from any other area. The Non Harris rate is shown for information.

The HU and BSA geocoding error rates presented here are promising, though they do underrepresent the true rates. More research needs to be done to evaluate and improve the geocoding process. The geocoding operation is important in that it places a housing unit or basic address in the correct tract and block, allowing the Census Bureau to collect, control, and tabulate data by geographic area.

It is difficult to speculate how different any of our results would have been if our study had included areas with noncity-style addresses. The matching process certainly would have been more difficult. The coverage provided with these noncity-style addresses will need to be evaluated in future tests and studies.

Previous research. Research presented in August, 1995 evaluated the 1995 Census Test precanvass operation to assess the accuracy and completeness of the MAF used for that test (Barrett, August 1995). The evaluation looked at the coverage for Paterson County, NJ and Oakland County, CA, both of which are urban areas. The precanvass operation added 3.4 and 2.2% for Paterson and Oakland counties, respectively. Our evaluation found 2.4% HU undercoverage in Harris, 1.42 overall. The precanvass operation deleted 7.8 and 5.7% of the housing units. Our evaluation found 4.9% overcoverage in Harris, 4.0 overall. Future research. Currently, the Census Bureau is conducting a Quality Improvement

Program in 6 counties. This evaluation is very similar to the one we did, but includes independent listings done just for the evaluation. There will be approximately 2,500 housing units in each county for this evaluation. In January of 1998, there should be another similar evaluation done using 200,000 housing units. These two projects should give more detailed information about the quality of the MAF and will allow data to be compared better across geographic area and address types.

An evaluation of the accuracy and timeliness of the DSF using building permit data from current surveys is also being planned.

Future efforts will focus on methods to add addresses to the MAF and improve its accuracy. This evaluation was done on the basic MAF before any quality steps were taken, so these rates will most likely decrease as the MAF is updated through these improvement plans.

There is a lot of data in the full paper that was not reflected in the presentation, or in this abbreviated version. The numbers were not significant, but may be interesting. Also, detailed information about the procedures, NHIS listings, and data breakdowns on size of structure are included in the full paper. You may contact Kathy Ott at the above address for a copy.

Definitions of terms used in weighting:

- The *NHIS sampling interval* represents the rate at which NHIS sample was selected within the NHIS stratum within a Primary Sampling Unit (each block is entirely within an NHIS stratum within a PSU).
- The *combined block measure of size* was the number of measures in the block. This number, based on the 1990 census figures, was used in sampling for NHIS and other demographic surveys. The expected size of each measure is 4, hence, the multiplication by 4 in the denominator of the non sub-block weight calculation.
- If a sample block for NHIS was too large for a single Lister to canvass, it was partitioned into sub-blocks to create smaller assignments. Sampling was done at the block level; we calculated no "sub-block measure of size" at time of sampling. Therefore, we used the only measure available to use, the *number of units listed for NHIS*, to represent the size of the sub-block.

Reference:

Barrett, Diane (August 1995). "An Evaluation of the Precanvass Operation to Measure the Completeness of the Master Address File". 1995 Census Test Results, Decennial Management Division Memorandum No. 5.