

# AN ASSESSMENT OF THE DECENNIAL MASTER ADDRESS FILE: CONSISTENCY OF AGGREGATED HOUSING UNIT COUNTS WITH DEMOGRAPHIC BENCHMARKS

Kirsten K. West, J. Gregory Robinson and Antonio Bruce, U. S. Bureau of the Census<sup>1</sup>  
Kirsten K. West. U.S. Bureau of the Census, Washington, DC 20233

**Key Words:** Census coverage, demographic benchmarks, housing units

## 1. Introduction

A good census starts with a good addresses list. Without the list, the questionnaires would never reach the respondents. The list helps the Census Bureau control and track field and data capture operations. It follows that the better the address list, the better the chance of a complete coverage of housing units in the census and hence, the better the coverage of the population. In the 1990 Census, an estimated one third of the not matched population could be attributed to a housing unit or a building being missed in the census (Hogan, 1993).

The Master Address File (MAF) has been built over the last decade. The starting point was the addresses in the 1990 Census address control file. This list was combined with the U.S. Postal Service Delivery Sequence File (DSF). Throughout the decade, the list has been updated and supplemented with address information provided by census programs. During the census, field and partnership operations also provide updates to the file. The Decennial Master Address File (DMAF) that is used to conduct Census 2000 is an extract of the MAF.

This paper describes and illustrates a methodology to assess the completeness of the decennial address list. Completeness is assessed by comparing the number of housing units in the address file to independently derived estimates of housing units. The estimates are benchmarks. As such they have inherent shortcomings. However, we deem the results of the comparisons to be of sufficient quality to be used as warning signs of potential coverage errors early in the census process. The earlier information is available about the quality of the data, the better the chance that corrective steps can be taken to remedy such shortcomings. It is acknowledged that operational considerations severely limit the ability to enact remedies for Census 2000.

The assessment of the address list is the first of a series of demographic analysis phases for Census 2000. A central feature of the Demographic Analysis (DA) Program for Census 2000 is its sequential and cumulative nature. As in previous censuses, traditional demographic analysis benchmarks are used to evaluate coverage of population at the national level. The new program extends the scope of the estimates to subnational areas and evaluates census coverage at several points in time during the census process--each point building upon the other. When the Accuracy and Coverage Evaluation (A.C.E.) results are available in January of 2001, the coverage patterns will be examined and compared to the direct DA estimates. The principal focus will be at the national level, though broad geographic patterns indicated by the subnational DA benchmarks will also be used to check the A.C.E. results (Robinson et al., 2000).

## 2. Background

The use of demographic benchmarks to evaluate census data is not new to the Census Bureau. Demographic Analysis (DA) is a well known evaluation tool with a history that dates back to the 1950's (Coale, 1955). It has accompanied every census since then (Siegel and Zelnik, 1966; U.S. Bureau of the Census, 1974; 1988; Himes and Clogg, 1993; Robinson et al., 1993; Robinson and West, 2000). Historically, the demographic approach measures census-to-census changes in coverage of the population (Robinson, 1994). In recent years, plans have evolved for expansion of the program. The vision is to produce coverage estimates on a timely basis and to extend the scope of demographic coverage indicators below the national level. The vision also includes the use of demographic benchmarks, such as housing unit estimates, as a tool to provide assessment of coverage early in the census process. This goal has become more attainable with the automation of data collection and processing ensuring earlier availability and accessibility of the data. The 1995 Census Test provided the first opportunity to demonstrate the utility of an evaluation program expanded to the subnational level (see Robinson, 1996a;

---

<sup>1</sup>This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

1996b; and Kohn, 1996). The Census 2000 Dress Rehearsal offered the opportunity to focus on housing unit estimates as well (Robinson et al., 1999; West, 1999).

### 3. Methodology

This section briefly describes the units of comparison (the DMAF (3a) and the housing unit benchmark (3b)), and the stratification of the analysis results (by type of enumeration area (3c)).

#### 3a. The Decennial Master Address File

The MAF was constructed differently for the different questionnaire delivery areas. In mailout/mailback areas, the MAF was prepared by matching the 1990 Address Control File to the U.S. Postal Delivery Sequence File (DSF). In these areas, the United States Postal Service (USPS) delivers the questionnaires, and the respondents mail them back. In update/leave areas, where many housing units have rural-style mailing addresses, enumerators deliver the questionnaires and update maps and the list of addresses. Again, the respondents mail back the questionnaires. In these areas, the MAF was created by rural address listing operations. Census enumerators canvassed assigned geographic areas and recorded the address and geographic location of all living quarters within the areas. Also, as part of the process to improve the MAF, a Local Update of Census Addresses (LUCA) operation was used to assess its completeness. Local and tribal officials were invited to review and update the pre-census MAF of their jurisdiction.

In areas that are remote and sparsely populated, the address list is created in an operation called list/enumerate. The address listing and the completion of the census questionnaires are done concurrently.

In this paper, the DMAF counts are analyzed at two points in time: July 1999 (the initial DMAF), and January 2000. Later analyses will focus on the April and June extracts. For the purpose of analysis, we start with the July 1999 DMAF count and create a January 2000 DMAF count by adding and subtracting units. Several operations could result in a unit being added. Similarly, updates (removal of an address or change to the address) could come from several sources.

Additions come from: new addresses from the November 1999 DSF; new congressional addresses (trivial number); any mailout/mailback addresses that are now DMAF worthy, such as newly geocoded addresses, addresses that are now in mailout/mailback areas, and LUCA98

supplemental addresses; and addresses that were verified and corrected in the LUCA98 operation.

The following outcomes are types of updates: addresses visited by LUCA98 Field Verification (LUCA98 FV); addresses merged with existing July DMAF worthy addresses; addresses now flagged as congressional addresses (trivial number); block and type of enumeration area changes since the July delivery; and block (only) changes. This universe is treated as the universe of potential deletes.

Thus, the January DMAF count is derived as follows:

$$\text{Jan00DMAF} = \text{July99DMAF} + \text{ADDS} - \text{UPDATES}$$

This is a conservative approach which will overstate the actual number of deletes between July 1999 and January 2000 because not all updates will result in an address being deleted.

The aggregate counts from the DMAF are compared to independent estimates of housing units developed by the Population Estimates Branch.

#### 3b. The independent housing unit benchmarks

The housing unit estimates are produced from the 1990 Census housing unit count, estimated annual residential construction from building permits, annual estimates of nonpermitted residential construction, annual estimates of new mobile home placements, annual estimates of housing loss from demolition permits, and annual estimates of nonpermitted housing loss. Several adjustments are made to ensure that the housing unit estimates are consistent with the independently derived county population estimates.

Although the county housing unit estimates are not released to the public, they represent the best available independent benchmark at the time of analysis (July 1999 and January 2000) and are consistent with other estimates released by the Population Division. The housing unit estimates will be one of the many estimates subject areas that will receive careful review once Census 2000 data are available.

#### 3c. Types of counties (enumeration areas)

In the Census 2000, there are nine types of enumeration areas (TEA):

- TEA 1 - Block canvassing and mailout/mailback
- TEA 2 - Address listing and Update/Leave

- TEA 3 - List/Enumerate
- TEA 4 - Remote Alaska
- TEA 5 - "Rural" Update/Enumerate
- TEA 6 - Military
- TEA 7 - "Urban" Update/Leave
- TEA 8 - "Urban" Update/Enumerate
- TEA 9 - Additions to address listing block universe

The comparisons between the DMAF and the housing unit benchmark are made at the county level. Four types of counties are analyzed: counties that are mailout/mailback only (TEA=1, i.e., there is no other TEA in the county); counties that are update/leave only (TEA=2), i.e., there is no other TEA in the county; counties with a mixture of these two types of enumeration (Tea=1+2) and finally, counties that are mailout/mailback and some other TEA combination.

It is the expectation that the DMAF count will exceed the independent estimate. It is estimated that nationwide, 3 percent of the housing stock is lost every ten years (Prevost, 1998).

#### 4. Limitations

Counties with any portion of list/enumerate (TEA=3) and remote Alaska (TEA=4) are excluded from the analysis. In these areas, the address list is created at the time of the enumeration. There are 201 such counties (out of 3,142 US counties). It is expected that these areas will account for approximately 500,000 units.

#### 5. Results

The housing unit counts on the July 1999 DMAF and the January 2000 DMAF are presented in Table 1 below. The counts are shown by type of enumeration area: counties that have only mailout/mailback (TEA=1); counties that have only update/leave (TEA=2); counties that have mailout/mailback and update/leave and no other types of enumeration (TEA=1+2); and counties with a mixture (including TEA=1, 2, 5, 6, 7, 8 and 9).

In the universe of counties selected for this analysis, there is a total of 114.5 million DMAF housing units in July 1999. This total is 7.7 million higher (7.3 percent) than the total based on the housing unit estimate. The percent difference between the DMAF and the estimate ranges from 8.9 for TEA=1 only counties to 4.6 for TEA=2 only counties (Table 1).

If the DMAF has less housing units than the estimated number for a county, the difference is categorized as a negative difference (DMAF shortage). If the DMAF has

more units than the estimated number, the difference is categorized as a positive difference (DMAF excess).

The DMAF is lower than the housing unit estimate in 518 counties. These counties tend to be small and rural. The DMAF has a total of close to 4.4 million housing units in these counties. The difference amounts to approximately 130,000 housing units (a shortage of 3.0 percent). In comparison, in the remaining 2,407 counties, the DMAF has 7.9 million more housing units than the estimated number (an excess of 7.7 percent). These counties represent approximately 110 million DMAF housing units (Table 1).

DMAF update operations occurred between July 1999 and January 2000. The updates involve adding new units and flagging existing units for future processing. As a result of these operations, there are approximately 834,000 new adds to the July DMAF count—an increase of less than 1 percent (0.7 percent). Some updates may result in deletions to the July 1999 count. Others reflect changes in geography or corrections to address information. The updates amount to more than 5 million units nationwide, or 4.4 percent.

If it is assumed that the updates result primarily in deletes, then it appears that the new January 2000 DMAF counts and the housing unit estimates are becoming more concordant. For all counties in the analysis, the difference is 3.2 percent, down from a difference of 7.3 percent based on the July 1999 file. For TEA=1 only, TEA=1 + 2 only, and for TEA="Other," the percent difference is 3.3, 3.1 and 3.1 respectively (see Table 1, next to last row). For TEA=2 only (update/leave) no change occurred between July 1999 and January 2000. The update/leave operation occurs during the month of March. Thus, the difference remains 4.6 percent.

Table 2 shows the July 1999 to January 2000 change in the average percent differences between the DMAF and the housing unit estimates for three different types of enumeration areas. The distribution of counties in TEA=2 by the average percent DMAF difference is also shown for reference.

As stated previously, and shown in Table 1, overall, the percent difference declines ("improves") between July and January for all types of areas, but primarily for TEA=1, where the percent difference is reduced from 8.9 to 3.2 percent. For the mixture of areas, TEA=1 + 2, the percent difference is reduced from 6.7 to 3.1 percent. For the counties that are TEA=1 and combinations of other types of areas, the percent changes from 7.8 to 3.1.

**Table 1. Difference between the DMAF Housing Unit Count and Housing Unit Estimates by Type of Enumeration Area in County: July 1999 and January 2000**

Difference: DMAF-98HU Est.	Total HUs	Type of Enumeration Area			
		TEA=1 only	TEA=2 only	TEA=1+2 only	TEA=other
<u>July 1999 DMAF</u>	114,480,926	19,247,951	5,299,753	60,009,990 <sup>1</sup>	29,923,232 <sup>2</sup>
Difference	7,732,358	1,576,369	245,110	3,743,873	2,167,006
Percent <sup>3</sup>	7.3	8.9	4.6	6.7	7.8
<u>Negative</u>					
Total DMAF	4,452,455	132,331	1,163,186	2,322,858	834,080
Diff. (Low)	129,644	3,623	48,550	41,149	36,322
Percent <sup>3</sup>	3.0	2.7	4.0	1.7	4.6
<u>Positive</u>					
Total DMAF	110,028,471	19,115,620	4,136,567	57,687,132	29,089,152
Diff. (High)	7,862,002	1,579,992	293,660	3,785,022	2,203,328
Percent <sup>3</sup>	7.7	9.0	7.7	7.0	8.2
<u>Jan. 2000 DMAF</u>					
Adds	833,792	158,541	n.a.	484,095	191,156
Percent	0.7	0.8	n.a.	1.0	0.8
Updates	5,065,603	1,157,744	n.a.	2,421,925	1,485,934
Percent	4.4	6.0	n.a.	4.9	6.0
New Total <sup>4</sup>	110,249,115	18,248,748	5,299,753	58,072,160	28,628,454
Difference	3,501,951	577,166	245,110	1,807,447	872,228
Percent <sup>3</sup>	3.2	3.3	4.6	3.1	3.1
Total counties	2,925	148	818	1,499	460

<sup>1</sup> Adds and updates are for the TEA=1 portion only. TEA=1 portion = 49,297,150 housing units.

<sup>2</sup> Of the 460 counties, 111 had no TEA=1. The TEA = 1 portion = 24,650,990

<sup>3</sup> Based on percent difference in the independent housing unit estimates (not shown)

<sup>4</sup> Updates are assumed to be deletes, thus Jan DMAF = July DMAF + Adds - Updates

**Table 2. Average Percent Difference between the DMAF Count and the Housing Unit Estimate (July 1999 and January 2000) by Type of Enumeration Area and by Categories of Percent Difference**

Percent Diff. DMAF-98HU Est.	Number of Counties and Average Percent DMAF Difference											
	TEA=1 only			TEA=2 only			TEA=1+2			TEA=Other		
	N	July	Jan.	N	July	Jan.	N	July	Jan.	N	July	Jan.
<b>Total</b>	<b>148</b>	<b>8.9</b>	<b>3.3</b>	<b>818</b>	<b>4.6</b>	<b>n.a.</b>	<b>1499</b>	<b>6.7</b>	<b>3.1</b>	<b>460</b>	<b>7.8</b>	<b>3.1</b>
>20	10	27.5	4.1	26	27.4	n.a.	37	25.0	14.9	12	24.1	15.0
10-20	27	14.1	5.0	137	13.7	n.a.	264	13.2	8.3	76	13.2	9.8
5-10	64	7.3	3.8	182	7.2	n.a.	492	7.1	4.1	145	7.3	4.8
0-5	41	3.3	1.8	198	2.4	n.a.	543	2.8	1.0	153	2.9	1.4
-0 to -5	3	-1.6	-3.8	166	-2.3	n.a.	143	-1.8	-3.1	52	-2.2	-3.2
-5 to -10	3	-7.1	-8.5	80	-7.1	n.a.	18	-6.7	-7.2	16	-7.3	-7.9
< -10	0	-----	-15.7	29	-16.0	n.a.	2	-12.2	-14.1	6	-13.0	-13.2

Within the TEA=1 only counties, the negative difference grows larger as expected. The number of additions is not sufficient to offset the number of updates which are assumed to represent deletes. Large shifts occur in the positive categories, especially in the 10 to 20 percent and the over 20 percent categories, where the percent differences change from a DMAF excess of 14.1 to 5.0 percent and from 27.5 to 4.1 percent, respectively.

Within the TEA=1 + 2 only counties, the shifts are not as pronounced. It should be kept in mind that at this point in time DMAF updates affect only the areas that are TEA=1. As expected, the negative average percent differences grow larger. On the positive side, the counties with an average of 25.0 percent difference in the July 1999 DMAF file, have an average difference of 14.9 percent in the January 2000 file.

Finally, for counties with a combination of type of enumeration areas, substantial improvements are observed for the counties with a 20 percent excess of DMAF housing units. For these counties, the average excess is 15.0 percent in January 2000 compared to 24.1 percent in July 1999. In the 10 to 20 percent category, the excess is reduced from 13.2 to 9.8 percent.

## 6. Discussion and Conclusion

This paper focuses on the DMAF extracts from July 1999

and January 2000. The January 2000 DMAF count is created by applying updates and new additions to the initial July 1999 DMAF extract. In particular, it is assumed that all updates from the 1998 Local Update of Census Addresses Field Verification (LUCA FV) represent potential deletes. This approach overstates the number of deletes. Thus, the reductions ("improvements") reported here in the number of excess DMAF units may be overstated. Similarly, increases in the DMAF shortage may be overstated, if too many units are subtracted in counties where the DMAF count is already low. At the national level, about 5.5 million addresses were involved in the field verification.

Given these parameters, it is the conclusion that there are substantial gains between July 1999 and January 2000 in making the DMAF count and the housing unit estimates more concordant in counties where the initial DMAF had a substantial excess of housing units. On the other hand, new additions to the January 2000 DMAF do not offset the number of updates in counties where the July DMAF showed deficiencies. We will monitor these trends in the future DMAF extracts.

When reviewing the results, it should be kept in mind that operations occurring between July 1999 and January 2000 impacted only the status of existing housing units in areas designated as mailout/mailback (TEA=1). For update/leave areas there are no changes between July and

January. (Later DMAF extracts will reflect the results of the field operations in the update/leave areas).

## 7. References

Coale, Ansley J. 1955. "The Population of the United States in 1950 Classified by Age, Sex and Color-A Revision of Census Figures." Journal of The American Statistical Association, Vol. 50 (269):16-54.

Himes, Christine L. and Clifford C. Clogg. 1992. "An Overview of Demographic Analysis as a Method for Evaluating Census Coverage in the United States." Population Index 58 (4):587-607.

Hogan, Howard. 1993. "The 1990 Post-Enumeration Program: Operations and Results." Journal of the American Statistical Association, Vol.88, No. 423, September, pp1047-1061.

Kohn, Felipe. 1996. "Evaluation of Reduction in Differential Undercount Based on the Analysis of Sex Ratios." Integrated Coverage Measurement (ICM) Evaluation Project 13. U.S. Bureau of the Census.

Prevost, Ron. 1998. "Master Address File (MAF) Enhancements." Internal memorandum. February 20.

Robinson, J. Gregory and Kirsten K. West. 200 "Demographic Analysis." In the Encyclopedia of the U.S. Census. Congressional Quarterly. Forthcoming.

Robinson, J. Gregory, Arjun Adlakha and Kirsten K. West. 2000. "The Demographic Analysis Program for Census 2000." Paper presented at the Annual Meeting of the Population Association of America. Los Angeles, California, March 23-25.

Robinson, J. Gregory, Kirsten K. West and Arjun Adlakha. 1999. "Assessment of Consistency of Census Results with Demographic Benchmarks." Census 2000 Dress Rehearsal Evaluation Memorandum C7.

Robinson, J. Gregory. 1996a. "Evaluation of CensusPlus and Dual System Estimates Results with Independent Demographic Benchmarks," Integrated Coverage Measurement (ICM) Evaluation Project 15. U.S. Bureau of the Census.

\_\_\_\_\_. 1996b. "Demographic Review of the Housing and Population Results of the 1995 Test Censuses." Memorandum to Arthur J. Norton, Chief, Population Division, U.S. Bureau of the Census.

\_\_\_\_\_. 1994. "Use of Analytical Methods for Coverage Evaluation in the 2000 Census." Paper presented at the Annual Meeting of the Population Association of America, Miami, Florida, May 5-7.

Robinson, J. Gregory, Ahmed, B.; Das Gupta, P.; and Woodrow, K. A. 1993. "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis," Journal of the American Statistical Association, Vol. 88, No. 423, p. 1061-1071.

Siegel, Jacob S. and Melvin Zelnik. 1966. "An Evaluation of Coverage in the 1960 Census of Population by Techniques of Demographic Analysis and by Composite Methods." In Proceedings of the Social Statistics Section of the American Statistical Association. Washington, D.C.: American Statistical Association, pp.71-85.

West, Kirsten K. and J. Gregory Robinson. 1999. "An Assessment of Census 2000 Dress Rehearsal Results: Consistency of Housing Unit Data With Demographic Benchmarks." Paper presented at the Annual Meeting of the American Statistical Association. Baltimore, MD, August 8-12.

U.S. Bureau of the Census. 1988. "The Coverage of Population in the 1980 Census." Evaluation and Research Reports, PHC80-E4. Washington, D.C.: U.S. Government Printing Office.

U. S. Bureau of the Census. 1974. Estimates of Coverage of Population by Sex, Race, and Age: Demographic Analysis. Census of Population and Housing: 1970 Evaluation and Research Program, No. PHC(E)4. Washington, D.C.: U.S. Government Printing Office.