AN ASSESSMENT OF CENSUS 2000 DRESS REHEARSAL RESULTS: CONSISTENCY OF HOUSING UNIT DATA WITH DEMOGRAPHIC BENCHMARKS

Kirsten K. West and J. Gregory Robinson, U. S. Bureau of the Census[1]
Kirsten K. West. U.S. Bureau of the Census, Washington, DC 20233

KeyWords: Census coverage, demographic benchmarks, housing units

1. Introduction

Demographic analysis has a long history with the Census Bureau. In the 1950's, Ansley Coale used the balancing equation of demography to create an estimate of the true population against which the 1950 Census could be evaluated (Coale, 1955). In every census since then demographic methods have played an important role in evaluations of data quality and in assessments of completeness of coverage (Siegel and Zelnik, 1966, U.S. Bureau of the Census, 1974, Fay et al., 1988, Himes and Clogg, 1993, Robinson et al., 1993).

Traditionally, the demographic method has produced estimates of completeness and accuracy at the national level. After the census, the results have also served as validations for estimates of undercount derived from post-enumeration surveys and dual system estimation. However, in recent years, plans have evolved for expansion of the demographic evaluation program (Robinson et al, 1993, Robinson, 1994). The vision is to produce coverage estimates on a timely basis and to extend the scope of demographic coverage indicators below the national level. More recently, 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 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, 1996b, and Kohn, 1996). The Census 2000 Dress Rehearsal offered the opportunity to focus on housing unit estimates as well (Robinson et al., 1999).

In the sections below, we draw on Census 2000 Dress Rehearsal data to discuss the utility of an expanded demographic program that includes housing unit estimates. It is estimated that in the 1990 Census, one third of the not matched population could be attributed to a housing unit or a building being missed in the census

(Hogan, 1993). It follows that the more complete the coverage of housing units in the census, the more complete the coverage of the population. Section 2 provides background on the dress rehearsal. The methodology is discussed in Section 3, followed by the limitations in Section 4, and the recommendations in Section 5.

2. Background

The Census 2000 Dress Rehearsal made it possible to assess the housing unit count in different geographic and socioeconomic settings in three locations across the country: 1) Sacramento City, California, 2) Columbia, South Carolina and surrounding counties, and 3) Menominee County, Wisconsin, including the Menominee American Indian Reservation. The Sacramento City site represents a typical urban environment with a diverse population. All housing units have a city-style residential address. The South Carolina site contains the city of Columbia in its entirety, Irmo which is in Lexington and Richland Counties, and the following 11 contiguous counties in north central South Carolina: Chester, Chesterfield, Darlington, Fairfield, Kershaw, Lancaster, Lee, Marlboro, Newberry, Richland, and Union. The site has a mixture of address types. The Menominee site is rural.

3. Methodology

Coverage is assessed by comparing the Decennial Address Master File (DMAF)[1] housing unit counts to independently derived housing unit estimates.[2] It is the expectation that the DMAF count exceeds the independent housing unit estimates. This is expected, because the unduplication operation in the Master Address file (MAF) building process may not catch all duplicate addresses and some housing units lost through demolition and disasters may not have been deleted from the file. It is estimated that over a ten year span, 3 percent of the housing stock is lost nationwide (Prevost, 1998).

The assessments are performed at the site level. Additionally, for South Carolina, comparisons are reported at the county, place and tract levels. The final

---

census housing unit counts are shown to demonstrate the utility of the demographic approach.

For the site and county level analysis, the focus is on the January 21, 1998 DMAF extract. For the place level analysis, the focus is on the January 26, 1998 extract and for the tract level, the extract was dated April 29, 1998. These releases include some, though not all revisions made to the file as a result of local review. It incorporates the "refresh" of the MAF with the November 1997 Delivery Sequence File (DSF) and the results of the Local Update of Census Address (LUCA) Field Verification Operation.

The independent housing unit estimates used in the assessments are also of different vintage. The early assessments used estimates for July 1, 1996 and July 1, 1997 extrapolated to the dates of the DMAF extracts. The final assessment was based on estimates produced for July 1,1998 interpolated to April 18, 1998.

4. Limitations

It is recognized that differences may be due to errors in the housing unit estimates methodology. For example, discrepancies may be caused by inconsistency between the definition of place employed in the housing unit estimation methodology and in the way the DMAF aggregation of housing units was performed. Thus, areas which experience boundary changes, such as annexations, new incorporations, and mergers since the last census could be especially prone to show large discrepancies. Discrepancies may also result if collection and tabulation geography differ for an area.

For the tract level analyses, the 1998 tract level housing unit estimates were made available by a vendor (Claritas, Inc.). The Census Bureau's population estimates program does not produce housing unit estimates for this level of geography. We have no independent assessment of the quality of the Claritas estimates. The base year for the estimates is the 1990 census.

5. Results

First, we show the results for the site and county level. Next, we focus on place and tract levels, and then the final census outcome is discussed.

5.1 Site and county level results

Table 1 presents the percent difference between the housing unit estimates and the housing unit counts (last column). In the Sacramento and the Menominee sites the DMAF count exceeds the independent estimates by 7.6 and 8.5 percent, respectively. As stated previously, it is the expectation that the DMAF count will exceed the independent estimates by at least 3 percent. Thus, the observed differences are larger than expected.

In the South Carolina site overall, the DMAF and the estimates are almost in agreement (DMAF is higher by 0.8 percent). For Richland County which includes Columbia, the DMAF is lower than the estimates by 2.3 percentage points. In comparison, for the total of the other ten counties in the South Carolina site, the DMAF exceeds the estimates by 3.2 percent. The DMAF is much higher than expected in Darlington and Kershaw, with the DMAF exceeding the estimates by 10.1 and 9.7 percent, respectively. In contrast, the DMAF for Marlboro and Newberry Counties are lower than the estimates by almost 6 percentage points. Also, for Union County the DMAF is lower (3.9 percent). Since we expected the DMAF to contain addresses that will be deleted during the census, this shortfall relative to the estimate was a concern.

We examined 1990 Census characteristics such as county population size, percent group quarters population, vacancy rates and percent trailers to look for patterns that would explain the discrepancies between the DMAF count and the estimates. Overall, the data showed no discernible patterns at the county level.

5.2 Place level results

To understand the site and county level differences better, we then looked at the consistency between the DMAF count and the housing unit estimate at the subcounty or place level. The analysis was limited to the 11 counties in the South Carolina Dress Rehearsal site.

There were a total of 49 places in the dress rehearsal site. In 32 of these places the DMAF exceeded the estimate as expected. In nine places the DMAF was *lower* than the housing unit estimates by more than 10 percentage points. Table 2 lists these places. Seven of the places identified as outliers are in the four counties previously identified as counties with low DMAF counts: Richland, Marlboro, Newberry and Union. In addition, our analysis identified Heath Springs in Lancaster County and Patrick town in Chesterfield County as outliers. In eight of the nine places, the DMAF implied an unlikely decline in the number of housing units since 1990.

5.3 DMAF tract level results

To make a clearer assessment of where the discrepancies are occurring, we next looked at tract level data from the South Carolina dress rehearsal site. Forty four tracts showed differences in excess of 10 percentage points (DMAF *lower* than the estimate). The four counties identified previously as counties of concern contained 36 of these tracts. Richland County had 27 tracts, Marlboro three, Newberry two and Union County accounted for four. Lancaster and Chesterfield Counties also appeared with outlier tracts, consistent with our place level results. Finally, our analysis revealed that Chester, Darlington and Kershaw contained outlier tracts.

We examined the outlier tracts in more detail. Tract size in 1990, percent growth since 1990, and difficulty of enumeration in 1990 as measured by Hard-to-Count (HTC) scores[1] were factors we considered. The most extreme outliers tended to have high HTC scores, but otherwise the differences between the DMAF and the estimates did not show a discernable pattern by size of tract or by HTC scores. However, the tracts with the largest discrepancies were tracts that grew by more than 40 percent since 1990. The tract in Darlington County and some of the outlier tracts in Richland County were fast growing tracts.

5.4 Final census results

The results reported in the previous sections were available between May and October 1998, i.e. before the final census results. Thus, they have the potential to provide a heads-up of final outcomes and potential coverage problems. To illustrate this point, Table 3 shows a comparison of the Census 2000 Dress Rehearsal results with independent housing unit estimates produced by the Census Bureau and by the State agency. Percent differences for housing units are shown in column 5. The South Carolina results are reported for the site and for the counties. For reference, results from the 1990 Census are provided in column 1. (Note that the data in Table 3 represent "census level" estimates and do not include adjustment for net undercount in either 1990 or 1998.)

In Sacramento, the Census 2000 Dress Rehearsal housing total is within the range of the independent estimates. The housing unit total falls short of both the Census Bureau and the California Agency estimate (by 0.5 and 1.9 percent)--but the margin of error in the independent estimates could be this large.

In Menominee County, the Census 2000 Dress Rehearsal housing total is higher than the independent estimates. The housing unit count is higher than expected (6.9%), however, given the imprecision in the independent estimate for such a small site we cannot make any reliability statements.

In the South Carolina site, the Census 2000 Dress Rehearsal housing totals fall consistently below the independent estimates, with the housing shortfall exceeding 10 percent in one county. For the total site, the census housing total is 5.6 percent below the independent estimate. All counties have fewer housing units than estimated; the shortage exceeds 7 percent in three counties (Marlboro, Newberry, and Union). These counties were in fact the counties where the early evaluation of the DMAF indicated the greatest shortfall (see Table 1).

6. Discussion and conclusion

The first part of the evaluation assessed the consistency of the DMAF for Sacramento City, Menominee County, and the eleven counties in the South Carolina site. The DMAF counts exceeded the independent housing estimates for Sacramento, Menominee, and seven of the South Carolina counties--this relationship (higher DMAF) is consistent with our expectation. The DMAF counts were lower than the housing benchmarks for four counties in South Carolina, and the shortfall was almost 6 percentage points for two of these (Marlboro and Newberry). It was the expectation that at the site level the DMAF count would exceed the independent estimate. In the MAF building process, the unduplication operation may not have caught all duplicate addresses and some housing units lost through demolition and disasters may not have been deleted from the file. It is estimated that nationwide, 3 percent of the housing stock is lost every ten years. Given the findings from the dress rehearsal this assumption may need to be revisited.

Next, the evaluation of the place and tract level data for the 11 counties in the South Carolina site identified specific areas where the DMAF was considerably lower than the independent estimate. The analyses also showed that the DMAF tends to be low for smaller tracts and tracts that experience rapid growth over the decade. The methodology is able to identify tracts that are not conforming to the overall pattern for the site, and extreme outlier tracts can easily be spotted.

It is recommended that in Census 2000, local knowledge be used to resolve discrepancies identified through this methodology. For example, the Federal-State Cooperative Program for Population Estimates (FSCPE) might be of assistance in reviewing the results. The FSCPE produces some subnational estimates in cooperation with the Population Division's Population Estimates Program. It is further recommended that the local knowledge be solicited in time for input to the Geography Division's MAF building operation for the Census 2000. Finally, areas showing large discrepancies should be checked for annexation activity. If no such activity occurred since the last census, there is little chance of inconsistency due to this source.

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.

Fay, Robert E., Passel, J. S., and J.G. Robinson. 1988. "The Coverage of Population in the 1980 Census. U.S. Bureau of the Census Evaluation and Research Report, No. PHC80-E4. Washington, D.C.: U.S. Government Printing Office.

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, H. 1993. "The 1990 Post-Enumeration Program: Operations and Results." Journal of the American Statistical Association, Vol. 88 (423):1047-1061.
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.

Long, John L. 1993. "Postcensal Population Estimates: States, Counties and Places." Technical Working Paper No. 3. U.S. Bureau of the Census: Population Division.

Prevost, Ron. 1998. "Master Address File (MAF) Enhancements. Draft Proposal." Internal Census Bureau Memorandum dated February 20.

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. U.S. Bureau of the Census.

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.

Robinson, J. Gregory. 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 (423):1061-1071.

Siegel, Jacob S. and M. 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, pp. 71-85. Washington, D.C.: American Statistical Association.

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.

8. Endnotes

1. The Decennial Master Address file (DMAF) is a computer file of addresses that serves as the basis for the Census 2000 address list which is used to deliver census questionnaires to households. The DMAF is an extract from the larger Master Address File (MAF).

2. The housing unit estimates are produced by the Census Bureau by adding new construction (permit and non-permit) and new mobile homes to the previous years housing stock and subtracting demolitions (permit and non-permit). For a complete description of the housing unit methodology see Long (1993).

3. The index is created from 1990 Census data using 16 variables related to difficulty of census enumeration. These variables include percent renter, percent multi-units, percent crowded units, percent households in poverty, percent unemployment, percent linguistic isolated households, and percent without a high school degree. The higher the score on this index, the higher the prevalence of such characteristics in the tract. For this site, the highest score on the index is 117, the lowest is zero.

Table 1. Comparison between Decennial Master Address File Counts and Housing Unit Estimates in the Census 2000 Dress Rehearsal Sites

Location	DMAF Count (1)	Housing Unit Estimate (2)	Absolute Difference (3=1-2)	Percent Difference (4=3/2)
SACRAMENTO, CA				
Sacramento City	171,569	159,500	12,069	7.6
MENOMINEE, WI				
Menominee County	2,071	1,909	162	8.5
SOUTH CAROLINA				
Site Total*	285,510	283,289	2,221	0.8
Richland County	121,037	123,872	-2,835	-2.3
10-Counties	164,473	159,417	5,056	3.2
Chester	13,410	13,274	136	1.0
Chesterfield	18,181	17,842	339	1.9
Darlington	28,960	26,293	2,667	10.1
Fairfield	10,317	9,657	660	6.8
Kershaw	22,132	20,183	1,949	9.7
Lancaster	24,751	23,396	1,355	5.8
Lee	7,513	7,503	10	0.1
Marlboro	11,803	12,462	-659	-5.3
Newberry	14,774	15,669	-895	-5.7
Union	12,632	13,138	-506	-3.9

* Excludes Lexington County (Irmo part only)
Source: Col. 1: Housing unit estimate, July 1, 1997 extrapolated to January 23, 1998. Col. 2: DMAF extract, January 23, 1998.

Table 2. Comparison of Decennial Master Address File Counts and Housing Unit Estimates for Places in the South Carolina Census 2000 Dress Rehearsal Site Ranked by Percent Difference

Place and County	1990 Census Count (1)	1998 DMAF Count (2)	1998 HU Estimates (3)	Absolute Difference (4=2-3)	Percent Difference (5=4/3)
Heath Springs town, Lancaster	356	71	371	-300	-80.9
Eastover town, Richland	355	150	393	-243	-61.8
Bleinheim town, Marlboro	91	69	107	-38	-35.5
Pomoria town, Newberry	110	80	120	-40	-33.3
McColl town, Marlboro	1038	850	1207	-357	-29.6
Silverstreet town, Newberry	82	62	88	-26	-29.5
Blythewood town, Richland	71	62	75	-13	-17.3
Patrick town, Chesterfield	151	152	181	-29	-16.0
Jonesville town, Union	528	487	552	-65	-11.8

Source: Col. 1: Revised data from 1990 Census (April 1, 1990). Revisions include post-1990 census corrections of political geography or geographic misallocations and boundary updates. Col. 2: DMAF extract as of January 26, 1998. Col. 3: Housing unit estimate as of January 26, 1998. The estimates are extrapolations from July 1, 1997.

Table 3. Comparison of Housing Unit Results: 1990 Census, Census 2000 Dress Rehearsal Results, and Independent Housing Unit Estimates

Site and source of Independent Estimate	HOUSING UNITS				
	1990 Census Count	Final Dress Rehearsal	1998 Housing Unit Estimate	Difference: Dress Rehearsal - Estimate	
				Amount	Percent
	(1)	(2)	(3)	(4=2-3)	(5=4/2)
Sacramento City, CA					
Census Bureau	153,362	158,281	159,058	-777	-0.5%
California Agency	153,362	158,281	161,348	-3,067	-1.9%
Menominee County, Wi					
Census Bureau	1,742	2,046	1,914	132	6.9%
Wisconsin Agency	n.a.	n.a.	n.a.	n.a.	n.a.
South Carolina Site					
Census Bureau:					
Site Total	253,285	273,497	289,848	-16,351	-5.6%
Richland County	109,555	119,214	126,615	-7,401	-5.8%
Other Counties (Total)	143,730	154,283	163,233	-8,950	-5.5%
Chester	12,293	12,677	13,393	-716	-5.3%
Chesterfield	15,100	17,316	18,258	-942	-5.2%
Darlington	23,601	26,108	26,686	-578	-2.2%
Fairfield	8,730	9,607	9,782	-175	-1.8%
Kershaw	17,479	20,453	20,591	-138	-0.7%
Lancaster	20,929	22,396	23,694	-1,298	-5.5%
Lee	6,537	7,128	7,650	-522	-6.8%
Marlboro	10,955	10,908	12,646	-1,738	-13.7%
Newberry	14,455	14,503	15,848	-1,345	-8.5%
Union	12,230	12,014	13,237	-1,223	-9.2%
South Carolina Agency	n.a.	n.a.	n.a.	n.a.	n.a.

Sources:

Col. 1: Revised data from 1990 census. Revisions include post-1990 census corrections of political geography or geographic mis-allocations and boundary updates. Three counties in the South Carolina site are affected: Chesterfield, Marlboro, and Richland.

Col. 2: From dress rehearsal results available on Census Bureau's Internet site (www.census.gov)

Col. 3: Independent housing estimates for dress rehearsal Census Day (4-18-98).

Note: The housing data for Irmo town in Lexington County is included in the South Carolina site total and 'Other County' total, but is not shown separately.