Census 2000 Evaluation O.20
August 18, 2003

Assessment of Consistency of Census Data with Demographic Benchmarks at the Subnational Level

FINAL REPORT

This evaluation study reports the results of research and analysis undertaken by the U.S. Census Bureau. It is part of a broad program, the Census 2000 Testing, Experimentation, and Evaluation (TXE) Program, designed to assess Census 2000 and to inform 2010 Census planning. Findings from the Census 2000 TXE Program reports are integrated into topic reports that provide context and background for broader interpretation of results.

Arjun L. Adlakha, J. Gregory Robinson, Kirsten K. West, and Antonio Bruce
Population Division

USCENSUSBUREAU
Helping You Make Informed Decisions • 1902-2002

CONTENTS

TABLES

FIGURES

EXECUTIVE SUMMARY

The Demographic Analysis approach has been used at the Census Bureau to evaluate the completeness of census coverage since 1960. Up until 1990, the evaluations were limited mainly to the national level. In this paper, we have implemented a systematic approach of using demographic benchmarks of housing and population to draw inferences about the Census 2000 coverage at the subnational level.

Demographic Benchmarks for Housing Units

Independent estimates of housing units, developed by updating the 1990 census housing unit counts for the loss of units due to demolition and the additions of new units since the census, are used to evaluate the number of housing unit addresses at the early stages and the housing unit counts at the final stages of Census 2000. The analysis shows that at the start of the census operations the Decennial Master Address File had about 5 percent more addresses than expected for mailout/mailback areas. The address file was used to mailout questionnaires and control field and processing operations. However, at the completion of the census operations, the housing unit count was in line with the estimated count. This finding is consistent with the national demographic analysis results which show the Census 2000 population to be very close to the expected population overall.

Demographic Benchmarks for Population

Three different types of demographic benchmarks are used to assess the census coverage of three

age segments of the population at the subnational level. These include: independent benchmarks of the population of school age provided by school enrollment data for 1990 and 2000; independent benchmarks of the population ages 65 and older in 1990 and 2000 based on Medicare data; and independent estimates of the population of children 0 to 9 based mainly on intercensal births. In addition, an analysis of census sex ratios is carried out to infer improvement in coverage of males relative to females.

The school enrollment ratios and Medicare enrollment ratios for 1990 and 2000 indicate that net population coverage improved from 1990 to 2000 for the nation, its regions and groupings of counties, and that the coverage improvement was greater in counties that had higher concentration of minorities or higher hard-to-count scores. This finding is consistent with the national DA results that show greater reduction in net undercount between 1990 and 2000 for Blacks than for non-Blacks.

The independent estimates of population age 0 to 9 show: (i) regions varied with regard to undercount rates, with net census undercount below the national average of 1.7 percent in the Midwest and Northeast regions and higher in the South and West regions, (ii) between 1990 and 2000 coverage improved in all regions; but the pattern of regional differences remained unchanged from 1990 to 2000, (iii) coverage improvement was greater for Blacks than non-Blacks in every region, (iv) consequently, coverage differentials between the Black and non-Black children narrowed, and (v) despite this narrowing, Black children continued to have higher undercount rates than non-Black children in every region except the West. The West region anomaly perhaps is a result of the fact that a large proportion of non-Black children under age 10 in the region is Hispanic. Hispanic children, according to our preliminary analysis, had higher undercount rates than non-Hispanic children, which is consistent with other findings.

The sex ratio analysis shows lower sex ratios for Blacks than for non-Hispanic Whites across all regions, revealing a higher undercount rate for Black men in each region. Also, the Black and White sex ratio gap for 2000 is about the same as in 1990, indicating about equal coverage improvements for Black males and females in Census 2000.

Recommendations

The findings reported above for the subnational areas are plausible, consistent with the national DA results, and add geographic context to the national results. Our recommendations are: (1) further develop subnational demographic analysis as a census evaluation tool; (2) integrate housing benchmark analysis with other evaluations of Master Address file throughout the decade; (3) use demographic benchmark analysis to evaluate coverage in the test censuses among other measures; and (4) expand the current demographic analysis to include subnational benchmarks in the 2010 census evaluation. Also, we recommend that basic research and development activity be undertaken to expand the coverage assessment using demographic benchmarks at the individual state or county level.

1. BACKGROUND

Demographic Analysis (DA) is an analytic approach that has been used extensively at the Census Bureau to evaluate the consistency of census results and completeness of coverage at the national level in every census since 1960 (see Siegel and Zelnik, 1966; U.S. Bureau of the Census, 1974, 1988; Robinson et al., 1993 for the demographic evaluations of the 1960-1990 censuses; and U.S. Bureau of the Census, 2001, for the initial DA evaluation of Census 2000). For Census 2000, an initial evaluation was carried out in March 2001, and subsequently the DA was revised to incorporate recommendations from an in-depth research program on the DA components conducted during March-September 2001 (see U.S. Bureau of the Census, 2001a).

In the 1990s, the demographic analysis program was expanded to include census coverage evaluations at the subnational level by employing a number of housing and population benchmarks. An application of the demographic analysis methodology was successfully used to evaluate (1) the consistency of the Census 2000 Dress Rehearsal census estimates in Sacramento City, CA, Menominee County, Wisconsin, and the Columbia, South Carolina site, and (2) the effectiveness of the Integrated Coverage Measurement (ICM) in achieving a reduction in the differential undercount in these sites (Robinson, West, and Adlakha, 1999). The expanded program was also used in a comprehensive evaluation of the quality of the 1995 test censuses in Oakland, CA, and Paterson, NJ, and six parishes in Louisiana (see Robinson, 1996, 1996a, and Kohn, 1996). In order to present a better perspective on the subnational findings a summary of the national DA findings is presented below.

1.1 The 2000 National DA: A Summary of the Findings

As reported at the outset, Demographic Analysis has been used at the Census Bureau to assess coverage in every census since 1960. The 2000 Demographic Analysis estimates for the population under age 65 and ages 65 and over are produced using different techniques. The DA estimates for the population under age 65 are based on the compilation of historical estimates of the components of population change: births (B), deaths (D), immigration (I), and emigration (E). The population estimates ($P_{0\text{-}64}$) are derived by the basic demographic accounting equation of population components applied to each single year birth cohort: $P_{0\text{-}64} = B - D + I - E$. The DA estimates for population ages 65 and over are based on aggregate Medicare enrollment data adjusted for under-enrollment.

The DA estimates of population coverage in Census 2000 were carried out in two stages. First, in March 2001, early results of the national DA were produced for the Executive Steering Committee on Accuracy Coverage Evaluation Policy (ESCAP). Due principally to the uncertainty in the estimates of unauthorized immigration, the DA used a range for the number of unauthorized immigrants and produced two sets of the national DA: (1) the "Base" DA set of estimates–which was at the low end of the range–assumed that the net increase in the number of unauthorized immigrants during the intercensal 1990-2000 period was 2.77 million; and (2) the "Alternative" set–the high end of the range–doubled the assumed increase in unauthorized immigrants to 5.53 million (yielding an implied total of 8.86 million of unauthorized immigrants in the United States in 2000).[1]

The "base" DA and "alternative" DA total population estimates for the year 2000 were 279.6 million and 282.3 million, respectively, compared with the Census 2000 count of 281.4 million. Thus the "base" DA estimated respectively a net overcount of 1.8 million or a net overcount rate of 0.65 percent in Census 2000. The "alternative DA" which assumed additions to the flows of unauthorized immigration in the 1990s gave a net undercount of 0.9 million, or 0.32 percent.

Between March and October 2001, an extensive DA research program was undertaken which examined historical levels of the components of population change with special emphasis on the international migration component. Based on the results of this investigation, revised DA estimates were produced in October 2001.

The revised October DA estimates of population changed little from the March "alternative" DA results (from 282.3 million based on the March DA estimate to 281.8 million based on the revised September DA estimate). In fact, the revisions of the DA components of change lowered the estimated DA net undercount rates–from 1.85 percent to 1.65 percent for 1990 and from 0.32 percent to 0.12 percent for 2000. The revisions did not alter the DA finding that net undercount rates in 2000 were substantially lower than in 1990–or that a differential undercount continued to exist between Blacks and the rest of the population.

---

[1]The higher number of assumed unauthorized immigrants (5.53 million) was reasonably consistent with the foreign-born totals from the March CPS (for details see U.S. Bureau of the Census, 2001)

The largest numerical revision to the components of change was for unauthorized immigrants. The revised estimate of the residual foreign-born population–a category comprised primarily of the unauthorized population–was 10.24 million, or 1.38 million higher than the implied estimate used in the March Alternative DA population estimate of 282.3 million. However, the estimate of legal immigration decreased by 0.88 million and the estimate of births was lowered by 0.72 million. The net effect of all the revisions, including revisions to the other components such as deaths, legal emigration, other legal immigration, and revisions to population ages 65 and over, was to lower the DA estimate of the population by 0.58 million.

Some of the important findings about the census coverage from the national DA are as follows:

1. The DA estimates show a lower net undercount rate for the Census 2000 than for the 1990 Census, (see Table 1). The net undercount rate was lowered by 1.53 percentage points from 1.65 percent in 1990 to 0.12 percent in 2000.

2. The DA estimates show reduced net undercount rates for the Census 2000 compared to the 1990 Census among all demographic categories-- males and females, Blacks and non-Blacks, and all broad age groups.

3. The DA estimates show a narrowing of differential undercount between Blacks and all other groups from 4.4 percentage points in 1990 to 3.1 percentage points in 2000.

Table 1. Demographic Analysis Estimates of Percent Net Undercount,

by Race, Sex, and Age: 1990 and 2000

(a minus sign denotes a net overcount)

	Demographic Analysis		
Category	1990 (1)	2000 (2)	Change (1-2)
Total	1.65	0.12	1.53
Black	5.52	2.78	2.74
0-17	5.27	1.30	3.97
Male, 18+	9.57	7.15	2.42
Female, 18+	2.05	0.07	1.98
Non-Black	1.08	-0.29	1.37
0-17	1.12	0.54	0.58
Male, 18+	1.74	0.17	1.57
Female, 18+	0.44	-1.27	1.71

Source: U.S. Census Bureau, Population Division.

1.2 Objectives

The primary purpose of this evaluation report is to compare subnational census data on housing and population with a number of demographic benchmarks to draw inferences about the coverage at the subnational level.

Specific objectives of the study are as follows:

1. Compare the total number of addresses in the Decennial Master Address File (DMAF) with independent housing benchmarks to draw inferences, early in the census process (July 1999), about the completeness of the housing address list.

2. Compare April 2000 census housing counts with independent housing benchmarks to draw inferences about the magnitude of housing coverage and its geographic differentials.

3. Use subnational demographic tools (such as school enrollment ratio, Medicare enrollment ratio) to study relative improvement in coverage between the 1990 and 2000 censuses by selected groupings of counties.

4. Compare independent estimates of children age 0-9 by race (Black and non-Black) based mainly on births during 1990-2000 with the census population age 0-9 for the regions to draw inferences about the coverage of children age 0-9 in Census 2000 and change in coverage from 1990. Develop illustrative demographic coverage benchmarks for Hispanic children using demographic data on births, deaths, and migration estimates.

5. Compare sex ratios in 1990 with sex ratios in 2000 by age and race (Black and non-Black) to assess changes in the coverage of males relative to coverage of females.

2. METHODS

We use selected demographic benchmarks to assess the Census 2000 results at the subnational level. The demographic benchmarks were obtained from different data sources: housing estimates; birth registration data; Medicare data; and school enrollment data. Sex ratios are an example of a demographic measure obtained through statistical manipulation of the data.

2.1 Description of Methodology for Demographic Benchmarks for Housing Units

The early assessment of housing unit coverage was obtained from a tabulation of the number of residential addresses on the first extract of the Decennial Master Address File (DMAF). The last assessment examined the final housing unit count in Census 2000.

The Population Division's housing unit estimate served as an independent benchmark. The housing unit estimates were developed from the 1990 Census housing unit count plus annual estimates of residential construction from building permits, of non-permitted residential construction, of new mobile home placements, of housing unit loss from demolition permits, and of non-permitted housing loss. Furthermore, adjustments were made to ensure that the housing unit estimates were consistent with the independently derived population estimates. The benchmarks for the early assessments were the housing estimate for July 1, 1999. For the final assessment the July 1, 1999 estimates were extrapolated to April 1, 2000.

July 1999 represents the first extract of decennial addresses from the Master Address File (MAF). The extract file is known as the Decennial Master Address File (DMAF). If an address is not in the DMAF at the start of the census, it can be added during decennial operations. Similarly, if an address is believed to be a duplicate or nonexistent, it can be flagged so that it is not included in the final results. Thus, throughout the census operations the DMAF is updated to reflect the outcomes of previous operations and new extracts are used to control the subsequent operations. The April 2000 count, representing the final census housing unit count, is derived from the hundred percent edited file (HCEF).

For the analysis purposes, counties were aggregated based on the type of enumeration within the county. County groups were created according to the Type of Enumeration Area (TEA): counties that were enumerated by the mailout/mailback mode only (TEA=1), counties that had update/leave only as enumeration mode (TEA=2), and counties with a mix of only these two types of enumeration modes (TEA=1+2). Counties with a mix of list/enumerate (TEA=3) or remote Alaska (TEA=4) were treated separately. (See Appendix 2 for a complete list of enumeration areas and the distribution of counties by type of enumeration area)

The initial number of DMAF addresses was expected to exceed the housing unit estimates. An estimated three percent of the housing unit stock is lost every ten years (Prevost, 1998). This loss may not yet be reflected on the initial DMAF. Also, the housing unit estimates have inherent uncertainties (nonsampling errors) -especially for smaller geographic areas- that lead us to believe that a difference within plus/minus five percent from the DMAF is acceptable. In the evaluation, differences outside this range would lead to recommendations of review. A review consists of an in depth examination of the components used to create the independent housing unit estimate. It might also involve field checks and expert review of both the DMAF and the benchmark for the area in question.

2.2 Description of Methodology for Demographic Benchmarks for Population

Three types of demographic benchmarks are used to assess the census coverage of the population at the subnational level. These data include: independent benchmarks of the population of school age provided by the school enrollment data; independent benchmarks of the population 65 and older in 1990 and 2000 based on Medicare data; and independent estimates of population of children under age 10 based mainly on the intercensal births. In addition, an analysis of census sex ratios is carried out to infer improvement in coverage of males relative to females.

<u>Independent benchmarks of the population age 7 to 14 based on school enrollment data</u>

Administrative data on school enrollment provide independent benchmarks for evaluating coverage of the school-age population. The school enrollment data, which are quite complete (especially for public schools), can be compared with the enumerated school age population to provide coverage ratios at the time of the 1990 census and Census 2000. The enrollment ratios from 1990 to 2000 provide an effective means to measure <u>change</u> in completeness of census coverage between the two points in time (1990 and 2000). See Appendix 1 for a detailed discussion of the use of school enrollment data to assess coverage of the population age 7 to 14.

Independent benchmarks of the population age 65 and older in 1990 and 2000 based on Medicare data

Medicare data have been used extensively in the development of Demographic Analysis coverage estimates for the nation and in the production of postcensal population estimates for states and counties. Medicare tabulations provide independent benchmarks for assessing the 1990 census and Census 2000 results of the population 65 and over. The change in the census to Medicare ratios from 1990 to 2000 provides an effective means to measure change in completeness of coverage between the two points in time (1990 and 2000), in a manner similar to the use of school enrollment data. See Appendix 1 for a detailed discussion of the use of Medicare data to assess coverage.

Independent benchmarks of population age 0-9 based on registered births

The independent estimates for the population under 10 years of age are derived as follows: First, birth cohorts in each intercensal year are survived to April 1, 2000 by the use of national life table survival ratios. Second, an appropriate allowance is made for the estimates of net intercensal migration of young children. The population estimates are produced for each state by race (Black and non-Black). Illustrative estimates are developed for Hispanic children age 0 to 9. The difference between the census count and the population estimate is an indicator of net census undercount. This technique provides demographic-type undercount estimates of census coverage for 2000 at the subnational level for young children - a group which has been disproportionately undercounted in censuses.

Sex ratio

The sex ratio is defined as a ratio of males per 100 females. Sex ratios by age have been used effectively in the census evaluation studies at the national level to draw inferences about the differential coverage of adult Black males (U.S. Census Bureau, 2001). Sex ratios can be calculated easily for any geographic area and used to infer coverage differential of males relative to females (Robinson, 1994).

3. LIMITATIONS

The aggregate administrative data and estimates that are incorporated in the demographic benchmarks are subject to various types of errors. Many assumptions go into this estimation process, some of which can be validated and some of which are based on quite limited information. The limitations of the housing units are discussed first, followed by a discussion of the limitations of the demographic benchmarks.

3.1 Housing Unit Estimates

The housing unit estimates are developed for the county level. Within a county, the estimated number of housing units in an area can be below or above the DMAF. Such differences are masked at the higher level of geography. A tract or block level analysis would be a better tool to pinpoint where geographically the differences exist within the county.

The methodology for developing the housing unit estimates has inherent uncertainties. The data on mobile homes, the numbers of homes constructed without permits and those of demolished housing units are particularly subject to uncertainty. We have developed data quality indicators to examine the components of the independent estimates for each county to assess if there are signs that the estimates over -or understate the number of units (Devine, 2000). The quality indicators reflect percent mobile homes, and percent non-permitted construction and demolition in a county.

In spite of these uncertainties, the county housing unit estimates represent the best available current estimates and are consistent with other estimates released by the Population Division during the 1999 production year.

3.2 Population Benchmarks

The overall accuracy of the demographic estimates depends on the quality of the demographic data and the quality of any modifications applied to the data. The internal consistency of the demographic estimates permits the trends and changes in coverage patterns over time to be estimated more precisely than the exact level of net coverage in any given census (see Robinson et al., 1993). Currently we have no accepted statistical basis to assess the accuracy of the independent population estimates or other benchmarks. Until some uncertainty models are developed, the demographic results can be considered only "face validity" indicators. The consistency of the findings across a variety of different demographic benchmarks (e.g., population estimates from different sources, use of school enrollment data, Medicare data) provides some check on the reasonableness of the estimates.

The major DA estimates are available only at the national level and only for two broad race categories: Black and All other Races Combined. The latter is referred to as "non-Black" in this report.[2] Finally, the demographic benchmarks for states and counties are not available by all population characteristics, that is, we do not have reliable independent demographic benchmark estimates at the subnational level for race groups, sex, and tenure (owner, renter).

---

[2]Throughout this report the term Black is used to refer to the Black or African American population.

4. RESULTS

4.1 Housing Units: July 1999 DMAF Comparison with Demographic Benchmarks for Housing Units

The purpose in this section is to compare the total number of addresses in the July 1999 Decennial Master Address File (DMAF) with independent housing benchmarks to draw inferences early in the census process about the completeness of the housing address list.

The July 1999 DMAF extract served as the control file for the field and processing operations associated with the census. As shown in Table 2, nationwide, the file started out with about 120.2 million addresses at the national level. This was 5.3 percent more addresses than expected given our benchmark of 114.2 million housing units. The differences vary by type of enumeration in the county.

Most addresses in the U.S. are in the mailout/mailback universe (the respondent receives a questionnaire in the mail and sends it back in the mail). For this universe, (Type of Enumeration, or TEA, equals 1) the difference amounts to 6.2 percent (Table 2), which is above the established range of plus/minus 5 percent. In the update/leave areas (TEA=2), a smaller percent difference is observed (3.4 percent). In these areas, the enumerator updates the address around the time of the census and leaves a questionnaire for the respondent to mail back. It is expected that the update process makes address corrections, deletes non-existent addresses, and adds new addresses. In

July 1999, the benchmark does not indicate problems of concern for these areas. Similarly, the percent difference for the counties with a mix of mailout/mailback and update/leave falls close to the acceptable range (5.1 percent). Finally, the difference (2.6 percent) between the DMAF and the benchmark estimate for counties with other mode (TEA = Mix), including list/enumerate and remote Alaska is within the established range of plus/minus 5 percent.

Similar benchmark analyses were conducted for extracts of the DMAF in January, April and June 2000. Overall, the benchmark analyses pointed to an initial excess of units regardless of type of enumeration area. It was the assessment that unless a substantial number of units were deleted during the census operation, the census would end up with a higher than expected housing unit count. Many counties showed positive differences in excess of 10 percent. Many of these counties were large, i.e., they had more than 50,000 housing units in 1990. For a full account of the findings, reference is made to West, 2000, 2000a, 2000b, 2000c, 2000d, and 2000e.

	National	**≥ 95% TEA=1**	**TEA=2**	**TEA=1+2**	**TEA=Mix**
DMAF	120,244,120	67,077,859	5,299,753	29,731,561	18,134,947
Benchmark	114,226,276	63,140,457	5,123,093	28,281,404	17,681,322
Diff.	6,017,844	3,937,402	176,660	1,450,157	453,625
% Difference	5.3	6.2	3.4	5.1	2.6
Counties	**3,142**	**391**	**818**	**1,302**	**631**

4.2 Housing Units: Census 2000 Counts Compared with Demographic Benchmarks for Housing Units

In this section we compare April 2000 census housing counts with independent housing benchmarks to draw inferences about the magnitude of housing coverage and its geographic differentials.

The hundred percent census edited file (HCEF), which became available internally in December of 2000, represents the final universe of housing units (vacant and occupied). For the nation, the HCEF count of housing units is 0.4 percent above the benchmark (Table 3). This implies an excess or overcount of around 500,000 units in the census, relative to the benchmark. Also, relative coverage varies by type of enumeration area. In counties that are almost entirely mailout/mailback, there is a negative difference of 0.5 percent, or an implied undercount of around 300,000 units relative to the benchmark. In update/leave counties, the difference of 5.1 percent is slightly outside the acceptable range of plus/minus 5 percent. The relative small size of the update/leave universe (about 5 million housing units) should be noted. Also, the benchmark is considered less accurate for small, rural areas. Many counties designated as update/leave fall in this category. In the counties with both mailout/mailback and update/leave areas (TEA 1 and 2), and in counties with mixed types of enumeration mode, the difference is 1.2 percent–well within the expected range of difference.

In summary, the DMAF started out with 5 percent more addresses than expected but the census count ends up being very close to the estimate overall, which is consistent with the small net undercount of population reported by the national DA. When examined by type of enumeration area, it is the finding that the HCEF count and the housing unit estimate are in concurrence for mailout/mailback areas. Most of the nation’s housing stock is in this type of enumeration area. In the update/leave areas, no improvements were registered between the initial count and the HCEF count–the HCEF count exceeds the benchmark by about 5 percent.

In Mailout/mailback areas (TEA = 1), counties in the Northeast have an excess of units (1.2 percent) compared to observed shortages in counties in the other three regions (Table 3). In update/leave areas (TEA = 2), we find that counties in the South have more overcount (5.8 percent) than counties in the other regions. It should be noted that the update/leave counties are concentrated in the South. For counties with a combination of mailout/mailback and update/leave, negative differences are seen in the Northeast region (the proportion of mailout/mailback within the counties dominates as the enumeration method). Excesses are observed in the other regions. Shortages are observed in the Midwest for the counties with mixed enumeration mode. It should be noted that the benchmark estimates are less reliable for small, rural counties which comprise a majority of counties in the mixed mode type. However, the observed shortages are within the 5 percent range.

Table 3. Percent Difference Between the HCEF Housing Unit Count and the Housing Unit Estimate for the Nation, and Regions by Type of Enumeration Area: April 2000					
	National (1)	≥95% TEA=1 (2)	TEA=2 (3)	TEA=1+2 (4)	TEA=Mix (5)
National					
HCEF	115,904,641	63,465,541	5,443,378	28,888,644	18,107,078
Benchmark	115,397,326	63,763,369	5,181,458	28,555,864	17,896,635
Difference	507,315	-297,828	261,920	332,780	210,443
Percent Diff.	0.4	-0.5	5.1	1.2	1.2
Northeast					
HCEF	22,180,440	13,925,434	97,968	5,299,958	2,857,080
Benchmark	22,013,342	13,758,459	95,776	5,328,064	2,831,043
Difference.	167,098	166,975	2,192	-28,106	26,037
Percent Diff.	0.8	1.2	2.3	-0.5	0.9
Midwest					
HCEF	26,963,635	15,370,927	789,178	9,176,092	1,627,438
Benchmark	26,977,660	15,446,980	769,283	9,122,972	1,638,425
Difference	-14,025	-76,053	19,895	53,120	-10,987
Percent Diff.	-0.1	-0.5	2.6	0.6	-0.7
South					
HCEF	42,382,546	20,073,346	4,168,654	13,168,994	4,971,552
Benchmark	41,967,711	20,248,031	3,938,765	12,884,024	4,896,891
Difference	414,835	-174,685	229,889	284,970	74,661
Percent Diff.	1.0	-0.9	5.8	2.2	1.5
West					
HCEF	24,378,020	14,095,834	387,578	1,243,600	8,651,008
Benchmark	24,438,613	14,309,899	377,634	1,220,804	8,530,276
Difference	-60,593	-214,065	9,944	22,796	120,732
Percent Diff.	-0.2	-1.5	2.6	1.9	1.4
Number of counties	**3,142**	**391**	**818**	**1,301**	**632**

Note: One county changed type of enumeration code between the initial DMAF and final Census.
The category ≥ 95% TEA = 1 indicates counties in which at least 95 percent of the housing units were enumerated by mailout/mailback mode.
TEA = 2 indicates counties that had update/leave only as mode of enumeration.
TEA = 1+2 indicates counties with a mix of only TEA = 1 and TEA = 2 mode of enumerations.
TEA = Mix indicates that there were a number of enumeration procedures within each county.
(see Appendix 2 for details).

4.3 Demographic Benchmarks

This section uses the available demographic benchmarks (school enrollment data, Medicare enrollment data, and independent estimate of children 0-9), and sex ratios for subnational areas to study relative improvement in coverage between the 1990 and 2000 censuses.

During the 1990s, we developed a set of illustrative coverage indicators for subnational areas and used these to broadly evaluate coverage in the 1995 Test Census and 1998 Dress Rehearsal (Robinson, 1996; Robinson et al, 1999). The indicators used here include independent benchmarks for the three age segments of the population. The benchmarks for the population age 7 to 14 are based on school enrollment data, the benchmarks for the population age 65 and over are based on Medicare data, and benchmarks for children under age 10 are based largely on registered births. We examine the overall consistency of these benchmarks with the census results on the assumption that the coverage of the benchmarks has not changed over time. As such, if inconsistences are found (e.g., relative change in an indicator is different from the change in the population count) they are assumed to result from the changes in the census coverage. In addition, we analyze sex ratios to draw inferences about the coverage differentials of males in relation to females.

<u>School and Medicare enrollment data</u>

Table 4 presents coverage ratios in 1990 and 2000 for county aggregations implied by two different data sets; school enrollment data to infer relative coverage of the population age 7 to 14; and Medicare data to assess relative coverage of the population ages 65 and over. The counties are classified by region, by the percent of their population that is not non-Hispanic White (reflecting minority concentration), by percent of population that live in rented units, by hard-to-count score (HTC), and for the population classified by metropolitan statistical areas and nonmetropolitan statistical area (MSA and non-MSA). The following findings emerge.

The comparison of ratios between 1990 and 2000 of each age segment indicate an increase in population coverage from 1990 to 2000 for the nation and its regions and all other groupings of counties. For example, for the nation as a whole the ratio of enumerated children to the school enrolled population increased by 4.5 percentage points, from 96.9 in 1990 to 101.4 in 2000, and the ratio of enumerated population ages 65 and over to the Medicare-enrolled population increased by 2.2 percentage, from 103.0 in 1990 to 105.3 in 2000. All county groupings show an increase in the ratios between 1990 and 2000 although the amount of increase varies, indicating universal improvements in coverage for all areas but not alike in all areas. The finding of "universal" coverage improvements from 1990 to 2000 is consistent with the national DA estimates that indicate a decline in net undercount rates–for non-Blacks as well as Blacks and for broad age groups -older ages and for younger ages.

Second, the ratios indicate that coverage improvement was greater in counties with higher concentration of minorities or with high hard-to-count scores. For example, Table 4 shows that the ratio of enumerated population ages 65 years and older to the Medicare enrolled population increased from 103.0 in 1990 to 105.3 in 2000, an increase of 2.2 percentage points. The increase was greatest for the 159 counties with more than 50 percent minority concentration (103.7 to 108.9). For the 1,673 counties with less than 10 percent minority population, the 1990 to 2000 increase is only 1.7 percentage points. The pattern of change is similar for the measures based on the school enrollment data. The ratios increased for all county groupings–but tended to increase more for the areas with greater minority concentrations and with higher hard to count scores. This finding is also consistent with the national DA results that show greater reduction in net undercount between 1990 and 2000 for Blacks than for non-Blacks.

Table 4. Census Counts Relative to Demographic Benchmarks (per 100) for County Groupings: 1990 and 2000							
		Ages 7-14			Ages 65+		
		Census Compared to school enrollment			Census Compared to Medicare enrollment		
Category	Counties	1990 Census	2000 Census	Diff	1990 Census	2000 Census	Diff
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Region							
All Counties	3,141	96.9	101.4	4.5	103.0	105.3	2.2
Northeast	217	98.3	100.8	2.5	101.9	104.6	2.7
Midwest	1,055	98.4	101.8	3.4	101.4	103.3	1.8
South	1,424	95.7	100.9	5.2	104.5	106.7	2.2
West	445	96.0	102.4	6.4	103.9	105.8	1.8
Minority Concentration							
Minority>=50%	159	93.7	100.2	6.5	103.7	108.9	5.2
Minority 25-50%	557	96.0	100.5	4.6	102.9	105.1	2.2
Minority 10-25%	752	97.9	101.3	3.5	103.2	105.2	2.0
Minority<10%	1,673	98.5	103.4	4.9	102.7	104.4	1.7
Tenure							
Renters>=30%	729	95.8	100.5	4.7	102.2	104.6	2.4
Renters<30%	2,412	98.6	102.9	4.3	104.0	106.0	2.0
Hard-to-Count (HTC) Scores							
HTC>=70	258	94.3	100.4	6.2	103.3	108.3	5.0
HTC 50-70	535	95.9	100.1	4.3	102.0	103.8	1.8
HTC 30-50	813	96.8	101.2	4.3	102.9	105.2	2.3
HTC<30	1,535	99.1	103.0	3.9	103.6	104.9	1.3
MSA/Non-MSA							
MSA	848	96.9	101.1	4.1	103.4	105.5	2.1
Non-MSA	2,293	96.9	102.9	6.0	101.9	104.4	2.5

Note: Numbers may differ in last digit due to rounding.
The total number of counties in Tables 2 and 3 differ from that in Table 4 for the following reason. The housing unit estimates in tables 2 and 3 were developed from the 1990 census when there were 3142 counties on the file. In 2000 there were 3141 counties on the file. The Yellow Stone National Park county was incorporated into Yellow Stone county.

Minority concentration is based on 2000 Census population of counties, and includes all groups other than Non-

Hispanic Whites. HTC Scores are based on 12 specific demographic, housing, and socioeconomic variables that are associated with nonresponse and undercount (including percent renter, multi-units, lack of telephone, vacancy rates, poverty, not high school graduate, mobility, and language isolation). For a description of the HTC scores, see Bruce et al., 2001.

Estimated undercount rate for children ages 0-9

Black/Non-Black

Table 5 shows estimated percent undercount of children under age 10 for 1990 and 2000 by region and races, Black and non-Black. These are derived by comparing the expected population under age 10, based on registered births for 10 years prior to the census, and estimated migration data, with the enumerated population under age 10. The analysis was carried out at the state level but the results are provided only at the regional level. This is done to minimize the effect of errors in the migration component which is estimated indirectly (see Appendix 1). Since births, which are believed to have complete coverage, are the primary basis for these estimates, they could be viewed as relatively more robust than the measures discussed in the previous section. Furthermore, a problematic aspect of this analysis is the method used to allocate the race of birth. Demographic analysis classified births by the race of father. This is not necessarily consistent with the race allocation for the children enumerated in the census for whom more than one race was reported (see Appendix 3 for converting multiple race into a single race).

The national estimate of undercount rate for children age 0-9 using this technique is 1.7 percent in 2000. This is lower than the undercount rate for children age 0-9 of 2.6 percent estimated by the national DA. Some difference is expected because the migration component is estimated differently between the national DA and the subnational DA. However, some caution should be exercised when discussing regional and race differences.

Some of the broad findings are:

1. The regions vary with respect to undercount rate of children age 0-9 in Census 2000 (Figure 1). The West region has the highest estimated net undercount rate of 4.3 percent for the population age 0-9 in 2000 and the Midwest has the lowest estimated net undercount of -1.0 percent. The estimated net undercount rates in the South and Northeast rank second and third, respectively. A similar pattern of net undercount is noted for the 1990 Census. The regional variations in net undercount rates reflect largely the differential coverage pattern of non-Black children.

2. Between 1990 and 2000, population coverage for ages 0-9 improved in every region (Table 5). The largest improvement was in the South region where the undercount rate declined by nearly four percentage points, from about six percent in 1990 to about two percent in 2000. Coverage improved appreciably in the Northeast and West regions as well, where the undercount rates declined by about two percentage points. The Midwest region, the region with the measured overcount of population in both the 1990 census and Census 2000, had the least coverage change.

3. Between 1990 and 2000, the net undercount rate for ages 0-9 declined in every region for both Blacks and non-Blacks, but the decline was substantially greater for Blacks (varied between 5 percentage points for the Midwest and South regions to 9 percentage points for the West region) than non-Blacks (varied between 0 percentage points for the Midwest region to about 3 percentage points in the South region). Consequently, this benchmark analysis suggests that the

coverage differential between Blacks and non-Blacks narrowed in all regions (compare Black and non-Black differences for 1990 and 2000 on the bottom panel of Table 5). These findings add geographic context to the earlier national results that measured improvement in coverage nationally.

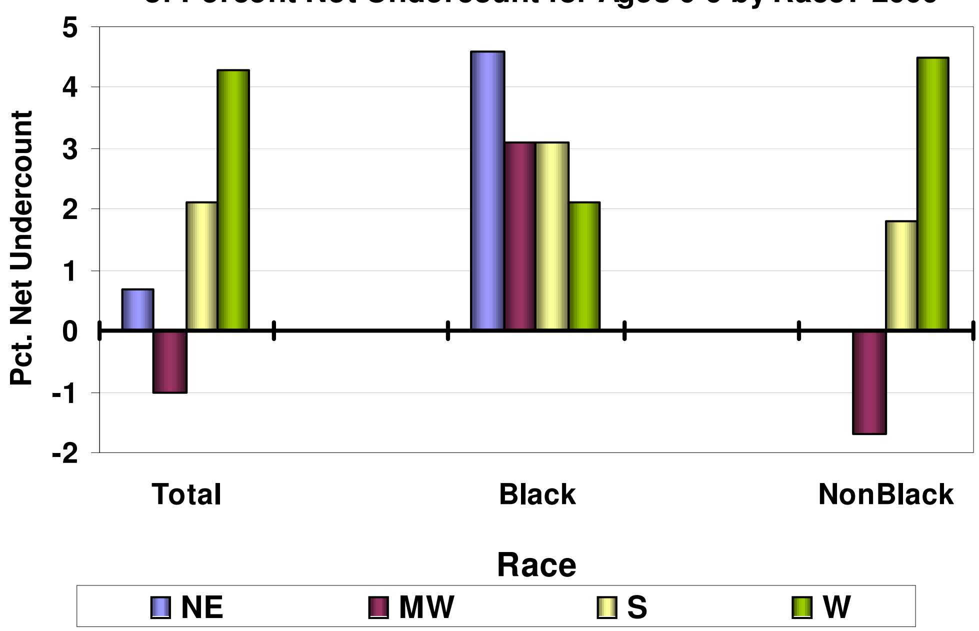
4. Despite narrowing of the coverage differential between races, Blacks continue to have higher undercount rates than non-Blacks in 2000 in every region except the West. This unusual finding in the West region perhaps arises due to the fact that nearly 38 percent of non-Black children under age 10 in the region are Hispanic who, as shown below, may have higher net undercount in the census than the non-Hispanic children and Black children.

Table 5. Demographic Benchmark Analysis Estimates of Percent Net Undercount for Children Ages 0-9: 1990 and 2000

Region	1990	2000	Diff	% Hispanic 2000
Total	**4.0**	**1.7**	**2.3**	**18.5**
Northeast	2.9	0.7	2.2	13.9
Midwest	-0.2	-1.0	0.8	7.9
South	5.9	2.1	3.8	15.9
West	6.4	4.3	2.1	35.9
Black	**8.8**	**3.3**	**5.6**	**5.5**
Northeast	10.9	4.6	6.3	13.5
Midwest	7.9	3.1	4.8	2.7
South	8.1	3.1	5.0	5.0
West	10.8	2.1	8.7	12.1
Non-Black	**3.1**	**1.4**	**1.7**	**21.0**
Northeast	1.4	-0.1	1.4	13.9
Midwest	-1.5	-1.7	0.2	8.7
South	5.2	1.8	3.4	20.2
West	6.1	4.5	1.6	37.6
Black / Non-Black Diff	**5.7**	**1.9**	**X**	**X**
Northeast	9.6	4.7	x	x
Midwest	9.3	4.8	x	x
South	2.8	1.3	x	x
West	4.7	-2.4	x	x

x - not applicable

Figure 1. Demographic Benchmark Analysis Estimates of Percent Net Undercount for Ages 0-9 by Race: 2000



Hispanic/Non-Hispanic

Table 6 shows the coverage estimates by Hispanic/non-Hispanic origin of children age 0-9 for Census 2000. These are produced using the same methodology as employed for the Black/non-Black children discussed above. However, the estimates in Table 6 are illustrative mainly because they are very sensitive to the way the births are assigned to origin (Hispanic or non-Hispanic). The National Center for Health Statistics (NCHS) classifies births by the mothers' origin. Based on special tabulations from NCHS, we have developed an alternative classification of births which uses origin of both parents, i.e., a birth is classified Hispanic if either or both parents are Hispanic and a birth is classified non-Hispanic when both parents are non-Hispanic.

Table 6 shows the coverage estimates based on the both classifications. The Census 2000 coverage estimates based on the "mother only" classification in Table 6 show a lower undercount rate for the Hispanic children (national of 0.4 percent) than the rate for the non-Hispanic children (national rate of 2.5 percent). These results are surprising and clearly inconsistent with the coverage studies using sample survey techniques, which have consistently shown much higher net undercount rates for Hispanics than for non-Hispanics. On the other hand, the estimates based on the "either or both parents are Hispanic" classification provide an expected coverage pattern - in every region the census coverage of Hispanic children age 0-9 is markedly less complete than non-Hispanic children - for the United States the estimated undercount rate for the Hispanic children is 12 percent compared with almost no undercount for the non-Hispanic children.

These estimates are, however, subject to classification error which would depend on the magnitude of the disagreement between the assigned origin to births based on the "either or both parents" origin and the way the origin is reported in the census for children of mixed origin parentage. More work is needed to determine the appropriate birth tabulations on Hispanics to compare to the census.

Region	Mother Only	Either or Both Parents
U.S.	**2.1**	**2.3**
Northeast	1.0	1.1
Midwest	-0.6	-0.4
South	2.9	3.2
West	4.1	4.4
Hispanic	**0.4**	**11.9**
Northeast	-1.5	10.4
Midwest	-18.7	2.7
South	2.6	12.4
West	2.8	13.6
Non-Hispanic*	**2.5**	**-0.1**
Northeast	1.4	-0.6
Midwest	0.7	-0.7
South	3.1	1.2
West	4.8	-1.7
Hispanic / Non-Hispanic Diff	-2.1	12.0
Northeast	-2.9	11.0
Midwest	-19.5	3.3
South	-0.5	11.2
West	-1.9	15.3

Sex Ratios

The sex ratio, defined as the ratio of males per 100 females, is frequently used in census evaluation studies to assess the quality of census data on sex composition, particularly classified by age. Sex ratios are affected by differential census coverage, by sex ratio at birth, and sex differential in mortality and migration rates.

Sex ratios have been used to evaluate the relative counts of males and females by race (Black and non-Black) in the U.S. censuses at the national level. The evaluations have been carried out by developing expected sex ratios based on data on births, deaths, and net migration and comparing them with census sex ratios (Robinson, 2001). For Census 2000, the results are shown in Figures 2 and 3. The following observations are noted. First, there is a gap between the sex ratios for Blacks and non-Blacks in both the expected sex ratios based on Demographic Analysis(DA), (Figure 2) and census sex ratios (Figure 3). The Census 2000 sex ratio gap is much larger than the 2000 expected sex ratio gap based on DA. This indicates that the gap between the census sex ratios for the Black and non-Black population is not entirely a result of their differences in sexratio at birth, death rates, or migration rates but differential census coverage is also present. Second, the 2000 expected sex ratios are much higher in DA for adult Blacks than the corresponding sex ratios from Census 2000. This indicates a higher undercount rate of Black males relative to the undercount rate of Black females. For non-Blacks, the difference in the DA and census sex ratios is much smaller, reflecting the smaller male-female difference in undercount rates.

Figure 2. Sex Ratios by Race: 2000 DA

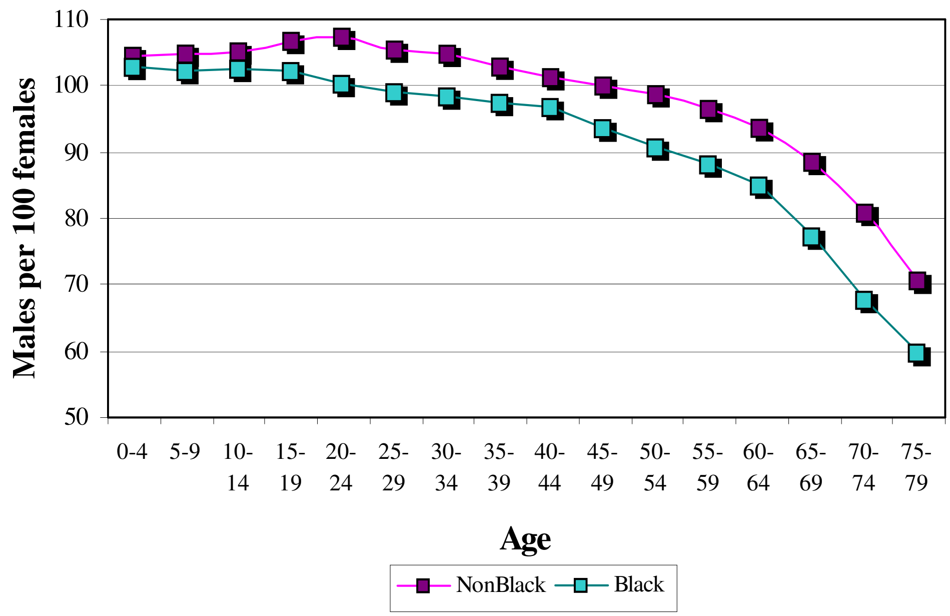
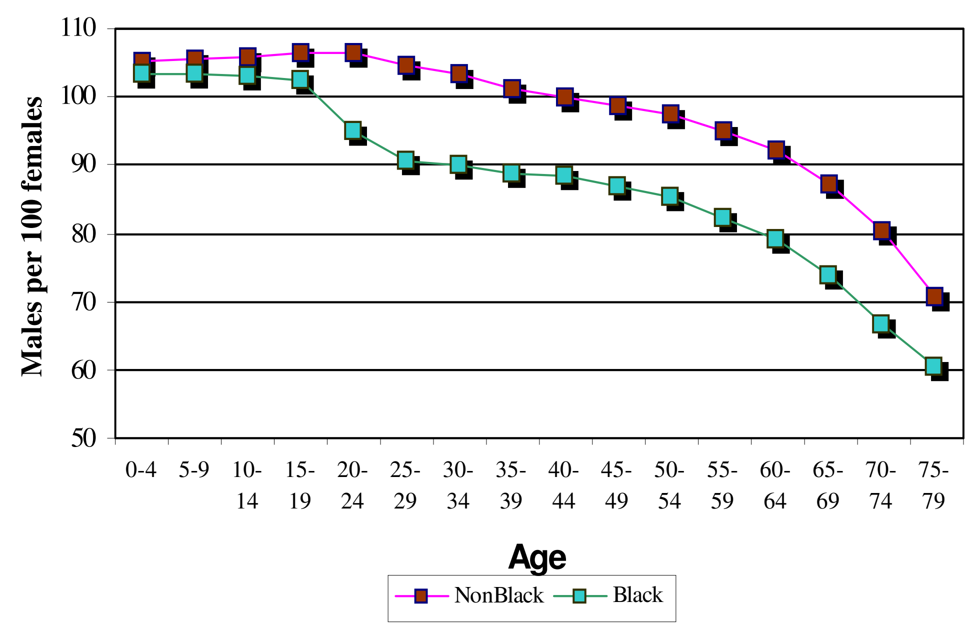


Figure 3. Sex Ratios by Race: Census 2000



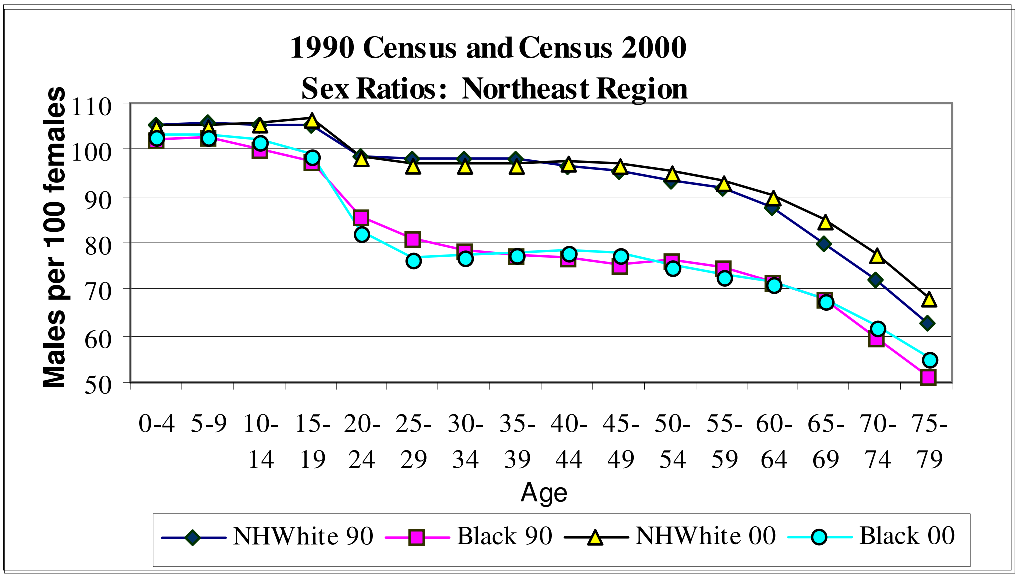
For subnational areas, expected sex ratios based on DA are not available. Consequently we rely on the gap between race groups' census sex ratios to infer coverage differentials. We compare the sex ratios of Blacks to the sex ratios of non-Hispanic Whites, a group believed to be well covered by the census for both males and females. We exclude the group quarters (GQ) population because it has greater concentration of males than females and the relative size of the GQ population varies dramatically by age.

In Figure 4, census sex ratios are shown for Blacks and non-Hispanic Whites in 1990 and 2000. The sex ratios are lower for Blacks than non-Hispanic Whites across all regions, a pattern similar to the national sex ratios by race shown in Figure 3. This difference, as is the case for the nation, is likely attributable in large part to the higher undercount rate of Black men in each region.

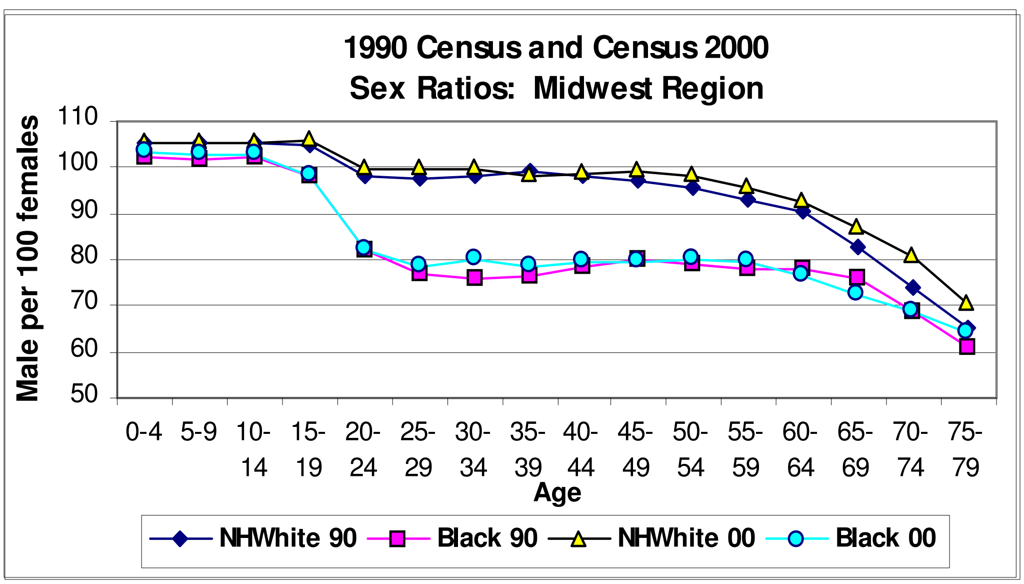
Second, in spite of the finding reported earlier that there was a greater improvement in the census coverage for Blacks, Figure 4 shows that the Black-White sex ratio gap for 2000 is about the same as the gap for 1990 across all regions. This indicates similar coverage improvements for Black males and females in the Census 2000, which would leave the sex ratio difference relatively unchanged. These findings also add geographic context to the national results which measure nearly equal improvements in the census coverage of Black adult males and females.

Figure 4 A-D. 1990 Census and Census 2000 Sex Ratios by Regions

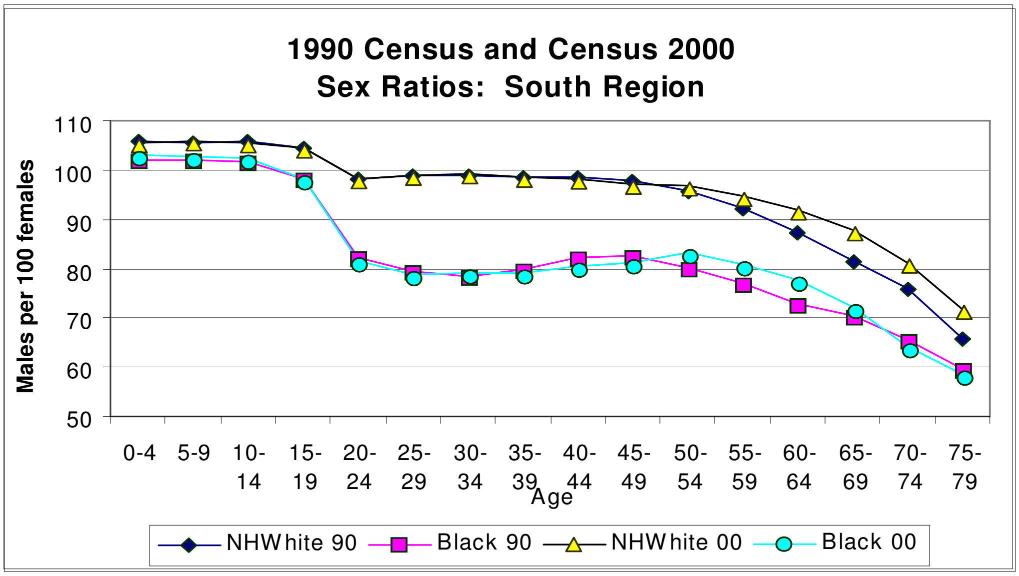
A



B

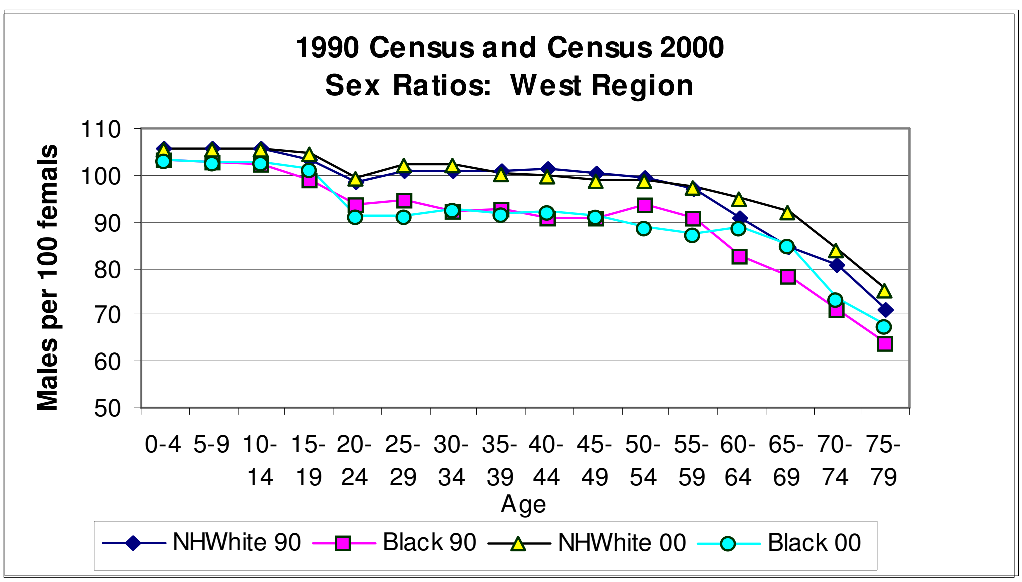


C



1990 Census and Census 2000
Sex Ratios: South Region
Males per 100 females
110
100
90
80
70
60
50
0-4
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
Age
NHWhite 90
Black 90
NHWhite 00
Black 00

D



1990 Census and Census 2000
Sex Ratios: West Region
Males per 100 femals
110
100
90
80
70
60
50
0-4
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
Age
NHWhite 90
Black 90
NHWhite 00
Black 00

5. SUMMARY AND RECOMMENDATIONS

5.1 Housing

The analysis of the housing data has led us to conclude that the demographic benchmark analysis framework is a useful tool to evaluate the completeness of the address list. If implemented throughout the decade, the tool will help assure the quality of the address list used for the 2010 Census. If the address list is still in a building stage, experiences from the Census 2000 Dress Rehearsal and Census 2000 suggest that an initial listing excess of 5 percent to 8 percent is acceptable. For geographic areas, where the current DMAF represents an update of the 2000 DMAF, areas that fall well above or below the benchmark should be flagged. These discrepancies should be investigated well in advance of the census. This would help assure the quality of the address list used for the 2010 Census.

At the onset of Census 2000, we used the demographic housing unit benchmarks to predict if a county would end up with an overage or a shortage. Counts of shortages and overages of housing units at the county level were produced by subtracting the number of estimated housing units from the count of addresses representing housing units on the DMAF. Interpreting a negative difference as a shortage and a positive difference as an overage assumes that the benchmark represents the "truth." Future research needs to test the validity of this assumption, especially for areas that are small and rural.

5.2 Demographic Benchmarks

Demographic evaluations of coverage have been an important part of every census since 1960. Up until 1990, the evaluations have mainly been limited to the national level. In this paper, we have implemented a systematic approach of using demographic benchmarks to draw inferences about the Census 2000 coverage at the subnational level. Overall results for the subnational areas are plausible, consistent with the national DA results, and add geographic context to the national results.

Despite the inherent limitations of the demographic estimates, the battery of demographic and analytic tools described here provides an important independent basis to assess the accuracy of the census results and underlying patterns of coverage. Our recommendations are: (1) further develop subnational demographic analysis as a census evaluation tool; (2) integrate housing benchmark analysis with other evaluations of Master Address file throughout the decade; (3) use demographic benchmark analysis to evaluate coverage in the test censuses among other measures; and (4) expand the current demographic analysis to include subnational benchmarks in the 2010 census evaluation.

In general, the coverage analysis has been carried out for aggregation of counties, because benchmark estimates have certain unmeasured deficiencies, the effect of which is dampened when data are aggregated for higher geographic levels. As such, to assess the coverage at the individual state or county level, several basic research and developmental activities would be required, including (1) investigating differences in the way race is reported in the census and vital statistics and developing models to account for such differences, (2) developing models to measure the effect of various factors on sex ratios for geographic areas, with the goal of isolating the effect of differential coverage, (3) investigating the use of alternative methods for measuring internal migration (e.g., school enrollment data, Medicare enrollment data, IRS data, census data on residence 5-years ago) to refine the indicator of coverage for subnational areas, and (4) developing models to assign measures of the uncertainty to the estimates. Together, these activities would enhance the utility of demographic benchmark analysis as a coverage analysis tool below the national level for future censuses.

6. REFERENCES

Bruce, Antonio, J. Gregory Robinson, and Monique V. Sanders. 2001. "Hard-to-Count (HTC) Scores and Broad Demographic Groups Associated with Pattern of Response Rates in Census 2000." Paper presented at the Annual Meeting of the American Statistical Association, August 5-9, Atlanta, Georgia.

Devine, J. 2000. "Results From the County Level Demographic Benchmark Analysis of the Decennial Master Address file—Extra: Using County Level Housing Unit Estimates as a Benchmark for the DMAF." Count Review Memorandum Series 00-02. U.S. Census Bureau, March 6, 2000.

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, R. 1998. "Master Address File (MAF) Enhancements." Internal Census Bureau Memorandum. February 20.

Robinson, J.G., B. Ahmed, P. Das Gupta, and K.A. Woodrow. 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, pp. 1061-1071.

Robinson, J.G., B. Ahmed and E. W. Fernandez. 1993a. "Demographic Analysis as an Expanded Program for Early Coverage Evaluation of the 2000 Census." Paper presented at the 1993 Annual Research Conference, March 21-24, Arlington, Va.

Robinson, J. G. and E. L. Kobilarcik. 1995. "Identifying Differential Undercounts at Local Geographic Levels: A Targeting Database Approach". Paper presented at the Annual Meeting of the Population Association of America. San Francisco

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

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

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

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

Robinson, J. Gregory, 1994. "Use of Analytic Method for Coverage Evaluation in the Census." Paper presented at the Population Association of America Meetings, Miami, May 5-7.

Siegel, J.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. Washington, D.C.: American Statistical Association, pp. 71-85.

________. 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 (4). Washington, D.C.: U.S. Government Printing Office.

_________. U.S. Bureau of the Census. 1988. "The Coverage of the 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. 2001. "Accuracy and Coverage Evaluation: Demographic Analysis results", by J. Gregory Robinson. *DSSD Census 2000 Procedures and Operations Memorandum Series B-4*, March 1.

__________. U.S. Bureau of the Census. 2001a. "ESCAP II. Accuracy and Coverage Evaluation: Demographic Analysis Results", by J. Gregory Robinson. *Executive Steering Committee For A.C.E. Policy II, Report No. 1,* October 13.

West, K.K. 2000. “Results From the County Level Demographic Benchmark Analysis of the Decennial Master Address File-Part A: Differences 5 Percent or Below for Selected Types of Enumeration Areas.” Count Review Memorandum Series 99-01. U.S. Census Bureau, January 10, 2000.

West, K.K. 2000a. “Results From the County Level Demographic Benchmark Analysis of the Decennial Master Address File-Part B: Differences in Excess of 10 Percent for Selected Types of Enumeration Areas.” Count Review Memorandum Series 99-02. U.S. Census Bureau, February 15, 2000.

West. K.K. 2000b. “Results From the County Level Demographic Benchmark Analysis of the Decennial Master Address File-Part C: Results From the January Decennial Master Address File.” Count Review Memorandum Series 00-01. U.S. Census Bureau. April 18, 2000.

West. K.K. 2000c. “Results From the County Level Demographic Benchmark Analysis of the Decennial Master Address File-Part D: Results From the January Decennial Master Address File.” Count Review Memorandum Series 00-03. U.S. Census Bureau. June 28, 2000.

West. K.K. 2000d. “Results From the County Level Demographic Benchmark Analysis of the Decennial Master Address File-Part E: Results From the April Decennial Master Address File.” Count Review Memorandum Series 00-04. U.S. Census Bureau. August 11, 2000.

West. K.K. 2000e. “Results From the County Level Demographic Benchmark Analysis of the Decennial Master Address File-Part E: Results From the April Decennial Master Address File.” Count Review Memorandum Series 00-05. U.S. Census Bureau. October 23, 2000.

West, Kirsten K. and J. Gregory Robinson. 1999. “What Do We Know About The Undercount of Children? U.S. Census Bureau, Population Division, Working Paper No. 39. August 1999.

7. APPENDIX

Appendix 1

Production of Housing Unit Estimates

The Population Estimates Branch in the Population Division produces the housing unit estimates in the following steps:

1. Obtain housing unit count from the 1990 Census.

2. Estimate residential construction from building permits compiled from internal data files developed by Manufacturing and Construction Statistics Division (MCSD).

3. Estimate non-permitted residential construction from the following procedure:

    a) Calculate the number of 1990 county housing units located in non-permit issuing jurisdictions within the county by subtracting the number of units located in permit issuing jurisdictions from the total units

    b) Sum the housing units located in non-permit issuing jurisdictions to the national level

c) Calculate the county's share of units located in jurisdictions not issuing permits by dividing the county's units by the nation's units -- (a)/(b).

d) Multiply the total number of units reported in the Survey of Construction as constructed without building permits by the county's share of units located in jurisdictions not issuing building permits.

4. Estimate new mobile home placements from shipment and production summary reports obtained from the National Conference of States on Building Codes and Standards. Ninety-eight percent of all mobile homes shipped to states are used for residential purposes. State mobile home information is distributed to counties based upon a county's proportion of the state's mobile homes as of the 1990 census.

5. Estimate housing loss from demolition permits from internal data files developed by MCSD. These files include imputed permits where a permit issuing jurisdiction did not report permit issuance for the entire year. No lag time is assumed for demolition permits. It should be noted that MCSD stopped collecting data on demolition permits in 1995. After 1995, all data on housing unit loss is estimated. Estimates of non-permitted housing loss are calculated from the county’s share of structures in region at risk of loss. The risk is based on the county’s share of the following types of structures: (1) Mobile homes and other; (2) Older units (pre-1939 construction); (3) Vacant for Seasonal and Recreational Use; (4) Boarded up units.

6. Adjust estimates. After the housing unit estimates are produced, several adjustments are made to ensure that the estimates are consistent with the county population estimates produced using the Tax Return method. County household estimates are produced using tax return data which are controlled to state household estimates. The state household estimates are based on formation rates from the Current Population Survey (CPS). Controlled county household estimates and county housing unit estimates are used to derive a vacancy rate. The county household population estimates from the Tax Return method and the controlled county household estimates are used to estimate persons per household (PPH). Adjustments are made to the housing estimates to ensure that these vacancy and PPH values fall within certain tolerances.

The use of Medicare data to assess coverage of the population 65 and older

We use Medicare data to assess coverage of the population 65 and older as follows:

$$\text{Ratio}_x = {}_{65+}P_x \,/\, {}_{65+}M_x \qquad (1)$$

where:

Ratio = Ratio of census population 65 and over to Medicare enrollment for the population(in 1990 or 2000)

${}_{65+}P_x$ = Census population 65 and over (in 1990 or 2000)

$_{65+}M_x$ = Count of the number of persons aged 65 and over enrolled in Medicare (in 1990 or 2000).

The ratios of the census population to Medicare enrollment in 1990 and 2000 are used to broadly assess change in coverage. If the ratio in 2000 is greater than the ratio in 1990 we infer an improvement in census coverage of the population 65 and older; if the ratio in 2000 is lower we infer a decline in coverage.

The ratios themselves cannot be used as direct measures of coverage because of known differences between the census and Medicare universes. First, no allowance is made for under enrollment in the Medicare files (estimated to be about 3 percent nationally). Second, the county of residence in the census could be different than that reported in the Medicare file (e.g., location of doctor's office address). As long as the under enrollment and residency reporting remain about the same in 1990 and 2000, the change in the ratios can be used as a rough indicator of change in coverage. Estimates of national under enrollment imply about the same relative under enrollment in 1990 and 2000.

The use of School Enrollment data to assess coverage of the population 7-14 years old

We utilize school enrollment data to infer the change in coverage as follows:

$$\text{Ratio}_x = {}_{7\text{-}14}P_x / SE_x \qquad (2)$$

where:

Ratio	=	Ratio of census population to school enrollment (in 1990 or 2000)
${}_{7\text{-}14}P_x$	=	Census population aged 7-14 (in 1990 or 2000)
SE_x	=	Count of the number of persons enrolled in grades 1 to 8 (in 1989-1990 or 1999-2000). The enrollment data refer to 1989-1990 (for comparison to the 1990 census) or 1999-2000 (for comparison to Census 2000). Both public and private enrollment data are used. Private enrollment data refer to 1988-1989 and 1997-1998 for comparison to the 1990 census and Census 2000, respectively.

If the ratio in 2000 is greater than the ratio in 1990, we infer an improvement in census coverage of the school-aged population; if the ratio in 2000 is lower we infer a decline in coverage. As with the discussion of the Medicare-based ratios, the ratios themselves cannot be used as direct measures of coverage because of known differences between the census and school enrollment universes. First, no allowance is made for children not enrolled in school (which would include those in institutions and schooled at home). Second, the county of residence in the census could be different than that reported in the school file (e.g., location of school's address). As long as the enrollment levels and school districts remain about the same in 1990 and 2000, the change in the ratios can be used as a rough indicator of change in coverage.

Independent demographic benchmarks for the population aged 0-9

The measurement of coverage for young children is singled out for three reasons:

a) Undercoverage is relatively high in these ages (West, 1999), and differentials by race are significant (mirroring that of the total population)

b) The development of subnational estimates for younger ages is more feasible than older ages because error in measuring net migration is reduced

c) The indicators can be produced for a series of censuses (e.g., 1960-2000), providing important historical measures of change in coverage at the subnational level.

To produce historical demographic indicators of population and coverage for States or counties, birth and death statistics are compiled from available data and net migration is estimated on the basis of changes in cohort size between successive censuses. The equations for the specific estimates from the 1990 census and Census 2000 are as follows:

$$ {}_{0\text{-}9}P'_{90} = B_{80\text{-}90} - D_{80\text{-}90} + M_{80\text{-}90} \qquad (3)$$

$$ {}_{0\text{-}9}P'_{00} = B_{90\text{-}00} - D_{90\text{-}00} + M_{90\text{-}00} \qquad (4)$$

where:

$_{0-9}P'_x$ = Demographic estimate for the population aged 0-9 in 1990 and 2000

$B_{x,x+10}$ = Births occurring in the intercensal period x to x+10 (1980-1990, 1990-2000)

$D_{x,x+10}$ = Deaths occurring to the birth cohort in the intercensal period x to x+10 (1990-2000). Deaths have been estimated with life table survival rates.

$M_{x,x+10}$ = Estimated net migration occurring in the intercensal period x to x+10 (1990-2000)

It should be noted that the actual calculations in equations 3-4 are carried out in single-year-of-age detail (0,1,....8,9).

The difference between the estimated population ($_{0-9}P'_x$) and the census count ($_{0-9}C_x$) is an indicator of the net census undercount and net undercount rate:

$$_{0-9}U_x = {}_{0-9}P_x - {}_{0-9}C_x \quad (5)$$

$$_{0-9}r_x = ({}_{0-9}U_x / {}_{0-9}P_x) * 100 \quad (6)$$

where:

U_x = indicator of net undercount of the population aged 0-9 in time x

${}_{0\text{-}9}r_x$ = indicator of net undercount rate

For the subnational demographic methodology for censuses 1990 and 2000, the net migration component (M) is estimated from the ratio of the census count for the population aged 10-14 in one census to the census count for ages 0-4 in the previous census. An adjustment is made for mortality and for relative change in coverage of the cohort between censuses. The equation is illustrated for the 1980-90 net migration estimate:

$$ {}_{10\text{-}14}MR_{80\text{-}90} = \frac{{}_{10-14}C_{90}}{{}_{0-4}C_{80} * SR_{80-90}} * \left(ru_{80-90}\right) \tag{7}$$

where:

${}_{10\text{-}14}MR_{80\text{-}90}$ = Ratio of the cohort aged 10-14 in 1990 to 0-4 in 1980

${}_{10\text{-}14}C_{90}$ & ${}_{0\text{-}4}C_{80}$ = Census count of the cohort aged 10-14 in 1990 and 0-4 in 1980

$SR_{80\text{-}90}$ = Survival rate from age 0-4 in 1980 to age 10-14 in 1990 (intercensal period)

$ru_{80\text{-}90}$ = Change in census coverage (1990 versus 1980 national coverage rates from DA are used) of the population aged 10-14 in 1990 and 0-4 in 1980

In application, if the ratio (MR) for age 10-14 in 2000 is greater than 1.0 we infer net in-migration and if the ratio is less than 1.0 we infer net out-migration. Migration "rates" and migration amounts (M's) for individual ages 0 to 9 are interpolated from the implied rate for age 10-14 (MR).

The migration estimates for subnational areas can be improved by using actual census data on net migration specific to the area (from the questions on State of birth and residence 5 years ago). For the estimates at the county level, historical trends can be supplemented with current administrative sources (e.g., school enrollments) to develop preliminary net migration estimates for the 0-9 population.

The subnational demographic indicators of coverage for the population aged 0-9 (as carried out in equations 3-4) are actually developed separately for Blacks and non-Blacks. Inconsistencies in the reporting of race affect the quality of the race estimates, however. The measures for Blacks still can provide crude indicators of racial differentials in coverage.

Although we focus on the periods since 1980 in this study, coverage measures for the population under age 10 since 1960 can be derived.

Appendix 2

Types of Enumeration Area and Distribution of Counties by Type of Enumeration Area

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

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 Universe of Blocks

The distribution of the counties by type of enumeration areas:

Counties Included in Analysis			
Type of Enumeration Area	Number of Counties	Type of Enumeration Area	Number of Counties
1	**148**	5	13
2	**818**	5+8	1
1+2	**1,499**	5+8+9	1
1+9	11	5+6	1
1+8	1		**16**
1+7	7		
1+7+8	1	4	6
1+5+7	1	3	26
1+5+6+9	1	3+5	1
1+2+9	117	1+3	14
1+2+8	17	1+3+5+8	1
1+2+8+9	1	1+2+4	1
1+2+7	19	1+2+3	22
1+2+7+9	3	1+2+3+9	2
1+2+6	57	1+2+3+8	1
1+2+6+9	3	1+2+3+7+9	1
1+2+6+7	1	1+2+3+6	4
1+2+5	82	1+2+3+5	13
1+2+5+9	10	1+2+3+5+8	5
1+2+5+8	17	1+2+3+5+7	1
1+2+5+7	1	1+2+3+5+7+8	1
1+2+5+7+8	1	1+2+3+5+6	3
1+2+5+6	10	1+2+3+5+8	3
1+2+5+6+9	2	2+3	55
1+2+5+6+8	2	2+3+9	1
1+2+5+6+7+9	1	2+3+7+9	1
2+9	14	2+3+6	1
2+7+8+9	1	2+3+5	26
2+6	8	2+3+5+8	1
2+5	67	2+3+5+6	2
2+5+9	2	2+3+4	1
2+5+6	2	2+4	8
	460		**201**
TOTAL	**2,925**	TOTAL	**217**

Appendix 3

Race Conversion Method from 31 Race Categories to 4 Race Categories

The 31 race categories of single and multiple race responses in Census 2000 were converted into 4 race categories using equal proportional assignment of race on aggregated data. Within each single year of age, sex, and Hispanic origin group, multiple race responses were assigned in equal proportions determined by the number of race groups reported. For example multiple race of two races were assigned one-half to each group, multiple race responses of three races were assigned one-third to each race group and so on (see Table 1). A rounding algorithm was used so that sum of the converted four race groups matches the population total before race conversion.

2000 Race Categories	White	Black	AIEA	
White alone	1.000			
Black alone		1.000		
American Indian and Alaska Native alone			1.000	
Asian alone				1.000
Native Hawaiian and Other Pacific Islander				1.000
White; Black or African American	0.500	0.500		
White; American Indian and Alaska Native	0.500		0.500	
White; Asian	0.500			0.500
White; Native Hawaiian and Other Pacific Islander	0.500			0.500
Black or African American; American Indian and Alaska Native		0.500	0.500	
Black or African American; Asian		0.500		0.500
Black or African American; Native Hawaiian and Other Pacific Islander		0.500		0.500
American Indian and Alaska Native; Asian			0.500	0.500
American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander			0.500	0.500
Asian; Native Hawaiian and Other Pacific Islander				1.000
White; Black or African American; American Indian and Alaska Native	0.334	0.333	0.333	
White; Black or African American; Asian	0.334	0.333		0.333
White; Black or African American; Native Hawaiian and Other Pacific Islander	0.334	0.333		0.333
White; American Indian and Alaska Native; Asian	0.334		0.333	0.333
White; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander	0.334		0.333	0.333
White; Asian; Native Hawaiian and Other Pacific Islander	0.334			0.666
Black or African American; American Indian and Alaska Native; Asian		0.334	0.333	0.333
Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander		0.334	0.333	0.333
Black or African American; Asian; Native Hawaiian and Other Pacific Islander		0.334		0.666
American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander			0.334	0.666
White; Black or African American; American Indian and Alaska Native; Asian	0.250	0.250	0.250	0.250
White; Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander	0.250	0.250	0.250	0.250
White; Black or African American; Asian; Native Hawaiian and Other Pacific Islander	0.250	0.250		0.500
White; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander	0.250		0.250	0.500
Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander		0.250	0.250	0.500
White; Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander	0.200	0.200	0.200	0.400
2000 Race Categories - AIAN = American Indian or Alaska Native, NHOPI = Native Hawaiian or other Pacific Islander				
1990 Race Categories - AIEA = American Indian, Eskimo or Aleut, API = Asian or Pacific Islander				