April 11, 2008

The Use of Population Estimates as Controls to the American Community Survey: An Evaluation

FINAL REPORT

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

VICTORIA A. VELKOFF
POPULATION DIVISION

THE USE OF POPULATION ESTIMATES AS CONTROLS TO THE AMERICAN COMMUNITY SURVEY: AN EVALUATION

Census Advisory Committee of Professional Associations
American Statistical Association/Population Association of America
April 11, 2008

Victoria A. Velkoff[1]
Assistant Chief for Population Estimates and Projections
Population Division
U.S. Census Bureau

Abstract:

The Population Estimates Program in the U.S. Census Bureau Population Division develops and disseminates the official estimates of the population of the United States. As part of this program, it produces annual estimates of the population of counties by age, sex, race, and Hispanic origin. These estimates are used to develop the population controls to the American Community Survey (ACS).

The Census Bureau employs population controls in its survey process both to reduce the variance and to reduce the bias caused by the differential coverage of population groups in the survey. Recently, members of the Census Advisory Committee of Professional Associations (CACPA) have questioned the use of these population controls in the ACS. More specifically, the members have asked about the accuracy of these population estimates by demographic characteristics at the county level. The accuracy of a set of population estimates is traditionally measured by comparing a set of population estimates prepared prior to a decennial census to the decennial census results. However, limitations such as differential coverage in censuses and changes in race categories need to be considered when evaluating the results of such comparisons.

To provide information about the accuracy of the population estimates by age, sex, race, and Hispanic origin, we compared a set of 1990-based population estimates by demographic characteristics to Census 2000 results. This paper presents the results of these comparisons for the population estimates for the nation, states, and weighting areas used in the Census 2000 Supplementary Survey (C2SS). The paper includes a discussion of the limitations of such an evaluation and outlines the issues that must be considered when using these results as a measure of the accuracy of the population estimates developed for the post-2000 decade.

[1] This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, or technical issues are those of the author and not necessarily those of the U.S. Census Bureau.

Questions for the Committee:

1. How can this evaluation of the 1990-based population estimates inform our future work both in terms of improving the population estimates and using the population estimates as survey controls?

2. What recommendations do you have for eliminating some of the limitations of the current evaluation?

3. What recommendations do you have for developing an evaluation plan for 2010?

Methodology:

1990-based population estimates

The 1990-based county population estimates by demographic characteristics were developed in a two-step procedure. First a set of state estimates by age, sex, race, and Hispanic origin were developed using a cohort-component technique. These estimates incorporated the most recent data with demographic detail at the state level on births, deaths, international migration and domestic migration. The county estimates by demographic detail were produced in the second step using a ratio method.

The detailed state estimates developed in step 1 along with the estimates of the total population of counties developed using the administrative record component of change technique served as the control totals for the detailed county estimates. The April 1, 1990 census counts for counties modified by age, sex, race, and Hispanic origin (MARS) were used as the starting point for each year of detailed county estimates produced. These county-level MARS data are raked to the state estimates developed in step 1 and the county total population estimates for a given year to produce the detailed county estimates for that year.

The process described above was used throughout the 1990s to develop the county population estimates by demographic detail through July 1, 1999. A small additional raking step was used to extend the estimates for July 1, 1999 to April 1, 2000. The July 1999 set of estimates was raked to a set of 1990-based national population estimates by age, sex, race, and Hispanic origin for April 1, 2000 and to a set of 1990-based estimates of the total population of counties for April 1, 2000. The 1990-based national population estimates and the 1990-based estimates of total population for counties had been previously developed as part of an earlier evaluation process.

Census 2000 population data in 1990 race categories

The 1990-based set of population estimates for counties by age, sex, race, and Hispanic origin was developed in the race categories used in the 1990 Census. The question on

race for Census 2000 was different from the one for the 1990 census in several ways. Most significantly, respondents were given the option of selecting one or more race categories to indicate their racial identities.[2] Because of these changes, the Census 2000 data on race are not directly comparable with data from the 1990 census or to race data in the 1990-based population estimates.

In an attempt to introduce some consistency in the race categories between 1990-based estimates and the Census 2000 data, algorithms developed by the National Center for Health Statistics were applied to the Census 2000 data to "bridge" the Census 2000 multiple-race population data to single-race categories used in the 1990 Census. Bridging has the most impact on the data for the American Indian and Alaska Native and the Asian and Pacific Islander populations. It has a small impact on the Black population and a negligible impact on the White population.[3]

Measures of Accuracy

For purposes of this analysis, we are defining measures of accuracy as the difference between the 1990-based population estimates for April 1, 2000 and the Census 2000 counts for the household population. The analysis was done for the estimates of the nation, states, and C2SS Weighting Areas. Although there is interest in the evaluations of county estimates, for these analyses, the C2SS weighting area was the smallest geographic level examined. This was done to provide consistency with the associated analysis being carried out on the impact of the population controls on the ACS.

Percent Difference: For each observation, we compute the percent difference between the 1990-based population estimate for April 1, 2000 and the April 1, 2000 census counts

$$PD = [(EHHP_{2000} - \text{Census } HHP_{2000})/\text{Census } HHP_{2000}] * 100$$

Where:

PD = Percent difference between the 1990-based population estimate for 2000 and the Census 2000 count
$EHHP_{2000}$ = 1990-based estimated household population for 2000
Census HHP_{2000} = Census 2000 count of household population

[2] Other changes included terminology and formatting changes, such as spelling out "American" instead of "Amer." for the American Indian and Alaska Native category and adding "Native" to the Hawaiian response category. In the layout of the Census 2000 questionnaire, the Asian response categories were alphabetized and grouped together, as were the Pacific Islander categories after the Native Hawaiian category. The three separate American Indian and Alaska Native identifiers in the 1990 census (i.e., Indian (Amer.), Eskimo, and Aleut) were combined into a single identifier in Census 2000. Also, American Indians and Alaska Natives could report more than one tribe.

[3] Ingram DD, Parker JD, Schenker N, Weed JA, Hamilton B, Arias E, Madans JH. United States Census 2000 population with bridged race categories. National Center for Health Statistics. Vital Health Stat 2(135). 2003.

Three Summary Measures of Accuracy

1. *Mean Algebraic Percent Error (MALPE)*
   The MALPE is simply the average of all the percent differences across all states or all weighting areas. It is a measure of mean bias and computed as follows:

   $MALPE = [Sum_{1 \ldots n} PD]/N$

2. *Mean Absolute Percent Error (MAPE)*
   The MAPE provides a summary measure of accuracy and differs from the MALPE in that the MAPE involves taking the absolute values of the percent differences.

A set of MALPEs and the MAPEs are developed using the state as the geographic level of analysis and the 607 C2SS weighting areas as the geographic level of analysis. The analysis is done for the following demographic characteristics:

Total population	
Sex	(Male, Female)
Age groups	(0-17, 18 to 34, 35 to 44, 45 to 64, 65 and over, 0 to 15, 16 to 64)
Race/Hispanic origin	(Hispanic, non-Hispanic White, non-Hispanic Black, non-Hispanic American Indian and Alaska Native, non-Hispanic Asian and Pacific Islander)

In addition to the univariate analyses, analysis was done for the multivariate characteristics of sex by age, and race/Hispanic origin. To eliminate the impact of small data cells in the multivariate analysis for weighting areas that included race/Hispanic origin, we imposed an additional set of criteria. The race/Hispanic origin sample count in a specific evaluation cell had to meet the threshold of at least 20 unweighted people in the C2SS and the race/Hispanic origin population group had to contain at least 10 percent of the total population of that weighting area.

For each geographic level and set of demographic characteristics, two MALPEs and MAPEs are developed – one set uses the numeric population count in each cell and a second set uses the proportion of the U.S. population in each cell.

In addition, for the evaluations using the C2SS weighting areas, additional analyses were done disaggregating the weighting area into 3 population size categories using the Census 2000 household population. The size categories are:

Greater than or equal to 750,000 total household population in 2000
250,000 to 749,999
Less than 250,000

3. Index of Dissimilarity

The index of dissimilarity is a measurement of the overall difference between two percentage distributions. It is calculated by adding together the absolute differences between the numbers in each pair of corresponding values and halving the total. The result shows the proportion of cases that would need to be reallocated in order to make the two distributions the same. The basic formula for the index of dissimilarity is:

$$ID = \frac{1}{2} \sum[(E_i/E) - (C_i/C)]$$

Where

ID = index of dissimilarity

E_i = the estimated population in 2000 for a specific demographic group in area i

E = the estimated population in 2000 for a specific demographic group in the nation

C_i = the Census 2000 population for a specific demographic group in area i

C = the Census 2000 population for a specific demographic group in the nation

Limitations:

Differential Coverage in the 1990 and 2000 Censuses

This analysis uses the Census 2000 results as the standard to measure the accuracy of the 1990-based population estimates. The 1990-based population estimates begin with the 1990 census as enumerated. The lack of consistency in coverage between the 1990 and 2000 census counts clouds this analysis. The 6.0 million difference between the 1990-based national household population estimate for 2000 and the Census 2000 count of the total household population of the United States is partially due to an undercoverage of about 4 million people in the 1990 census counts.[4]

Use of Census 2000 as Benchmark

To develop the annual population estimates, the process begins with the most recent census count and adds the estimated components of population change. The estimates for any given date represent the beginning census data and the cumulative estimate of the components of population change from the most recent census to the estimate date. Thus, the errors in the population estimates for one year out from a census are likely to be much smaller than the errors for 10 years out from the most recent census. The evaluations of estimates 10 years out from the last census date represent the "worse case scenario" for a set of population estimates.

Inconsistency in Race Groups

As noted above, the race categories used in the 1990-based population estimates differ from those used in Census 2000. Although an effort is made to introduce some consistency in the racial categories, there is no perfect algorithm to achieve consistency

[4]J. Gregory Robinson, ESCAP II: Demographic Analysis Results, October, 2001, Report 1 to Executive Steering Committee for A.C.E. Policy II.

nor is there a measure of the inconsistency. Thus the comparisons for race groups should be viewed with caution.

Improved Estimates Methodology

The methodology used post Census 2000 for population estimates differs substantially from that used during the 1990s. In other words, the population estimates, which are used currently as controls for the ACS, are based on a different methodology. The availability of new data sources has enabled the Population Estimates Program to adopt the cohort component approach to the development of the county estimates by demographic characteristics and to modify the estimation of domestic migration. Additionally, for the post-2000 period, the availability of annual data from the ACS has enabled the estimates area to modify its approach for estimating international migration. As pointed out, the intercensal estimates use the most recent census as it starting point and develops annual estimates of the components of population change. The estimates for post-2000 use Census 2000 as a starting point while the 1990-based estimates used the 1990 census as its starting point. Evaluations indicate that the measures of coverage are higher in Census 2000 compared to those in the 1990 census.

Results:

Table 1 provides the results of the national level evaluations for the household population. Given the known deficiency in the estimate of international migration and the undercount in the 1990 Census, it is not surprising that the largest differences are in the age group 18 to 34 and in the Hispanic origin populations. The rather large differences noted for the non-Hispanic American Indian, Eskimo, and Aleut and the non-Hispanic Asian and Pacific Islander populations must be viewed with the limitation noted for the inconsistencies in race groups between the 1990 and 2000 censuses. In contrast, note the relatively small differences for non-Hispanic Whites. This population is the least affected by the shift in census coverage and errors in the estimation of international migration. It is with this national perspective, that we can view the differences at lower geographic levels.

Table 2 provides the results of an evaluation of the state population estimates by demographic characteristics. The results by age and by sex indicate a relatively close agreement between the estimates and census results. A similar agreement is seen for the estimates for non-Hispanic Whites. For other race groups and Hispanic origin, the patterns in the results mirror those at the national level and should be viewed recognizing the limitations discussed above.

Table 3 provides the results of an evaluation of the population in C2SS weighting areas and Tables 4A, 4B, and 4C provide the same results disaggregated by the Census 2000 size of population in the C2SS weighting areas.[5] The average percent differences by age and by sex at this geographic level are also relatively small given the estimates are for a a

[5] The C2SS weighting areas are counties or aggregations of counties. This paper used all 607 weighting areas in the analysis.

date 10 years after the 1990 baseline. Although the differences are larger for Hispanic origin and race groups other than non-Hispanic Whites, it is difficult to draw conclusions about the accuracy of the estimates at this level given the limitations discussed above.

Next Steps:

The results presented here are the first steps in our evaluation of the 1990-based population estimates for counties by demographic characteristics. We began with states and C2SS weighting areas to be consistent with the analysis being done on the impact of the population controls in the ACS. Our next steps include an evaluation of the combined age, sex, race/Hispanic origin characteristics for weighting areas and a thorough analysis at the county level. Additional analysis will include an examination of the outliers to identify certain problem areas.

We will conduct the same analysis done is this paper but using an adjusted set of population estimates for 1990 that will compensate for some of the undercoverage of the 1990 census.

As we move forward, the next area we want to focus on is examining the uncontrolled ACS estimates and comparing them to the population estimates used as controls. We plan to do this comparison for the post-2000 data.

Table 1. Evaluation of the 1990-based Estimates of the Total Household Population of the United States by Demographic Characteristics Compared to Census 2000

Demographic Characteristic	1990-based Estimates	Census 2000	Difference (Estimate - Census)	Percent Difference
Total, All Ages	267,668,167	273,643,273	-5,975,106	-2.2
Age				
0 to 17 years	70,082,680	71,970,892	-1,888,212	-2.6
18 to 34 years	60,384,871	63,213,621	-2,828,750	-4.5
35 to 44 years	44,258,646	44,251,964	6,682	-
45 to 64 years	60,122,568	61,208,664	-1,086,096	-1.8
65 years and over	32,819,402	32,998,132	-178,730	-0.5
0 to 15 years	62,325,286	64,080,796	-1,755,510	-2.7
16 to 64 years	172,523,479	176,564,345	-4,040,866	-2.3
Sex				
Males	130,320,441	133,551,373	-3,230,932	-2.4
Females	137,347,726	140,091,900	-2,744,174	-2.0
Race and Hispanic Origin				
Hispanic (any race)	31,492,568	34,592,843	-3,100,275	-9.0
Non-Hispanic White	191,794,575	192,282,259	-487,684	-0.3
Non-Hispanic Black	32,044,743	33,381,417	-1,336,674	-4.0
Non-Hispanic American Indian, Eskimo, and Aleut	1,987,364	2,267,176	-279,812	-12.3
Non-Hispanic Asian and Pacific Islander	10,348,917	11,119,578	-770,661	-6.9

- Rounds to 0.0

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

Table 2. Evaluation of the 1990-based Estimates of State Population Estimates by Demographic Characteristics Compared to Census 2000

Demographic Characteristic	Population Level		Population Proportion		Index of Dissimilarity
	MALPE	MAPE	MALPE	MAPE	
Total, All Ages	-2.4	2.4	(X)	(X)	(X)
Age					
0 to 17 years	-2.8	3.0	-0.2	1.6	0.005
18 to 34 years	-4.1	4.9	0.4	3.9	0.017
35 to 44 years	-0.7	2.0	-0.7	2.0	0.009
45 to 64 years	-1.9	2.7	-0.1	2.3	0.009
65 years and over	-1.3	1.6	-0.8	1.4	0.005
0 to 15 years	-3.1	3.2	-0.4	1.5	0.005
16 to 64 years	-2.4	2.5	-0.1	1.8	0.008
Sex					
Males	-2.7	2.7	-0.3	1.3	0.006
Females	-2.2	2.2	-0.2	1.1	0.005
Race and Hispanic Origin					
Hispanic (any race)	-17.0	17.9	-8.9	13.2	0.033
Non-Hispanic White	-0.7	1.3	-0.4	1.2	0.005
Non-Hispanic Black	-8.7	9.0	-4.9	6.6	0.012
Non-Hispanic American Indian, Eskimo, and Aleut	-16.8	19.9	-5.1	14.5	0.001
Non-Hispanic Asian and Pacific Islander	-9.3	10.8	-2.5	5.7	0.020

(X) Not applicable

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

Table 3. Evaluation of the 1990-based Estimates of the Household Population in C2SS Weighting Areas by Demographic Characteristics Compared to Census 2000

Demographic Characteristic	Population Level		Population Proportion		Index of Dissimilarity
	MALPE	MAPE	MALPE	MAPE	
Total, All Ages	-1.8	2.5	0.4	2.1	0.010
Age					
0 to 17 years	-1.8	3.8	0.9	3.7	0.017
18 to 34 years	-2.7	5.6	1.8	5.6	0.028
35 to 44 years	-	4.0	-0.1	4.0	0.017
45 to 64 years	-2.1	4.5	-0.3	4.2	0.018
65 years and over	-1.2	2.9	-0.6	2.8	0.013
0 to 15 years	-1.9	4.2	0.9	4.0	0.019
16 to 64 years	-1.9	2.9	0.4	2.7	0.013
Sex					
Males	-2.0	2.7	0.4	2.4	0.012
Females	1.7	2.3	0.3	1.9	0.009
Race and Hispanic Origin					
Hispanic (any race)	-13.9	20.5	-5.5	19.8	0.061
Non-Hispanic White	-0.7	3.0	-0.4	3.0	0.014
Non-Hispanic Black	-9.0	18.6	-5.2	18.5	0.042
Non-Hispanic American Indian, Eskimo, and Aleut	-19.2	21.4	-7.8	16.6	0.056
Non-Hispanic Asian and Pacific Islander	-2.9	16.4	4.3	17.5	0.054

- Rounds to 0.0

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

Table 4A. Evaluation of the 1990-based Estimates of the Household Population in C2SS Weighting Areas by Demographic Characteristics Compared to Census 2000

Areas with Census 2000 Population Equal to or Greater than 750,000 (Number of areas = 84)

	Population Level		Population Proportion	
Demographic Characteristic	**MALPE**	**MAPE**	**MALPE**	**MAPE**
Total All Ages	-2.5	2.8	-0.4	2.0
Age				
0 to 17 years	-3.5	4.7	-0.9	3.7
18 to 34 years	-4.8	5.9	-0.3	4.9
35 to 44 years	0.6	4.9	0.6	4.9
45 to 64 years	-1.6	3.4	0.2	3.1
65 years and over	-0.1	2.1	0.4	2.0
0 to 15 years	-3.6	5.0	-0.9	4.0
16 to 64 years	-2.5	3.2	-0.2	2.7
Sex				
Males	-2.9	3.1	-0.5	2.4
Females	-2.2	2.5	-0.3	1.9
Race and Hispanic Origin				
Hispanic (any race)	-17.5	20.0	-9.3	18.0
Non-Hispanic White	0.3	3.9	0.6	4.0
Non-Hispanic Black	-6.7	11.0	-2.8	9.9
Non-Hispanic American Indian, Eskimo, and Aleut	-17.1	20.6	-5.5	14.2
Non-Hispanic Asian and Pacific Islander	-7.7	13.0	-0.8	11.7

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

Table 4B. Evaluation of the 1990-based Estimates of the Household Population by Demographic Characteristics for C2SS Weighting Areas Compared to Census 2000

Areas with Census 2000 Population 250,000 to 749,999 (Number of areas = 297)

	Population Level		Population Proportion	
Demographic Characteristic	**MALPE**	**MAPE**	**MALPE**	**MAPE**
Total All Ages	-1.9	2.4	0.3	1.9
Age				
0 to 17 years	-1.9	3.5	0.7	3.4
18 to 34 years	-3.3	5.5	1.2	5.3
35 to 44 years	-0.1	3.9	-0.1	3.9
45 to 64 years	-1.9	4.0	-0.2	3.8
65 years and over	-0.7	2.8	-0.2	2.7
0 to 15 years	-2.0	3.8	0.8	3.6
16 to 64 years	-2.1	2.9	0.2	2.5
Sex				
Males	-2.1	2.7	0.3	2.3
Females	-1.7	2.2	0.3	1.8
Race and Hispanic Origin				
Hispanic (any race)	-17.4	22.2	-9.2	20.3
Non-Hispanic White	-0.5	2.7	-0.2	2.7
Non-Hispanic Black	-7.4	14.8	-3.6	14.9
Non-Hispanic American Indian, Eskimo, and Aleut	-18.1	19.7	-6.6	14.2
Non-Hispanic Asian and Pacific Islander	-5.1	14.4	2.0	14.7

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

Table 4C. Evaluation of the 1990-based Estimates of the Household Population by Demographic Characteristics for C2SS Weighting Areas Compared to Census 2000

Areas with Census 2000 Population Less 250,000 (Number of areas = 226)

	Population Level		Population Proportion	
Demographic Characteristic	**MALPE**	**MAPE**	**MALPE**	**MAPE**
Total All Ages	-1.5	2.4	0.7	2.3
Age				
0 to 17 years	-0.9	3.9	1.7	4.1
18 to 34 years	-1.2	5.5	3.5	6.3
35 to 44 years	-0.2	3.6	-0.2	3.6
45 to 64 years	-2.4	5.4	-0.6	5.2
65 years and over	-2.1	3.4	-1.6	3.1
0 to 15 years	-1.1	4.3	1.7	4.4
16 to 64 years	-1.5	2.9	0.4	2.7
Sex				
Males	-1.5	2.6	1.0	2.6
Females	-1.5	2.4	0.5	2.2
Race and Hispanic Origin				
Hispanic (any race)	-8.1	18.6	0.9	19.9
Non-Hispanic White	-1.3	3.0	-1.1	2.9
Non-Hispanic Black	-12.0	26.5	-8.3	26.4
Non-Hispanic American Indian, Eskimo, and Aleut	-21.4	24.1	-10.3	20.6
Non-Hispanic Asian and Pacific Islander	1.7	20.4	9.3	23.3

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