AN EVALUATION OF THE 2000 CENSUS

Professor Eugene Ericksen

Temple University, Department of Sociology and Statistics

Introduction

Census 2000 has been marked by controversy and debate, both political and academic. On one side of the debate are those who favor the use of sampling and statistical models to adjust for the inevitable undercount of the population. Opposed are those who believe that any estimation will create more error than it removes, and that the goal of the census should be to reduce the undercount with better procedures.

Census 2000 was unprecedented in terms of its budget – more money was spent on it than on any previous census. Its achievements are impressive. Not only did the Census Bureau reduce the net undercount below the levels of previous censuses, but it also reduced the differentials between the non-Hispanic White and minority undercounts.

As a statistical consultant to the Presidential Members of the Census Monitoring Board, I have enjoyed the opportunity to observe the workings of Census 2000 and to analyze its results. In this report, my goal is to evaluate the accuracy of the census. To do this, I must first present criteria for defining accuracy, and discuss the merits of various alternatives. I will then provide my own analyses and draw conclusions.

As with many things that are large and complicated, the answer to the question of whether Census 2000 was more accurate than its predecessors depends on the criteria you use to judge. Census 2000 appears to have been successful by the key standard of allocating the population among states, counties, and places. It did not accomplish this feat, however, by observing and counting a greater share of the population than in 1990.

Goals of the Census

The Constitutional goal of the census is to allocate the population among states. This goal conceivably could be attained without a complete count of persons. For example, if every housing unit had two people living in it, we could simply count the dwellings to get a good estimate of the population distribution. This fanciful thought underlies a more serious point. When the census misses 5 percent of minority populations, but less than 1 percent of non-Hispanic Whites, its results are biased against the minorities.[1] It is better to have a consistent undercount of 3 percent for all groups than undercounts of 3 percent for minorities and 0.3 percent for Whites. The latter census is less fair, even though the net undercount of the national population is lower than it would be for the alternative. The Constitutional goal of the census is to fairly allocate the population distribution among local areas. The differential, rather than the national net undercount, is the key statistic.

[1] In 1990, the Bureau obtained the following estimates of undercount: American Indians – 12.2 percent; Hispanics – 5.0 percent; Blacks – 4.6 percent; Asians and Pacific Islanders – 2.4 percent; non-Hispanic Whites and others – 0.7 percent.

A second goal of the census is to describe the nature of the population. We need to know the racial and ethnic composition of the population, as well as observe the distributions of age, gender, income, and country of birth. For this, we need to obtain information directly from individuals, counting each person once, and only once.

Many of the methods used in Census 2000, such as whole person imputation, identifying duplicate cases by computer and then deleting them, and improving the address register with local information help to achieve the first goal. Some of these methods, however, detract from the second goal.

People Counted Directly and Correctly

In 1990, the official census count was 248.7 million, the adjusted estimate was 252.7 million, and the implied net national undercount was 4.0 million. The comparable results for Census 2000 are 281.4 million counted, 284.7 million estimated to exist, and an implied net undercount of 3.3 million. The net undercount was smaller in 2000 by 700,000 people, and in percentage terms the rate dropped from 1.6 to 1.2 percent.[2]

The Census Bureau estimated that 4.4 million of the counted people in the 1990 Census were "erroneous enumerations," people who were fabricated by enumerators or counted twice at the same location.[3] They also stated that 2.2 million were "whole person imputations," or people created by a computer program rather than counted directly.[4] Combining these two groups and dividing by the total count, we see that 2.65 percent of the official population was not counted directly and correctly. If we subtract the 6.6 million imputations and erroneous enumerations from the official count, we have 242.1 million people counted directly and correctly in 1990. Subtracting this figure from the estimated total, we estimate that (252.7 – 242.1 =) 10.6 million were omitted, or not counted directly. They comprise 4.19 percent of the estimated total.

For Census 2000, the Bureau informs us that there were 3.1 million erroneous enumerations and 5.7 million imputations for a total of 8.8 million.[5] They comprise 3.13 percent of the official count, an increase over 1990. There were 272.6 million people directly and correctly counted, and (284.7 – 272.6 =) 12.1 million omissions. They comprise 4.25 percent of the estimated total, and this percentage is about the same as obtained for 1990.

There was a shift in the nature of the undercount, though. In 1990, of the estimated 10.6 million persons not directly counted, 8.4 million were omissions and 2.2 million were imputations. In 2000, of the estimated 12.1 million persons not directly counted, 6.4 million were omissions and 5.7 million were imputations.

Omissions and Erroneous Enumerations

The Census Bureau's definition of erroneous enumerations is controversial, as it omits a substantial category of people counted at the wrong location. For example, if a person moved from New York to California on April 15, 2000, but was counted in California, (s) he would create two errors. New York would have one person too few and California would have one person too many. Because such an

2 *Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy (ESCAP)*, March 1, 2001.

3 U. S. Bureau of the Census, Memorandum from Ruth Ann Killion to John Thompson, dated May 15, 1997, "Gross Errors and Erroneous Enumerations in the 1990 Decennial Census."

4 John H. Thompson, "Census 2000 – Decision on Release of Statistically Corrected Redistricting Data," August 8, 2001 at the Joint Statistical Meetings of the American Statistical Association, Atlanta, Georgia.

5 Letter from William Barron to Representative Carolyn Maloney, dated April 9, 2001.

error does not affect the national net undercount, the Census Bureau does not include it in its definition of erroneous enumeration. I believe that such an error should be counted as an omission in New York and an erroneous enumeration in California.

Moreover, had the person been counted in both states, creating only one error, the Census Bureau would still not consider it to be an erroneous enumeration. This seems especially incorrect to me. Other examples of persons counted at locations other than their main residence, and not considered erroneous by the Bureau are:

- College students living away from home, counted both at home and school,
- Families counted at their main and vacation homes, and
- One or both members of a commuter marriage counted at both addresses.

Adding these incorrectly located people to the count of erroneous enumerations increases the Census Bureau estimate considerably. In 1990, 1.8 percent of the "E – sample" was a fictitious or duplicate case; 2.2 percent of the sample was counted at the wrong location.[6] Adding these in, the total estimate of people erroneously enumerated rose from 4.4 to 9.9 million. The estimated total of imputations and erroneous enumerations is 12.1 million, 4.87 percent of the official count.

For Census 2000, the parallel calculation provides an estimate of 6.3 million erroneous enumerations by the broader definition.[7] Adding these to the 5.7 million imputations gives a total of 12.0 million, 4.26 percent of the official count. The percentage of persons not directly counted is lower than in 1990, but the number of such cases is substantial in both censuses.

The Census Bureau estimates the number of omissions in the Census as the sum of erroneous enumerations and the net undercount.[8] In 1990, the net undercount was 4.0 million, and by the Bureau's definition 4.4 million were erroneously enumerated and 8.4 million were omitted. The comparable total for Census 2000 is 6.4 million omissions.

The gross error is defined to be the sum of omissions and erroneous enumerations. By the Bureau's calculations, this quantity fell from (8.4 + 4.4 =) 12.8 million in 1990 to (6.4 + 3.1 =) 9.5 million in 2000. By the expanded definition, there were 13.9 million omissions in 1990 for a gross error of 23.8 million.[9] There were 9.6 million omissions in 2000 for a gross error of 15.9 million.[10] Making no allowance for whole person imputations, the decline from 23.8 to 15.9 million indicates substantial improvement in Census 2000, relative to 1990. Even so, the number of omissions is very large.

It can be argued that each computer imputation represents one omitted person whom the Bureau could not directly count so the computer created his/her record. Adding these omissions to the previous totals, we obtain gross error estimates of 25.0 million in 1990 and 21.6 million in 2000.[11]

---

6 The "E-sample" is the sample of census records that the Bureau matched against the A.C.E. survey data to determine the percentages correctly and erroneously counted. The "P-sample" is the survey sample which is matched against the census records to determine the percentages included and omitted from the count.

7 Obtained from the Census Bureau data file entitled "E-Sample Person Dual System Estimation Output File," delivered to the Census Monitoring Board on February 16, 2001.

8 In making these calculations, the Census Bureau did not estimate the number of omissions directly. Noting that the net undercount is the difference between omissions and erroneous enumerations, it calculated the number of omissions as the sum of the net undercount and erroneous enumerations, e.g., for 1990, 4.4 + 4.0 = 8.4 million.

9 We have 9.9 million erroneous enumerations, 4.0 million net undercount, and 13.9 million omissions for a total gross error of 23.8 million.

10 We have 6.3 million erroneous enumerations, 3.3 million net undercount, and 9.6 million omissions for a total gross error of 15.9 million.

11 For 1990, the number of omissions is 16.1 million, which when added to 9.9 erroneous enumerations provides a total of 25.0 million; for 2000 the comparable sum is 15.3 + 6.3 = 21.6 million.

To summarize, the Bureau's achieved reduction in the net national and differential undercounts did not necessarily occur because they "counted" many more people directly and correctly. Depending on how one defines erroneous enumerations, omissions, and the gross error, the Bureau either did about as well in Census 2000 as it did in 1990, or moderately better in 2000. Even by the definitions most favorable to the Census Bureau, however, there was a substantial amount of indirect and erroneous counting in 2000.

Geographic Considerations

If an omission and erroneous enumeration occur on the same block, but to two different people, they cancel each other out at all meaningful levels of geography. To evaluate the effect of errors on population distributions, we need estimates of net undercount for each block. For example, if one block had 100 counts, 10 omissions and 5 erroneous enumerations, the adjustment would add 5 people even though the gross error was 15. An adjoining block might have 80 counts, 1 omission and 9 erroneous enumerations, and the adjustment would subtract 8 people even though the gross error was 10. The key statistic would be (5 + 8 =) 13 "changes."

In 1990, when the Census Bureau calculated adjustments to individual blocks, it added 5.45 million people and subtracted 1.46 million people for a total of 6.91 million changes.[12] This statistic is much smaller than the previously calculated estimates of gross error for two reasons: (1) many errors cancel out because they occur on the same blocks, and (2) the Bureau's adjustment procedure does not fully correct for the distribution of net errors across all blocks. I illustrate the point with two groups of blocks included in the 1990 Post Enumeration Survey.

The 1990 survey, as did the 2000 Accuracy and Coverage Evaluation (A.C.E.) survey, sampled entire blocks and then calculated direct dual systems estimates of each block's population. In the examples just given previously, the survey data would indicate census counts of 100 and 80 respectively, along with dual systems estimates of 105 and 72. In a perfect world, the adjusted estimates for our two blocks would equal the direct estimates, i.e., 105 and 72. The synthetic adjustment method used by the Bureau, since it could not "explain" all the variation in net undercount rates across sample blocks, considerably understated the block-level adjustments.[13]

In Table 1, I illustrate the point using estimates for 11 sample blocks in Manhattan and 8 sample blocks in Ulster County, New York. I present the percentage undercount as estimated by the direct and synthetic dual systems estimates for each block. For both Manhattan and Ulster, the direct estimates are more highly variable than are the synthetic estimates. This is demonstrated by the larger standard deviations for the direct (13.88 for Manhattan and 5.89 for Ulster) than the synthetic (4.81 for Manhattan and 1.28 for Ulster) estimates. More to the point, the synthetic estimate is usually between zero and the value of the direct estimate. For example, Block 1 in Manhattan has a direct estimate of –24.20 percent and a synthetic estimate of –4.93 percent while Block 11 has estimates of 21.49 and 7.74 percent respectively.

The synthetic adjustments are therefore smaller in absolute value than the direct estimates would be if they were available for all blocks. For Manhattan and Ulster combined, the ratio of the average direct to synthetic adjustment is about three.

---

[12] Howard Hogan, "The 1990 Post-Enumeration Survey: Operations and Results," *Journal of the American Statistical Association*, September 1993, p.1054.

[13] This is because the variables used to define post-strata in both 1990 and 2000 predict patterns of undercount in a general way for large aggregates. Different variables would be needed to predict which particular blocks in a neighborhood would have larger and smaller undercounts or overcounts. This point does not indicate errors on the part of Bureau; it merely points out the inevitable limitations of any adjustment model.

I repeated this calculation for the entire nation, and found that on average, the direct adjustment was 2.3 times larger than the synthetic adjustment. Therefore, if the synthetic adjustment created 6.91 million changes, as indicated above, the number of changes that needed to be made was larger, i.e., 2.3 * 6.91 = 15.9 million.

To explain it another way, had it been possible to sample all blocks in the United States, and calculate dual systems estimates for each one, I estimate that there would be 12.5 million additions and 3.4 million deletions to be made. Because the factors included in the adjustment model cannot fully predict the block-to-block variation in net undercount, the adjustments actually calculated only account for a share, about 43 percent, of the adjustments that need to be made. The Bureau's adjustments improve the estimated distribution of population, but not perfectly.[14] Moreover, the 15.9, rather than the 6.91 million, better indicate the extent of the undercount.

In Census 2000, the Bureau added 4.26 million and subtracted 1.00 million for a total of 5.26 million changes.[15] If we assume that the factor of 2.3 is appropriate for Census 2000, then the estimated number of changes that needed to be made would be 2.3 * 5.26 = 12.1 million. Looking at it another way, I estimate that there need to be 9.8 million additions and 2.3 million deletions across all blocks.

To summarize, when we define the gross error geographically, we see substantial progress in Census 2000, by (15.9 – 12.1 =) 3.8 million. To the extent, however, that the multiplier of 2.3 is too low, we should revise the estimated number of changes for 2000 upward from 12.1 million, and the actual improvement over 1990 would be smaller.[16]

Patterns of Undercount

A major story of Census 2000 is the reduction in the differential undercount (see Table 2). Both the Hispanic and the non-Hispanic Black undercounts in Census 2000 are about half of what they were in 1990, the non-Hispanic White undercount remained constant, and the differentials were cut sharply. This improvement in the estimated allocation of population among demographic groups reinforces the apparent reduction in gross error just discussed.

In designing the survey and estimation procedure for Census 2000, the Bureau defined post-strata not only by race and Hispanic origin, but also by tenure, metropolitan status, region (for White owners) and the mail return rate. Only the first three of these factors had a consistent effect on the estimated net undercount (Table 3).

Looking first at 16 groups of non-Hispanic White owners defined by geographic location, the rates cluster around zero. Six of the estimates are overcounts, seven of them are between 0 and 0.99 percent, and the remaining three are between 1.00 and 1.99 percent. Rates of undercount are slightly higher in non-metropolitan areas than elsewhere.

Rates of undercount are somewhat higher for White non-owners than owners, but the differences are not large. We also observe this pattern for other racial groups. For Hispanics, the undercounts were higher in smaller and non-metropolitan areas, while the opposite was true for the non-Hispanic Blacks. Indeed, the rates of net undercount for non-Hispanic Blacks and Whites living in smaller and

[14] The "imperfections" of the Bureau's method are likely to mean that remaining errors, after adjustment, exist at the block level. These block level errors largely cancel out within census tracts and legislative districts. The remaining errors for larger areas would be smaller on a percentage basis than they are for blocks.

[15] U. S. Census Bureau press release, "Statement by William G. Barron, Jr. on the Current Status of Results of Census 2000 Accuracy and Coverage Evaluation Survey," dated July 13, 2001.

[16] The associations between the variables used to define the post-strata and the pattern of net undercount appear to be weak in 2000, suggesting that the factor of 2.3 might be too low, and the estimate of 12.1 million therefore biased downward.

non-metropolitan areas were similar to each other. Finally, for non-Hispanic Asians, Pacific Islanders, and American Indians the rates of undercount were generally higher than average but substantially below their comparable estimates for 1990.

In general, the rates of undercount among different post-strata within the same racial category are not greatly different from each other, the one exception being owners versus non-owners. As different states typically have similar percentages of owner-occupied housing units, we would not expect to see large variations in undercount rates, once race has been taken into account. For smaller areas such as counties and places, concentrations of owner or renter occupied housing may have a more substantial impact.

Rates of Net Undercount for States

Relying on data provided by the Census Bureau, I have replicated their estimates of net undercount by state (Table 4). I have also calculated "synthetic" estimates of undercount, or estimates of undercount that you would get if you assumed that the national rates of undercount applied to each state. For example, if a state's population included 20 percent Hispanics, 30 percent non-Hispanic Blacks, and 50 percent non-Hispanic Whites, its synthetic estimate would be

$$.20 * 2.85 + .3 * 2.17 + .5 * 0.67 = 1.56 \text{ percent.}$$

Comparing the actual and synthetic estimates in Table 4, we see that they are quite similar. The estimates differ by more than a percentage point in only one state, Alaska. They differ by more than one-half of one percentage point in only eight states.

The synthetic and A.C.E. estimates for states are close because states are large and diverse areas including rich and poor, city, suburban and rural, and owner and non-owner areas. Moreover, the A.C.E. estimates themselves do not vary greatly, as the range extends only from 0.29 percent (Minnesota) to 2.67 percent (Alaska).

Variability on tenure and other indicators is greater for local areas within states. The Census Bureau defined minority post-strata by putting large and medium metro areas into one group, and smaller and non-metro areas into another group. This makes it possible to compare, within states, the minority-White differentials in more and less metropolitan districts. For example, in Georgia, is the Black-White differential in cities like Atlanta similar to the Black-White differential in more rural areas? In Tables 5 and 6, I present comparisons of undercount differentials, first for non-Hispanic Blacks and Whites and second for Hispanics and non-Hispanic Whites within more and less metropolitan areas in the same state.

To illustrate the method, we see in Alabama, that the Black – White differential for large and medium metro areas was (2.50 – 0.66 =) 1.84 percent. The corresponding differential for smaller and non-metro areas was (0.94 – 1.09 =) –0.15 percent, indicating a greater racial disparity in urban areas like Birmingham and Mobile than elsewhere.

Looking at the Black-White differentials in different states, they are consistently between 1.5 and 3 percent in the large/medium category. In these more metropolitan areas, the racial differential is consistently in the direction that we would expect from past censuses – the Black undercount is higher than the White. The story changes in the small/non-metro category, where there is no consistent difference in Black and White undercount rates, and they are usually close together. Indeed, the Black rate is lower than the White rate, though not by very much, in the rural and small city areas of 14 of the 34 states. This result has important implications for the South, as 85 percent of the non-Hispanic Black popula-

tion living in smaller and non-metropolitan areas is located there. There appears to have been no racially differential undercount in the less metropolitan South. While the Census Bureau has not offered any explanation for this result, my hypothesis would focus on the relative difficulties of building complete address lists, without duplications, in minority compared to White areas. I suspect that good address lists are most difficult to build in minority areas of large cities.

The pattern differs for the Hispanic – White comparison. As we would expect from past censuses, Hispanic undercounts are consistently higher, by a few percentage points, than non-Hispanic White undercounts. Both groups have higher undercounts in smaller and non-metropolitan areas, and the differential between Hispanics and Whites is somewhat larger there as well. As a result, in many areas of the West and Southwest the non-metropolitan rates of undercount are higher than those of large cities.

Undercount Rates for Counties

The preceding discussion suggests that the variation in undercount rates among counties may not be substantial. The racial differentials that we have just observed are typically less than three percentage points and are sometimes much less than this amount. To study this question, I calculated undercount rate estimates for approximately 1,500 counties located in 23 states. The states were selected by one or both of two criteria: (a) at least 25 percent of the population was something other than non-Hispanic White, or (b) it had a substantial share of its population located in large metropolitan areas as defined by the Census 2000 post-stratification plan. States with large minority, or "big city" populations are more likely to have variable rates of undercount among counties than the remaining less metropolitan states with smaller minority populations.

To calculate the undercount estimates, I first divided the non-group quarters populations of each county into eight categories – owners and non-owners among Hispanics, non-Hispanic Blacks, non-Hispanic Whites, and non-Hispanic others. I then obtained the ratios of adjusted to official populations for each state as provided by the Census Bureau. I display the individual county estimates in Appendix A and summarize the results in Table 7.

The variation in county rates of undercount is not substantial. A full 87 percent of all counties studied have rates of undercount between 0.00 and 1.99 percent. There are only four counties, all with small populations, that have rates of undercount above three percent. Six percent of counties, generally located in the Midwest, have overcounts but none of these is greater than one percentage point. As a general pattern, county rates of undercount are higher in the West and lower in the Midwest. The lack of a Black – White differential in less metropolitan areas had a substantial role in minimizing the variation among counties located in Southern states.

Effects of the Reduction in the Rate of Net Undercount

The ability of the Census Bureau to reduce both the national net and the racially differential undercounts is a major success story. Having counted a greater share of the minority population in Census 2000, we would expect to see the greatest improvement in areas with large minority populations. This could make it difficult, though, to know how much of the measured 1990 – 2000 population growth was real and how much was due to a reduced undercount.

I conducted a study of those counties located in large metropolitan areas, as defined by the A.C.E. post-stratification scheme. I selected these counties, whose collective Census 2000 official count is 86 million, because I believed these counties to be the ones where census-taking problems were most serious

in 1990. I sorted them into four groups defined by the percentage minority, as (a) 50 percent or more, (b) 25 to 49.9 percent, (c) 10 to 24.9 percent, and (d) less than 10 percent minority.[17]

I then obtained 1998 and 1999 population estimates from the Census Bureau website.[18] Because these estimates did not incorporate an adjustment for the undercount of the 1990 Census, they provided a good benchmark to evaluate Census 2000. In other words, the 1999 estimate added the 1990 – 99 growth to the unadjusted 1990 count. I calculated a "2000 Census Projection" by adding the 1998 – 99 change to the 1999 estimate. For example, if the 1998 estimate was 180,000 and the 1999 estimate was 185,000, I calculated the 2000 projection to be 190,000. This projection estimates what the Census 2000 count would have been had the level of undercount been the same. It is subject to the errors generally associated with population estimates, but there is no reason to expect these errors to be consistently positive or consistently negative.

I compared the projection to the official 2000 count. If the count was 200,000, and the projection 190,000, then the projection was short by 10,000 or 5 percentage points. I summed the relevant population counts and projections by category, and calculated the overall shortfalls by county group (Table 8).

Together, the counties with the largest minority populations had the largest shortfall. They were projected to grow by 5.24 percent, but actually "grew" by 9.25 percent. The shortfall of the population projection was 3.67 percentage points, and this shortfall may well reflect the effects of improved counting. It is larger than the shortfalls of 1.34, 1.10, and –0.55 percent found for the other three groups of counties. In other words, those counties with higher minority shares had larger shortfalls than did counties with lower shares.

New York City is an especially good example, as its projected growth rate was 1.73 percent compared to actual growth of 9.36 percent. Washington DC, Philadelphia, and Hudson County, New Jersey had similarly high and unexpected amounts of growth. The shortfall was positive in 15 of the 16 "high minority" counties. It was three percentage points or more in 10 of the 16 counties. While there is variation, we see consistently high and unexpected growth in urban areas with large minority populations. Some of this growth is undoubtedly due to improved counting.

It is tempting to believe that this improvement is due to the use of the Local Update of Census Addresses (LUCA) program. In this program, local governments were allowed to submit lists of addresses that they believed might not have been included in the master address file of the census. In all, the Bureau added just over 4 million addresses through LUCA.[19] One of the largest files of added addresses came from New York City, where the Census Bureau accepted over 280,000 added addresses. These comprised 8.88 percent of the eventual city total of 3.2 million housing units. LUCA's contribution compares to the 6.98 percent "shortfall" in New York City, suggesting that LUCA played no small part in creating a large amount of measured growth.

There is, however, no consistent pattern in other cities, some of whom such as Washington DC and Suffolk County, Massachusetts had large unexpected growth and a small LUCA contribution. Other counties had the opposite experience, i.e., they made a big LUCA contribution but did not observe unexpected growth. The overall correlation between the ratios of LUCA adds to all housing units and the shortfall of the population projection, measured among counties, is -.014.

[17] Minority is here defined as persons who did not self-identify as only one race, non-Hispanic White.

[18] See U. S. Bureau of the Census website, July 30, 2001

[19] Data file received July 13, 2001 by the Census Monitoring Board, Presidential Members, from U. S. Bureau of the Census, entitled "LUCA submissions and 'adds' by local government jurisdiction."

The Nature of Census Error

My emphasis on the improvement in Census 2000 in terms of the national net and differential undercounts is not intended to minimize the importance of remaining error, both for political representation and fund allocation. The Census Bureau decided not to adjust the results of Census 2000 on its Redistricting file, but continues to consider whether or not it should adjust these results for use in fund allocation and other purposes. In this section, I discuss some of the issues associated with that decision.

Perhaps the major reason for the Bureau's decision was the inconsistency between national population totals calculated by the A.C.E. survey and demographic analysis. As the Bureau put it in the March 1 ESCAP report,

> "Initial D[emographic] A[nalysis] results, however, presented a major inconsistency with the A.C.E. results – instead of confirming a net undercount, DA estimates that Census 2000 overcounted the national population by 1.8 million individuals . . .. substantially below the net undercount of 3.3 million shown by the A.C.E. (page 3)."

The most likely culprit, from the perspective of the A.C.E., is the underestimation of erroneous enumerations. In other words, the official count includes more duplications, fabrications, and persons counted in the wrong place than the A.C.E. indicated. Increasing the estimated number of erroneous enumerations would reduce the net undercount, but also increase the gross error and indicate that the quality of Census 2000 data was not as good as we originally thought.

There are good logical reasons to believe that the Census Bureau did underestimate erroneous enumerations. There were 16 million counted people excluded from the A.C.E., 8 million who lived in group quarters, 5.7 million who were whole person imputations in households, and 2.3 million "late adds" who were cases originally thought to be duplicates but who were added back into the count at the end of the census counting period.[20]

The group quarters population could include overcounts, for example, among people included in outdated lists of residents at places such as hospitals, dormitories, and prisons. The number of "whole person" imputations, 5.7 million, may be too large, and to my knowledge the Bureau has never studied the question of whether its computers created on average the correct number of records for addresses where whole person imputation occurred. Finally, we already have reason to suspect that many of the 2.3 million "late census adds" were duplicated cases. The Bureau may be studying these possibilities, along with their announced studies of subjects such as balancing error. We await its conclusions.

In general, problems of census taking arise due to the circumstances in which people live. There are neighborhoods where poverty is high, education is low, use of foreign languages may be common, housing is crowded or irregular, and crime rates high where it is especially difficult to count. Even where some but not all of these conditions exist in extreme forms, census taking may still be difficult. These difficulties lead not only to higher rates of omission, but also to higher rates of erroneous enumeration, whole person imputation, and records with incomplete and incorrect recording of characteristics such as race and Hispanic origin.

The focus on the racially differential undercount sometimes leads to a misplaced emphasis on racial identity itself, rather than the conditions in which many minority group members live, as an explanation for why the undercount exists. Just as we would expect counting for non-Hispanic Whites to be difficult when their living circumstances are difficult, we would expect the counting for Hispanics and non-Hispanic Blacks to be easier when their conditions were better.

---

[20] John H. Thompson, "Census 2000 – Decision on Release of Statistically Corrected Redistricting Data," August 8, 2001 at the Joint Statistical Meetings of the American Statistical Association, Atlanta, Georgia.

As part of its planning for Census 2000, the Bureau created a census tract planning file including information on the demographic and economic circumstances of local populations and tract level estimates of the undercount.[21] This file afforded me the opportunity to study the effects of poverty, as it combines with race, on the undercount. Because the poverty information is based on long-form data, it was not available for use in calculating actual adjustments to either of the 1990 or 2000 Censuses.

Working with the 1990 census tract data, I created five categories defined by race. One included those areas with Native American majorities. The second included that majority of tracts where the percentages non-Hispanic Black and Hispanic were each below 10 percent. I then identified tracts (a) where each percentage was between 10 and 29.9, (b) where one or both were between 30 and 49.9 but neither was as high as 50, and (c) where there was either a Black or Hispanic majority. I then subdivided the tracts a second way, depending on the poverty rate. My cross-classification made it possible to compare high and low poverty tracts where the racial composition was similar, as well as tracts of different racial composition where the poverty level was similar.

Table 9 shows that higher rates of poverty are associated with higher rates of undercount. Areas where the poverty rate is low and the population predominantly non-Hispanic White, have very low rates of undercount, 0.4 percent. Increases in this rate are associated both with increased percentages of poverty and racial minorities. It should be noted that the rate of undercount for predominantly White areas with a poverty rate over 50 percent is 3.6 percent, higher than the corresponding rate, 2.9 percent, for areas with concentrated minority populations but a low poverty rate. Many of the variables actually used to define post-strata, such as tenure and the mail return rate, are attempts to create proxies for the difficult counting conditions created in part by poverty. It is important, though, not to consider these proxies to be the same as the conceptual variables that best explain the variation in rates of undercount, but for which no data are available.

In the next step of my analysis, I attempted, for Census 2000, to demonstrate the manner in which the various forms of census error congregate in similar locations. I compared rates of omission, erroneous enumeration, and imputation for groups of post-strata (see Table 10) defined by the key proxy predictors – race, Hispanic origin, tenure, and metropolitan status.

There we see that those post-strata with higher rates of net undercount, generally those with minority non-owner populations, also have higher rates of non-matching,[22] erroneous enumeration, and imputation. Indeed the correlations between the net undercount and these three variables are, respectively, .88, .51, and .67. The correlation between non-match and erroneous enumeration rates is .80. In sum, conditions of poverty create difficult counting of all types. It is theoretically possible to imagine that the Bureau might solve the problem of differential undercount by increasing rates of erroneous enumeration and imputation in poor neighborhoods. This would offset the higher rates of omission and reduce the differential undercount. It would not mean, however, that a greater proportion of people were counted directly and correctly.

[21] This file is called the Planning Database "CD-DSSD-comm-7 1990 data for Census 2000"and was delivered to the Census Monitoring Board on November 14, 2000.

[22] The Census Bureau uses the term "non-match" to refer to persons in the P-sample whose record could not be found in the census.

The Limits of Improved Counting

The budget for Census 2000 is $6.55 billion, a large increase over the $2.6 billion budget for the 1990 Census even after inflation is taken into account.[23] Using constant fiscal year 2000 dollars, the "per household" cost rose from $36 in 1990 to $62 in 2000. There is little doubt that the added spending improved census data quality. At the same time, there are important types of census error that are impervious to budget size.

Better address lists are expensive to create, but they improve the count. While I was unable to demonstrate a direct link between LUCA investment and the improvement of the count in specific areas, it is intuitively logical that it should exist. This is especially true in a place like New York City with a concentration of older housing subdivided into apartments after originally being built for one family. Moreover, among the counties I studied, areas with a larger minority share were more likely to participate in LUCA, increasing the chance of reducing the differential undercount.

Paying enumerators more money also seems like a good return on investment, because enumerator mistakes are a major source both of omission and erroneous enumeration. Indeed, this investment may even lead to savings since the count may be completed more quickly. Similarly, money spent on advance publicity increases the mail return rate, and reduces the time and error of subsequent data collection. Finally, by investing in better and more extensive computer equipment, the Bureau can improve its ability to manage the entire data collection and estimation process.

An increased Census Bureau budget is not likely, though, to reduce very much the frequency of errors made by individuals filling out their census forms. The errors of people who enumerate themselves at two locations, add inappropriate people to their census forms, and/or mistakenly leave others off are usually honest mistakes. They occur frequently among people filling out and mailing back the forms. Once such errors have been made, there is no feature of the census process that can correct them. Within-household errors are probably the major component of omissions, and they are an important component of erroneous enumerations.[24]

Problems of obtaining correct enumerations within households lead me to believe that census error is inevitable, and is unresponsive to budget increases and design improvements. People will always be left off census questionnaires. This sort of omission is so prevalent, and impervious to census method, that we should always expect it to occur in the millions. The only way that the net undercount could ever be zero, or close to it, is to have the numbers of omissions and erroneous enumerations offset each other. The problem would then be that the geographic distributions of omissions and erroneous enumerations would differ, and the differences cause distortions to the census results. This is why some statistical adjustment is essential to correct the inevitable errors of the initial count. And, throughout the 1990s until this year, the Census Bureau agreed.

---

23 U. S. General Accounting Office, "2000 Census: Review of Partnership Program Highlights Best Practices for Future Operations," August 2001. In fiscal year 2000 dollars, the cost of the 1990 Census was $3.275 billion. Based on fiscal year 2000 dollars, the per officially counted person costs of the 1990 and 2000 Censuses were, respectively, $13.17 and $23.29. These calculations for 1990 are based on data given in "Bureau of the Census Federal Funds," Appendices to the Budgets of the United States Governments, submitted by the President of the United States, 1986 – 1996.

24 Howard Hogan, "The 1990 Post-Enumeration Survey: Operations and Results," *Journal of the American Statistical Association*, September 1993, p.1056. See also Elizabeth Martin, "Who Knows Who Lives Here?" *Public Opinion Quarterly*, Summer 1999, pp. 220-36 and Eugene Ericksen, Leo Estrada, John Tukey and Kirk Wolter, "Report on the 1990 Decennial Census and Post-Enumeration Survey," submitted to the U. S. Secretary of Commerce, June 21, 1991, Appendix A, Table 6.

Imputing and Deleting Records

Statistical estimation to adjust the census has been a controversial issue, especially when it appeared that the Bureau planned to adjust the results of Census 2000. Due in part to the political opposition to adjustment, the Bureau received a substantial budget increase, for a stated goal of "counting," as opposed to "creating people by computer." The Bureau did not adjust Census 2000, and it did reduce both the national net and the differential undercount. Yet it did not do these things simply by "counting more people." Moreover, if it had adjusted, it could have "eliminated" or corrected the remaining undercount.

Had an adjustment taken place, about 4.3 million records would have been added to the count and 1.0 million deleted. These changes are what the political opposition to adjustment prevented. Yet, a computer imputed 5.7 million persons. This imputation makes use of information about people who live in houses like those where the information was needed, but it is not based on direct observation. Critics of adjustment point out that people living in places like Midland, Texas may be used to change the populations of people living in New Haven, Connecticut.[25] Yet they are silent about the fact that donors and intended recipients of imputation are often very different. For example, it is very likely that information about a White male age 35 could be used for a Black female age 57. Imputation, like adjustment, improves the statistical estimate on average. For both methods, there are individual examples that appear to be incongruous.

The Bureau did not limit its use of the computer to imputation. Late in the census process, the Bureau used a complex computer program to identify about 6 million duplications in their data file. As the Bureau put it,

> "[A]nalyses of the April 2000 and June 2000 MAF extracts still indicated that there was an overcoverage problem. These concerns led the Census Bureau to identify and remove housing units (MAFIDs) from Census 2000. Housing units were identified as being included in error with a relatively high likelihood based on a set of person and address matching rules."[26]

It eliminated 3.64 million person records,[27] i.e., it took records of "real people" out of the census The Bureau returned the other 2.37 million people to the count, and they are referred to as "late census adds." Because the deletion and reinstatement operations took place late in the census process, the Bureau was not able to include the late census adds in the A.C.E.

A review of these materials makes it clear that the Bureau monitored the level of the count throughout the census data collection period, and took the appropriate action that it deemed necessary. When the count appeared to be too large, and therefore the rate of erroneous enumeration too high, the Bureau eliminated 6 million person records. Most of the information on these records was received from persons actually living in the affected households. Later, when it appeared that they might have reduced the count by too great an amount, they put about 2.4 million of the records back in. The net effect of these operations is that the eventual net undercount of 1.18 percent is substantially an artifact of the Bureau's decisions about the apparently duplicated housing.

---

25 U. S. Census Monitoring Board, Congressional members, "A Guide to Statistical Adjustment: How it Really Works," June 7, 2001.

26 U. S. Bureau of the Census, Memorandum from Howard Hogan to Susan Miskura, dated November 7, 2000, "Specification for Reinstating Addresses Flagged as Deletes on the Hundred percent Census Unedited File (HCUF)."

27 U. S. Bureau of the Census, Memorandum from Susan Miskura to Preston J. Waite, dated November 21, 2000, "Results of Reinstatement Rules for the Housing Unit Duplication Operations."

Left unasked is the question of how these 6 million (now 3.64 million) duplications occurred in the first place. Their inclusion in the census would have doubled the rate of erroneous enumeration by the Bureau's definition.[28]

Review of the Census Bureau procedure for removing duplicates reveals a complex method relying on the assumption that the census forms in question were filled out correctly. If a family filled out two forms, but did so inconsistently, the Bureau may not have recognized it as a duplication. Like adjustment, "duplication removal by computer" will improve census distributions on average, but make many individual mistakes. Duplication removal procedures are statistical in that they rely on prespecified rules applied consistently to actual census data. A rational census policy would apply the same criteria to duplication removal that they apply to statistical adjustment. The current policy, which depicts adjustment as a statistical procedure not to be used, but duplication removal as a permissible procedure, makes a very fine distinction. I am not certain that it is meaningful.

Conclusion

In this report, I have attempted to summarize, discuss and analyze the issues of Census 2000 that I considered to be most pertinent, with one exception. I have left the discussion of the possible fit between the demographic estimate and the census count to my colleague, Dr. Jeffrey Passel.

As I have said repeatedly, the Bureau improved Census 2000 substantially over its 1990 counterpart, helped no doubt by a doubling of the "per person" census budget. We must understand the limits to our ability to improve things by "better counting." The Bureau, even though it did not statistically adjust the census through the A.C.E. survey, did conduct an adjustment of sorts when it deleted 3.64 million apparent duplicates from the count. I believe that we need to broaden the discussion of census error and its possible remedies to include operations such as imputation and the deletion of possibly duplicate records.

While we note the improvement of Census 2000 over its predecessors, we must keep in mind that the differential undercount, especially in large cities, persists.

[28] For the Bureau, the increase is from 3.1 to 6.74 million, and if we add persons counted at the wrong location to the count of erroneous enumerations the increase is from 6.3 to 9.94 million.

Table 1: Comparison of Direct and Synthetic Dual Systems Estimates for 1990 in Manhattan and Ulster County, New York

Manhattan

				Subtractions		Additions	
	Direct	Synthetic		Direct	Synthetic	Direct	Synthetic
Block 1	-24.20%	-4.93%		11.69	2.82	-	-
Block 2	-17.11%	-4.59%		15.05	4.52	-	-
Block 3	-9.68%	-5.56%		14.92	8.9	-	-
Block 4	-6.89%	0.95%		7.09	-	-	1.05
Block 5	-5.10%	3.32%		1.99	-	-	1.41
Block 6	-4.14%	-4.75%		4.77	5.44	-	-
Block 7	1.34%	-0.71%		-	0.58	1.11	-
Block 8	3.39%	7.62%		-	-	2.74	6.43
Block 9	15.43%	4.21%		-	-	39.04	9.4
Block 10	18.68%	3.36%		-	-	8.73	1.32
Block 11	21.49%	7.74%		-	-	20.8	6.38
Standard Deviation	13.88%	4.81%	**Sum**	55.51	22.26	72.42	25.99

Ulster

				Subtractions		Additions	
	Direct	Synthetic		Direct	Synthetic	Direct	Synthetic
Block 1	-13.94%	2.15%		9.42	-	-	1.69
Block 2	-11.17%	0.10%		12.06	-	-	0.12
Block 3	-3.84%	-0.99%		1.96	0.52	-	-
Block 4	-0.83%	-0.45%		0.48	0.26	-	-
Block 5	0.00%	-0.27%		-	0.1	0	-
Block 6	0.35%	-1.90%		-	0.69	0.13	-
Block 7	1.11%	1.90%		-	-	0.73	1.26
Block 8	3.99%	-0.13%		-	0.11	3.49	-
Standard Deviation	5.89%	1.28%	**Sum**	23.92	1.68	4.35	3.07

Source: File of 5180 blocks from 1990 Census www.cmbc.gov/p/user/share/Census 2000/ FINAL REPORT/Tables/ Manhattan and Ulster

Table 2: Comparisons of Undercount Rates and Differences, 1990 and 2000 Census

Racial Group	1990	2000	Change
Hispanics (A)	4.99%	2.85%	-2.14%
Non-Hispanic Blacks (B)	4.57%	2.17%	-2.40%
Non-Hispanic Whites (C)	0.68%	0.67%	-0.01%
Hispanic- White Difference (A - C)	4.31%	2.18%	-2.13%
Black- White Difference (B - C)	3.89%	1.50%	-2.39%

Source: Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy, March 1, 2001, pages 3 and 4.

Table 3: Net Undercount Rates for Post-strata Groups[1]

GROUP	AREA	RATE
NH White Owners		
Northeast	Large	-0.09%
Northeast	Medium	-0.04%
Northeast	Small	0.06%
Northeast	Non-metro	1.52%
Midwest	Large	-0.23%
Midwest	Medium	-0.14%
Midwest	Small	0.33%
Midwest	Non-metro	-0.98%
South	Large	0.78%
South	Medium	0.51%
South	Small	0.95%
South	Non-metro	0.51%
West	Large	-0.18%
West	Medium	0.18%
West	Small	1.02%
West	Non-metro	1.26%
NH White Renters		
Total US	Large	1.58%
Total US	Medium	1.09%
Total US	Small	2.67%
Total US	Non-metro	2.46%
NH Blacks		
Owners	Large/Medium	0.91%
Owners	Small/Non-metro	0.17%
Non-Owners	Large/Medium	3.96%
Non-Owners	Small/Non-metro	2.32%
Hispanics		
Owners	Large/Medium	1.17%
Owners	Small/Non-metro	1.45%
Non-Owners	Large/Medium	3.90%
Non-Owners	Small/Non-metro	6.17%
NH Asians		
Total US	Owners	0.55%
Total US	Non-Owners	1.58%

Table 3: Net Undercount Rates for Post-strata Groups[1] (continued)

GROUP	AREA	RATE
Pacific Islanders		
Total US	Owners	2.71%
Total US	Non-Owners	6.58%
AI on Reservations		
Total US	Owners	5.04%
Total US	Non-Owners	4.10%
AI off Reservations		
Total US	Owners	1.60%
Total US	Non-Owners	5.57%
Total US	**Overall**	**1.18%**

[1]High and low return rates have been combined for the post-strata groups.
Poststrata are defined by the Census 2000 A.C.E. Methodology, vol. 3, tab 9, pg. 4.

Source: FinalDSEUS.dat Census files delivered to Census Monitoring Board by Census Bureau 02/26/01 and 02/16/01/neraphl/user/share/C200 Census 2000/Final Report/tables p/user/share/Census2000/FINAL REPORT/Tables/Net Undercount Rates for Post-strata

Table 4: Comparisons of A.C.E. Undercount Rates with Synthetic Estimates

State	A.C.E. Undercount	Synthetic Undercount	Difference
Alaska	2.67%	1.15%	1.53%
Vermont	1.51%	0.71%	0.80%
Idaho	1.64%	0.90%	0.74%
Wyoming	1.56%	0.91%	0.66%
Maine	1.33%	0.72%	0.61%
Montana	1.57%	0.96%	0.61%
West Virginia	1.19%	0.74%	0.45%
Washington	1.41%	0.98%	0.44%
Delaware	1.50%	1.08%	0.42%
Kentucky	1.23%	0.82%	0.41%
New Hampshire	1.13%	0.73%	0.40%
Nevada	1.68%	1.28%	0.39%
Utah	1.35%	0.96%	0.39%
Dist of Columbia	2.15%	1.78%	0.36%
Oregon	1.27%	0.93%	0.34%
Hawaii	2.16%	1.88%	0.28%
Arkansas	1.28%	1.00%	0.28%
Georgia	1.48%	1.23%	0.25%
Tennessee	1.23%	0.98%	0.25%
Virginia	1.33%	1.09%	0.24%
North Carolina	1.36%	1.14%	0.22%
Oklahoma	1.40%	1.19%	0.21%
Maryland	1.40%	1.21%	0.19%
Texas	1.76%	1.57%	0.19%
Colorado	1.26%	1.14%	0.12%
Louisiana	1.34%	1.23%	0.11%
Alabama	1.19%	1.11%	0.07%
South Carolina	1.20%	1.18%	0.02%
Mississippi	1.24%	1.26%	-0.02%
New Mexico	1.94%	1.96%	-0.02%
Arizona	1.43%	1.46%	-0.03%
California	1.52%	1.55%	-0.04%
Florida	1.24%	1.28%	-0.04%
Connecticut	0.97%	1.03%	-0.06%

Table 4: Comparisons of A.C.E. Undercount Rates with Synthetic Estimates (continued)

State	A.C.E. Undercount	Synthetic Undercount	Difference
Pennsylvania	0.82%	0.90%	-0.08%
Rhode Island	0.85%	0.95%	-0.10%
Indiana	0.77%	0.89%	-0.12%
Massachusetts	0.76%	0.92%	-0.16%
Wisconsin	0.70%	0.87%	-0.17%
New York	1.09%	1.26%	-0.18%
Michigan	0.71%	0.98%	-0.27%
Kansas	0.66%	0.95%	-0.29%
Iowa	0.48%	0.78%	-0.30%
Nebraska	0.56%	0.89%	-0.32%
Ohio	0.57%	0.90%	-0.33%
North Dakota	0.47%	0.89%	-0.42%
Missouri	0.46%	0.91%	-0.44%
South Dakota	0.56%	1.02%	-0.46%
Minnesota	0.29%	0.84%	-0.55%
New Jersey	1.15%	1.96%	-0.81%

Source: Census 2000 Summary File 100 Percent Data/CurrentStateSum.dat and FinalDSEUS.dat Census files delivered to Census Monitoring Board by Census Bureau 02/26/01 and 02/16/01/neraphl/user/share/
C200 Census 2000/Final Report/tables

Note: Synthetic Estimates are based on race and Hispanic origin only.

Table 5: State Undercount Rates

	Large/ Medium[1]			Small/ Non Metro[2]		
State	**Undercount**		**Difference**	**Undercount**		**Difference**
Alabama	NH - Black	2.50%	1.84%	NH - Black	0.94%	-0.16%
	NH - White	0.66%		NH - White	1.09%	
Arizona	NH - Black	2.84%	2.43%	NH - Black	1.26%	-0.09%
	NH - White	0.41%		NH - White	1.35%	
Arkansas	NH - Black	2.82%	2.18%	NH - Black	1.21%	0.09%
	NH - White	0.64%		NH - White	1.12%	
California	NH - Black	2.96%	2.53%	NH - Black	1.57%	0.10%
	NH - White	0.43%		NH - White	1.48%	
Colorado	NH - Black	2.99%	2.62%	NH - Black	1.40%	-0.23%
	NH - White	0.36%		NH - White	1.64%	
Connecticut	NH - Black	2.57%	2.18%	NH - Black	1.74%	0.58%
	NH - White	0.39%		NH - White	1.15%	
Delaware	NH - Black	2.58%	1.49%	NH - Black	1.20%	-0.25%
	NH - White	1.10%		NH - White	1.44%	
Florida	NH - Black	2.38%	1.80%	NH - Black	0.97%	-0.04%
	NH - White	0.58%		NH - White	1.01%	
Georgia	NH - Black	2.96%	2.25%	NH - Black	1.18%	0.00%
	NH - White	0.71%		NH - White	1.19%	
Illinois	NH - Black	2.29%	2.25%	NH - Black	1.32%	0.82%
	NH - White	0.04%		NH - White	0.51%	
Indiana	NH - Black	2.74%	2.61%	NH - Black	1.53%	0.77%
	NH - White	0.13%		NH - White	0.76%	
Kansas	NH - Black	2.76%	2.62%	NH - Black	1.57%	1.22%
	NH - White	0.14%		NH - White	0.35%	
Kentucky	NH - Black	2.80%	2.14%	NH - Black	1.39%	0.15%
	NH - White	0.65%		NH - White	1.24%	
Louisiana	NH - Black	2.48%	1.70%	NH - Black	1.09%	-0.16%
	NH - White	0.78%		NH - White	1.25%	
Maryland	NH - Black	2.44%	1.62%	NH - Black	1.04%	-0.22%
	NH - White	0.82%		NH - White	1.26%	
Massachusetts	NH - Black	2.56%	2.18%	NH - Black	1.52%	0.63%
	NH - White	0.38%		NH - White	0.89%	
Michigan	NH - Black	2.70%	2.49%	NH - Black	1.27%	0.92%
	NH - White	0.21%		NH - White	0.35%	

Table 5: State Undercount Rates (continued)

	Large/ Medium[1]			Small/ Non Metro[2]		
State	**Undercount**		**Difference**	**Undercount**		**Difference**
Minnesota	NH - Black	3.25%	3.19%	NH - Black	2.38%	2.45%
	NH - White	0.06%		NH - White	-0.07%	
Mississippi	NH - Black	2.46%	1.71%	NH - Black	1.12%	-0.13%
	NH - White	0.76%		NH - White	1.24%	
Missouri	NH - Black	2.48%	2.41%	NH - Black	1.77%	1.63%
	NH - White	0.07%		NH - White	0.14%	
Nevada	NH - Black	2.73%	1.96%	NH - Black	1.41%	-0.11%
	NH - White	0.78%		NH - White	1.52%	
New Jersey	NH - Black	2.27%	1.72%	NH - Black	0.52%	-1.25%
	NH - White	0.55%		NH - White	1.77%	
New York	NH - Black	2.17%	2.26%	NH - Black	1.56%	0.12%
	NH - White	-0.09%		NH - White	1.44%	
North Carolina	NH - Black	2.86%	2.12%	NH - Black	0.97%	-0.16%
	NH - White	0.73%		NH - White	1.13%	
Ohio	NH - Black	2.66%	2.54%	NH - Black	1.41%	0.98%
	NH - White	0.12%		NH - White	0.42%	
Oklahoma	NH - Black	3.17%	2.43%	NH - Black	1.29%	0.24%
	NH - White	0.74%		NH - White	1.05%	
Pennsylvania	NH - Black	1.99%	1.77%	NH - Black	1.42%	0.36%
	NH - White	0.22%		NH - White	1.06%	
South Carolina	NH - Black	2.52 %	1.76%	NH - Black	0.79%	-0.32%
	NH - White	0.76%		NH - White	1.10%	
Tennessee	NH - Black	2.40%	1.66%	NH - Black	1.05%	-0.08%
	NH - White	0.74%		NH - White	1.13%	
Texas	NH - Black	2.67%	1.57%	NH - Black	1.14%	0.02%
	NH - White	1.10%		NH - White	1.13%	
Virginia	NH - Black	2.91%	2.11%	NH - Black	0.82%	-0.35%
	NH - White	0.80%		NH - White	1.17%	
Washington	NH - Black	3.20%	2.46%	NH - Black	1.83%	0.15%
	NH - White	0.75%		NH - White	1.68%	
Wisconsin	NH - Black	2.93%	2.84%	NH - Black	2.30%	1.74%
	NH - White	0.09%		NH - White	0.56%	

[1] Large/ Medium MSA's as defined by Census poststrata found in the Census 2000 ACE methodology vol. 3 tab 9 pg. 4

[2] Small/ Non Metro MSA's and all other TEA's as defined by the Census poststrata found in the Census 2000 ACE methodology vol. 3 tab 9 pg. 4

Source: Census 2000 Summary File 100 Percent Data/CurrentStateSum.dat and FinalDSEUS.dat Census files delivered to Census Monitoring Board by Census Bureau 02/26/01 and 02/16/01/neraphl/user/share/C200 Census 2000/Final Report/tables

Table 6: State Undercount Rates

	Large/ Medium[1]			Small/ Non Metro[2]		
State	**Undercount**		**Difference**	**Undercount**		**Difference**
Arizona	Hispanic	2.54%	2.14%	Hispanic	3.12%	1.77%
	NH - White	0.41%		NH - White	1.35%	
California	Hispanic	2.57%	2.14%	Hispanic	3.60%	2.13%
	NH - White	0.43%		NH - White	1.48%	
Colorado	Hispanic	2.59%	2.23%	Hispanic	2.92%	1.28%
	NH - White	0.36%		NH - White	1.64%	
Connecticut	Hispanic	3.08%	2.69%	Hispanic	3.84%	2.69%
	NH - White	0.39%		NH - White	1.15%	
Florida	Hispanic	2.18%	1.61%	Hispanic	4.25%	3.23%
	NH - White	0.58%		NH - White	1.01%	
Georgia	Hispanic	3.82%	3.11%	Hispanic	5.05%	3.86%
	NH - White	0.71%		NH - White	1.19%	
Illinois	Hispanic	2.63%	2.59%	Hispanic	3.67%	3.16%
	NH - White	0.04%		NH - White	0.51%	
Indiana	Hispanic	2.66%	2.54%	Hispanic	4.33%	3.56%
	NH - White	0.13%		NH - White	0.76%	
Kansas	Hispanic	2.76%	2.62%	Hispanic	3.85%	3.51%
	NH - White	0.14%		NH - White	0.35%	
Louisiana	Hispanic	2.51%	1.73%	Hispanic	3.41%	2.16%
	NH - White	0.78%		NH - White	1.25%	
Maryland	Hispanic	2.79%	1.97%	Hispanic	3.83%	2.57%
	NH - White	0.82%		NH - White	1.26%	
Massachusetts	Hispanic	3.11%	2.73%	Hispanic	3.37%	2.48%
	NH - White	0.38%		NH - White	0.89%	
Michigan	Hispanic	2.57%	2.36%	Hispanic	3.06%	2.71%
	NH - White	0.21%		NH - White	0.35%	
Minnesota	Hispanic	3.15%	3.09%	Hispanic	3.45%	3.52%
	NH - White	0.06%		NH - White	-0.07%	
Missouri	Hispanic	2.56%	2.49%	Hispanic	3.84%	3.70%
	NH - White	0.07%		NH - White	0.14%	
Nevada	Hispanic	2.90%	2.13%	Hispanic	3.75%	2.22%
	NH - White	0.78%		NH - White	1.52%	
New Jersey	Hispanic	2.86%	2.31%	Hispanic	2.84%	1.07%
	NH - White	0.55%		NH - White	1.77%	
New Mexico	Hispanic	1.98%	1.72%	Hispanic	2.41%	0.85%
	NH - White	0.26%		NH - White	1.56%	

Table 6: State Undercount Rates (continued)

	Large/ Medium[1]			Small/ Non Metro[2]		
State	**Undercount**		**Difference**	**Undercount**		**Difference**
New York	Hispanic	3.09%	3.18%	Hispanic	3.62%	2.18%
	NH - White	-0.09%		NH - White	1.44%	
North Carolina	Hispanic	4.19%	3.46%	Hispanic	5.34%	4.21%
	NH - White	0.73%		NH - White	1.13%	
Ohio	Hispanic	2.63%	2.51%	Hispanic	2.88%	2.45%
	NH - White	0.12%		NH - White	0.42%	
Oklahoma	Hispanic	2.99%	2.25%	Hispanic	3.45%	2.40%
	NH - White	0.74%		NH - White	1.05%	
Oregon	Hispanic	3.23%	2.85%	Hispanic	3.90%	2.24%
	NH - White	0.38%		NH - White	1.67%	
Pennsylvania	Hispanic	2.39%	2.17%	Hispanic	4.49%	3.43%
	NH - White	0.22%		NH - White	1.06%	
Tennessee	Hispanic	4.02%	3.28%	Hispanic	4.57%	3.45%
	NH - White	0.74%		NH - White	1.13%	
Texas	Hispanic	2.60%	1.50%	Hispanic	2.85%	1.73%
	NH - White	1.10%		NH - White	1.13%	
Utah	Hispanic	2.61%	1.96%	Hispanic	3.77%	2.10%
	NH - White	0.65%		NH - White	1.67%	
Virginia	Hispanic	2.85%	2.06%	Hispanic	4.31%	3.15%
	NH - White	0.80%		NH - White	1.17%	
Washington	Hispanic	2.96%	2.21%	Hispanic	4.35%	2.66%
	NH - White	0.75%		NH - White	1.68%	
Wisconsin	Hispanic	2.99%	2.90%	Hispanic	4.14%	3.58%
	NH - White	0.09%		NH - White	0.56%	

[1] Large/ Medium MSA's as defined by Census poststrata found in the Census 2000 ACE methodology vol. 3 tab 9 pg. 4

[2] Small/ Non Metro MSA's and all other TEA's as defined by the Census poststrata found in the Census 2000 ACE methodology vol. 3 tab 9 pg. 4

Source: Census 2000 Summary File 100 Percent Data/CurrentStateSum.dat and FinalDSEUS.dat Census files delivered to Census Monitoring Board by Census Bureau 02/26/01 and 02/16/01/neraphl/user/share/C200 Census 2000/Final Report/tables

Table 7: Undercount Numbers and Rates by Region

Region and State	Less than 0.00%	0.00 to 0.99%	1.00 to 1.99%	2.00 to 2.99%	3.00% or Higher	Total
Northeast						**0**
Connecticut	0	5	3	0	0	8
Massachusetts	0	11	3	0	0	14
New Jersey	0	11	9	1	0	21
New York	3	20	38	1	0	62
Pennsylvania	0	33	34	0	0	67
Subtotal	3	80	87	2	0	172
Shares	1.7%	46.5%	50.6%	1.2%	0.0%	100.0%
Midwest						
Illinois	49	43	10	0	0	102
Michigan	38	41	4	0	0	83
Subtotal	87	84	14	0	0	185
Shares	47.0%	45.4%	7.6%	0.0%	0.0%	100.0%
South						
Alabama	0	40	27	0	0	67
Delaware	0	0	3	0	0	3
Florida	0	35	30	2	0	67
Georgia	0	59	93	7	0	159
Louisiana	0	23	40	1	0	64
Maryland	0	5	18	1	0	24
Mississippi	0	40	42	0	0	82
North Carolina	0	33	67	0	0	100
South Carolina	0	23	23	0	0	46
Texas	0	26	186	41	1	254
Virginia	0	31	99	5	0	135
Subtotal	0	315	628	57	1	1,001
Shares	0.0%	31.5%	62.7%	5.7%	0.1%	100.0%
West						
Arizona	0	1	6	7	1	15
California	0	5	41	12	0	58
Hawaii	0	0	0	4	1	5

Table 7: Undercount Numbers and Rates by Region (continued)

Region and State	Less than 0.00%	0.00 to 0.99%	1.00 to 1.99%	2.00 to 2.99%	3.00% or Higher	Total
Nevada	0	1	16	0	0	17
New Mexico	0	0	10	22	1	33
Subtotal	0	7	73	45	3	128
Shares	0.0%	5.5%	57.0%	35.2%	2.3%	100.0%
All States	90	486	802	104	4	1,486
Shares	6.1%	32.7%	54.0%	7.0%	0.3%	100.0%

Source: Census 2000 Summary File 100 Percent Data/CurrentStateSum.dat and FinalDSEUS.dat Census files delivered to Census Monitoring Board by Census Bureau 02/26/01 and 02/16/01/neraphl/user/share/C200 Census 2000/Final Report/tables

Table 8: Projected and Actual Growth in Large Metropolitan Counties, 1999-2000

Percent Minority	Official 1990 Population	Projected 2000 Population	Official 2000 Population	1990-2000 Percentage of Growth: Projected	Official	Shortfall[1]
50 or more	36,095,387	37,986,821	39,435,665	5.24%	9.25%	3.67%
25 to 49	22,725,781	25,224,826	25,568,405	11.00%	12.51%	1.34%
10 to 24.9	15,620,043	17,510,324	17,704,654	12.10%	13.35%	1.10%
0 to 9.9	2,807,202	3,242,678	3,225,047	15.51%	14.88%	-0.55%
Total	77,248,473	83,964,649	85,933,771	8.69%	11.24%	2.29%

[1] Defined as 1 minus the projected 2000 population over the official 2000 population.

Table 9: Percent of Population Living in Poverty in 1990

	Percent of Population Living in Poverty					
Racial Composition[1]	**0 to 9.9**	**10 to 29.9**	**30 to 49.9**	**50 or more**	**Total**	**Size of Population[2] (millions)**
Percentages Non-Hispanic Black and Hispanic each below 10	0.4%	1.1%	2.2%	3.6%	0.7%	146.6
Percentages Non-Hispanic Black and Hispanic each between 10 and 29.9	1.8%	2.2%	2.7%	3.9%	2.1%	49.3
One or both of the percentages Non-Hispanic Black and Hispanic between 30 and 49.9	2.3%	3.0%	3.3%	3.7%	2.9%	18.4
One or both of the percentages Non-Hispanic Black and Hispanic above 50	2.9%	3.8%	4.3%	4.7%	4.0%	31.7
Native American Majority	0.0%	3.7%	9.1%	11.5%	8.9%	0.5
Total	0.8%	2.0%	3.8%	4.9%	1.6%	246.5
Size of population (millions)	127.0	96.9	18.7	3.9	246.5	

[1] Except where specified, census tracts with Native American majorities are excluded from the calculations.

[2] An additional 2.2 million people live in census tracts with no poverty rate available.

Source: Census Bureau planning database issued September 20, 2000, CD-DSS D-COMM 7

Table 10: Undercount Rankings

Race[1]	Location	Tenure	Net Undercount	Nonmatch	Erroneous Enumeration[2]	Imputations[3]
Pacific Is.	All Areas	Non-owner	6.58%	17.62%	7.67%	4.70%
Hispanic	Small MSA/ Non-metro	Non-owner	6.17%	18.22%	7.74%	5.59%
American Indian	Off Reservation	Non-owner	5.57%	15.21%	6.84%	4.12%
American Indian	On Reservation	Owner	5.04%	14.57%	4.35%	6.00%
American Indian	On Reservation	Non-owner	4.10%	12.92%	3.85%	5.59%
NH Black	Lg/Med MSA	Non-owner	3.96%	17.05%	9.04%	5.07%
Hispanic	Lg/Med MSA	Non-owner	3.90%	14.91%	7.09%	4.81%
Pacific Is.	All Areas	Owner	2.71%	12.17%	6.21%	4.49%
NH White	Small MSA	Non-owner	2.67%	11.01%	6.09%	2.73%
NH White	Non-metro	Non-owner	2.46%	13.38%	6.73%	4.99%
NH Black	Small MSA/ Non-metro	Non-owner	2.32%	14.04%	8.21%	4.26%
American Indian	Off Reservation	Owner	1.60%	10.33%	5.45%	3.51%
NH Asian	All Areas	Non-owner	1.58%	12.67%	7.55%	4.10%
NH White	Large	Non-owner	1.58%	11.77%	6.92%	3.76%
Hispanic	Small/ Non-metro	Owner	1.45%	10.19%	3.88%	5.34%
Hispanic	Lg/Med MSA	Owner	1.17%	8.88%	3.71%	4.34%
NH White	Medium MSA	Non-owner	1.09%	11.08%	7.32%	3.04%
NH Black	Lg/Med MSA	Owner	0.91%	9.89%	5.68%	3.72%
NH White	Small MSA	Owner	0.63%	5.10%	3.16%	1.38%
NH Asian	All Areas	Owner	0.55%	7.66%	4.16%	3.13%
NH White	Non-metro	Owner	0.42%	7.44%	3.85%	3.32%
NH White	Medium MSA	Owner	0.17%	4.49%	3.01%	1.41%
NH Black	Small/ Non-metro	Owner	0.17%	9.82%	5.90%	3.99%
NH White	Large MSA	Owner	0.03%	4.99%	3.30%	1.73%
Total			**1.18%**	**8.42%**	**4.72%**	**2.93%**

[1] Whites and Blacks are non-Hispanic.
[2] Erroneous enumerations include people who are counted in the wrong area or people who are counted multiple times.
[3] Housing units that assumed to be occupied that the Census Bureau imputes.

Source: FinalDSEUS.dat Census files delivered to Census Monitoring Board by Census Bureau 02/26/01 and 02/16/01/neraphl/user/share/ C200 Census 2000/Final Report/tables p/user/share/Census2000/FINAL REPORT/Tables/Net Undercount Rates for Post-strata.