North Dakota Law Review

Volume 77 | Number 4 Article 4

2001

An Evaluation of the 2000 Census and Census Bureau Efforts to Measure the Gross and Net Rates of Undercount

Eugene Ericksen

AN EVALUATION OF THE 2000 CENSUS AND CENSUS BUREAU EFFORTS TO MEASURE THE GROSS AND NET RATES OF UNDERCOUNT*

EUGENE ERICKSEN**

I. INTRODUCTION

Census 2000 has been marked by controversy and debate, both political and academic.[1] 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.[2] Opposed are those who believe both 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.[3]

Census 2000 was unprecedented in terms of its budget—more money was spent on it than on any previous census.[4] Its achievements are impressive.[5] Not only did the Census Bureau reduce the net undercount below the levels of previous censuses, it also reduced the differentials between the non-Hispanic White and minority undercounts.[6]

This article presents my observations as a statistical consultant to the Presidential Members of the Census Monitoring Board of the workings of Census 2000 and my analysis of the census results. To evaluate the accuracy of the census, I present criteria for defining accuracy and discuss

---

* This article is substantially derived from the author's report to the 2000 U.S. Census Monitoring Board, Presidential Members. This report is available at http://www.cmbp.gov.

** Eugene Ericksen is a professor of Sociology and Statistics at Temple University. He received an M.S. degree in Mathematical Statistics and a Ph.D. degree in Sociology from the University of Michigan. He served as Co-Chair of the Special Advisory Panel to the Secretary of Commerce on the 1990 Census. He also served as a consultant to the 2000 Census Monitoring Board, Presidential Members. He would like to thank Frances Bourne of the Monitoring Board, whose advice and help in obtaining information was invaluable.

1. *See* U.S. CENSUS MONITORING BOARD CONGRESSIONAL MEMBERS, A GUIDE TO STATISTICAL ADJUSTMENT: HOW IT REALLY WORKS 5 (June 7, 2001) (addressing the debate between statisticians over statistical adjustment).

2. MODERNIZING THE U.S. CENSUS 75 (Barry Edmonston & Charles Schultze eds., 1995).

3. *Id.* at 78.

4. *See* U.S. CENSUS MONITORING BOARD PRESIDENTIAL MEMBERS, FINAL REPORT TO CONGRESS 10 (Sept. 1, 2001) (showing a cost of $6.5 billion), *available at*, www.cmbp.gov (last modified Sept.26, 2001).

5. U.S. CENSUS MONITORING BOARD CONGRESSIONAL MEMBERS, REPORT TO CONGRESS 1 (Sept. 1, 2001).

6. *Id.*

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.[7] Census 2000 appears to have been successful by the key standard of allocating the population among states, counties, and places.[8] It did not accomplish this feat, however, by observing and counting a greater share of the population than it did in 1990.[9]

II. GOALS OF THE CENSUS

The constitutional goal of the census is to allocate the population among states.[10] 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% of minority populations, but less than 1% of non-Hispanic Whites, its results are biased against minorities.[11] It is better to have a consistent undercount of 3% for all groups than undercounts of 3% for minorities and 0.3% 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.[12] The differential, rather than the national net undercount, is the key statistic.[13]

A second goal of the census is to describe the nature of the population.[14] We need to know the racial and ethnic composition of the

---

7. *See id.* at 3-5 (showing five things that worked well) *and* at 7-12 (showing six things to improve).

8. *See id.* at 1 (stating that communities will benefit from fair representation because of Census 2000).

9. *See id.* at 13 (stating that "severely undercounted neighborhoods remain severely undercounted").

10. *See* U.S. CONST. art. I, § 2, cl. 3 ("Representatives and direct Taxes shall be apportioned among the several states . . . according to their respective Numbers . . .").

11. In 1990, the Census Bureau obtained the following estimates of undercount: American Indians, 12.2%; Hispanics, 5.0%; Blacks, 4.6%; Asians and Pacific Islanders, 2.4%; non-Hispanic Whites and others, 0.7%. U.S. CENSUS BUREAU, U.S. DEP'T OF COMMERCE, REPORT OF THE EXECUTIVE STEERING COMM. FOR ACCURACY AND COVERAGE EVALUATION POLICY 5 tbl.1b (Mar. 1, 2001) [hereinafter ESCAP Report March 2001].

12. U.S. CONST. art. I, § 2, cl. 3.

13. *See* ESCAP Report March 2001, *supra* note 11, at 2 (stating both net and differential undercoverage were reduced by Census 2000).

14. Barry Edmonston, *Composition of the Population, in* ENCYCLOPEDIA OF THE U.S. CENSUS 75 (Margo J. Anderson ed., 2000).

population, as well as observe the distributions of age, gender, income, and country of birth.[15] For this, we need to obtain information directly from individuals, counting each person once and only once.[16]

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.

A. 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.[17] The comparable results for Census 2000 were 281.4 million counted, 284.7 million estimated to exist, and an implied net undercount of 3.3 million.[18] The net undercount was smaller in 2000 by 700,000 people, and in percentage terms, the rate dropped from 1.6% to 1.2%.[19]

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.[20] The Census Bureau also stated that 2.2 million were "whole person imputations," or people created by a computer program rather than counted directly.[21] Combining these two groups and dividing by the total count, we see that 2.65% 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% of the estimated total.

---

15. *Id.*

16. *Id.*

17. Letter from William G. Barron, Jr., Acting Director of U.S. Census Bureau, to Carolyn B. Maloney, U.S. House of Representatives (Apr. 9, 2001) (on file with author) [hereinafter Barron letter].

18. *Id.*

19. ESCAP Report March 2001, *supra* note 11, at 4 tbl.1a *and* at 5 tbl.1b.

20. Memorandum from Ruth Ann Killion, Chief, Decennial Statistical Studies Div., U.S. Census Bureau, to John Thompson, Assistant to the Assoc. Dir. For Decennial Census 4 tbl.1 (May 15, 1997) (regarding Gross Errors and Erroneous Enumerations in the 1990 Decennial Census) (on file with author) [hereinafter Killion memo].

21. John H. Thompson, Census 2000–Decision on Release of Statistically Corrected Redistricting Data, Address Before the Joint Statistical Meetings of the American Statistical Association, Atlanta, Georgia (Aug. 8, 2001) (on file with author) [hereinafter Thompson Address].

For Census 2000, the Census Bureau informs us that there were 3.1 million erroneous enumerations and 5.7 million imputations for a total of 8.8 million.[22] They comprise 3.13% of the official count, an increase over 1990. Subtracting, there were 272.6 million people directly and correctly counted, and (284.7 - 272.6 =) 12.1 million omissions.[23] They comprise 4.25% 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.[24] In 2000, of the estimated 12.1 million persons not directly counted, 6.4 million were omissions and 5.7 million were imputations.[25]

B. 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.[26] For example, if a person moved from New York to California on April 15, 2000, but was counted in California, he or she would create two errors. New York would have one person too few and California would have one person too many. Because such an 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 Census Bureau, include: 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.[27]

---

22. Barron letter, *supra* note 17, at 3.

23. *Id.*

24. *See* Killion memo, *supra* note 20, at tbl.1 (showing the omissions) *and* Thompson Address, *supra* note 21 (showing the imputations).

25. *See* Barron letter, *supra* note 17 (showing the Census 2000 omissions) *and* Thompson Address, *supra* note 21 (showing the Census 2000 imputations).

26. *See* Killion memo, *supra* note 20, at tbl.1 & n.4 ("Erroneous Enumerations include duplicates and fictitious. They do not include geocoding errors, unmatchables, unresolveds, nor other people counted at the wrong address.").

27. *Id.*

Adding these incorrectly located people to the count of erroneous enumerations increases the Census Bureau estimate considerably.[28] In 1990, 1.8% of the "E-sample" was a fictitious or duplicate case; 2.2% of the sample was counted at the wrong location.[29] 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% of the official count.

For Census 2000, the parallel calculation provides an estimate of 6.3 million erroneous enumerations by the broader definition.[30] Adding these to the 5.7 million imputations gives a total of 12.0 million, 4.26% 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.[31] In 1990, the net undercount was 4.0 million, and by the Census Bureau's definition, 4.4 million were erroneously enumerated and 8.4 million were omitted.[32] The comparable total for Census 2000 is 6.4 million omissions.[33]

The gross error is defined to be the sum of omissions and erroneous enumerations.[34] By the Census 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.[35] By the expanded definition, there were 13.9 million omissions in 1990 for a gross error of 23.8 million.[36] There were 9.6 million omissions in 2000 for a gross error of 15.9 million.[37] Making no allowance for whole person imputations, the decline from 23.8 to 15.9 million indicates

---

28. *Id.*

29. The "E-sample" is the sample of census records that the Census Bureau matched against the Accuracy and Coverage Evaluation (A.C.E.) survey data to determine the percentages correctly and erroneously counted. Howard Hogan, *The 1990 Post-Enumeration Survey: Operations and Results*, 88 J. AM. STAT. ASS'N 1047, 1048 (1993) [hereinafter Hogan 1990 PES]. The "P-sample" is the survey sample that is matched against the census records to determine the percentages included and omitted from the count. *Id.* at 1047-48.

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

31. 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. Thus, for 1990 the net undercount is (4.4 + 4.0 =) 8.4 million.

32. Barron letter, *supra* note 17.

33. *Id.*

34. *Id.*

35. *Id.*

36. With 9.9 million erroneous enumerations, 4.0 million net undercount, and 13.9 million omissions, the total gross error equals 23.8 million.

37. With 6.3 million erroneous enumerations, 3.3 million net undercount, and 9.6 million omissions, the total gross error equals 15.9 million.

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 Census Bureau could not directly count, so the computer created his or 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.[38]

To summarize, the Census Bureau's achieved reduction in the net national and differential undercounts did not necessarily occur because it "counted" many more people directly and correctly. Depending on how one defines erroneous enumerations, omissions, and the gross error, the Census 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.

C. 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, ten omissions and five erroneous enumerations, the adjustment would add five people even though the gross error was fifteen. An adjoining block might have eighty counts, one omission and nine erroneous enumerations, and the adjustment would subtract eight people even though the gross error was ten. 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.[39] 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 Census 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.

---

38. For 1990, adding the 2.2 million imputations to the 13.9 million actual omissions provides a revised estimate of 16.1 million omissions. Adding this amount to the 9.9 million erroneous enumerations provides a total of 25.0 million; for 2000 the comparable sum is (15.3 + 6.3 =) 21.6 million.

39. *See* Hogan 1990 PES, *supra* note 29, at 1054.

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.[40] 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 Census Bureau, since it could not "explain" all the variation in net undercount rates across sample blocks, considerably understated the block-level adjustments.[41]

In my evaluation of the 2000 census for the Census Monitoring Board, I illustrate the point using estimates for eleven sample blocks in Manhattan and eight sample blocks in Ulster County, New York.[42] I present the percentage undercount as estimated by the direct and synthetic dual systems estimates for each block.[43] For both Manhattan and Ulster, the direct estimates are more highly variable than are the synthetic estimates.[44] 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.[45] More to the point, the synthetic estimate is usually between zero and the value of the direct estimate.[46] For example, block 1 in Manhattan has a direct estimate of -24.20% and a synthetic estimate of -4.93%, while block 11 has estimates of 21.49% and 7.74% respectively.[47]

---

40. *Id.*; *see also* HOWARD HOGAN, DEP'T OF COMMERCE, U.S. CENSUS BUREAU, ACCURACY AND COVERAGE EVALUATION: THEORY AND APPLICATION 5-8 (2000) (report prepared for February 2-3, 2000 DSE Workshop of the National Academy of Science Panel to Review the 2000 Census); Stephen E. Fienberg, *Capture-Recapture Methods, in* ENCYCLOPEDIA OF THE U.S. CENSUS 49-54 (Margo J. Anderson ed., 2000) (describing dual systems estimation as a form of systematic coverage improvement, based on the idea that using two sources of information will provide a more accurate estimate of the population).

41. This is because the variables used to define post-strata, or sub-groups of the population, in both 1990 and 2000 predict patterns of undercount in a general way for large aggregates. Additional 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 the Census Bureau; it merely points out the inevitable limitations of any adjustment model.

42. Eugene Ericksen, *An Evaluation of the 2000 Census, in* U.S. CENSUS MONITORING BOARD PRESIDENTIAL MEMBERS, FINAL REPORT TO CONGRESS 28 tbl.1 (Sept. 1, 2001) (obtained from data file of 5,180 blocks from the 1990 census found at www.cmbc.gov) [hereinafter Ericksen Evaluation]. Table 1 of this evaluation is a comparison of Direct and Synthetic Dual Systems Estimates for 1990 in Manhattan and Ulster County, New York. *Id.*

43. *Id.*

44. *Id.*

45. *Id.*

46. *Id.*

47. *Id.*

The synthetic adjustments are therefore smaller in absolute value than the direct estimates would be if they were available for all blocks.[48] For Manhattan and Ulster combined, the ratio of the average direct to synthetic adjustment is about three.[49]

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%, of the adjustments that need to be made. The Census Bureau's adjustments improve the estimated distribution of population, but not perfectly.[50] Moreover, the 15.9, rather than the 6.91 million, better indicate the extent of the undercount.

In Census 2000, the Census Bureau added 4.26 million and subtracted 1.00 million for a total of 5.26 million changes.[51] Assuming 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.[52]

---

48. *Id.*

49. *Id.*

50. The "imperfections" of the Census 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.

51. Press Release, Statement by William G. Barron, Jr. on the Current Status of Results of Census 2000 Accuracy and Coverage Evaluation Survey, (July 13, 2001) (on file with author) [hereinafter Barron Press Release].

52. 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.

III. PATTERNS OF UNDERCOUNT

A major story of Census 2000 is the reduction in the differential undercount.[53] 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.[54] 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 Census Bureau defined post-strata, the creation of demographic subgroups, not only by race and Hispanic origin, but also by tenure, metropolitan status, region (for White homeowners) and the mail return rate.[55] Only the first three of these factors had a consistent effect on the estimated net undercount.[56]

Looking first at sixteen groups of non-Hispanic White homeowners defined by geographic location, the rates cluster around zero.[57] Six of the estimates are overcounts, seven of them are between 0% and 0.99%, and the remaining three are between 1.00% and 1.99%.[58] Rates of undercount are slightly higher in non-metropolitan areas than elsewhere.[59]

Rates of undercount are somewhat higher for White non-homeowners than for White homeowners, but the differences are not large.[60] We also observe this pattern for other racial groups.[61] For Hispanics, the undercounts were higher in smaller and non-metropolitan areas, while the opposite was true for the non-Hispanic Blacks.[62] Indeed, the rates of net undercount for non-Hispanic Blacks and Whites living in smaller and non-metropolitan areas were similar to each other.[63] Finally, for non-Hispanic

---

53. *See* ESCAP Report March 2001, *supra* note 11, at 4 tbl.1a *and* at 5 tbl.1b (showing undercount rates and differences for the 1990 and 2000 censuses); *see also* Ericksen Evaluation, *supra* note 42, at 29 tbl.2 (showing comparisons of undercount rates and differences for the two censuses).

54. *Id.*

55. *See* Memorandum from Donna Kostanich, Assistant Div. Chief, Sampling and Estimation Decennial Statistical Studies Div., to Howard Hogan, Chief, Decennial Statistical Studies Div. 4 tbl.1a (Apr. 19, 2000) (defining post strata for Census 2000) (on file with author) [hereinafter Kostanich memo].

56. Ericksen Evaluation, *supra* note 42, at 30-31 tbl.3 (showing net undercount rates for post-strata groups).

57. *Id.*

58. *Id.*

59. *Id.*

60. *Id.*

61. *Id.*

62. *Id.*

63. *Id.*

Asians, Pacific Islanders, and American Indians the rates of undercount were generally higher than average but substantially below their comparable estimates for 1990.[64]

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.[65] As different states typically have similar percentages of owner-occupied housing units, it would not be expected 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.

A. RATES OF NET UNDERCOUNT FOR STATES

Relying on data provided by the Census Bureau, I have replicated its estimates of net undercount by state.[66] 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.[67] For example, if a state's population included 20% Hispanics, 30% non-Hispanic Blacks, and 50% non-Hispanic Whites, its synthetic estimate would be:

$$(0.20 * 2.85) + (0.30 * 2.17) + (0.50 * 0.67) = 1.56\%.$$

Comparing the actual and synthetic estimates found in my evaluation report to the Census Monitoring Board, they are quite similar.[68] The estimates differ by more than a percentage point in only one state, Alaska.[69] They differ by more than one-half of one percentage point in only eight states.[70]

The synthetic and Accuracy and Coverage Evaluation (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% (Minnesota) to 2.67% (Alaska).[71]

---

64. *Id.*
65. *Id.*
66. *Id.* at 32-33 tbl.4 (showing comparison of A.C.E. synthetic estimates of undercount for states).
67. *Id.*
68. *Id.*
69. *Id.*
70. *Id.*
71. *Id.*

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.[72] This makes it possible to compare, within states, the minority-White differentials in more and less metropolitan districts.[73] For example, in Georgia, is the Black-White differential in cities like Atlanta similar to the Black-White differential in more rural areas? In my evaluation report for the Census Monitoring Board, I present comparisons of undercount differentials, first for non-Hispanic Blacks and Whites and second for Hispanics and non-Hispanic Whites within areas more and less metropolitan in the same state.[74]

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%.[75] The corresponding differential for smaller and non-metro areas was (0.94% - 1.09% =) -0.15%, indicating a greater racial disparity in urban areas like Birmingham and Mobile than elsewhere.[76]

Looking at the Black-White differentials in different states, they are consistently between 1.5% and 3% in the large/medium category.[77] In these more metropolitan areas, the racial differential is consistent with the pattern found in past censuses—the Black undercount is higher than the White.[78] 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.[79] Indeed, the Black rate is lower than the White rate, though not by very much, in the rural and small city areas of fourteen of the thirty-four states.[80]

This result has important implications for the South, as 85% of the non-Hispanic Black population living in smaller and non-metropolitan areas is located there. There appears to have been no racially differential undercount in the less metropolitan South.[81] While the Census Bureau has not offered any explanation for this result, my hypothesis would focus on the

72. *See* Kostanich memo, *supra* note 55, at 4 tbl.1a.

73. *Id.*

74. Ericksen Evaluation, *supra* note 42, at 34-35 tbl.5 (showing state undercount rates) *and* at 36-37 tbl.6 (showing comparison of Hispanic and Non-Hispanic White rates of undercount in large and small metropolitan areas).

75. *Id.* at 34 tbl.5.

76. *Id.*

77. *Id.* at 34-35 tbl.5.

78. *Id.*

79. *Id.*

80. *Id.*

81. *Id.*

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.[82] As we would expect from past censuses, Hispanic undercounts are consistently higher, by a few percentage points, than non-Hispanic White undercounts.[83] Both groups have higher undercounts in smaller and non-metropolitan areas, and the differential between Hispanics and Whites is somewhat larger in those areas as well.[84] As a result, in many areas of the West and Southwest the non-metropolitan rates of undercount are higher than those of large cities.[85]

B. 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.[86] To study this question, I calculated undercount rate estimates for approximately 1500 counties located in twenty-three states. The states were selected by one or both of two criteria: (a) at least 25% 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 summarize the results of my calculations in my evaluation report.[87]

The variation in county rates of undercount is not substantial. A full 87% of all counties studied have rates of undercount between 0.00% and 1.99%.[88] There are only four counties, all with small populations, that have

82. *Id.* at 36-37 tbl.6.
83. *Id.*
84. *Id.*
85. *Id.*
86. *Id.* at 34-35 tbl.5 *and* at 36-37 tbl.6.
87. *Id.* at 38-39 tbl.7 (showing rates of net undercount for counties sorted by state and region).
88. *Id.*

rates of undercount above 3%.[89] Six percent of counties, generally located in the Midwest, have overcounts but none of these is greater than one percentage point.[90] As a general pattern, county rates of undercount are higher in the West and lower in the Midwest.[91] 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.

C. 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.[92] 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, however, 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 Census Bureau.[93] 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% or more, (b) 25% to 49.9%, (c) 10% to 24.9%, and (d) less than 10% minority.[94]

I then obtained 1998 and 1999 population estimates from the Census Bureau web site.[95] 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 1990-1999 growth to the unadjusted 1990 count. I calculated a "2000 Census Projection" by adding the 1998-1999 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

89. *Id.*

90. *Id.*

91. *Id.*

92. U.S. CENSUS MONITORING BOARD CONGRESSIONAL MEMBERS, REPORT TO CONGRESS 1 (Sept. 1, 2001).

93. Kostanich memo, *supra* note 55.

94. Minority in this context is defined as persons who did not self-identify as being only one race, non-Hispanic White.

95. *See* U.S. Bureau of the Census, *available at* http://www.census.gov/population/www/estimates/countypop.html (July 30, 2001).

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 five percentage points. I summarized the relevant population counts and projections by category and calculated the overall shortfalls by county group.[96]

Together, the counties with the largest minority populations had the largest shortfall.[97] They were projected to grow by 5.24%, but actually "grew" by 9.25%.[98] The shortfall of the population projection was 3.67 percentage points, and this shortfall may well reflect the effects of improved counting.[99] It is larger than the shortfalls of 1.34%, 1.10%, and –0.55% found for the other three groups of counties.[100] In other words, those counties with higher minority shares had larger shortfalls than did counties with lower minority shares.[101]

New York City is an especially good example, as its projected growth rate was 1.73% compared to actual growth of 9.36%. Washington, D.C., Philadelphia, and Hudson County, New Jersey had similarly high and unexpected amounts of growth. The shortfall was positive in fifteen of the sixteen "high minority" counties. It was three percentage points or more in ten of the sixteen counties. While there is variation, we see consistently high and unexpected growth in urban areas with large minority populations.[102] 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.[103] In all, the Census Bureau added just over 4 million addresses through LUCA.[104] One of the largest files of added addresses came from New York City, where the Census Bureau accepted over 280,000 added

---

96. *See* Erickson Evaluation, *supra* note 42, at 40 tbl. 8 (showing projected and actual growth in large metropolitan counties, 1999-2000).

97. *Id.*

98. *Id.*

99. *Id.*

100. *Id.*

101. *Id.*

102. *Id.*

103. U.S. CENSUS MONITORING BOARD CONGRESSIONAL MEMBERS, REPORT TO CONGRESS 4 (Sept. 1, 2001).

104. U.S. CENSUS MONITORING BOARD PRESIDENTIAL MEMBERS, LUCA SUBMISSIONS AND "ADDS" BY LOCAL GOVERNMENT JURISDICTION (July 13, 2001).

addresses.[105] These comprised 8.88% of the eventual city total of 3.2 million housing units.[106] LUCA's contribution compares to the 6.98% "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 which such as Washington, D.C., 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 -0.014.

IV. 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 the 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 it continues to consider whether or not it should adjust these results for use in fund allocation and other purposes.[107] This section discusses 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.[108] As the Census Bureau put it in the March 1 ESCAP report, "[i]nitial 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."[109]

The most likely culprit, from the perspective of the A.C.E., is the underestimation of erroneous enumerations.[110] In other words, the official count includes more duplications, fabrications, and persons counted in the wrong place than the A.C.E. indicated.[111] Increasing the estimated number of erroneous enumerations would reduce the net undercount, but also

---

105. *Id.*
106. *Id.*
107. U.S. CENSUS MONITORING BOARD CONGRESSIONAL MEMBERS, A GUIDE TO STATISTICAL ADJUSTMENT: HOW IT REALLY WORKS 3-4 (June 7, 2001).
108. ESCAP Report March 2001, *supra* note 11, at ii.
109. *Id.* at 3.
110. *Id.* at 15.
111. *Id.*

increase the gross error and indicate that the quality of Census 2000 data was not as good as originally thought.[112]

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.[113]

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 Census 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 Census 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 are high; in these types of neighborhoods it is especially difficult to count.[114] Even where some but not all of these conditions exist in extreme forms, census taking may still be difficult.[115] These difficulties lead to higher rates of omission as well as higher rates of erroneous enumeration, whole person imputation, and records with incomplete and incorrect recording of characteristics such as race and Hispanic origin.[116]

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.

---

112. *Id.* at 22.

113. Thompson Address, *supra* note 21.

114. U.S. CENSUS MONITORING BOARD CONGRESSIONAL MEMBERS, REPORT TO CONGRESS 8-9 (Sept. 1, 2001).

115. *Id.*

116. *Id.*

As part of its planning for Census 2000, the Census 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.[117] 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 of non-Hispanic Black and Hispanic were each below 10%. 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.

My evaluation report to the Census Monitoring Board shows that higher rates of poverty are associated with higher rates of undercount.[118] Areas where the poverty rate is low and the population predominantly non-Hispanic White have very low rates of undercount, 0.4%.[119] Increases in this rate are associated both with increased percentages of poverty and racial minorities.[120] It should be noted that the rate of undercount for predominantly white areas with a poverty rate over 50% is 3.6%, higher than the corresponding rate, 2.9%, for areas with concentrated minority populations but a low poverty rate.[121] 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

---

117. 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.

118. *See* Ericksen Evaluation, *supra* note 42, at 41 tbl.9 (showing effects of poverty and racial composition estimated rates on undercount for census tracts-1990 census).

119. *Id.*

120. *Id.*

121. *Id.*

congregate in similar locations. I compared rates of omission, erroneous enumeration, and imputation for groups of post-strata defined by the key proxy predictors—race, Hispanic origin, tenure, and metropolitan status.[122]

This demonstrated that those post-strata with higher rates of net undercount, generally those with minority non-owner populations, also have higher rates of non-matching, erroneous enumeration, and imputation.[123] Indeed the correlations between the net undercount and these three variables are, respectively, 0.88, 0.51, and 0.67. The correlation between non-match and erroneous enumeration rates is 0.80. In sum, conditions of poverty create difficult counting of all types. It is theoretically possible to imagine that the Census 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 had been counted directly and correctly.

V. THE LIMITS OF IMPROVED COUNTING

The budget for Census 2000 was $6.55 billion, a large increase over the $2.6 billion budget for the 1990 census even after inflation is taken into account.[124] Using constant fiscal year 2000 dollars, the "per household" cost rose from $36 in 1990 to $62 in 2000.[125] There is little doubt that the added spending improved census data quality.[126] At the same time, there are important types of census error that are impervious to budget size.[127]

Better address lists are expensive to create, but they improve the count.[128] 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

---

122. *Id.* at 42 tbl.10 (showing rates of undercount and other indicators of census error by post-stratum group).

123. 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.

124. U.S. GENERAL ACCOUNTING OFFICE, REPORT TO CONGRESSIONAL REQUESTERS, 2000 CENSUS: REVIEW OF PARTNERSHIP PROGRAM HIGHLIGHTS BEST PRACTICES FOR FUTURE OPERATIONS 12 tbl. (Aug. 2001) (comparing dollars associated with the 1970-2000 censuses). 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.

125. *Id.*

126. U.S. CENSUS MONITORING BOARD CONGRESSIONAL MEMBERS, REPORT TO CONGRESS 13 (Sept. 1, 2001).

127. *Id.*

128. *Id.* at 4.

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.[129] Indeed, this investment may even lead to savings since the count may be completed more quickly.[130] Similarly, money spent on advance publicity increases the mail return rate, and reduces the time and error of subsequent data collection.[131] Finally, by investing in better and more extensive computer equipment, the Census Bureau can improve its ability to manage the entire data collection and estimation process.[132]

An increased Census Bureau budget is not likely, however, to reduce by very much the frequency of errors made by individuals filling out their census forms.[133] 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.[134] They occur frequently among people filling out and mailing back the forms.[135] 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.[136]

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.[137] This sort of omission is so prevalent and impervious to

---

129. *Id.* at 5.

130. *Id.*

131. *Id.*

132. ESCAP Report March 2001, *supra* note 11, at 12.

133. *See* Edmonston & Schultze, *supra* note 2, at 102 (advocating respondent-friendly forms).

134. ESCAP Report March 2001, *supra* note 11, at 12.

135. Eugene P. Ericksen, *Errors in the Census, in* ENCYCLOPEDIA OF THE U.S. CENSUS 207 (Margo J. Anderson ed., 2000) [hereinafter *Errors in the Census*].

136. Hogan 1990 PES, *supra* note 29, at 1056; *see also* Elizabeth Martin, *Who Knows Who Lives Here? Within-Household Disagreements as a Source of Survey Coverage Error*, 63 PUB. OPINION Q. 220, 220-36 (1999); Eugene Ericksen et al., Report on the 1990 Decennial Census and Post-Enumeration Survey, submitted to the U. S. Secretary of Commerce, June 21, 1991, app. A, tbl.6.

137. Hogan 1990 PES, *supra* note 29, at 1056.

census method that we should always expect it to occur in the millions.[138] 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.[139] The problem would then be that the geographic distributions of omissions and erroneous enumerations would differ, and the differences would cause distortions to the census results.[140] This is why some statistical adjustment is essential to correct the inevitable errors of the initial count.[141] And, throughout the 1990s until this year, the Census Bureau agreed.[142]

A. IMPUTING AND DELETING RECORDS

Statistical estimation to adjust the census has been a controversial issue, especially when it appeared that the Census Bureau planned to adjust the results of Census 2000.[143] Due in part to the political opposition to adjustment, the Census Bureau received a substantial budget increase, for a stated goal of "counting," as opposed to "creating people by computer."[144] Although it did not adjust Census 2000, the Census Bureau did reduce both the national net and the differential undercount.[145] 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.[146] These changes are what the political opposition to adjustment prevented. Yet, a computer imputed 5.7 million persons.[147] This imputation makes use of information about people who live in houses similar to those houses 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

---

138. *Id.*

139. *Errors in the Census, supra* note 135, at 207.

140. *Id.*

141. Hogan 1990 PES, *supra* note 29, at 1057.

142. U.S. CENSUS MONITORING BOARD CONGRESSIONAL MEMBERS, A GUIDE TO STATISTICAL ADJUSTMENT AND HOW IT REALLY WORKS 1 (June 7, 2001).

143. U.S. CENSUS MONITORING BOARD PRESIDENTIAL MEMBERS, FINAL REPORT TO CONGRESS 5 (Sept. 1, 2001), *available at*, www.cmbp.gov (last modified Sept. 26, 2001).

144. MARGO J. ANDERSON & STEPHEN E. FIENBERG, WHO COUNTS? THE POLITICS OF CENSUS TAKING IN CONTEMPORARY AMERICA 211-13 (paperback ed. 2001).

145. U.S. CENSUS MONITORING BOARD CONGRESSIONAL MEMBERS, REPORT TO CONGRESS 1 (Sept. 1, 2001).

146. Barron Press Release, *supra* note 51.

147. Thompson Address, *supra* note 21.

change the populations of people living in New Haven, Connecticut.[148] Yet they are silent about the fact that donors and intended recipients of imputation are often very different.[149] For example, it is very likely that information about a White male, age thirty-five could be used for a Black female, age fifty-seven. Imputation, like adjustment, improves the statistical estimate on average. For both methods, there are individual examples that appear to be incongruous.

The Census 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 its data file.[150] As the Census 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.[151]

It eliminated 3.64 million person records, or in other words, it took records of "real people" out of the census.[152] The Census Bureau returned the other 2.37 million people to the count, and they are referred to as "late census adds."[153] Because the deletion and reinstatement operations took place late in the census process, the Census Bureau was not able to include the late census adds in the A.C.E.[154]

A review of these materials makes it clear that the Census 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 Census Bureau eliminated 6 million person records.[155] Most of the information on these records was received from

---

148. U.S. CENSUS MONITORING BOARD CONGRESSIONAL MEMBERS, A GUIDE TO STATISTICAL ADJUSTMENT: HOW IT REALLY WORKS 15 (June 7, 2001).

149. *Id.* at 18.

150. Memorandum from Howard Hogan, Chief, Decennial Statistical Studies Div., to Susan Miskura, Chief, Decennial Mgmt. Div. (Nov. 7, 2000) [hereinafter Hogan memo]; Memorandum from Susan M. Miskura, Chief, Decennial Mgmt. Div., to Preston J. White, Assistant to the Assoc. Dir. for Decennial Census (Nov. 21, 2000) [hereinafter Miskura memo].

151. Hogan memo, *supra* note 150.

152. Miskura memo, *supra* note 150.

153. *Id.*

154. ESCAP Report March 2001, *supra* note 11, at 25-26.

155. Miskura memo, *supra* note 150.

persons actually living in the affected households.[156] 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.[157] The net effect of these operations is that the eventual net undercount of 1.18% is substantially an artifact of the Census Bureau's decisions about the apparently duplicated housing.

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 Census Bureau's definition.[158]

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 Census Bureau may not have recognized it as a duplication. Like adjustment, "duplication removal by computer" will improve census distributions on average, but it will also 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 are applied 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.

B. ADJUSTMENT POSTSCRIPT

On October 17, 2001 the Census Bureau revised its undercount estimates and made its final decision on the possibility of adjustment of Census 2000.[159] After further research, it drew the following conclusion:

> After assessing considerable new evidence, [The Executive Steering Committee for A.C.E. policy] now recommends that unadjusted Census 2000 data also be used for non-redistricting purposes. The effect of this new evidence is that the Accuracy and Coverage Evaluation (A.C.E.) overstated the net undercount by at least 3 million persons. The cause of this error was that the A. C. E.

156. *Id.*

157. *Id.*

158. For the Census 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.

159. REPORT OF THE EXECUTIVE STEERING COMM. FOR ACCURACY AND COVERAGE EVALUATION POLICY ON ADJUSTMENT FOR NON-REDISTRICTING USES 1 (Oct. 17, 2001).

failed to measure a significant number of census erroneous enumerations, many of which were duplicates.[160]

In other words, the Census Bureau increased the estimated number of erroneous enumerations by at least 3 million.[161] Assuming that the number was exactly 3 million, then the A.C.E. estimate of the national population was reduced to 281.7 million, for a net undercount of (1 – 281.4 / 281.7 =) 0.1%.

The gross error, however, was increased. Adding the 3 million to the original Census Bureau estimate of 3.1 million erroneous enumerations provides a revised total of 6.1 million. Adding these to the estimated 6.4 million omissions leads to a revised gross error estimate of 12.5 million, about the same number that was observed in 1990.[162]

We see, therefore, that the reduced net undercount does not indicate an improved census. For the constitutional purpose of allocating political representation to states and local areas, the larger gross error indicates a more substantial problem than appeared to exist before. The extent of the problem depends on the tendency of omissions and erroneous enumerations to occur in different types of areas.[163]

VI. CONCLUSION

In this article, I have attempted to summarize, discuss, and analyze the issues of Census 2000 that I consider to be most pertinent. The Census Bureau improved Census 2000 substantially over its 1990 counterpart, helped no doubt by a doubling of the "per person" census budget. Even so, we must understand the limits of our ability to improve things by "better counting."[164] The Census Bureau, 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.[165] 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.

160. *Id.*
161. *Id.*
162. Hogan 1990 PES, *supra* note 29, at 1056.
163. *Errors in the Census, supra* note 135, at 207.
164. Edmonston & Schultze, *supra* note 2, at 56.
165. Miskura memo, *supra* note 150.

***