U.S. CENSUS MONITORING BOARD
Congressional Members

Unkept Promise: Statistical Adjustment Fails to Eliminate Local Undercounts, as Revealed by Evaluation of Severely Undercounted Blocks From the 1990 Census Plan.

Report to Congress
September 30, 1999

U.S. Census Monitoring Board
Congressional Members
4700 Silver Hill Road
P.O. Box 610
Suitland, MD 20752
(301) 457-5080
www.cmbc.gov
email: feedback@cmbc.gov

THE U.S. CENSUS MONITORING BOARD

In November 1997, Congress established the eight-member Census Monitoring Board: four members appointed by Congress, four by the President, charged "to observe and monitor all aspects of the preparation and implementation of the 2000 Decennial census." The issue is the Census Bureau's plan to count over 275 million Americans in 2000 – a process which will determine both Congressional apportionment and the allocation of billions of federal dollars. This is the third in a series of reports to Congress. Earlier reports were released in February and April 1999. Additional reports are scheduled through September 2001.

The Congressional members of the Census Monitoring Board welcome your input. We can be reached at our offices in Suitland, Maryland, or via e-mail at *feedback@cmbc.gov*. For more information on the census, or to download this or other reports, visit our web site at *www.cmbc.gov*.

A NOTE ON THE REPORT

This is the third in a series of reports to Congress. Earlier reports were released in February and April 1999. Additional reports are scheduled through September 2001. Associate Director of the 1990 Decennial Census Charles Jones, Robert Lerner, Ph.D. and Michael Miguel contributed to this report. The data used, and instructions to replicate our analysis, are available upon request.

September 30, 1999

The Hon. Albert Gore
President
United States Senate
Washington, DC 20510

The Hon. J. Dennis Hastert
Speaker
United States House of Representatives
Washington, DC 20515

Dear Mr. President and Mr. Speaker:

As Congressional Members of the U.S. Census Monitoring Board, charged with overseeing the Census Bureau's preparations for the 2000 Decennial Census, and particularly to review the degree to which those preparations "shall achieve maximum possible accuracy at every level of geography," we write to alert you to a number of serious concerns over the proposed statistical adjustment raised by a new study of Census Bureau data.

This report, the third in a series required by our statute, shows that statistical adjustment, heralded as a kind of statistical remedy, will fail in its main charge: to prevent traditionally undercounted communities from receiving less than their fair share of representation and funding. In fact, the claimed remedy will do little if anything to correct the severe undercounts that often affect predominantly minority neighborhoods.

According to our extensive evaluation of 1990 Census data – only recently made available to the public – heavily undercounted areas will remain heavily undercounted despite statistical adjustment, and overcounts in many areas will actually be increased.

The purpose of this report is to warn that the benefits of statistical adjustment have been overstated. Our concern is that statistical adjustment will fall short of its goal – correcting severe undercounts in traditionally undercounted areas. Therefore, we believe it is imperative that Members of Congress, state and local officials and community leaders understand the inadequacies of adjustment and take appropriate steps <u>locally</u> to ensure their constituents are counted.

Relying on statistical adjustment to fix the problem of undercounts would be a tragic mistake that could have serious repercussions for the people living in severely undercounted areas. A critical failing is that statistical adjustment conveys the illusion of correcting the undercount, when in fact actual people's needs are not met.

This report provides an explanation of the statistical problems with adjustment revealed by our study, and the methodology we employed.

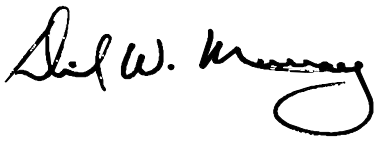
Clearly, the only certain way for any neighborhood or local area to overcome the differential undercount is to get the best initial count during the census. As a result, we also include a series of steps that Members of Congress and local leaders, who are likely to face undercount problems in spite of statistical adjustment, can take to help ensure that their communities are counted properly.

Without the active involvement of informed local leaders in the census process, traditionally undercounted communities will once again be left with empty promises and unrealistic expectations.

All of us want the same outcome – the fairest and most accurate census possible, including a dramatic decrease in the differential undercount of African American, Latino, Asian and American Indian populations. No one denies the census will miss people. In a nation as large, diverse and mobile as ours, that is unavoidable. However, we believe it is unacceptable for the 2000 Census to systematically miss the same people in the same communities that depend so heavily on the census for a fair share of crucial services and political representation.

Sincerely,

J. Kenneth Blackwell
Co-Chairman



Dr. David W. Murray
Congressional Member

A. Mark Neuman
Congressional Member

Joe Whitley
Congressional Member

Table Of Contents

Section I

INTRODUCTION

THE MYTH: STATISTICAL ADJUSTMENT WILL CORRECT SEVERE UNDERCOUNTS IN PREDOMINANTLY MINORITY NEIGHBORHOODS

For more than 10 years, the country has debated the use of "statistical adjustment" to resolve the problems of undercounting endemic to the decennial census – particularly in minority neighborhoods. The January 1999 Supreme Court ruling in large measure settled this debate by requiring the Census Bureau to attempt a full (100 percent) enumeration for the apportionment of seats in the U.S. House of Representatives.

The Administration and the Census Bureau, however, have insisted that statistical adjustment be used for all other purposes, including the allocation of federal and state funds for a variety of uses, from roads and schools, to health care and community development. The primary purpose of statistical adjustment is "to measure and correct overall and differential coverage of U.S. residents in Census 2000."[1] Presumably, this includes correcting the disproportionate undercount of African Americans, Asians, Latinos and Native Americans.

However, a new study, undertaken by the Congressional members of the U.S. Census Monitoring Board, reveals the ability of statistical adjustment to correct the undercount has been wildly overstated. The evidence clearly indicates that:

1. Statistical adjustment will not correct large undercounts or overcounts in small areas such as blocks and neighborhoods.

2. Heavily undercounted areas will remain heavily undercounted.

3. Statistical adjustment will add people to many overcounted areas (areas where the Bureau mistakenly counts too many people).

4. Until the census is improved in local areas that are heavily undercounted, the differential undercount will persist at the local level.

Note, statistical adjustment may well "correct the undercount" in a global sense – that is, improve the overall, national "count." But the practical point of the census is to apportion the count correctly. That is, "the count"

[1] 23 February 1999, Updated Summary: Census 2000 Operational Plan, p. 2.

is a numerical product, which may well improve through even random additions to population. But "the count" also represents real people, with real political rights and needs. A statistical adjustment that "corrects" the count nationally may do nothing palpable for the needs of actual people living in undercounted areas. Unless statistical adjustment is distributed correctly, such that people are added where and only where they are missed, the Bureau will not have remedied the failure of fairness that statistical adjustment is intended to address.

The Board's review of Bureau data shows the local benefits of adjustment are more myth than methodology – a statistical promise to cure the problem of local undercounts which cannot be kept. Even worse, for local leaders who believe in the promise of an accurate census, the illusion of a proper adjustment could encourage a false sense of security concerning the ability of statistical adjustment to correct local undercounts.

BACKGROUND ON THE U.S. CENSUS MONITORING BOARD STUDY

Statisticians debate whether statistically adjusting the census will provide a better picture of the population at the national or state level. Many statisticians and demographers have published extensive analysis and criticism of the Census Bureau's attempts to measure the undercount, questioning the reliability of the Bureau's methodology even at the national or state level (Appendix E).

Everyone agrees, however, that statistical adjustment gets increasingly less accurate at lower levels of geography. Everyone also agrees that political power and money are distributed among local areas defined by geographic or political boundaries, and that large undercounts are distributed unevenly in these areas throughout the country. Many blocks and neighborhoods have large undercounts – undercounts greater than 10 percent.

The Bureau proposes to correct these undercounts through statistical methodology – adjustment. Adjusting the census using statistical methods involves two distinct operations: measuring the undercount of different groups of people, and adding people in the right places to correct such undercounts. There is strong reason to believe that both parts of the proposed methodology will fail.

This study conducted by the Board focuses on the second of these two tasks: adding people in the right places. Thus, for the purposes of this analysis, the survey's direct measurements of undercount are treated as if they were accurate. This analysis focuses on whether statistical adjustment using the so-called "synthetic method"[2]would have succeeded in correcting undercounts and overcounts identified by the sample survey at the lowest levels of geography – blocks and neighborhoods.

The Congressional members of the Board requested detailed data in late 1998, which were only provided by the Bureau after four months' delay. To our disappointment, the Bureau resisted the release of these data to public and academic review for almost a year. Only recently, after repeated requests from the Congressional members of the Board and the U.S. House of Representatives' Subcommittee on the Census, did the Bureau agree to release most of these data.

We analyzed the 5,170 local areas surveyed nationwide in the Bureau's 1990 post-enumeration survey (PES).[3] The PES is virtually the same as the Accuracy and Coverage Evaluation (ACE) survey the Bureau plans to use for Census 2000.

This study found that statistical adjustment does not perform adequately at the local level.. Regardless of how "accurate" the national picture may appear, statistical adjustment adds far too few people to heavily undercounted **local** areas to assure these communities get their fair share of representation or public funding. In addition, adjustment adds people to many overcounted areas – making overcounts worse.

[2] There is a distinction between "direct" estimates and "synthetic" estimates of population. Direct estimates of an area's population rely on data collected exclusively from that area. Synthetic estimates use data from an entire region to make estimates about smaller areas within that region. Therefore, a direct estimate of population in one of the 1990 PES block clusters uses observations made only in that block cluster. A synthetic estimate of population in one of the 1990 PES block clusters may use observations from hundreds or thousands of clusters in a region. For a more detailed discussion of the Bureau's synthetic estimation, see Section V: Methodology.

[3] Although the file provided to the Board contains data for 5,180 block clusters, there were 10 block clusters where the PES sample failed to find any of the people counted in the census. In this case the match variable "M" (see page 17) is zero. Since this variable is the denominator in the standard direct DSE calculation, the calculation cannot be performed. In mathematical terms, the result of a division by zero is "undefined." In addition, the Bureau previously reported 5,290 block clusters included in the 1990 PES. Bureau staff informed the Board that the clusters not included in the file had no recorded population.

Section II

KEY FINDINGS AND DISCUSSION

The Congressional members of the Census Monitoring Board examined the results of the 1990 PES, the model for next year's Accuracy and Coverage Evaluation (ACE), to predict the effectiveness of the ACE in correcting local undercounts. The Board examined the 5,170 block clusters surveyed in the 1990 PES. For a detailed description of the analysis, see Section IV: Methodology and Section V: Limitations.

- Areas with undercounts greater than two percent would have received small additions: an average increase of only 1.36 percent in 2,200 clusters where the PES indicated there should have been an average increase of 8.97 percent to remedy the undercounts.
- Areas with accurate counts (within two percent of a full count) would have received small additions: an average increase of 0.6 percent in 1,591 clusters.
- Most areas with overcounts also would have received small additions: an average increase of 0.17 percent in 996 clusters with overcounts between two and 10 percent, where the PES indicated there should have been an average *decrease* of 4.89 percent to remedy the overcounts.

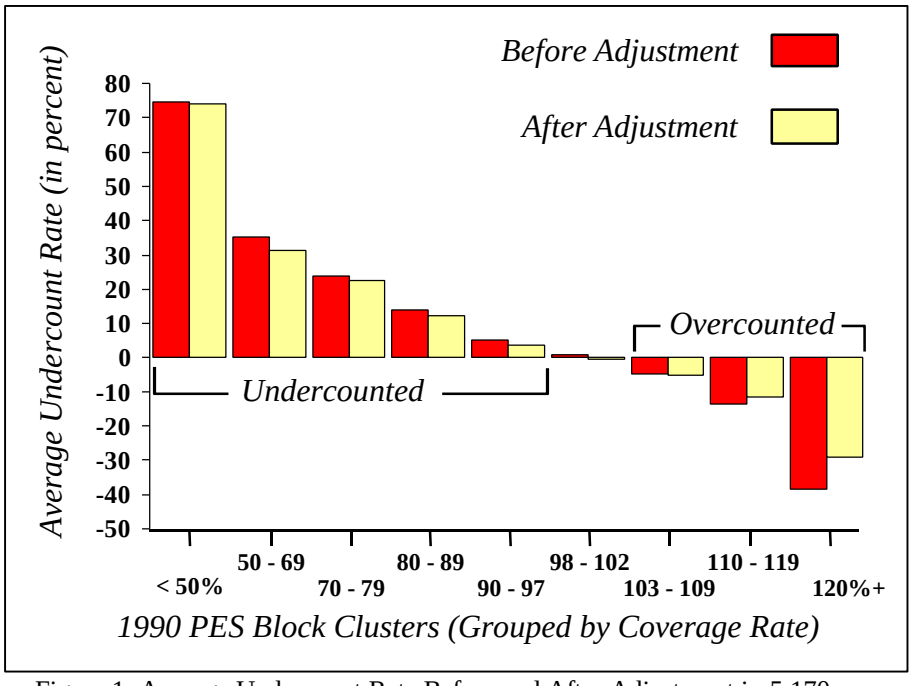


Figure 1: Average Undercount Rate Before and After Adjustment in 5,170 1990 PES Block Clusters.

The dark bars in Figure 1 show the undercount rates measured by the PES and the light bars show the undercounts remaining after statistical adjustment. If statistical adjustment had succeeded, the light bars would be much smaller than the dark bars. Instead, however, the two sets of bars are almost identical. Where the survey measured large undercounts, statistical adjustment tended to make just a small improvement. Where the survey measured overcounts, statistical adjustment also tended to make only small improvements – or to increase overcounts.

1. STATISTICAL ADJUSTMENT CANNOT CORRECT LARGE UNDERCOUNTS IN LOCAL AREAS.

Analysis of the 1990 PES adjustments reveals that statistical adjustment is insensitive to local undercount rates. The Bureau's adjustment would have added people marginally to undercounted areas and to many overcounted areas, instead of adding large numbers to areas with large undercounts, and lowering numbers from areas with overcounts. This approach is inadequate to overcome the unfair under-representation of heavily undercounted neighborhoods (Figure 2).

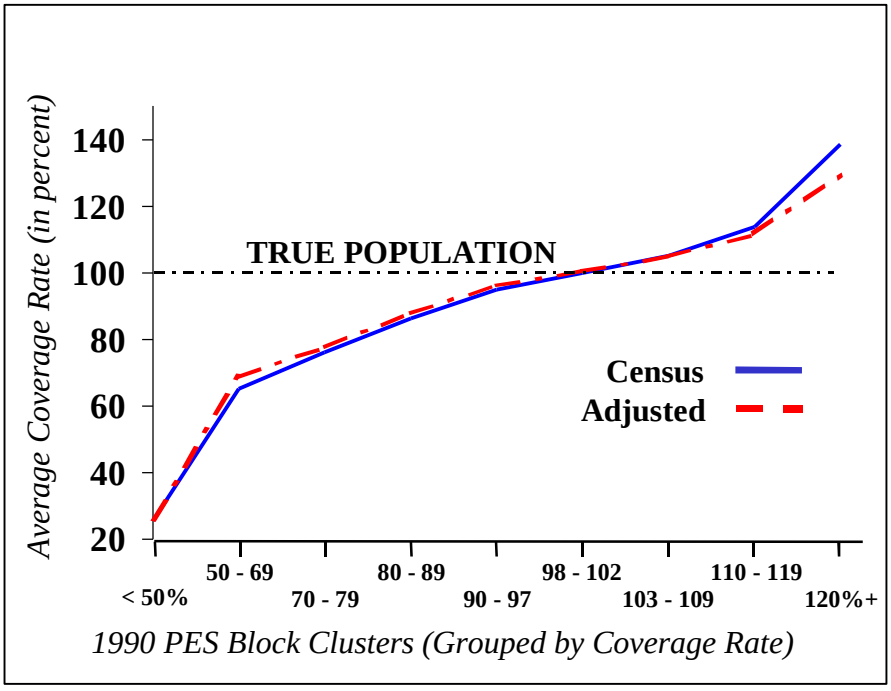


Figure 2: Average Coverage Rate Before and After Adjustment in 5,170 1990 PES Block Clusters.

If statistical adjustment succeeded, the line in Figure 2 representing adjusted coverage rates would be very similar to the horizontal line labeled "True Population." Instead, the line representing the adjusted coverage rates is almost identical to the line representing the census coverage rates.

Unfortunately, the insensitivity to local undercount rates is not unique to the 1990 PES. Any use of "synthetic estimation" to distribute undercount over a large region will likely have the same result – it will not correct large undercounts in local areas.

Poststratification Fails: By design, statistical adjustment using synthetic estimation does not correct extreme errors – large undercounts in certain minority areas, for example. This design depends on a false assumption, the homogeneity assumption: that people meeting certain demographic criteria have the same probability of being counted in the census, regardless of where they live in a region.

To try to satisfy the homogeneity assumption, the Bureau categorizes all people in each region by criteria such as age, race, gender, and tenure. Each combination of these categories forms a demographic *poststratum.* Poststrata cut across all the blocks in a region (that is, each block is divided into several poststrata, and each poststratum contains people from many blocks across a broad region). The Bureau multiplies the count of all members of a poststratum in each block by a common adjustment factor.

For example, in 1990, African American women between 18 and 29 paying rent in cities in a particular region formed one poststratum. In each block, each woman in that poststratum was multiplied by an adjustment factor. The same factor was used for that poststratum in every block in the entire region. An adjustment factor was applied to each person in each poststratum in each region. The adjusted poststrata totals in each block were summed, to determine adjusted population in each block.

The logic is that the census misses similar people at similar rates, so a common adjustment factor applied to people fitting a certain description should correct large undercounts. A high adjustment factor for the young, urban African American women in the example above is intended to make large additions to areas where many women fitting that description live.

However, real-life examples confound this logic. For example, African-American women renting in Chicago's affluent "Gold Coast" are not undercounted at the same rate as African American women renting a few miles away, in or around the Robert Taylor Homes public housing project.

If statistical adjustment in poststrata worked, one would expect to see many people added to areas where many people were missed. This did not occur.

Instead, heavily undercounted local areas surveyed in the 1990 PES would have received only marginal increases, on average (Figure 3). Correlation analysis shows that poststratification and synthetic estimation, as applied in the 1990 PES and planned for the 2000 ACE, are ineffectual in correcting large local undercounts. In the 1990 PES block clusters, there was practically no correlation between the size of the directly estimated undercount and the size of the proposed adjustment using synthetic estimation. The computed correlation was 0.097. If adjustment were effective, we would expect the correlation to be much stronger.

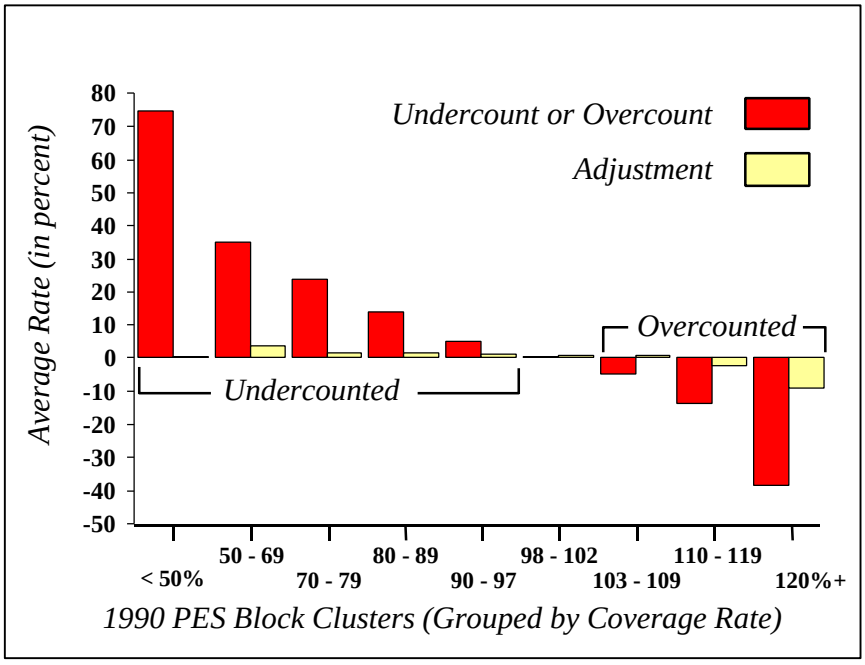


Figure 3: Average Undercount Rate vs. Average Adjustment Rate in 5,170 1990 PES Block Clusters.

In Figure 3, the two sets of bars would be very similar if statistical adjustment did a good job of correcting for the undercounts and overcounts measured by the PES. But the two sets of bars are radically different. The bars representing statistical adjustment are much smaller, indicating that statistical adjustment corrects only a negligible portion of the apparent net undercount measured by the PES. In fact, the bars representing statistical adjustment are all just about the same size. One bar, representing the average adjustment in 996 clusters with census overcounts between two and 10 percent, even points in the wrong direction: population is increased even where *overcounts* were measured.

Minority Neighborhoods Disproportionately Affected: The evidence clearly indicates that statistical adjustment will not correct large undercounts or overcounts in small areas such as blocks and neighborhoods. Heavily undercounted areas will remain heavily undercounted. Overcounts in many areas will be increased. Until the actual enumeration of people in the census is improved in heavily undercounted local areas, the differential undercount will persist at the local level.

- **A Case in Point: Robert Taylor Homes**

Robert Taylor Homes is the picture of a hard-to-count neighborhood: low-income residents in a dense urban area who do not trust government. Covering 15 city blocks in Chicago, Robert Taylor has 22 sixteen-story buildings. Ninety-nine percent of the residents are African American. Eighty-four percent earn less than $10,000 a year.

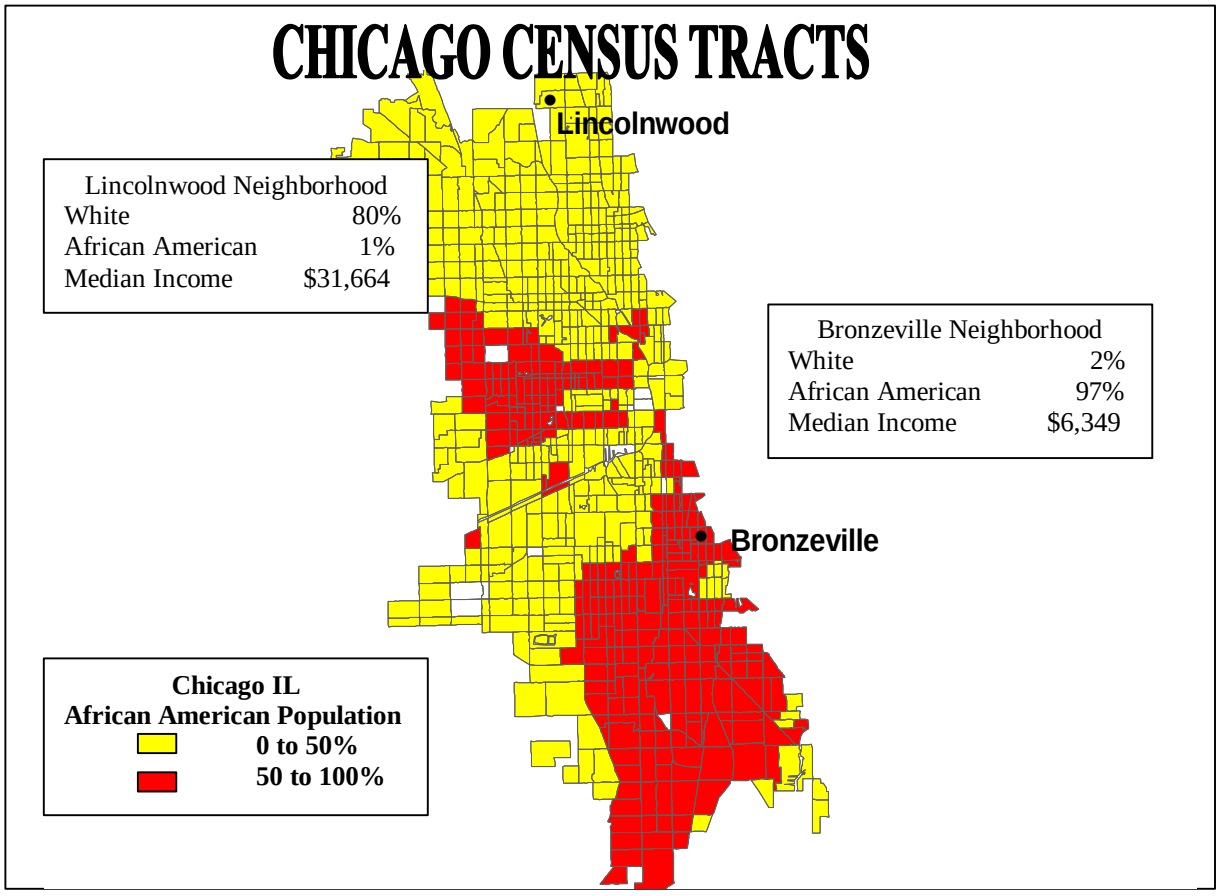
After the 1990 census, Chicago Housing Authority records showed the census missed 3,500 people in Robert Taylor – a local undercount of 29 percent.[4] Statistical adjustment would have added 673 – only a five percent increase, leaving one out of four Robert Taylor residents (counted by the Chicago Housing Authority) uncounted by the census.

Statistical adjustment would have added over 100,000 people to Illinois. But only 673 would have been added where they were needed most – in the 15 blocks of Robert Taylor Homes. While Illinois would have supposedly shown little or no undercount after adjustment, one in four Robert Taylor residents would have remained uncounted and without their fair share of representation or funding.

[4] Chicago Housing Authority, Statistical Profile: The Chicago Housing Authority, 1991 to 1992, prepared by CHA Management of Information Systems Department and the Department of Research and Program Development, (Chicago, IL, 1992), 5.

- **A 2nd Case in Point: Lincolnwood and Bronzeville**

The 1990 PES surveyed block clusters in Lincolnwood and Bronzeville, two Chicago neighborhoods. The cluster surveyed in Bronzeville was heavily *undercounted* by about 17 percent. The cluster surveyed in Lincolnwood was heavily *overcounted* by about 29 percent. Yet statistical adjustment would have added people to the neighborhoods surrounding *both* surveyed clusters. According to the Bureau's official figures, statistical adjustment would have added only 4.4 percent to the neighborhood surrounding the heavily undercounted Bronzeville cluster. Statistical adjustment would have added 1.1 percent to the neighborhood surrounding the heavily *overcounted* Lincolnwood cluster[5] (see map).



[5] Adjustments calculated as a percentage of the official census count, as reported at www.census.gov.

In a nation as large, diverse and mobile as ours, some people will be missed. That is unavoidable. But statistical adjustment consistently misses the same people in the same communities – the very communities that depend on the census for the health care, schools and child care that come with being counted. The solution is to actually count people better, not to rely upon a statistical adjustment incapable of correcting the problem.

2. STATISTICAL ADJUSTMENT WILL ADD PEOPLE TO MANY OVERCOUNTED AREAS.

In the 1990 PES block clusters, 1,379 (out of 5,170) were overcounted.[6] The PES indicated the population as reported in these areas should have been reduced by an average of 8.9 percent. However, statistical adjustment would have only reduced the population by an average of one (1.0) percent. In addition, the majority of overcounted PES block clusters (996) would have received an average *increase* of 0.17 percent (Figure 1).

Credibility Gap: Local leaders (such as city council members, mayors, city planners, social service providers, school boards, county commissioners, etc.) who expect substantial gains in representation and funding through statistical adjustment are likely to be disappointed. The data show that blocks and neighborhoods will experience only marginal change, regardless of the severity of their undercount or overcount (Figure 3).

On average, statistical adjustment marginally reduces undercounts in undercounted blocks. In addition, statistical adjustment adds to overcounts in many overcounted blocks. Statistical adjustment does not alleviate the essential unfairness to heavily undercounted communities. By asserting the adjustment will correct undercounts in 2000, proponents of statistical adjustment have raised unrealistic expectations likely to create a "credibility gap" that may compromise future census efforts and condemn future generations to continued undercounts.

Director of the Census Bureau, Dr. Kenneth Prewitt said, "I don't want to create a census in 2000 that creates a problem for 2010. We could over promise and then come in 2010 and the communities will say, 'My God, you told us that if we filled this form out, we were going get a new school, so where is it?'"[7]

[6] That is, census coverage rate was equal to or higher than 102 percent.
[7] U.S. Census Monitoring Board, *Young & Rubicam briefing*, 8 July 1999.

- **A Case in Point: Sacramento City Council Districts**

Sacramento City Council member Lauren Hammond testified that the severe undercount in her district – a predominantly African American area – resulted in insufficient resources for a much-needed middle school: "I have four high schools and eight elementary schools and not one darn middle school because there aren't the numbers to reflect that."[8]

Logic and school enrollment suggest Councilwoman Hammond's district needs a middle school. She believes statistical adjustment will help get one. The evidence suggests that it won't.

Marginal increases in the population of heavily undercounted blocks in the Councilwoman's district would be unlikely to strengthen her case for a middle school. In fact, her argument might be weakened, since increases in her district would be offset by comparable increases to undercounted and overcounted blocks in neighboring districts.

Aggregation of adjustments at higher levels of geography (big cities, states), is likely to produce a harmful result: Councilwoman Hammond's reports of egregious undercounts in her district would be contradicted by the appearance of only minor undercounts in Sacramento, California and the nation. Legitimate complaints about census undercount would be precluded by the notion that the undercount had been officially eliminated. Adjustment can appear to substantially reduce the undercount at the big-city and state level, without correcting undercounts in local areas.

[8] U.S. Census Monitoring Board, *Public Hearing*, 16 December 1998.

Section III

RECOMMENDATIONS

THE TRUTH: LOCAL DATA DELIVER

Clearly, statistical adjustment will fail the very communities that depend on the census most for the schools, health care and child care that come with being counted. Local and state leaders' efforts to get constituents counted will do much more to ensure a fair share than statistical adjustment ever could.

The Congressional members of the Board, therefore, recommend a concerted effort on the part of Congress, the Administration, the Census Bureau and state and local leaders to frankly acknowledge that the Bureau's proposed statistical adjustment will not correct large undercounts that often affect minority neighborhoods.

We recommend Congress, the Administration and the Census Bureau aggressively pursue the following solutions to improve the census in these neighborhoods:

1. Re-direct census funding to the Census Bureau's Regional Census Offices for discretionary spending;
2. Re-direct support, including funding, to the Bureau's community partners in state and local Complete Count Committees;
3. Shift attention, planning and resources to count neighborhoods the Bureau has identified as Hard-to-Count;
4. Remove barriers to census employment in hard-to-count areas (for example, suspend testing requirements for Spanish-fluent enumerators in Latino colonias along the Mexican border);
5. Support aggressive use of Be Counted forms in hard-to-count areas.

Additionally, we recommend state and local governments take the following steps:

1. Join or form a Complete Count Committee (CCC);
2. Work with the Census Bureau's Local Census Office (LCO) and the city planning department to identify and direct extra efforts to hard-to-count neighborhoods;
3. Meet with community leaders in hard-to-count neighborhoods for specific ideas and suggestions about how to get complete counts in those areas;
4. Work with community leaders and the Bureau to ensure census workers are hired from hard-to-count neighborhoods. Ask your LCO to recruit trusted third parties who could serve as hard-to-count "guides" for enumerators;
5. Work with state government to grant a waiver(s) to those for whom Census 2000 employment may offset eligibility for government assistance programs such as TANF or other federally funded programs.
6. Work aggressively with the Bureau to ensure there are enough foreign-language census materials to meet the language needs of people in your community.

Finally, recommendations made to Congress and the Bureau in previous reports may still improve the census in local areas:

1. Reinstate the Parollee/Probationer Coverage Improvement Program;
2. Reinstate Post Census Local Review of census counts.

Section IV

METHODOLOGY

CENSUS BUREAU'S PLAN FOR STATISTICAL ADJUSTMENT

- **Measuring the Undercount: Dual System Estimation**

In 2000 (as in 1990), the Bureau will attempt to measure the undercount by means of dual system estimation (DSE). DSE is a method for deriving an estimate of true population by combining the results of two separate surveys.

In 1990, the Bureau used the census and the 1990 post enumeration survey (PES) in its DSE. In 2000, the Bureau will use the census and the Accuracy and Coverage Evaluation (ACE), a revision of the 1990 PES.

The experts cited in Appendix E provide ample evidence that this manner of measuring undercount is seriously flawed. In the case of the 1990 PES, the Census Bureau's own senior statisticians and demographers concluded, "About 45% of the revised estimated undercount is actually measured bias and not measured undercount."[9] In other words, in the 1990 PES, many of the people reported as missed or counted in error really just represented problems in the PES itself, such as faulty information, errors in matching between the census and the survey, etc. Although concern over local undercounts is valid, the survey's measurement of those undercounts is not.

Proponents of the Bureau's adjustment plan consistently dismiss scientific criticism, and often attempt to personally discredit the critics. Nevertheless, the scientific literature casts serious doubt on the ability of the Bureau's plan to accurately measure the undercount at any level. A bibliography of relevant scientific criticism is found in Appendix E.

- **Distributing the Adjustment: Synthetic Estimation**

After measuring the 2000 undercount through DSE, the Bureau proposes to distribute adjustment via synthetic estimation. The same fundamental approach was proposed and rejected in the 1990 Census.

Synthetic estimation uses estimates about the undercount of a demographic group in a large region to make estimates about that same

[9] Department of Commerce, Bureau of the Census, Committee on Adjustment of Postcensal Estimates (CAPE Committee), *Assessment of Accuracy of Adjusted Versus Unadjusted 1990 Census Base for use in Intercensal Estimates* (Washington, DC, 7 August 1992), p. 15.

demographic group in smaller areas, such as blocks or neighborhoods. Synthetic estimation assumes the undercount rate for this demographic group, in every block, is the same as the undercount rate for the whole region.

In the 1990 PES, the Bureau used information from regions containing several states to estimate the number and characteristics of people in all blocks within those regions.

For example, the 1990 synthetic estimation applied a single adjustment factor to all male Asian and Pacific Islanders between the ages of 18 and 29 who owned their home. In other words, the census assumed young, male homeowners of Chinese, Japanese, Philippine and Korean descent, in every neighborhood from Honolulu, Hawaii, to Bangor, Maine, had the same likelihood of being counted in the census.

Although the proposed 1990 statistical adjustment was rejected, the Bureau proposes to adjust Census 2000 in essentially the same way.

- **Method of Analysis**

To determine the local effects of the Census 2000 adjustment, the Congressional members of the Board examined the 5,170 block clusters surveyed in the 1990 PES.[10] Using data only recently made available to the public, census counts in the 1990 PES block clusters were compared to their synthetically adjusted counts.[11]

The ideal analysis would compare the census count and the synthetically adjusted count in each block to the true population in that block. However, the true population is not known. Therefore, we compare the census count and the synthetically adjusted count to a third number: a direct estimate of the population for each block cluster, based on the data from that block cluster. The third number is the "direct DSE."

The direct DSE is an estimate of the population of each block cluster, calculated by the Bureau, based on the data collected from that block

[10] A block cluster is a single block or a group of two, three or four blocks, or a portion of a large block. Although the file provided to the Board contains data for 5,180 block clusters, there were 10 block clusters where the PES sample failed to find any of the people counted in the census. In this case the match variable "M" (see page 16) is zero. Since this variable is the denominator in the standard direct DSE calculation, the calculation cannot be performed. In mathematical terms, the result of a division by zero is "undefined."

[11] The data used is the revised 1990 PES file, with 357 poststrata. This file contains information on those areas surveyed in the 1990 PES. To our knowledge, its release is the first time the Bureau has identified the exact blocks used in the 1990 PES to an agency outside the Bureau.

cluster in both the actual census, and in the PES. The direct DSE is compared to the synthetic adjustment for each block cluster, to determine how well the synthetic adjustments "fix" the apparent undercounts measured by the PES in each block cluster.[12]

The Bureau calculated the direct DSE for each of the PES block clusters, but has resisted its release to the public. Data necessary to calculate the direct DSE exist only for the block clusters surveyed in the 1990 PES.

Several statisticians and demographers have criticized the accuracy of data from the 1990 PES and the reliability of the resulting DSE, citing lack of independence, errors in matching, etc. For a bibliography of these criticisms, see Appendix E.

Nevertheless, it is pertinent to compare statistical adjustments to the local measurements of undercount upon which they are based. Such a comparison shows whether statistical adjustment succeeds in adding people to the areas where the survey determined they had been missed.

If the adjustments do not correct the undercounts and overcounts in the sample area, they can hardly be counted upon to correct undercounts and overcounts in the rest of the nation. If statistical adjustment fails to add a large number people to block clusters where large undercounts are supposedly measured, and if they add people even to areas where *overcounts* are supposedly measured, then they are obviously not the solution to the problem of faulty census counts. Even if the measurements of undercount were accurate, such adjustments would be grossly inaccurate.

U.S. CENSUS MONITORING BOARD STUDY

The Census Bureau provided the Census Monitoring Board with detailed data on the 5,170 block clusters included in the 1990 Post Enumeration Survey (PES). A description of the data and the Board's analysis follows.[13]

Each of 5,170 PES block clusters was located and identified by Federal Information Processing Standard (FIPS) codes denoting its state and

[12] We performed identical analysis, substituting the "Census Plus" estimate for the direct DSE. Census Plus is an alternative population estimate calculated by the Census Bureau. Use of this substitute did not yield appreciably different results or conclusions. Here we report our results using the direct DSE as a standard.
[13] Data is the revised 1990 PES file, with 357 poststrata.

county, as well as Bureau codes identifying census tract and cluster number. The data included these figures for each 1990 PES block cluster:

1. The **E-Sample** (E). The number of people counted in the cluster during the 1990 census. **This count excludes whole person imputations.** Used by the Bureau for analysis, and to generate synthetic estimates.

2. The **erroneous enumerations** (EE). The estimated number of persons in the E-Sample who were erroneously enumerated or for whom there was not sufficient information for matching.

3. The **P-Sample** (P). The number of persons counted in the 1990 Post-Enumeration Survey (PES).

4. **Matches** (M). The estimated number of P-Sample persons who could be matched to census persons.

5. The **adjusted count** (SynDSE). The synthetic estimate generated by the 1990 PES. It is the E-Sample, adjusted by the Bureau's proposed statistical adjustment using synthetic estimation.[14]

6. The **direct dual system estimate** (DirDSE). An estimate of the population of each block cluster based on the E-Sample and P-Sample from only that cluster. It has been generated internally by the Bureau, defined as:

$$DirDSE = E \times \frac{E - EE}{E} \times \frac{P}{M}$$

In addition, the Bureau provided a separate file, the PES Block file, with block-level data about the areas surveyed in the PES. Specifically, the PES Block file provided the number of persons *imputed* into each cluster. Although the PES Block file overstates the number of imputations,[15] the Board computed **E2**: E, plus the number of imputations in the cluster (according to the PES Block file). E2 approximates the census count

[14] The Bureau adjusts population counts via synthetic estimation using poststrata. Poststrata are subpopulations that result from crossclassification of the population by age, sex, race or origin, tenure (owner or renter), and geographic type (urbanized area). In the revised 1990 PES data used in this analysis, there were 357 such poststrata. Poststrata may span thousands of blocks and several states in a region. Each census block generally contains members from many poststrata. Synthetic estimation multiplies the census count of all members of each poststrata in each block by a common adjustment factor. In a given block, the resulting adjusted counts for each poststrata are then added together, to determine an adjusted total for each block.

[15] Due to subsampling (surveying only parts of some blocks), the PES Block file slightly overstates the number of imputations in the sample. In addition, the PES Block file appears to overstate the total population of the sample. For details, see Section V: Limitations.

reported for each area surveyed in the 1990 PES. The 5,170 block clusters were indexed by *i* where *i* = 1 to 5,170.

$E2_i = E_i + Imputations_i$

The census coverage rate for the i^{th} block cluster was defined as the census count, E2, divided by DirDSE, and then multiplied by 100 for expression as a percentage. This is the percentage of each block cluster's estimated true population reported in the census.

$$Census\ Coverage\ Rate = \frac{E2_i}{DirDSE_i} \times 100$$

For comparison, the adjusted coverage rate for the i^{th} block cluster was also calculated. It was defined as the adjusted count, SynDSE, divided by the DirDSE, and multiplied by 100 for expression as a percentage. It is the percentage of each block cluster's estimated true population reported after adjustment.

$$Adjusted\ Coverage\ Rate = \frac{SynDSE_i}{DirDSE_i} \times 100$$

Comparison of these two variables in local areas – census coverage rate and adjusted coverage rate – is the basis of this analysis. Specifically, this analysis compares census coverage before and after adjustment in block clusters with varying coverage rates (and, therefore, varying undercount rates). To do so, PES clusters were sorted according to coverage rate.[16] The groups were denoted as J.

Coverage Group (J)	Census Coverage Rate	Number of Clusters
1	<50%	42
2	50 – 69.9	46
3	70 – 79.9	111
4	80 – 89.9	463
5	90 – 97.9	1,538
6	98 – 101.9	1,591
7	102 – 109.9	996
8	110 – 119.9	240
9	120%+	143

Groups 1 through 5 include all PES block clusters with less than 98 percent coverage: undercounted clusters. Groups 7 through 9 include all PES block clusters with 102 percent coverage or more: overcounted

[16] The estimates for any given block cluster are subject to random variation. By sorting the block clusters into these groups, that random variation is reduced to a minimum.

clusters. Group 6 includes clusters with coverage rates between 98 and 102 percent: defined for the purposes of this analysis as accurately-counted clusters.

To compare the local undercount to the local adjustment, an average undercount rate, as a percent of the estimated true population, was calculated for all clusters in each group. (All the clusters in the j^{th} group are indexed by I (sub j)). The average census undercount was defined as the sum of all census counts, E2, in a group, divided by the sum of all DirDSE in the group, and subtracted from one. The value was multiplied by 100 for expression as a percentage. A positive value indicates an undercount. A negative value indicates an overcount.

$$Average\ Undercount = \left[1 - \frac{\sum_{i \in I_j} E2_i}{\sum_{i \in I_j} DirDSE_i}\right] \times 100$$

For example, an aggregated average of the coverage rates of the 463 block clusters with coverage rates between 80 and 89.9 percent was obtained, subtracted from 1.0 and the result multiplied by 100 to yield an average undercount rate of 13.76 percent. The same procedure was repeated for each group (Appendix D).

Finally, the average local adjustment was calculated as a percent of the estimated true population. That is, the adjustment for the i^{th} block cluster via synthetic adjustment was defined as the value of SynDSE minus E2, divided by DirDSE. Again, the values were grouped and averaged, and the value multiplied by 100 for representation as a percentage.

$$Average\ Adjustment = \left[\frac{\sum_{i \in I_j} (SynDSE_i - E2_i)}{\sum_{i \in I_j} DirDSE_i}\right] \times 100$$

This is the average addition to block clusters in a group, expressed as a proportion of the estimated true population. For example, in the 996 block clusters with coverage rates between 102 percent and 109.9 percent (overcounted areas), the average addition is 0.17 percent.

Comparing the coverage rate before and after adjustment in the 1990 PES block clusters, as well as the local undercount and the local adjustment, clearly illustrates statistical adjustment's failure to correct large undercounts in local areas.

Section V

LIMITATIONS

- **MEASURING "TRUE" POPULATION.** To measure relative undercount and adjustment, a standard of true population must be set for each block cluster. For the purposes of this analysis, we accept the Census Bureau's direct dual system estimate (direct DSE) for each block cluster as this standard.

Statisticians and demographers have provided ample evidence that the DSE is seriously flawed (Appendix E). Therefore, the validity of using the direct DSE as a standard may be debated.

However, the statistical adjustment proposed by the Bureau is based on a fundamental assumption that the dual system estimate methodology is valid. Therefore, this analysis examines an "even if" question: Even if the direct DSE accurately measures local undercounts, would the proposed statistical adjustment using synthetic estimation correct those undercounts?

- **RESULTS ARE UNWEIGHTED.** This analysis examines all but 10 of the 5,180 block clusters surveyed in the 1990 PES: an excellent basis for conclusions about the effect of statistical adjustment **in those block clusters.**[17] It is also appropriate to draw some conclusions about the efffects of statistical adjustment in general. For example, the results in these clusters clearly indicate statistical adjustment fails to correct large undercounts in local areas.

However, without weighting, these data should not be used to make inferences about larger areas such as states and the nation as a whole. Heavily undercounted areas appear in the 1990 PES block clusters with greater frequency than if the sample were drawn randomly. For example, 13 percent of the block clusters in the sample have undercounts higher than 10 percent. In the nation as a whole, the percent of heavily undercounted clusters would be somewhat lower. Therefore, without the proper weighting of the sample, it is impossible to compare the proportion of heavily undercounted areas in the 1990 PES block clusters to the proportion of heavily undercounted areas in the nation.

[17] There were 10 block clusters where the PES sample failed to find any of the people counted in the census. In this case the match variable "M" (see page 17) is zero. Since this variable is the denominator in the standard direct DSE calculation, the calculation cannot be performed. In mathematical terms, the result of a division by zero is "undefined."

This over-sampling of heavily undercounted areas results in a similar over-sampling of racial and ethnic minorities. As noted by the Bureau,[18] heavily undercounted areas tend to have high minority populations. African Americans, Latinos, Asians and American Indians also appear in the 1990 PES block clusters with greater frequency than if the sample were a simple random sample.

While over-sampling limits the use of this sample as a proportionate reflection of the nation, the relatively large sample of undercounted areas allows for a better picture of undercounted neighborhoods with predominantly minority residents. At publication, the Board did not have the weights for the 1990 PES block clusters. We have requested these data, and will continue analysis.

- **DIFFERENCES DUE TO IMPUTATION.** At publication, the whole-person imputations reported in the official census counts in the block clusters surveyed in the 1990 PES were not available. Although the Bureau provided whole-person imputations for every block included in the PES, some of those blocks were *subsampled*. That is, portions of some blocks were not included in the survey. At publication, the Bureau was unable to separate imputations made in block portions included in the PES, from imputations made in block portions not included in the PES. Therefore, data provided to the Board overstate the number of imputations in the areas surveyed in the PES.

The Bureau does not include imputations in the calculation of the block-level direct DSE, because imputations in the census cannot be matched to responses from the PES. This analysis uses the exact value for the direct DSE provided by the Bureau.

However, we added imputations to the E-Sample counts for the purposes of this analysis. Adding imputations to the E-Sample counts approximates the official census counts in those areas. We believe the approximations will be very close. The error from this source is small, and does not affect the conclusions of this paper.

[18] Department of Commerce, Bureau of the Census, *Report to Congress – The Plan for Census 2000* (Washington, DC, August 1997), 6.

DISCREPANCIES IN BUREAU DATA. In the course of this analysis, the Bureau provided the Board with two data files. One file, with block cluster-level data for those areas surveyed in the PES, is the **XYZ file**. The XYZ file was delivered on April 6, 1999, in response to a request submitted by the Congressional members of the Board in December 1998. The second file, the **PES Block file**, provides block-level data, including counts by race and Hispanic origin, for those same areas. It was provided upon request in July 1999.

When analysis uncovered discrepancies between the population counts in the XYZ file and the PES Block file, the Congressional members of the Board asked for an explanation. In August 1999, Bureau staff reported that an error in creating the PES Block file had resulted in inaccurate data, and undertook a revision. On September 10, 1999, the revised PES Block file was delivered to the Board, with assurances that the data had been corrected, and accompanying documentation. Again, the counts did not match.

The XYZ file reported that, in the 5,170 block clusters included in this analysis, the E-Sample count (not including imputations) was **392,543 persons**. The revised PES Block file reported that, in the 5,170 block clusters included in this analysis, the E-Sample count (not including imputations) was **446,099 persons**. From discussions with Bureau staff, and pertinent scientific literature, we surmise the XYZ file is accurate.

Rather than wait indefinitely on a corrected PES Block file, we simply noted the discrepancy. One set of data from the PES Block file was used in this report: the number of imputations in each cluster (a total of 5,905, which did not change in the Bureau's revision). We have requested the corrected file.

In addition, the Bureau previously has reported that 5,290 block clusters were included in the 1990 PES. Through discussions with Bureau staff, we surmise that clusters with no recorded population were not included in the XYZ file or the PES Block file.

Our analysis of the effects of adjustment on people of various race and Hispanic origin has been seriously compromised by these data discrepancies. We will complete this analysis when the Bureau provides correct data on race and Hispanic origin, and PES weights for the block clusters.

Section
VI

APPENDIX

A. Average Coverage Rate Before and After Adjustment

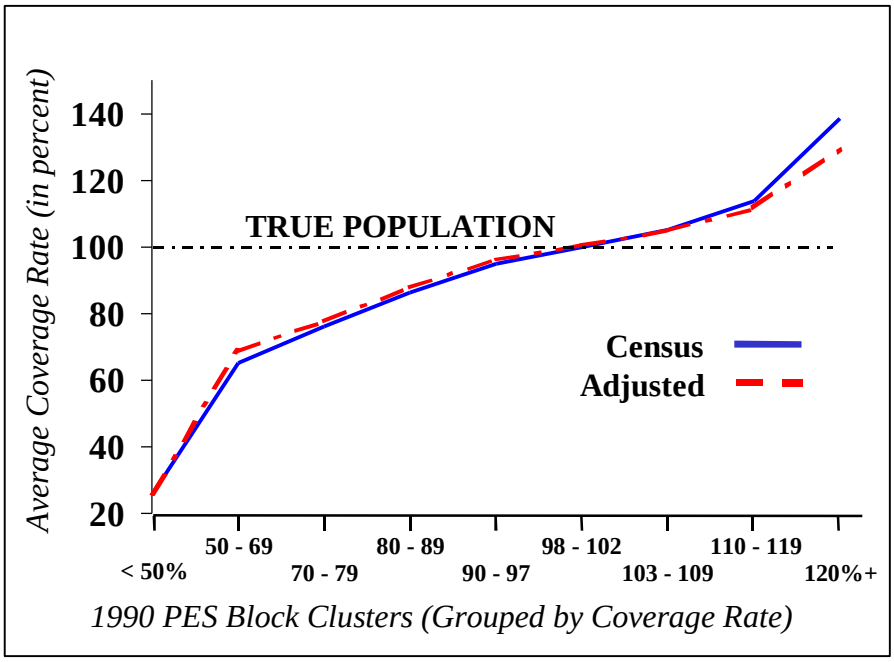
B. Average Undercount Rate Before and After Adjustment

C. Average Undercount vs. Average Adjustment

D. Equations

E. Scientific Criticism

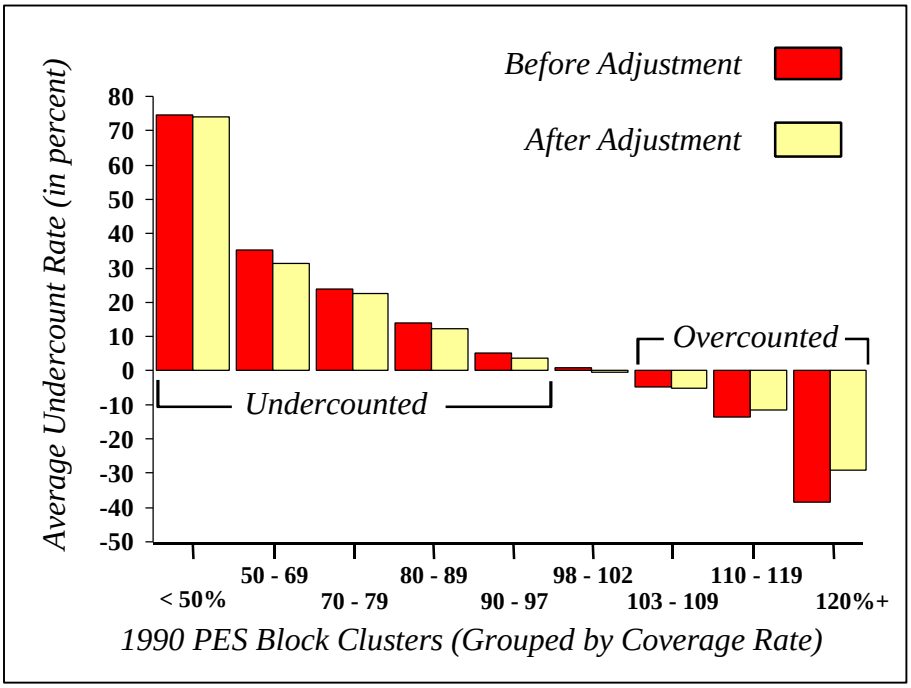
APPENDIX A
Average Coverage Rate
Before and After Adjustment
5,170 1990 PES Block Clusters



Coverage Rate Group	Number of Clusters	Average Census Coverage Rate (percent)	Average Adjusted Coverage Rate (percent)
<50%	42	25.3426	25.9006
50 – 69.9	46	64.8909	68.7447
70 – 79.9	111	76.0130	77.6127
80 – 89.9	463	86.2357	87.7499
90 – 97.9	1,538	94.9515	96.1709
98 – 101.9	1,591	99.9261	100.5239
102 – 109.9	996	104.8856	105.0538
110 – 119.9	240	113.6864	111.3537
120%+	143	138.3281	129.3282

<50% – 89.9	662	81.6930	83.4089
<50% - 97.9	2,200	91.0327	92.3989
102 – 120%+	1,379	108.9231	107.9685

APPENDIX B
Average Undercount Rate
Before and After Adjustment
5,170 1990 PES Block Clusters

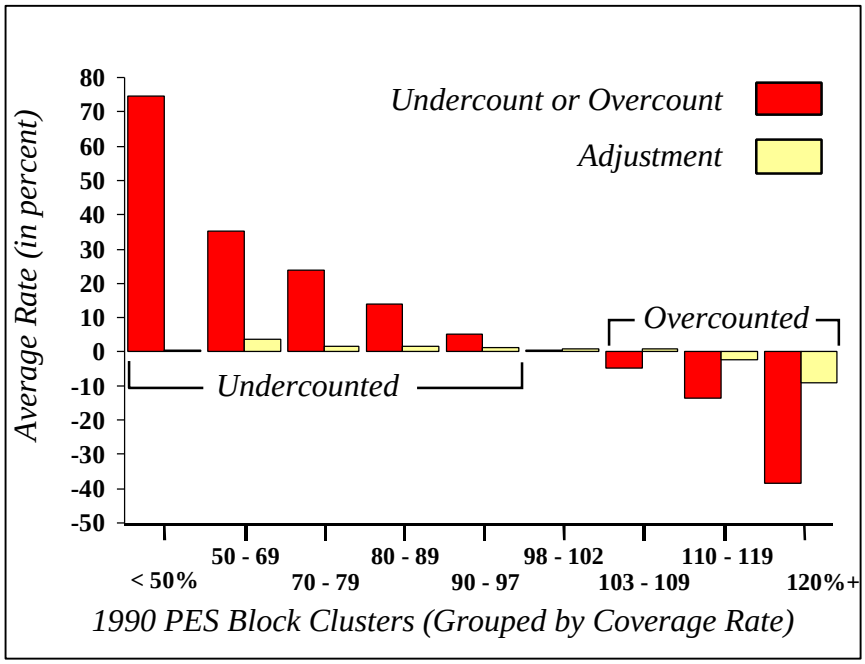


Data Shown as Number of Persons in 5,170 PES Block Clusters

Coverage Rate Group	Number of Clusters	Total Census Count	Total Adjusted Count	Total Direct DSE	Average Undercount Rate Before Adjustment (percent)*	Average Undercount Rate After Adjustment (percent)*
<50%	42	238	243.24	939.13	74.6574	74.0994
50 – 69.9	46	3,056	3,237.49	4,709.44	35.1091	31.2553
70 – 79.9	111	6,432	6,567.36	8,461.71	23.9870	22.3873
80 – 89.9	463	34,191	34,791.36	39,648.33	13.7643	12.2501
90 – 97.9	1,538	121,657	123,219.30	128,125.40	5.0485	3.8291
98 – 101.9	1,591	124,337	125,080.90	124,429.00	0.0739	-0.5239
102 – 109.9	996	78,007	78,132.09	74,373.40	-4.8856	-5.0538
110 – 119.9	240	20,431	20,011.79	17,971.37	-13.6864	-11.3537
120%+	143	10,099	9,441.94	7,300.76	-38.3281	-29.3282
<50% – 89.9	662	43,917	44,839.45	53,758.61	18.3070	16.5911
<50% - 97.9	2,200	165,574	168,058.75	181,884.01	8.9673	7.6011
102 – 120%+	1,379	108,537	107,585.82	99,645.53	-8.9231	-7.9685

* Negative values indicate overcounts.

APPENDIX C
Average Undercount Rate vs. Average Adjustment Rate 5,170 1990 PES Block Clusters



Data Shown as Number of Persons in 5,170 PES Block Clusters

Coverage Rate Group	Number of Clusters	Total Census Count	Total Adjusted Count	Total Direct DSE	Average Undercount Rate Before Adjustment (percent)*	Average Adjustment (percent)
<50%	42	238	243.24	939.13	74.6574	0.5580
50 – 69.9	46	3,056	3,237.49	4,709.44	35.1091	3.8537
70 – 79.9	111	6,432	6,567.36	8,461.71	23.9870	1.5997
80 – 89.9	463	34,191	34,791.36	39,648.33	13.7643	1.5142
90 – 97.9	1,538	121,657	123,219.30	128,125.40	5.0485	1.2194
98 – 101.9	1,591	124,337	125,080.90	124,429.00	0.0739	0.5979
102 – 109.9	996	78,007	78,132.09	74,373.40	-4.8856	0.1682
110 – 119.9	240	20,431	20,011.79	17,971.37	-13.6864	-2.3327
120%+	143	10,099	9,441.94	7,300.76	-38.3281	-8.9999
<50% – 89.9	662	43,917	44,839.45	53,758.61	18.3070	1.7159
<50% - 97.9	2,200	165,574	168,058.75	181,884.01	8.9673	1.3661
102 – 120%+	1,379	108,537	107,585.82	99,645.53	-8.9231	-0.9546

* Negative values indicate overcounts.

APPENDIX D
Equations

$$DirDSE = E \times \frac{E - EE}{E} \times \frac{P}{M}$$

$$E2 = E + Imputations$$

$$Census\ Coverage\ Rate = \frac{E2_i}{DirDSE_i} \times 100$$

$$Adjusted\ Coverage\ Rate = \frac{SynDSE_i}{DirDSE_i} \times 100$$

$$Average\ Undercount = \left[1 - \frac{\sum_{i \in I_j} E2_i}{\sum_{i \in I_j} DirDSE_i} \right] \times 100$$

$$Average\ Adjustment = \left[\frac{\sum_{i \in I_j} \left(SynDSE_i - E2_i \right)}{\sum_{i \in I_j} DirDSE_i} \right] \times 100$$

Coverage Group (J)	Census Coverage Rate	Number of Clusters
1	<50%	42
2	50 – 69.9	46
3	70 – 79.9	111
4	80 – 89.9	463
5	90 – 97.9	1,538
6	98 – 101.9	1,591
7	102 – 109.9	996
8	110 – 119.9	240
9	120%+	143

APPENDIX E
Scientific Criticism

Committee on Adjustment of Postcensal Estimates. Assessment of accuracy of adjusted versus unadjusted 1990 Census base for use in intercensal estimates (C.A.P.E. Report). Technical report, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, 1992.

L. Breiman. The 1991 Census Adjustment: Undercount or bad data? *Stat. Sci.*, 9:458 – 537, 1994.

L.D. Brown, M.L. Eaton, D.A. Freedman, S.P. Klein, R.A. Olshen, K.W. Wachter, M.T. Wells, and D. Ylvisaker. Statistical controversies in Census 2000. Technical Report 537, Dept. Statistics, Univ. of Calif., Berkeley, January 1999.

U.S. Department of Commerce, Office of the Secretary. Decision of the Secretary of Commerce on Whether a Statistical Adjustment of the 1990 Census of Population and Housing Should Be Made for Coverage Deficiencies Resulting in an Overcount or Undercount of the Population. ACTION: Notice of final decision. *Federal Register*, 56 FR 140, 22 July 1991.

U.S. Department of Commerce, Economics and Statistics Administration Bureau of the Census. Decision of the Director of the Bureau of the Census on whether to use information from the 1990 Post-Enumeration Survey (PES) to adjust the base for intercensal population estimates produced by the Bureau of the Census. ACTION: Notice of final decision. *Federal Register*, 58 FR 69, 1993.

K. Darga. Fixing the Census Until It Breaks. Washington, D.C., publication planned for 2000.

K. Darga. Sampling and the Census: A Case Against the Proposed Adjustments for Undercount. AEI Press, Washington, D.C., 1999.

M.C. Davis and P. Biemer. 1990 Post-Enumeration Survey Evaluation Project P7 Series #H-2, Estimates of P-sample clerical matching error from a rematching evaluation. Technical report, U. S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, 1991.

D. Freedman. Adjusting the 1990 Census. *Science*, 252:1233 –1236, 1991.

D. Freedman. Adjusting the Census of 1990. *Jurimetrics*, 34:99 – 106, 1993.

D. Freedman and K. Wachter. Heterogeneity and Census Adjustment for the Intercensal Base. *Stat. Sci.*, 9: 458 – 537, 1994.

D. Freedman and W.C. Navidi. Should we have adjusted the U.S. Census of 1980? Survey Methodology, 18:3 – 74, 1992.

D. Freedman, K. Wachter, R. Cutler, and S. Klein. Adjusting the Census of 1990: Loss functions. Evaluation Review, 18:243 – 280, 1994.

Report to the Committee on Governmental Affairs, U.S. Senate. 2000 Census: Preparations for dress rehearsal leave many unanswered questions. Technical Report GAO/GGD-98-74, U.S. General Accounting Office, Washington, D.C., 1998.

P.B. Stark. Differences Between the 1990 and 2000 Census Adjustment Plans, and their Impact on Error. Technical Report 550, Dept. Statistics, Univ. of Calif., Berkeley, March 1999.

K. Wachter. Recommendations on 1990 Census adjustment, report to the Secretary of Commerce as a member of Special Advisory Panel. Technical report, U.S. Department of Commerce, 1991.

K. Wachter. Comment: Ignoring nonignorable effects. J. Am. Stat. Assoc., 88:1161 - 1163, 1993.