A FOLLOWUP FORUM ON CENSUS UNDERCOUNTS AND PREPARATIONS FOR THE 1990 CENSUS

NEW YORK STATE

ADVISORY COMMITTEE

TO THE UNITED STATES

COMMISSION ON CIVIL RIGHTS

This summary report of the New York State Advisory Committee to the United States Commission on Civil Rights was prepared for the information and consideration of the Commission. Statements and viewpoints in the report should not be attributed to the Commission or to the Advisory Committee, but only to individual participants in the community forum where the information was gathered.

A SUMMARY REPORT

JANUARY 1990

The Philadelphia Inquirer

SUNDAY
February 4, 1990

BUSINESS

SECTION C

Would an estimate make census count more accurate?

By Neill A. Borowski
Inquirer Staff Writer

Will the 1990 Census be the customary head count of everyone living in the United States?

Or will it also include an estimate of those it has traditionally missed?

The issue isn't just one of dry statistics. It strikes at the very heart of equal representation and the financial well-being of America's cities, say advocates of the estimates.

A traditional count would miss millions of blacks, of the poor and people who live in cities, say those who favor including an estimate.

Opponents, however, say such estimates would undermine the reliability of a count considered important enough to be mandated in the Constitution.

"Adjusting the figures would only undermine the credibility of what has become the standard for counting America's population," said South Dakota Gov. George S. Mickelson, in objecting to adding the estimates.

But Carnegie Mellon University statistician Joseph B. Kadane terms the need for adjusting the count "a civil rights issue."

"When the Constitution was written, a black person, then almost certainly a slave, counted for 60 percent of a white person. By 1980, a black person would count for about 94 percent of a white person," said Kadane. About 6 percent of the black population was missed that year, according to the Census Bureau.

"The 1990 census might be the first in which every American, black or white, counts the same, if a reasonable adjustment is done," Kadane said.

The issue crystallizes around controversial guidelines, proposed in early December by the Commerce Department, to adjust the count. The guidelines grew out of a lawsuit filed by New York City and other cities to force an adjustment for any undercount.

Friday was the deadline for filing public comments on the guidelines, setting the stage for a fierce battle in the coming months.

"If the final guidelines look anything like the proposed guidelines, we will be back in court," Lorna B. Goodman, a division chief in New York City's Law Department, said Friday.

When the population last was counted, in 1980, between 2 million and 5 million people were overlooked, according to a Census Bureau estimate. The undercount missed an estimated 100,000 people in Philadelphia and 450,000 in New York City, acording to the bureau.

While less than 1 percent of the whites were undercounted, nearly 6 percent of the black population was missed, the Census Bureau estimates.

At issue is whether to adjust the final 1990 census results to compensate for the estimated undercount. If the Census Bureau statistically adds in those estimated to have been missed by the tally, it would be a major departure from the way the nation counts its population.

"The census process consistently undercounts minorities and the disadvantaged," Judith A. Sanders-Castro, a lawyer with the Mexican American Legal Defense and Educational

(See CENSUS on 11-C)

Debate rages over adding an estimate to census count

CENSUS, from 1-C

Fund in Texas, said in comments on the proposed guidelines. "That process must be changed."

At stake are political clout and government funding.

Cities say the undercount will cost them political representation because new political boundary lines are drawn after every decade's census. When a city such as Philadelphia loses population, it can lose seats in the U.S. House of Representatives as well as the state House.

Also threatened are millions of federal dollars that are allocated according to population.

New York and Los Angeles have accused the Commerce Department with writing proposed guidelines that in themselves rule out the prospect of an adjustment.

"It's just clear they're biased" against an adjustment, Goodman said. The guidelines show a "total lack of understanding," she said.

Others contend that an adjustment would be unconstitutional.

Adjusting the census "seems to me to be an obvious ploy by urban areas to be over-represented in Congress and state legislatures because their liberal welfare-state politics have been pushed aside by the American people," John D. Rogers, Republican leader of the Kentucky Senate, told the Commerce Department.

Rogers and more than 100 other politicians, statisticians and special-interest groups have filed lengthy comments about the guidelines.

"I believe that the guidelines are written in a manner that many outside observers would interpret as being biased in favor of the partisan political interests of the current administration," said Harvard University statistician Thomas R. Belin.

Belin said there was a widespread perception that Democrats, who often have large urban constituencies, favor adjustment, while Republicans are opposed.

The proposed guidelines are one-sided in favor of not adjusting, contended Eugene P. Ericksen, a Temple University sociology professor who co-chairs a special census advisory panel that will make a recommendation on whether to adjust.

Ericksen said those who were often overlooked in the census count included people who were not regular members of a household; they might include an uncle or a boarder living with a family. Black children also are undercounted, possibly because their parents are confused about how to fill out the census form, he added.

The reasons people don't show up in the census range from their concern over the confidentiality of their responses to a misunderstanding of the questions. Most of the census is taken through the mail, so missed housing units on the Census Bureau's mailing list also could result in an undercount.

The undercount is most severe among black males. The Census Bureau estimated that a little more than 80 percent of black males between 40 and 44 years old made the 1930 census.

A "dress rehearsal" of the 1990 census in St. Louis in 1988 showed the undercount would again plague cities. Census Bureau estimates after the St. Louis census showed about 5 percent of the city's population was missed.

Communities across the country have teamed up with the Census Bureau to promote the census. Philadelphia, for instance, has named hundreds of community leaders to a complete-count committee to urge residents to fill out their census forms completely.

Yet even massive publicity campaigns will never eliminate an undercount, said Temple's Ericksen.

"Most people who get missed don't get missed because the publicity campaign didn't get to them," he said last week.

Many who responded to the Commerce Department's invitation for comments on the proposed guidelines took the opportunity to oppose an adjustment for those missed.

If adjusted, the census would be biased in favor of states with vast numbers of "undercounted population," said South Dakota's Mickelson, noting that Midwestern states such as his "do not have large numbers of what have generally been considered the 'undercounted population.'"

In Texas, with its large Hispanic population, adjusting for an undercount has found official favor.

"A decision against adjustment is a decision against a more accurate census," Javier P. Guajardo, assistant attorney general of Texas, wrote in response to the proposed guidelines. "A decision against a more accurate census is a decision against fair representation of the American public, in particular those who need to be counted most — minorities, the poor and the homeless."

Opposing adjustment was state Rep. John M. Perzel of Philadelphia, the Republican policy chairman of the Pennsylvania House of Representatives. Perzel said in a letter to the Commerce Department that even if the current way to count the population was proven inadequate, "at least it has been consistently statistically inadequate for 200 years."

The minority leader of the New Jersey Senate, Republican John H. Dorsey of Morris County, noted that the Constitution required an "actual enumeration" and not an estimate of the population every 10 years.

Even if an adjustment is approved, it might already be too late. Some state political leaders told the Commerce Department they fear a disruption of the political redistricting process if final numbers arrive too close to redistricting deadlines.

Carol M. Straus, manager of the Kentucky Economic Information System at the University of Kentucky, said if the census in 2000 was to be adjusted, preparations should begin immediately. "It is too late for an adjustment to the 1990 census to constitute an improvement."

A FOLLOWUP FORUM ON CENSUS UNDERCOUNTS AND PREPARATIONS FOR THE 1990 CENSUS

NEW YORK STATE

ADVISORY COMMITTEE

TO THE UNITED STATES

COMMISSION ON CIVIL RIGHTS

This summary report of the New York State Advisory Committee to the United States Commission on Civil Rights was prepared for the information and consideration of the Commission. Statements and viewpoints in the report should not be attributed to the Commission or to the Advisory Committee, but only to individual participants in the community forum where the information was gathered.

A SUMMARY REPORT

JANUARY 1990

UNITED STATES
COMMISSION ON
CIVIL RIGHTS

Eastern Regional Division
1121 Vermont Avenue, N.W. Rm. 710
Washington, D.C. 20425

LETTER OF TRANSMITTAL

New York State Advisory Committee
to the
U.S. Commission on Civil Rights

MEMBERS OF THE COMMISSION

William B. Allen, Chairman
Murray Friedman, Vice Chairman
Mary Frances Berry
Esther Gonzalez-Arroyo Buckley
Sherwin T.S. Chan
Robert A. Destro
Francis S. Guess
Blandina Cardenas Ramirez

Melvin L. Jenkins, Acting Staff Director

The New York State Advisory Committee submits this summary report to advise the Commission on preparations for the 1990 census and on the issue of adjusting for decennial census undercounts. The report summarizes information received at a community forum conducted by the Committee in New York City on April 27, 1989.

The forum was a followup to the Committee's November 19, 1987, forum on the same topic. Although initially agreeing to participate in the first forum, the Census Bureau of the U.S. Department of Commerce ultimately attended neither forum but submitted an 11-page critique of statements by the panelists who did appear at our first forum. The followup forum was intended to afford the panelists an opportunity to respond to the Bureau's critique and to report upon more recent developments.

As you know from our prior report, persistent undercounts during decennial censuses disproportionately affect minority communities and thereby have an adverse effect on the jurisdictions in which undercounted minorities reside. This is because census figures determine congressional reapportionment and how funds in various public programs are allocated to States and localities.

By a unanimous vote of its 8 incumbent members, the Committee approved this summary report. We hope that it will be useful as you follow developments related to whether the U.S. Department of Commerce Secretary eventually adjusts the census tallies to correct for the anticipated undercount.

Respectfully,

Walter Y. Oi, Chairman
New York State Advisory Committee

Acknowledgments

The New York State Advisory Committee wishes to thank the staff of the Commission's Eastern Regional Division for its help in the preparation of this summary report. The forum and report were the principal assignment of Tino Calabia with support from Tina James Martin, Edna Nicholson, and Juanique Caldwell. The project was carried out under the overall supervision of John I. Binkley, Director of the Commission's Eastern Regional Division.

TABLE OF CONTENTS

Page

Preface 1
Forum Participants and Bureau Commentary 3
First Forum--November 1987 4
Followup Forum--Responding to Bureau's Comments 5
Commerce Department Reverses Census Bureau 7
Reporting Deadline Not Absolute 8
Prediction That Undercount Will Worsen 9
Issues Affecting Asian Americans 10
Opposition to Undercount Adjustment 11
Effects of Various Adjustment Procedures 13
Postscript: Commerce Department Reverses Itself 14

Summary 15

APPENDICES

"Statement on Census Adjustment" 16
--D.A. Freedman, P. Diakonis, et al (Appendix A)

"Use of Statistical Methods to Produce 19
Accurate Census Coverage"
--Barbara A. Bailar (Appendix B)

"Effects of Census Adjustment" 38
--William P. O'Hare (Appendix C)

Preface

On July 17, 1989, an agreement in the U.S. District Court in Brooklyn closed a chapter in a long-standing dispute over how to resolve the problem of undercounts in decennial censuses.[1] Such undercounts of the population have been acknowledged and, since the 1950's, even been measured by the Bureau of the Census of the U.S. Department of Commerce. Many experts in census matters, public officials, and others familiar with the issue claim that undercounts have serious and adverse effects on the apportionment of congressional seats in the U.S. House of Representatives and on the allocation of public funding to the States and localities for various tax-based programs.[2]

In the early 1980's, both the New York City and New York State governments filed suit against the Commerce Department to have the 1980 decennial count corrected in order to alleviate the perceived problems of undercounting. The city and the State, along with other jurisdictions which joined in the suit, failed to prevail in their suit. However, with prospects of continued undercounts in 1990, the city and State returned to court in November 1988, with a new suit calling for the Department and

---

[1] City of New York v. U.S. Department of Commerce, No. 88 Civ. 3474 (E.D.N.Y. July 17, 1989) (stipulation and order).

[2] See, e.g., Dr. William P. O'Hare, "Introduction: the 1990 Census and Political Power for Minorities," Redistricting in the 1990s: a Guide for Minority Groups, Population Reference Bureau, Inc., July 1989, who writes, "People not counted by the Census Bureau cause their community to be deprived of its rightful share of public money. In short, the census is connected with money and power." P. 7.

its Census Bureau to lay the groundwork for corrections for the anticipated undercount after the 1990 census.

The Department attempted to convince the court to dismiss the new suit, but finally in mid-July, an agreement was reached. By that agreement, the Bureau will carry out a postenumeration survey (PES).[3] Based upon the PES results and guidelines to be published by March 10, 1990, and implemented with the assistance of a panel of 8 outside advisors, the Commerce Secretary may opt to correct for the undercount that many experts believe will occur during the 1990 census.[4]

The New York State Advisory Committee to the U.S. Commission on Civil Rights began looking into the undercount and related issues starting with a forum on November 19, 1987,[5] and a followup forum on April 27, 1989. In June 1989, the U.S. House Subcommittee on Census and Population invited the Committee to summarize its work during a July 24, 1989, hearing of the subcommittee in New York City. A week prior to that hearing, the agreement between the

---

[3]U.S. Census Bureau, Department of Commerce, "Agreement Announced! 1990 Census Agreement Litigation Settled," **Census and You**, vol. 24, no. 9, September 1989, p. 4. See also Richard Levine, "Accord on Census May Bring Change in Minority Data; Undercounting at Issue; Random Home Survey in U.S. Is Agreed to in New York With Trial Due in Suit," New York Times, July 18, 1989, p. 1, and Dennis Hevesi, "Census Weighing a Second Survey in '90 to Correct Any Undercounting," New York Times, July 24, 1989, p. A-10 (hereafter cited as July 24, 1989 Times article).

[4]Whether either side has gained through the settlement is an open question. See, e.g., "New Yorkers Have No Cause Yet to Celebrate Census Settlement," a letter to the editor from U.S. Representative Charles E. Schumer, New York Times, Aug. 7, 1989, p. A-14.

[5]New York State Advisory Committee to the U.S. Commission on Civil Rights, Census Undercounts and Preparations for the 1990 Census, (December 1988).

Commerce Department, its Census Bureau, and the plaintiffs in the suit was reached. However, the hearing remained scheduled, and Advisory Committee Chairman Walter Y. Oi shared the highlights of the Committee's two forums with the House subcommittee.[6]

Forum Participants and Bureau Commentary

On November 19, 1987, the Committee held its first forum on the undercount issue, and a summary report of the first forum, Census Undercounts and Preparations for the 1990 Census, was approved by the Commissioners the following December. At the forum, elected and appointed officials of the State and the city of New York and their staff advisors, the volunteer chairman of the U.S. Census Bureau's Asian/Pacific Islander Census Advisory Committee, and an independent scholar-researcher were panelists.[7] The Bureau was also invited, initially indicated it would be represented, but then declined. Nonetheless, after reviewing a draft of the forum report, the Bureau submitted an 11-page commentary on the remarks of the panelists, and it was published in the Committee's report.

---

[6]Walter Y. Oi, Chairman, New York State Advisory Committee to the U.S. Commission on Civil Rights, "Testimony Offered at Invitation of House Subcommittee on Census and Population," during the U.S. House Subcommittee's hearing in New York City on July 24, 1989. Dr. Oi's testimony is on file at the Eastern Regional Division office in Washington, D.C.

[7]The panelists were: State Assemblyman Angelo del Toro, cochairperson of the State Legislative Task Force on Demographic Research and Apportionment; Jeffrey M. Wice, special counsel to the State Assembly Speaker; New York City Corporation Counsel Peter L. Zimroth and Charles N. Weinstock and John Low-Beer of Mr. Zimroth's staff; Charles P. Wang, chairman of the Census Bureau's Asian-Pacific Islander Census Advisory Committee and executive director of the Chinatown Planning Council; and Erol R. Ricketts, assistant division director at the Rockefeller Foundation. Only Dr. Ricketts did not attend or provide a statement for the second forum.

In order to afford the panelists an opportunity to respond to the Bureau's extensive critique, the panelists and the Bureau were invited to a followup forum held on April 27, 1989. Once again the Bureau declined. Four of the original panelists agreed to participate, and the Pennsylvania Advisory Committee, which had provided testimony to the same House subcommittee prior to the 1980 census, sent a representative to take part in the April forum. What follows are highlights of what the panelists said in the first forum and more detailed highlights of their responses to the Bureau's critique which the panelists made in the second forum.[8] Where appropriate, other documents are cited to report on subsequent developments or to amplify on the matter under discussion.

First Forum--November 1987

The first forum was convened in New York City less than 3 weeks after the U.S. Department of Commerce announced its decision not to adjust the 1990 decennial census figures. Estimating an undercount of 500,000 New York City residents in the 1980 census, the State and city officials decried the disenfranchisement and the loss of public funding that decennial undercounts may cause and called for an adjustment. Indeed, Jeffrey M. Wice, special counsel to the State assembly speaker, urged that the adjustment be referred to as a "correction" made necessary to compensate for the undercount.

The city of New York's corporation counsel, Peter L. Zimroth, circulated a speech and documents by former or current Bureau officials indicating that adjustment is feasible, and Charles N. Weinstock, a member of the corporation counsel's staff, pointed

[8]This report is based on the December 1988 summary report and the official transcript of the April 27, 1989, forum. The latter is on file at the Eastern Regional Division office in Washington, D.C.

out that the Bureau added 5 million people to the 1970 count and 3.3 million to the 1980 count through an "imputation" process. This form of adjustment is used, for example, when the Bureau is faced with inconsistent answers or no responses to the census questionnaires. Each State and city official appearing at the forum called for adjustment on the grounds of fairness in terms of voting representation and the allocation of public funds to the jurisdictions where the uncounted persons reside.

Charles P. Wang, chairman of the Bureau's Asian/Pacific Islander Census Advisory Committee, expressed fears over the format of the race question to be asked of Asians in the 1990 census questionnaire. This new format differed from that used in 1980; the former called for some write-in responses which, Mr. Wang feared, could result in a drop in accurate returns and a loss in timely reporting of data. He also agreed on the necessity of a postenumeration survey, with corrections based upon the PES to compensate for undercounts.

Erol R. Ricketts, a Rockefeller Foundation official who had carried out demographic studies down to the census tract level prior to serving at the foundation, noted that the 1990 census is basically a large survey. He then explained that one type of undercount adjustment would require a second survey, saying that how well you adjust the first one depends on how well you do the second one. The same point was made by Barbara A. Bailar, when a top official of the Bureau and while president of the American Statistical Association. After resigning from the Bureau, Dr. Bailar became executive director of the association and furnished the Committee with a statement advocating undercount adjustment.

Followup Forum--Responding to Bureau's Comments

In June 1988 the Bureau submitted to Commission staff an 11-page letter commenting on many of the remarks made by the panelists,

plus 17 enclosures. This past April, at the followup forum in New York City, 5 of the Committee's original panelists provided statements and/or discussed the Bureau's comments. The cochair of the State Legislative Task Force on Demographic Research and Reapportionment, Assemblyman Angelo del Toro, expressed his appreciation to the Bureau for its recent "attempts to help those of Hispanic origin differentiate on the census forms between Hispanic, Spanish and other backgrounds." In addition, he thanked the Bureau for its strong opposition to proposals intended to exclude undocumented persons from the 1990 count.[9]

However, Assemblyman del Toro remained concerned that the Bureau has not extended the 14-day period given to local jurisdictions to review the Bureau's preliminary count. That review is to take place in June 1990, and he believed that 14 days is too short a time for an adequate review. On undercount adjustment, he noted that the Bureau reduced its postenumeration survey from 300,000 to 150,000 households, "a number which may prove insufficient to properly analyze the census undercount."

Assemblyman del Toro then reported that on the previous Friday, April 21, 1989, the U.S. district court ordered that the suit calling for an undercount adjustment be heard on the merits in July, thus ruling against the Federal Government's motion to

---

[9]See Spencer Rich, "Suit to Block Illegal Aliens From Census Count Voided," Washington Post, May 10, 1989, p. A-17, also "Census Suit Thrown Out," Hispanic Link Weekly Report, vol. 7, no. 19, May 15, 1989, p. 1. The lawsuit aimed at barring the Bureau from counting undocumented immigrants in the 1990 census for the purpose of reapportioning congressional seats. The U.S. District Court in Pittsburgh ruled that the plaintiffs had no standing to sue since it could not be determined that any particular plaintiff would be harmed.

dismiss the suit.[10] Both the assemblyman and special counsel Wice speculated that, if the State and other plaintiffs were to prevail, the Federal Government would appeal. Because an appeal would cut into the time required for the preparations needed to adjust for an undercount, Assemblyman del Toro said that efforts in Albany are shifting towards increasing work with the Bureau on ways of achieving the best count possible. However, here, again, problems have cropped up; maps of New York City prepared by the Bureau have omitted or improperly identified streets in some neighborhoods which have been in existence for over 200 years, Assemblyman del Toro reported.[11]

Commerce Department Reverses Census Bureau

Mr. Wice cited a March 10, 1989, publication by the Congressional Quarterly that indicates that up until June 1987, the Bureau was prepared to announce a decision to make an adjustment aimed at correcting the count.[12] He added that in the early 1980's, the Bureau had appointed a blue-ribbon panel of officials from the National Academy of Sciences that eventually supported efforts aimed at undercount adjustment.

---

[10] City of New York v. U.S. Department of Commerce, No. 88 Civ. 3474 (E.D.N.Y. Apr. 21, 1989). See also, Constance L. Hays, "New York Wins First Step in Effort to Adjust Census," New York Times, May 22, 1989, p. B-4.

[11] See also a May 3, 1989, letter from New York City Mayor Eward I. Koch to U.S. Secretary of Commerce Robert A. Mosbacher in which the Mayor states that "These errors . . . include misaligned streets, missing, misnamed, or nameless street segments, extra and redundant street segments, census tract boundary misalignments, and census block numbering problems."

[12] Robert K. Landers, "1990 Census: Undercounting Minorities," Editorial Research Reports, Congressional Quarterly, Mar. 10, 1989 (hereafter cited as March 10, 1989 Editorial Research Reports.)

However, with the Commerce Department's decision against any adjustment, the Bureau has reduced its internal research on undercount methods and reduced the size of the postenumeration survey by half, thereby adversely affecting the Bureau's ability to estimate the undercount, according to Mr. Wice. He concluded that debating a decision now to adjust or not to adjust the count may well be 2 years too late; nevertheless, even conceding the possibility that "the overall national implications of overcount [sic] adjustment might not shift congressional seats," at the local level--focusing on New York City in particular--"you're likely to see more of the 150 State legislative assembly seats within New York City than you would have without an adjustment. . . ."

On the timing of a decision on undercount, Mr. Wice also noted that a former Census Bureau official has reportedly stated "that a decision could be made administratively to still correct in time by August of this year."

Reporting Deadline Not Absolute

John Low-Beer, an attorney with the office of the New York City Corporation Counsel, agreed with the State officials. He said that the Bureau acknowledges that, if by August 1989, the court orders the Bureau to take administrative steps towards making an adjustment, the Bureau could adjust the 1990 count. He further indicated that on April 21, 1989, the court suggested that the statutory deadlines for reporting the decennial count are not absolute, "at least to the extent that they conflict with Constitutional rights," and that the court believed that it could grant a reasonable extension of time for the Bureau to accomplish an undercount adjustment, if an extension were necessary.

Mr. Low-Beer also presented the corporation counsel's statement in which Mr. Zimroth asserts that "In early 1987, the [Bureau

director] himself announced to the Commerce Department that the Bureau had the technical ability to correct the census and it was the Bureau's goal to carry out that correction by December 31, 1990." Mr. Low-Beer added that on July 30, 1987, in a statement in Indiana, the Bureau's Deputy Director publicly acknowledged that adjustment of the census is technically feasible.

Mr. Low-Beer then pointed out that, were the postenumeration survey of 300,00 households to take place as previously planned, the results would still have been subject to a review by panels of experts. These experts would be called upon to judge whether the data measured up to preset standards. If the data met those standards, an adjustment would be done; however, if they did not, an adjustment would not be done.

Prediction That Undercount Will Worsen

In his formal statement, Mr. Zimroth, the city's corporation counsel, responded to the Bureau's criticism of the estimate he used in the first forum, a 500,000-person undercount affecting New York City. In the followup forum, Mr. Zimroth explained that this estimate and other percentages he had cited did not differ significantly from one of the Bureau's own sets of estimates, the Bureau's "3-8 Series" estimates. He pointed out that the "3-8 Series" estimates are relied upon almost exclusively by the Bureau in its own internal memoranda and added that "There's broad consensus among knowledgeable experts that these figures are generally accurate."[13]

---

[13]Temple University professor Eugene Ericksen is quoted as estimating "the range of undercounting in most large cities in 1980" as being in the range of 3 to 7 percent. A Los Angeles official estimated Los Angeles' undercount at 4.6 percent, and a Houston official estimated that the 1990 undercount in Houston would run between 6 to 8 percent. July 24, 1989 Times article.

On the Bureau's reliance on census improvement procedures other than adjustment, Mr. Zimroth stated that the city of New York actively supports the Bureau's outreach programs, but he added that "Bureau officials have been unable to come up with even a single piece of evidence that any of the programs . . . will reduce this differential. In fact, it is very likely that both the overall undercount and the differential undercount will be significantly worse. . . ." His prediction was based on the premise that the national population of blacks and Hispanics has increased over the numbers present in 1980, and, since these minority communities contain the hardest-to-count individuals, the undercount problem will increase for the population as a whole and particularly for these 2 communities.[14]

Issues Affecting Asian Americans

Charles P. Wang, chairman of the Asian/Pacific Islander Census Advisory Committee, reported that, with the help of Members of the U.S. Congress and others, a bill was passed in both Houses to restore the question on Asians to the format used in 1980, a goal he had pressed for in the first forum. He stated that, although President Reagan vetoed that bill, the Bush administration later approved the desired format.[15]

Despite such progress, other problems persist including delays in the publication of census data on Asians and the scarcity of Asians in the Bureau's top management and in some regional work forces. He also pointed out that the census questionnaire would appear in English and Spanish, but not in any Asian language,

---

[14]Public officials in Los Angeles and Dade County, Florida are reported to believe that the 1990 census undercount will be larger in their jurisdictions. July 24, 1989 Times article.

[15]See William Dunn, "Census to Split Asian Count," USA Today, Jan. 16, 1989.

although an instruction booklet will reportedly be of assistance to those speaking any of 30 or so languages.
He continued to urge an adjustment of the count, hoping that "the Voting Rights Act would benefit Asians as well" through an accurate count enabling Asians to gain election districts in the reapportionment for the 1992 elections. He further noted that he had just returned from a forum involving the American Statistical Association and the American Marketing Association, both of which apparently maintain that there are acceptable ways of adjusting the count so that the margin of error approaches zero.

Opposition to Undercount Adjustment

As mentioned at the outset, the Committee was unable to engage the Bureau in its second forum. However, reacting to the first draft of the summary report of that forum, the Bureau submitted an 11-page letter for the record. The Bureau's major points included one statement that the Bureau does not have a single official estimate for the undercount of the total population but a series of estimates based on different assumptions and a second point to the effect that the Bureau does not know how many blacks and Hispanics were undercounted in central cities since it has not produced net undercount rates for them in central cities.

The Bureau also observed that not all statisticians agree about the Bureau's ability to make census counts more accurate through adjustment and that, regarding the New York lawsuit calling for an adjustment of the 1980 census figures, "the Court finds as a matter of fact that the Census Bureau correctly determined that an adjustment of the census is not technically feasible or warranted and that no such adjustment should be made."[16]

---

[16]Roland H. Moore, Associate Director for Field Operations, U.S. Bureau of the Census, letter to Tino Calabia, June 30, 1988, p. 3.

While the Bureau declined to appear at the followup forum, the Committee did identify 13 university professors who had submitted a brief joint "Statement on Census Adjustment" to the U.S. House Subcommittee on Census and Population for its hearing of March 3, 1988.[17] (See appendix A.) Their "Statement" opposed the adjustment method which employs statistical techniques and a postenumeration survey, concluding that "real data (with real flaws) would be replaced by complicated and poorly tested mathematical models of data."[18] Three of the "Statement" signatories were contacted but were unable to attend the forum. However, the organizer and chief signatory of the "Statement" shared the document with the Committee.

Commenting on the "Statement" during the followup forum, the panelists generally agreed that experts can be found to differ on many complex issues. They then again cited the support for the PES method given by the American Statistical Association, the Bureau's own blue-ribbon task force known as the Panel on Decennial Census Methodology, and former and current officials and technicians in the Bureau. In that regard, the Committee has recently received a new paper by Dr. Bailar, executive director of the American Statistical Association. (See appendix B.) She had been the association's chairperson and also the Bureau official in charge of adjustment methodologies until 1987, when

---

[17]David A. Freedman, professor of statistics, University of California/Berkeley, P. Diakonis, professor of mathematics, Harvard University, et al, "Statement on Census Adjustment," unpublished paper submitted to the U.S. House of Representatives' Subcommittee on Census and Population for the Subcommittee's Mar. 3, 1988, hearing.

[18]Ibid., p. 2.

no adjustment would be made.[19]

Effects of Various Adjustment Procedures

The Committee has also received for its record an article by the Policy Studies director of the Washington, D.C.-based Population Reference Bureau, William P. O'Hare.[20] (See appendix C.) His article outlines the various results which would be yielded by the application of eight "scenarios," including one involving no adjustment. The author concludes that none of the "scenarios" would result in "a big difference in the overall apportionment of Congress following the 1990 Census." Apparently there would be no change at all from the three congressional seats already projected to be lost in New York State. Pennsylvania could gain one seat if 6 of the 8 "scenarios" were implemented; however, Pennsylvania would be one of only 2 States that might possibly gain from an adjustment, California being the second State.

Mr. Low-Beer, the attorney with the city corporation counsel's office, observed that there is no unanimity among the experts on the results described in the article; in fact, he asserted that many others believe that a loss of congressional seats would occur if no adjustment is made of the 1990 count. He also mentioned that the city of New York has commissioned a study of the matter by an independent scholar, and it may be that 2 or

---

[19]Barbara A. Bailar executive director, American Statistical Association, "The Use of Statistical Methods to Produce Accurate Census Coverage," an undated 36-page paper sent to Tino Calabia, U.S. Commission on Civil Rights, with a letter of Apr. 20, 1989. (See appendix B.) Some details surrounding Dr. Bailar's resignation from the Bureau are described in March 10, 1989 Editorial Research Reports.

[20]William P. O'Hare, director, Policy Studies Department, Population Reference Bureau, "Effects of Census Adjustment," appearing in the March 1989 issue of Population Today, pp. 6-8.

2 1/2 congressional seats would shift if no adjustment were made.

While electoral politics would certainly be affected by the outcome of the 1990 census, Mr. Low-Beer emphasized that the work of the Bureau needs to be insulated from politics. For example, he hoped that "as a first step that President Bush [would] appoint a career professional, a statistician or a research scientist as director of the Census Bureau. . . . The Bureau itself has been dealing in a vacuum without a new director since [Bureau director Dr. John G.] Keane left at the end of last year."[21]

Postcript: the Commerce Department Reverses Itself

As mentioned at the outset, on July 17, 1989, the Department of Commerce agreed to lay the groundwork for possible adjustment of the 1990 census tallies. But what prompted the Department to do so despite its past refusal? According to Eileen Shanahan, writing in Governing,

> Speculation is that Secretary of Commerce Robert A. Mosbacher and Undersecretary Michael R. Darby simply decided their Reagan administration predecessors were wrong. Reportedly, there were also fears that U.S. District Court Judge Joseph M. McLaughlin might issue a flat order to make the adjustment, without review of its quality, if the case came to trial. And the Census Bureau desperately wanted the litigation out of the way

---

[21]Ann Devroy and Spencer Rich, "Californian May Take Next Census; Democrats Express Caution About Redistricting Expert Heslop," Washington Post, Apr. 26, 1989, p. A-25. See also "Pollster Is Likely Choice to Head Census Bureau," Washington Post, Aug. 16, 1989, p. A-19, which reported that the vice president of "a top Republican polling firm" appeared to be the White House's nominee to head the Bureau.

for now, so it could devote its attention to the basic 1990 head count.[22]

S U M M A R Y

The Committee held 2 forums on decennial census undercounts and on proposals to make adjustments to correct for such undercounts. In the course of its inquiry, the Committee invited proponents of adjustments and also the Bureau and nongovernmental specialists opposed to adjustments. Neither the Bureau nor other opponents of adjustment found it possible to meet with the Committee. But the Bureau eventually reacted in writing to the statements made by each participant during the first forum, and a nongovernment census specialist opposed to adjustment furnished the Committee with a document explaining the opposition views he shares with 12 other academics.

Those who did appear in the forums included elected and appointed officials of the State and the city of New York as well as the volunteer chairman of the Bureau's Asian/Pacific Islander Census Advisory Committee and an independent scholar and researcher. The Committee also benefited from receiving articles and other documents from several expert sources reflecting either arguments for or against adjustment or simply estimating what the results would be depending on which version of adjustments is made.

Only 5 months remain before 400,000 census workers take to the field. The results of the 1990 count--adjusted or unadjusted--will shortly thereafter determine electoral reapportionment and many funding allotments affecting public life until the close of the 20th century.

---

[22]Eileen Shanahan, "Census Will Try to Fix Undercount," Governing, September 1989, pp. 11-12.

To the US House of Representatives
Subcommittee on Census and Population

For the Hearing of March 3, 1988

STATEMENT ON CENSUS ADJUSTMENT

by

RJ Beran
Professor of Statistics
University of California, Berkeley

L Breiman
Professor of Statistics
University of California, Berkeley

P Diaconis
Professor of Mathematics
Harvard University

ML Eaton
Professor of Statistics
University of Minnesota, Minneapolis

DA Freedman
Professor of Statistics
University of California, Berkeley

S Geisser
Professor of Statistics
University of Minnesota, Minneapolis

L Gordon
Associate Professor of Mathematics
University of Southern California

JL Hodges, Jr
Professor of Statistics
University of California, Berkeley

W Kruskal
Professor of Statistics
University of Chicago

D Lane
Professor of Statistics
University of Minnesota, Minneapolis

P Meier
Professor of Statistics
University of Chicago

KW Wachter
Professor of Demographics & Statistics
University of California, Berkeley

A Zellner
Professor of Economics and Statistics
Graduate School of Business
University of Chicago

Census counts are used to apportion seats in Congress, and to allocate billions of dollars in tax moneys. There is some evidence to show that there has been a small differential undercount. And there is a proposal to adjust the 1990 Census for this undercount, using statistical techniques and a "Post Enumeration Survey."

These methods are open to serious question. Yet the Bureau is under considerable pressure to adjust the counts-- from some newspapers, from some politicians, and most regrettably from some statisticians. One egregious tactic is to assert that there is a consensus of technical opinion favoring adjustment.

Like any large-scale statistical enterprise, the Census makes mistakes. It misses some people, and includes others who should not be counted. Demographic analysis (itself an imperfect instrument) suggests that on balance there is an undercount, particularly among minorities.

For 1980, the undercount has been reported as about 1% overall, and 5% among blacks and hispanics. If the undercount can be estimated with sufficient reliability by statistical methods, it can be corrected. To help with the allocation of tax money and apportionment of congressional seats, the corrections would have to be done in fine geographical detail; probably at the block level.

The statistical methodology being proposed involves two large samples: an "E-sample" of census records to check for erroneous enumerations, and a "P-sample" of small geographical areas to estimate the gross undercount using capture-recapture techniques. Proposed samples range up to hundreds of thousands of households.

The procedure uses computer matching to see whether persons in the P-sample were previously captured in the census. At best, this is a complex and error-prone process, especially when some of the data are bound to be wrong. Properly identifying persons who moved between census day and the time of the Post Enumeration Survey is a special difficulty. Furthermore, sample weights have to be estimated, to extrapolate from the sample blocks to the rest of the country. If different persons have different response probabilities, these weights are subject to bias; post-stratification might (or might not) mitigate the problem; the magnitude of the bias seems hard to quantify.

Statistical modeling techniques would be needed to smooth the estimates. Such techniques are helpful in theory, if certain assumptions hold. Again, it is hard to assess the degree to which these assumptions would be violated in practice, or the impact of failures in the assumptions. However, adjustment can easily introduce more mistakes than it fixes: for example, if the total undercount is 1%; and the overall error rate in the adjustment process exceeds 1%, as seems likely.

The last element of the proposal: adjust the statistical adjustments themselves, to agree with the totals from the demographic analysis. The latter would have to be augmented by quite speculative estimates for the numbers of illegal aliens.

This entire process needs to be fitted into an already tight Census schedule, and would reduce the time available for field work, degrading the quality of the data.

In sum, real data (with real flaws) would be replaced by complicated and poorly tested mathematical models of data. We do not see that as progress. We are sympathetic to the goal of funneling additional tax money to cities. But we would prefer a cleaner separation between the technical issues and the political ones. So far, the technical case for adjustment is weak.

American Statistical Association
1429 Duke Street, Alexandria, Virginia 22314-3402
(703) 684-1221

April 20, 1989

Mr. Tino Calabia
U.S. Commission on Civil Rights
Eastern Regional Division
1121 Vermont Avenue, NW
Room 710
Washington, DC 20425

Dear Mr. Calabia:

I'm enclosing a copy of a paper I wrote describing Census adjustment. Also, for your interest, I'm enclosing a paper done by Congressional Quarterly's Editorial Research Reports that I think is very good.

Sincerely,

Barbara A. Bailar, Ph.D.
Executive Director

Enclosures

THE USE OF STATISTICAL METHODS TO PRODUCE ACCURATE CENSUS COVERAGE

Barbara A. Bailar

1. Introduction

The United States has taken a population census every ten years beginning in 1790. We are now preparing to take the 1990 census. Taking a census is not a matter of choice; it is written into the Constitution. One of the primary uses of census data, as stated in the Constitution, is to reapportion the House of Representatives. Thus, it was seen at an early date that political power would be dependent on census results. Over the years, the use of census data to define political jurisdictions has grown. Within states, census population counts are used to determine the boundaries of congressional election districts as well as districts for state legislative offices. The census counts are used to insure that these districts are as equal in size as possible. Census counts are also used to classify local governments by size class. In more recent years, census data have been used by both Federal and state governments to allocate funds. Because of the large amounts of money allocated, political jurisdictions are very concerned about the accuracy of census data.

No census is perfect, even though the final tabulation may suggest perfect accuracy. For example, the U.S. census count in 1980 was 226,545,805, with no plus or minus following to indicate a reasonable amount of uncertainty. This is because the census population total is based on counting, not on a sample, so that no sampling error is involved. Yet there are errors involved because a census depends on the work on a great many people, most of them temporary workers, and people make errors. In this paper, I discuss only one error -- that

of imperfect coverage. In other words, the census does not count perfectly all the people who should be counted. This would probably not be too disturbing an event provided that the undercoverage is small. However, the undercoverage is not evenly distributed over all population groups. It is much larger among minority groups than the White population, among men rather than women, and among younger people rather than older people. In 1980, though the White population may have been slightly overcounted, an estimate of undercount for the Black population was 5.3 percent and 8.0 percent for Black males. See Table 1.

It is because of the differential nature of the undercount that legal action was taken against the Census Bureau after the 1980 census. Many cities, such as New York City, with large proportions of Blacks, sued the Census Bureau. They believed that they had been undercounted and were thus losing political power and millions of dollars. No final legal decisions have been made. Yet the Census Bureau has had to move forward to prepare for 1990. The remainder of this paper describes research done to develop, implement, and test statistical methodology that could be built in as part of the 1990 census to reduce the undercount.

2. History of the Undercount and Undercount Research

Though the U.S. Census achieves high levels of coverage, there has always been an undercount. After the first census in 1790, George Washington and Thomas Jefferson each stated, in letters to friends, that there had been an undercount. They fully expected the population in 1790 to be over 4 million and it was 3.9 million. Similar historical references to an undercount occurred over the years, but, until 1950, there was no way of measuring the undercount.

TABLE 1. Net Undercount Rates by Race and Sex
1950 to 1980 Decennial Censuses

Population Category	1950	1960	1970	1980
Total population	3.3	2.7	2.2	0.5-1.4[a]
Male	3.8	3.3	3.1	NA
Female	2.8	2.2	1.4	NA
Black population	9.7[b]	8.0	7.6	5.3
Male	11.2	9.7	10.1	8.0
Female	8.2	6.3	5.3	2.7
White and other races population	2.5[c]	2.1	1.5	-0.2
Male	2.8	2.5	2.1	0.6
Female	2.1	1.7	0.9	-0.9

A minus sign indicates net overcount.

NA = Not available.

[a]Lower percentage assumes presence of 2 million undocumented aliens in estimated population; higher percentage assumes presence of 4 million undocumented aliens.

[b]Blacks and other nonwhites.

[c]Whites only.

SOURCE: The Bicentennial Census, Panel on Decennial Census Methodology, Committee on National Statistics, National Academy Press, Washington, D.C. 1985.

Two methods were developed and used with the 1950 census and those two methods, after much testing and refinement, are with us today.

The first of these methods is called demographic analysis, a demographic accounting method developed by Ansley Coale (1). Basically, demographic analysis depends on using birth records, death records, and estimates of migration into and out of the U.S. Since 1950, estimates of the completeness of the census have been made by using demographic analysis. Estimates from this method are shown in Table 1. Though there are many assumptions made in producing the estimates, the method and the resulting estimates have achieved credibility. However, the estimates are not suitable to use for adjusting census population for an undercount. The reasons are as follows:

- the estimates are available only at the national level and for no sub-national areas.
- the estimates are available only for Whites, Blacks, and a combined residual group. There are no estimates for any other ethnic groups.
- the absence of data on illegal immigrants.

The second method developed was based on case-by-case matching of records. This method requires two samples to estimate net coverage error. One sample is from a source other than the census. It is generally called a Post-Enumeration Survey (PES). It may be another survey or a special coverage measurement survey. It provides an estimate of gross underenumeration. The second sample is a sample selected from the census itself. This sample, usually called the Enumeration Sample, is revisited to determine which census persons were correctly enumerated and which were erroneously enumerated, thus

providing an estimate of gross overenumerations. The two samples together provide estimates of net coverage error. It is this method that I will describe in more detail in the remainder of this paper.

In 1950, this matching method was tried for the first time. The assumption underlying its use was that coverage errors were caused by a failure to carry out census procedures correctly. Thus, specially trained enumerators who were paid higher rates were instructed to re-enumerate sample areas. After the re-enumeration, the records were matched into the 1950 census records. The estimate of undercount from this study was 1.4 percent, about 2 percentage points lower than the estimates from demographic analysis. This downward bias was called "correlation bias", and was described as the tendency for the PES to miss the same types of persons missed in the census.

A similar study was carried out in 1960 with no major changes in methodology. In 1970, the Current Population Survey, the labor force survey carried out monthly by the Census Bureau, was matched to the census, but no Enumeration Sample was selected to measure gross overenumerations. In 1980, the Census Bureau also used the Current Population Survey, selected an Enumeration Sample, and prepared to provide estimates of net census error for the nation, states and large cities, and for major race and ethnic groups. A description of the plan for the program appears in the Conference on Census Undercount, Proceedings of the 1980 Conference (2). As the study data became available, several problems emerged. There were large amounts of missing data, matching errors, problems in getting correct addresses for people who had moved, and other such difficulties. By varying the treatment of the difficult cases, the Census Bureau derived 12 sets of estimates, as shown in Table 2. Because all 12 sets were based on

TABLE 2. 1980 Post Enumeration Program Estimates of Percentage Undercoverage for Demographic Groups at National Level

National	Black	Nonblack Hispanic	Other
1.4	6.7	5.6	0.3
1.3	6.3	5.3	0.2
1.0	5.6	4.4	0.0
.8	5.2	4.1	-0.1
1.6	4.3	6.4	0.8
2.0	5.4	7.6	1.1
0.2	2.7	3.6	-0.4
1.6	6.9	5.5	0.4
1.7	7.2	5.8	0.6
-0.3	2.5	1.2	-0.8
-1.0	0.7	-0.2	-1.4
-1.1	2.0	1.0	-0.6

A minus sign indicates net overcount.

SOURCE: The Bicentennial Census, The Panel on Decennial Census Methodology, Committee on National Statistics, National Academy Press, Washington, D.C. 1985.

assumptions that could not be verified, the Census Bureau was unable to choose among them. That did not mean that others did not. In the trial of the lawsuit brought by the City and State of New York, the city selected a particular estimate as its favorite. Interesting extensions of methodology used by the city appeared in the paper by Ericksen and Kadane (4).

One thing clear from this table is that no matter what the true 1980 census undercount was, there was a differential undercount. Clearly more Blacks and Hispanics were not counted in the census. Given those results, and realizing that the differential undercount was unlikely to disappear before 1990, the Census Bureau began an extensive program of research to develop more robust methods for measuring the undercount in 1990. That research is the focus of the remainder of this paper.

3. Undercount Research Leading to 1990

Every aspect of the measurement of undercount has been under scrutiny since 1980. Testing of methodology was done in the test censuses in 1985 and 1986 and will continue in the test census in 1987 and the dress rehearsal in 1988. Research results, findings, and recommendations refer to five different areas:

- the two surveys that provide estimates of over- and underenumeration.
- the matching methodology.
- the handling of nonresponse.
- the use of a capture-recapture model.
- the use of other models for indirect estimation for smaller geographic areas.

Each of these will be described in turn.

3.1 Surveys to Measure Net Census Error

In 1970 and 1980, the Current Population Survey (CPS) was used to measure gross underenumeration. The CPS is a large survey, about 70,000 housing units and 185,000 persons, carried out monthly by the Census Bureau to measure labor force participation. The samples have a partial overlap from month to month so the Bureau selected the April and the August sample which have no overlap of units. The April sample was seen to be advantageous since Census Day was April 1. Thus, the problem of people moving between Census Day and the CPS interview was minimized. August was the next sample that could be used where no units were in common with the April CPS sample. Additional movers were expected. It was hoped that estimates from the two samples could be combined to produce an estimate with smaller sampling error.

The CPS is a multistage probability sample. In April 1980, it was spread over 626 primary sampling units (PSU's). Clusters of four housing units were selected within these PSU's. These four housing units were close together to reduce interviewer travel.

The Enumeration Sample in 1980 consisted of 110,000 census questionnaires selected in clusters of 10 housing units. For 50 percent of the sample there was a search for duplicate enumerations within the same geographic area. A reinterview was attempted for the full sample. Questions were asked to find any other addresses at which the people in these housing units may have been counted. The interviewers also confirmed the correct geographic location of the housing units.

After the CPS sample data were returned to the Bureau, the census files were searched for persons included in the CPS. However, the search was restricted

to a limited geographic area. Because the April 1980 CPS was based on 1970 census geography, the CPS cases had to be coded to the correct geography to be available for searching. Also, since all searching was done by clerks, there was a limit to the area of research. Because of this, it was possible for a person counted in the census, but within an area outside the CPS segment, to be tabulated as missed in the census. Since the CPS sample was based on segments of size 4, there was no way of searching the CPS files for people counted in the census. It was, thus, a one-way match.

In 1990 we will not depend on a sample selected for another purpose. Instead, we will first stratify the country into 100 strata. The strata will be defined by such things as percentage of minority group population, percentage of owners and renters, urban or rural location, and other such factors. Each block will be placed into one of these strata. Thus, we expect to see a stratum that will contain blocks from the inner cities of New York, Chicago, Detroit, Los Angeles, and so forth. Similarly, we expect to find strata that contain rural, unpopulated areas of Maine, Nebraska, etc. This method of stratification does not force geographical stratification, another weakness of the 1980 design.

Within these strata, blocks will be selected with probability proportional to a measure of size for the Post Enumeration Survey. We expect to select approximately 300,000 housing units altogether which will contain about 750,000 persons.

We intend to produce estimates of total population for each stratum for various demographic categories. The estimate of the total population is

$$\hat{N} = \frac{(N_c - \hat{E})\hat{N}_p}{\hat{N}_m}$$

where N_c denotes the census count for the stratum

$\hat{E}$ denotes the estimates of erroneous enumerations

$\hat{N}_p$ denotes the estimate of the stratum population from the PES

and

$\hat{N}_m$ denotes the estimate of the matched persons between the PES an the census

The variance of this estimator was approximated as follows:

$$V(\hat{N}) = \frac{NP_u[1-P_u]\ [1+(\bar{m}-1)\rho]}{b\bar{m}} + \frac{NP_e[1-P_e]\ [1+(\bar{m}-1)\rho]}{b\bar{m}}$$

where P_u is the proportion of persons expected to be missed by the census

P_e is the proportion of persons erroneously enumerated

$\bar{m}$ is the average block size

b is the number of sample blocks

ρ is the intraclass correlation that arises from a block sample for measuring over and underenumeration.

This estimator assumes no correlation between the estimates of over and underenumeration. The estimator then is conservative, in the sense that any positive correlation would reduce the variance of $\hat{N}$. We also assumed that we wanted a coefficient of variation for $\hat{N}$ of about 1 percent in each stratum.

Since we must provide for a variety of situations in the PES, we assumed the following situations:

		P_u	P_e	ρ	$\bar{m}$	Value of Resulting Sample Size
Situation	1	0.05	0.03	0.1	80	100
	2	0.07	0.03	0.1	80	100
	3	0.05	0.04	0.1	80	100

Thus if 10 of the 100 strata were situation 2, 10 were situation 3 and the remaining 80 situation 1, the required sample size is about 300,000. If 10 were situation 1, 10 were situation 2, and 80 were situation 3, the sample size would be about 333,000. These two examples seem to be the extremes of what could happen in 1990.

In the Los Angeles test census of 1986, we tried a small-scale version of this technique. Of course, in an area of less than 400,000 persons, it is difficult to simulate the 1990 census activity. Nonetheless, if there were difficulties in the smaller area, they would be important to solve before the next test.

The Los Angeles test site had three major racial or ethnic groups: Hispanics, Asians, and Whites. Very few Blacks lived in the area. Sampling strata were defined as follows:

Hispanics in large multiunit structures

Hispanics in small multiunit structures

Hispanics in single units

Asians

Non-Hispanic, Non-Asian

Blocks with 2 or fewer housing units

After the data were collected, we post-stratified the sample in order to carry down the estimates of the undercount to the block level. This work is described by Diffendal (3). It had been shown in earlier work that people who rent their units rather than own them are more likely to be undercounted.

Finally, all post-strata were crossed by age and sex. Thus, there were 8 stratification variables x 2 sex variables x 5 age variables, giving 80 post-strata estimates. The post-strata were as follows:

Hispanic renters in multiunit structures in blocks having 50 percent or more Hispanics

Hispanic renters in single unit structures in blocks having 50 percent or more Hispanics

Hispanic owners in blocks having 50 percent or more Hispanics

Hispanics in all other blocks

Asian renters in all blocks

Asian owners in all blocks

Non-Hispanic, Non-Asian renters in all blocks

Non-Hispanic, Non-Asian owners in all blocks

Eighty adjustment factors, the estimates of the population divided by the census count, were derived. Some of these were based on small samples and had large sampling variances. To reduce the sampling error of the adjustment factors, a Bayesian regression model was fit to the 80 adjustment factors. Indicator variables for post-strata, and age, sex, and race within post-strata were the independent variables for the Bayesian regression model. The adjustment factors were averaged with the Bayesian regression estimates to produce the final adjustment factors. Table 3 shows the final results.

Putting together all these considerations, we feel we have made considerable progress since 1980 in designing a sample, showing effective use of

TABLE 3. Results of Smoothing Adjustment Factors in Los Angeles Test Census

Post Stratum	Sex	Age	Adjustment Factors		Standard Error of Adjustment Factors	
			Original	Smoothed	Original	Smoothed
Hispanic renter in Hispanic block	Male	0-14	1.131	1.130	.020	.016
"	"	15-29	1.247	1.211	.030	.021
"	"	30-44	1.165	1.144	.029	.020
"	"	45-64	1.099	1.114	.043	.024
"	"	65+	1.055	1.110	.044	.023
Hispanic renter in Hispanic block	Female	0-14	1.124	1.126	.023	.018
"	"	15-29	1.234	1.203	.032	.022
"	"	30-44	1.084	1.098	.017	.015
"	"	45-64	1.125	1.121	.040	.024
"	"	65+	1.099	1.122	.045	.024
Hispanic owner in Hispanic block	Male	0-14	1.056	1.050	.018	.015
"	"	15-29	1.078	1.084	.018	.015
"	"	30-44	1.087	1.072	.016	.014
"	"	45-64	1.031	1.031	.012	.011
"	"	65+	1.073	1.054	.028	.019
Hispanic owner in Hispanic block	Female	0-14	1.059	1.051	.020	.016
"	"	15-29	1.088	1.090	.016	.014
"	"	30-44	1.033	1.034	.012	.011
"	"	45-64	1.020	1.022	.012	.011
"	"	65+	1.033	1.035	.019	.015
Hispanic in non-Hispanic block	Male	0-14	1.105	1.051	.052	.023
"	"	15-29	1.154	1.106	.054	.025
"	"	30-44	1.131	1.050	.065	.024
"	"	45-64	1.063	1.036	.050	.023
"	"	65+	0.999	0.999	.000	.000
Hispanic in non-Hispanic block	Female	0-14	1.137	1.059	.047	.023
"	"	15-29	1.033	1.060	.022	.017
"	"	30-44	1.079	1.051	.037	.021
"	"	45-64	1.033	1.031	.028	.019
"	"	65+	0.947	1.013	.040	.022
Asian renter	Male	0-14	1.059	1.076	.041	.026
"	"	15-29	1.127	1.137	.044	.028
"	"	30-44	1.195	1.093	.077	.031
"	"	45-64	1.004	1.063	.057	.030
"	"	65+	0.999	0.999	.000	.000

TABLE 3. Results of Smoothing Adjustment Factors in Los Angeles Test Census - continued

Post Stratum	Sex	Age	Adjustment Factors		Standard Error Adjustment Factors	
			Original	Smoothed	Original	Smoothed
Asian renter	Female	0-14	1.067	1.079	.047	.028
		15-29	1.215	1.153	.055	.029
		30-44	1.173	1.087	.105	.032
		45-64	1.012	1.065	.061	.030
		65+	1.212	1.087	.127	.032
Asian owner		0-14	1.045	1.041	.030	.019
		15-29	1.059	1.085	.038	.022
		30-44	1.091	1.053	.040	.022
		45-64	1.035	1.033	.020	.016
		65+	1.031	1.037	.051	.023
Asian owner		0-14	1.040	1.039	.041	.022
		15-29	1.052	1.086	.046	.024
		30-44	1.035	1.037	.036	.021
		45-64	1.038	1.035	.019	.015
		65+	1.051	1.041	.045	.022
Non-Hispanic, non-Asian renter		0-14	1.037	1.049	.059	.027
		15-29	1.252	1.115	.114	.031
		30-44	1.144	1.062	.066	.028
		45-64	1.055	1.047	.031	.022
		65+	1.068	1.054	.056	.027
Non-Hispanic, non-Asian renter		0-14	1.148	1.064	.062	.027
		15-29	1.126	1.112	.054	.028
		30-44	1.134	1.064	.057	.027
		45-64	1.068	1.049	.041	.025
		65+	0.948	0.992	.021	.018
Non-Hispanic, non-Asian owner		0-14	1.044	1.040	.037	.021
		15-29	1.148	1.103	.064	.025
		30-44	1.006	1.032	.048	.023
		45-64	1.036	1.034	.017	.014
		65+	1.017	1.025	.019	.016
Non-Hispanic, non-Asian owner		0-14	1.159	1.052	.068	.024
		15-29	1.081	1.092	.042	.023
		30-44	0.997	1.011	.017	.014
		45-64	1.025	1.026	.012	.011
		65+	0.997	1.004	.012	.011

Hispanic blocks are blocks in which 50 percent or more of the population is Hispanic.

post-stratification, using regression methods to take care of outliers, and other techniques.

3.2 Matching Records

The blocks selected into the PES are assigned for independent listing. W do not depend on the lists of housing units assembled for the census. This may be one of the strong points of the PES, since unusual living arrangements that do not fit neatly into census definitions abound. For example, we found in Los Angeles that in certain parts of the city, the census had listed a single family house at an address. On closer inspection, one could find out that the garage had been made into a living quarters and that a recreational vehicle such as a van was set up in the backyard as another living quarters. Why do people do this? An explanation offered us was that the rents for the original house are high and to help pay the rent, additional people are brought in. These new living quarters are not revealed to the city, so they would never appear on original census lists. However, in a PES, the listers are in the block, frequently see lights on in the garages and vans, and list these units. Even with this, we do not get them all.

After the listing, PES enumerators go into the sample blocks and interview at all the housing units. They do not repeat the census questionnaire. Instead they concentrate on coverage. They ask additional questions only to help us locate or match individuals.

Matching in 1980 as well as earlier years was the operational stumbling block to success. In 1980, there were 150,000 housing units containing 350,000 people to be matched. This was all done clerically. Every case was handled

in a laborious, time-consuming way. Clerks were trained to look in the census for the housing unit within a geographical area that matched the CPS housing unit and then to match each of the people within the housing unit. It was often difficult for the clerks to know whether they were in the right geographic area. When they matched people, they were instructed to match names, sex, race, and relationship within housing unit. Unless everything matched with no discrepancies, the case was given to another set of clerks to review and make decisions about. Finally, a group of statisticians in Washington reviewed all cases that did not match right away. This operation took months and was prone to much error. Some of the difficulty was caused by the lateness of the operation. The match operation started in the fall of 1980. The followup began in the winter of 1981, almost a full year after the census. Matching continued until the end of 1981, with some sporadic work in 1982.

When planning began for 1990, we decided that computerizing the match should have a high priority. If the computer could match the cases that were simple, a clerical staff could spend their time on more "interesting", more difficult cases. Also, the computer could be used to assist the clerical matching process.

The lists we will be matching in 1990 are from two census processes. One of them is the census itself and the other the PES, compiled a few months after the census. The records on these two files are compared to see how similar they are. The success of matching is dependent on the quality of the records on the file. When two records are identified for which the characteristics match, they are assigned a match code. A record on one file without a record on the other file with enough similar information is assigned a not matched code or, sometimes, a possible match code.

The range of the area over which the searching for like records takes place must be limited. We cannot search the entire census file for a record on the PES file. The census file is blocked into subsets depending on selected characteristics. A probabilistic model, based on theory developed by Fellegi and Sunter (5), is the basis for the matching within these subsets.

3.3 Treatment of Nonresponse

In the 1980 census, nonresponse was a serious problem. Because we used an existing survey, the CPS, we had to accept the nonresponse from that survey. It was 4.4 for April and 5.3 for August. Then, after the matching of records occurred, cases that did not match were sent back to the field for followup. That got more nonresponse. In addition, there were geographical coding problems. So the final nonresponse in 1980 was 8.4 for April and 9.7 for August. Those rates were high for trying to measure something at the level of 1 to 2 percent. Even worse, the nonresponse was not spread evenly over the population. It was difficult to see how the population distribution of the states was going to be improved when the quality of the data was so variable over the states and cities.

In the 1980 study, if the CPS case could not be matched to a census case with certainty, it was sent out to be contacted again by an enumerator. About 13 percent of the cases were sent out for this further follow-up. Among these cases were the cases that could not be resolved. These latter cases were imputed a match status based only on the group of cases that went to follow-up. Thus, only 13 went to follow-up and 15-20 percent of those were never resolved. Those never resolved were imputed from those that were resolved. For that reason, a very high proportion of the unresolved cases were assigned to a

"missed" category. In fact, the percentage of the missed category that came from imputations was over 40 percent.

This method of imputing for the unresolved cases came under a lot of scrutiny. Those who defended the imputation said that the unresolved were difficult cases since they had gone to follow-up in the first place and thus were more likely to be missed cases. However, at least seven percent of those sent to followup were sent because of timing problems. A match to the census had never been attempted. Those who opposed this imputation said the donor pool was too limited. Alternative methods of imputation were tried. Using the entire sample for the donor pool was tried, and so was pulling out parts of the unresolved and treating them differently. The results from these various treatments gave very different results as was shown in Table 2. Some of them showed net overcounting and some net undercounting. It was clear that not only was there too much missing data, but that the strategy for handling it was not robust.

In addition, there was missing data for the Enumeration Sample. About 4.7 of the cases were unresolved and were imputed for. Approximately 30 percent of the erroneous enumerations came from imputation. Again, questions were raised on what the donor pool should be.

With that in mind, we devised a strategy for 1990 and tested it in Los Angeles. First of all, we monitored the nonresponse rate in the field so that steps were taken in a timely way to keep nonresponse at a low level. In 1980, we knew what it was only at the end of the processing. Second, we allowed 3 weeks in the field to complete the PES interviewing compared to only one week in 1980.

In addition, we developed a questionnaire in the PES that had several questions on it to help us locate individuals who moved. Using this strategy in the Los Angeles test, we were able to keep nonresponse rates to 4.5 for the PES sample, about half of what they were in 1980. The Enumeration Sample cases that did not match were sent for follow-up.

The other thing that will be different in 1990 is the imputation strategy. We also developed and tested this in Los Angeles. There were some housing units for which there was no response. We used a weighting adjustment for those units. That means that within a sampled block, the sampling weight, which was identical for every person, was inflated by the inverse of the completed-interview rate for the block. This kind of weighting adjustment is based on the assumption that the households not interviewed are the same as those interviewed. This is probably not the case, but it is a conservative treatment.

In post-stratifying the PES samples, certain key variables need to be present. In the Los Angeles test census, those variables were tenure of housing unit and size of structure, sex, age, and race. When these variables were missing, imputation was used. Table 4 shows the missing data rates for these items in Los Angeles.

TABLE 4. Percent of Missing Data in Los Angeles Test for Characteristics Needed for Post-Stratification

Characteristic	Post-Enumeration Sample	Enumeration Sample
Tenure	3.5 %	0.7%
Structure type	2.3	1.6
Sex	2.1	0.4
Age	0.7	2.1
Race	0.8	7.0

Missing data for these characteristics in both the P and E samples were imputed by means of a hot-deck computer system in which data from recently processed cases were used for imputation. A hot-deck procedure is standard practice at the Census Bureau for the decennial census and many household surveys. The only added feature was that the imputation was done in two passes. On the first pass, tenure, structure, and race were imputed using the most recently observed data. On the second pass, sex and age were imputed at random from distributions tabulated during the first pass using all observed data. The results from the imputation confirmed other experiences. More males were imputed, as were rented housing units, and multi-unit structures.

Another very important characteristic that was missing was the match status for the PES sample and the enumeration status for the Enumeration Sample. This was missing in the PES sample usually because there was not enough information to match or there were movers for which there was trouble finding a Census Day address.

In 1980, a match status was imputed. A person was assigned as matched or not matched. In 1990, we will impute a match probability. This was tested in Los Angeles. A logistic regression approach was used to impute the match probabilities. If X denotes a vector of predictor variables, Y = M or N (matched or not matched) and p = Pr(Y = M|X), then the parameter vector β of the logistic regression model

$$\text{logit}\ (p) = \log\,[p/(1-p)] = X'\beta$$

was estimated from the data for the resolved cases using Bayesian techniques. Then, for an unresolved case j, with $X = x_j$, the imputed match probability was

$$\hat{p}_j = \text{logit}^{-1}\ (x'_j\hat{\beta}) = \frac{\exp\ (x'_j\hat{\beta})}{[1 + \exp(x'_j\hat{\beta})}$$

where $\hat{\beta}$ is the estimate of β.

Using these techniques in the Los Angeles test, we found the match rate for resolved PES sample cases to be 87.8 percent and the imputed match rate for the unresolved PES sample cases to be 77.4 percent.

For the Enumeration Sample, there will be cases that come back after followup for which there will not be a clearly defined status of correct or erroneous enumeration in the census. This can happen when the respondents in the followup says they have no knowledge of the person in question, a potential indication of fabrication in the census; when the followup is a noninterview; and when not enough information is provided to make a determination.

As with the PES sample, a probability of erroneous enumeration was imputed for each unresolved case in Los Angeles. Since missing correct or erroneous enumeration status resulted solely from followup, only the resolved cases from followup were used in estimating the logistic regression. In the Los Angeles test, the percent erroneous enumeration for the non-followup cases was 1.6 percent. For the imputed cases, it was 2.2 percent.

3.4 Dual-System Estimation

There are alternatives in the ways the data from the PES can be used in a model. For several years, a version of a dual-system estimator based on capture-recapture models has been discussed. Such a model was used in 1980 and was subject to considerable criticism of the underlying assumptions. However, the 1986 paper by Wolter (9) clearly lays out a variety of alternative models with their underlying assumptions and develops very clearly the model the Census Bureau will use in 1990. Again, this model was tested and refined in the Los Angeles Test Census.

The model is as follows:

There is a census and a survey, both used to provide information on the size of the total population. For simplicity, assume first that the sample is a complete enumeration. The assumptions underlying the model are:

1. The population is closed and of fixed size N. Given the long enumeration period in the United States, and the fact that people move in and out of the country as well as within the country, this assumption is incorrect. We use estimates of erroneously enumerated from the Enumeration Sample to correct for this.
2. The joint event that the i-th individual is in the census or not, and in the sample or not, is correctly modeled by the multinomial distribution with the following parameters.

Survey

Census		in	out	
	in	p_{i11}	p_{i12}	p_{i1+}
	out	p_{i21}	p_{i22}	p_{i2+}
		p_{i+1}	p_{i+2}	1

3. The census and survey population estimates are created as a series of N mutually independent trials, where N is the fixed but unknown size of the population. There is a multinomial distribution that

is the basis of each trial, where each trial corresponds to a member of the population. The resulting data are

Survey

Census		in	out	
	in	x_{11}	x_{12}	x_{1+}
	out	x_{21}	x_{22}	x_{2+}
		x_{+1}	x_{+2}	$x_{++} = N$

where $x_{ab} = \Sigma x_{iab}$ and x_{iab} is an indicator random variable signifying whether the i-th individual is in cell (a,b), for a,b = 1, 2, +. In the PES situation, x_{11}, x_{12}, x_{21} are observable after the matching operation, and x_{1+} is the census count. Unobserved is x_{22}.

4. Matching of cases between the census and survey can be done without error.
5. Both the census and the sample survey contain no spurious elements. Again, the Enumeration Sample is of value here.
6. Nonresponse can be coped with so that exact matching can proceed.
7. Since post-stratification is desirable, the variables used are completely and correctly recorded in the Census and sample survey.
8. The event of being included in the Census is independent of the event of being included in the sample. That is, the cross-product ratio, θ_i, satisfies

$$\theta_i = \frac{p_{i11}\, p_{i22}}{p_{i12}\, p_{i21}} = 1 \quad \text{for } i=1, \ldots N.$$

Wolter has called this causal independence. This assumption has been the focus of much scrutiny. It seems unlikely that this assumption holds for some groups of people -- for example, those that fear the government and any contact with it are likely going to be missed in the census and the survey.

9. The capture probabilities satisfy $p_{i1+} = p_{1+}$ and $p_{i+1} = p_{+1}$ for i=1, ... N.

The traditional estimator used in this model is $\hat{N} = \frac{x_{1+}x_{+1}}{x_{11}}$. When a sample is used, sample-based estimates of x_{+1} and x_{11} are used. Wolter (9) went on to describe the statistical properties of the sample-based estimator of N. An important result was that showing that the estimator has two sources of variability: sampling variability and model variability.

This model was used to estimate the undercount in the Los Angeles test. To meet census conditions, we used the estimator

$$\hat{N} = \frac{(N_C - \hat{E})\hat{N}_P}{M}$$

where these terms are defined on page 10.

3.5 Indirect Estimation

At the completion of the estimation stage, there will be direct estimates at a post-stratum level where a post-stratum might be composed of Black renters in multi-unit structures in cities of over 1 million persons, crossed by age, race, and sex. To be useful in the census, that level of undercount must be distributed to all the components of the stratum, that is, to all the blocks in the stratum. Some questions arise concerning the level at which an improvement over census data will be made and what method should be used

to distribute the undercount. Work on this at the Census Bureau has been going on using synthetic populations to measure any improvement. Use of synthetic populations can always be criticized since they may be different from the real population, but there is no alternative. To cope with this problem, Isaki, Diffendal, and Schultz (6) have constructed three different artificial populations, using different assumptions about the number of illegal aliens in the country. Within the 1980 Census, there are people imputed into the census. The imputations are made by taking people already counted and replicating them again. These cases are called substitutions and they occur when: no census questionnaire was completed but people may have lived in a housing unit, only the number of people who lived in a housing unit was known but not the characteristics, for machine failure, and when the field counts for an area were larger than the counts after machine processing. In all of the artificial populations, these substitutions were used as a proxy for the undercount. An analysis using state data in 1980 showed that the census substitution rate was the most important explanatory variable of several types of nonmatch rates. Since the nonmatched rates are the basis of the missed rates, substitutions were used as a proxy for the undercount. Artificial population 1 (AP1) uses census minus substitutions as the census count and substitutions as the undercount. AP2 and AP3 were constructed so that the totals at the national level by age, race, and sex equaled an independent estimate of the total population provided by the method of demographic analysis, assuming 3.5 million illegal aliens in the U.S. In both AP2 and AP3, the substitution counts are adjusted by factors F_D, the ratio of the difference between

the demographic analysis estimate N_D and the census, to the total substitutions. Thus

$$F_D = \frac{N_D - \text{Census}}{\text{Substitutions}}$$

Thirty factors were required -- five age categories, two sex, and three race or ethnic categories. The three race or ethnic factors were Black, Non-Black Hispanic, and Rest.

The difference between AP2 and AP3 is in the treatment of the Hispanic population. Demographic analysis does not provide an estimate for that population. For AP2, the Hispanics were assumed to be like the Non-Hispanics and those factors were used. For AP3, the Hispanics were assumed to be like the Black population, and those factors were used.

Three different types of synthetic estimators were used. One used age, race, and sex groups by geography, emphasizing urban-rural differences. This estimator, called syn 1, had 90 adjustment factors. A second estimator emphasized census divisions and size of place within division and was called syn 2. This estimator had 96 factors. The third used more detailed age categories in an age-race-sex stratification, but no geographic substrata below the U.S. level. This was called syn 3.

Each of the three estimators were used to estimate the total population and population by race for states and counties for each artificial population. Several summary measures suggested by Preston and Schirm (8), were used to evaluate the performance of the synthetic estimators.

Three of these measures relate to counts of areas with certain characteristics. Table 5 shows these three measures for the three artificial populations and the three synthetic estimators compared to the census. The first measure compares the absolute relative errors of the standard count for the area as represented by the artificial population. So

$$ARE(c_i) = \left| \frac{c_i - s_i}{s_i} \right|$$

where C_i is the census count for the i-th area and s_i is the standard. Similarly,

$$ARE(e_i) = \left| \frac{e_i - s_i}{s_i} \right|$$

where e_i is the estimated count from the synthetic estimator.

In the first measure in Table 5, we are comparing the number of states for which the absolute relative error is less for the census than for the synthetic estimate. Notice that for the total population, this is a small number.

The second measure is the absolute proportional error, ADP, where

$$ADP(c_i) = \left| \frac{c_i}{N} - \frac{s_i}{N} \right|$$
$$\sum_{i=1} c_i \qquad \sum_{i=1} s_i$$

and

$$ADP(e_i) = \left| \frac{e_i}{N} - \frac{s_i}{N} \right|$$
$$\sum_{i=1} e_i \qquad \sum_{i=1} s$$

TABLE 5. Counts of Areas with Certain Characteristics using Synthetic Estimators Compared to the Census at the State Level Using Three Artificial Populations for Total Population

	AP1				AP2				AP3			
	Syn 1	Syn 2	Syn 3	Census	Syn 1	Syn 2	Syn 3	Census	Syn 1	Syn 2	Syn 3	Census
Number of states where $ARE(c_i) < ARE(e_i)$	9	4	7	-	9	5	8	-	9	7	6	-
Number of states where $ADP(c_i) < ADP(e_i)$	14	12	13	-	13	15	14	-	9	8	8	-
Apportionment	2	2	2	2	2	0	2	6	4	2	4	8

Then the second measure in Table 5 shows the number of states for which

$$ADP(c_i) < ADP(e_i)$$

Again, for all three synthetic populations, the results show that this is a small number of states. Both absolute relative errors and proportional errors are made in substantially fewer states when a synthetic estimator is used.

The third measure pertains to the apportionment of seats in the House of Representatives and shows how many are erroneously assigned. For AP1, the census and all three synthetic estimators behaved equally in error. On AP2, all three synthetic estimators were better than the census, with syn 2 performing best. AP3 showed similar results with higher levels of error.

The next types of summary evaluation measures involve error assessment of the absolute level of the adjustment estimates. These measures are compared in Table 6. The first of the three is just the mean absolute relative error

$$MARE = \frac{1}{N} \sum_{i}^{N} \left| \frac{e_i - s_i}{s_i} \right|$$

The next two are also based on the absolute relative error, one being the maximum, and the other the median. The final measure in their group in the weighted squared relative error, α, where

$$\alpha = \sum_{i=1}^{N} \frac{(e_i - s_i)}{s_i}^2$$

TABLE 6. Error Measurements Related to Absolute Level of Adjustment Estimates Using Synthetic Estimation Compared to the Census at the State Level Using Three Artificial Populations

	AP1				AP2				AP3			
	Syn 1	Syn 2	Syn 3	Census	Syn 1	Syn 2	Syn 3	Census	Syn 1	Syn 2	Syn 3	Census
Absolute relative error	.0054	.0042	.0052	.0134	.0052	.0044	.0053	.0147	.0050	.0045	.0047	.0136
Maximum ARE(e)	.0169	.0147	.0190	.0398	.0183	.0200	.0297	.0771	.0184	.0228	.0300	.0773
Median ARE(e)	.0050	.0028	.0048	.0121	.0048	.0026	.0047	.0113	.0040	.0026	.0032	.0092
Weighted squared relative error	8336	4504	8533	55221	9074	6179	9925	77513	8979	5866	9344	82339

This latter measure is a favorite of many, because it gives greater weight to errors in larger places.

From Table 6, we see that each of the three synthetic estimators outperforms the census for all three artificial populations. Generally, syn 2 did the best.

The final three quantities measure the error in the proportionate shares of the population derived from the adjustment estimates. Some people are more interested in proportional distributions rather than absolutes. Table 7 shows these results for the three artificial populations. The first of these measures is

$$SADP = \sum_{i=1}^{N} \left| \frac{c_i}{\sum_{i=1}^{N} c_i} - \frac{s_i}{\sum_{i=1}^{N} s_i} \right|$$

the sum of the absolute proportional errors.

The second is:

$$PI = \sum_{i=1}^{N} \frac{IMPV_i}{M}$$

where $M = \sum_{i=1}^{N} s_i$

and

$$IMPV_i = s_i \text{ if } \left| \frac{e_i}{\sum e_i} - \frac{s_i}{\sum s_i} \right| < \left| \frac{c_i}{\sum c_i} - \frac{s_i}{\sum s_i} \right|$$

$$= 0 \text{ otherwise}$$

When this measure is greater than 1/2, then the adjustment provides better data for over half of the population.

TABLE 7. Error Measurements Related to Proportionate Shares of the Population Using Synthetic Estimation Compared to the Census at the State Level Using Three Artificial Populations

	AP1				AP2				AP3			
	Syn 1	Syn 2	Syn 3	Census	Syn 1	Syn 2	Syn 3	Census	Syn 1	Syn 2	Syn 3	Census
Sum of absolute proportional errors SADP	.0048	.0031	.0048	.0052	.0048	.0037	.0049	.0067	.0048	.0033	.0047	.0078
Proportion of population improved PI	.622	.830	.654	-	.757	.703	.694	-	.701	.872	.715	-
Weighted squared relative error differences	8332	4501	8211	9735	9073	6179	9758	17368	8923	5810	9266	22048

Finally, there are weighted squared relative error differences,

$$\Phi = \sum_{i=1}^{N} s_i \left\{ \left[\frac{e_i - s_i}{s_i} \right] - \left[\frac{\sum e_i - \sum s_i}{\sum s_i} \right] \right\}^2$$

With these measures, too, the three synthetic estimators produced better results than the census. Syn 2 almost always indicated smaller error than syn 1 and syn 3. There was some difference in behavior of the estimators over the three artifical populations, especially for syn 2. This is because syn 2 treated Blacks and Hispanics alike which favors its performance under AP3 but not AP2.

All of these results are available by race group and show similar patterns. From these data, it seems that states would be improved by use of synthetic estimation and that the synthetic estimates are generally superior to the census. The same sort of analysis was also done for counties. For counties, syn 3 performed better. The universe of counties, 3137 of them, were then separated by size into three groups. Group 1 included counties with population of 10,000 or less; Group 2 included counties with population between 10,000 and 50,000; and Group 3 included counties with population greater than 50,000. These groups included 25, 50, and 25 percent of all counties, respectively. It turned out that syn 3 did best for Group 1, syn 2 for Group 3 and there was no clear picture for Group 2. The absolute relative error was reduced in about two-thirds of the counties. At the present time, we are examining data for smaller geographic areas than counties.

Regression models were also tested, using one of the twelve 1980 PEP estimates as the standard. All evaluations were made at the state level. The regression models were formed at the district office level of aggregation and were used to predict the district office population counts. These were summed to the state level to compare to the assumed standard.

Three models of net undercount using unweighted linear regression were compared. The assumed model for all three was

$$\underline{Y} = \underline{X}\,\underline{\beta} + \underline{\varepsilon} \quad \text{where} \quad \varepsilon \quad N\,(o,\ \sigma^2 I).$$

The variables, $\underline{X}$, that predict the percent net undercount, $\underline{Y}$, are census variables. One model used only one variable -- percent of non-vacant renter occupied housing that is lived in by minority populations. The second regression used two variables, adding the variable of the percent of total population that had not attended high school. Model 2 was slightly better than model 1, but neither was too impressive.

In forming the next model, the district offices were divided into three groups and each group had its own model. The three groups were based on the kind of census conducted there. One group was in large cities and had the mailout/mailback census with enumeration followup. A second group was in rural areas with enumerators taking the census. The third group was in surburan areas and for which the mailout/mailback census worked best. This group contained the largest part of the district offices.

Using the three models, each group had different explanatory variables. These were combined into one model using indicator variables. All three models were compared with the census, and were superior to it.

Other approaches, such as weighted regression and Bayesian hierarchical regression models, have also been developed and tested. All approaches seem to provide an improvement, but with many differences among models.

4. Conclusions

The issue of accurate census coverage became a statistical problem when we learned how to measure the undercount. The kinds of statistical issues are many and cut across many areas: conceptualization, measurement and measurement error, modeling, validation of models, and operationalizing statistical procedures in a census. Besides being an extremely interesting statistical problem, the undercount measurement and adjustment program has significant public policy implications.

REFERENCES

(1) Coale, Ansley J., March 1955. "The Population of the United States in 1950 Classified by Age, Sex, and Color - A Revision of Census Figures," Journal of the American Statistical Association, Vol. 50: pp 16-54.

(2) Conference on Census Undercount, Proceedings of the 1980 Conference, July, 1980, Washington, D.C., Government Printing Office.

(3) Diffendal, Gregory, 1987. "1986 Test of Adjustment Related Operations, Procedures and Methodology," presented at the April 9-10, 1987 meeting of the Census Advisory Committee of the American Statistical Association.

(4) Ericksen, Eugene and Joseph Kadane, "Estimating the Population in a Census Year: 1980 and Beyond" (with discussion), Journal of the American Statistical Association, Vol. 80: pp 98-131.

(5) Fellegi, Ivan P. and Alan B. Sunter, 1969. "A Theory for Record Linkage," Journal of the American Statistical Association, Vol. 64: pp 1183-1210.

(6) Isaki, Cary, Greg Diffendal, and Linda Schultz, 1986. "Survey Based Synthetic Estimates of Undercount for Small Areas" in Proceedings of the Second Annual Research Conference, Washington, D.C., U.S. Bureau of the Census.

(7) Panel on Decennial Census Methodology, 1985. The Bicentennial Census: New Directions for Methdology in 1990, C.F. Citro and M.L. Cohen, editors, Washington, National Academy Press.

(8) Preston, Samuel H. and Allen L. Schirm, 1986. "Census Undercount Adjustment and the Quality of Geographic Population Distributions," to be published in the Journal of the American Statistical Association.

(9) Wolter, Kirk M., 1986. "Some Coverage Error Models for Census Data", Journal of the American Statistical Association, Vol. 81: pp 338-346.

Population Reference Bureau, Inc.
777 14th St., N.W., Suite 800, Washington, D.C. 20005
Phone: (202) 639-8040 Fax: (202) 347-1690 Telex: 4900010456(PRB UI)

PRB

April 5, 1989

Mr. Tino Calabia
U.S. Commission on Civil Rights
Eastern Regional Division
1121 Vermont Ave. NW
Room 710
Washington, D.C. 20425

Dear Tino,

Thanks for your letter of March 31, 1989, inviting me to join you at the New York State Advisory Committee meeting on Census Undercounts and Preparations for the 1990 Census: Part II.

Unfortunately, I will not be able to attend the meeting in New York on April 27th, but I am enclosing a couple of items that you might want to share with your colleagues who share our interest in the topic of census adjustment. The first item is a recent edition of Population Today which contains a short article on the potential impact of several types of adjustments to the Census data. The second item is a PRB Working Paper which I completed last fall. This working paper provides more detailed information about the impact of potential adjustment.

I am sorry I won't be able to join you at the meeting on April 27th. If it is possible, please share the enclosed material with those at the meeting. Let's keep in touch.

Sincerely,

Bill

William P. O'Hare
Director of Policy Studies

DEMOGRAPHER'S PAGE

Effects of Census Adjustment

by William P. O'Hare

With the 1990 Decennial Census a little over a year away, many political observers are beginning to think about the impact the next census will have on the apportionment of Congress. With large population shifts from the Northeast and Midwest to the Sunbelt states, many seats in Congress will change.

These major shifts will unquestionably change the makeup of Congress. But another, separate issue is whether the decennial census count should be *adjusted*, with the theoretical goal of making it more accurate. Such an adjustment has never been done, but several powerful interest groups are now lobbying for it.

Not only whether to adjust, but *how* to adjust are important questions. The potential impact of various adjustment scenarios on the apportionment of Congressional seats following the 1990 Decennial Census follows.

Proposed Adjustments

The change most widely discussed is the possible adjustment of the census count to correct for the undercount of minorities. A long series of studies by the Census Bureau show that over the last several censuses, minorities have been missed at a much higher rate than whites. About 5.9 percent of blacks were missed in the 1980 Census, compared to about 1 percent of whites and other races.[1] While the data are less reliable, it is generally believed that Hispanics are missed at about the same rate as blacks.[2] Following the 1980 Census, many large cities sued the Census Bureau hoping that the courts would compel the Bureau to adjust its figures to reflect the uncounted minorities. After protracted litigation, the Census Bureau position prevailed and counts were not adjusted.

The issue of undercount adjustment in the 1990 Census is already being raised by powerful interest groups. For example, a House Bill (HR 3511) which would require the Census Bureau to adjust census figures based on the undercount of minorities was seriously entertained by the House Post Office and Civil Service Committee before going down to defeat, in large part because of the amendments that were added to it. A suit has already been filed in federal court asking that the Bureau be required to make statistical adjustments. No trial date has been set for this case.

The second adjustment issue that has received popular attention involves undocumented aliens. Just prior to the 1980 Census, the Federation for American Immigration Reform (FAIR) sued the Census Bureau in an attempt to have the courts make the Bureau eliminate any illegal aliens from the numbers used in Congressional apportionment. That suit was thrown out on technical grounds.[3]

There is a long history of legislative efforts[4] to have illegal aliens removed from the census counts that goes back as least as far as 1929. During the last Congress, five separate bills were introduced which would exclude illegal aliens from the population count used to apportion the House of Representatives.[5] A suit was filed in February 1988 which would accomplish the same thing if successful.

The third type of adjustment that resulted in litigation following the 1980 census involves imputation – a technical procedure employed by the Census Bureau. In a few cases, Census Bureau enumerators are unable to make contact with people in a housing unit, even though neighbors have indicated that the housing unit is definitely inhabited. If enumerators are unable to make contact after several callbacks, the Census Bureau notes that the housing

Table 1. Eight Scenarios for Adjustments to the 1990 Census

Scenario 1.	No adjustment to Census figures.
Scenario 2.	The minority population (blacks plus Hispanics) in each state is increased by 5 percent.
Scenario 3.	The number of undocumented aliens is subtracted from each state's figure.
Scenario 4.	The number of imputed persons is subtracted from each state's figure.
Scenario 5.	Two adjustments are made simultaneously: 1) The minority population in each state is increased by 5 percent, and 2) The number of undocumented aliens is subtracted from each state's figure.
Scenario 6.	Two adjustments are made simultaneously: 1) The minority population in each state is increased by 5 percent, and 2) The number of imputed persons is subtracted from each state's figure.
Scenario 7.	Two adjustments are made simultaneously: 1) The number of undocumented aliens is subtracted from each state's figure, and 2) The number of imputed persons are subtracted from each state's figure.
Scenario 8.	All three adjustments are made simultaneously.

William O'Hare is Director of Policy Studies at PRB

unit is inhabited but the number of inhabitants is unknown. The Bureau then uses a computerized procedure for assigning, or "imputing," a number of inhabitants for the housing unit based on information about surrounding housing units.

This issue is less well known than the issues of minority undercount adjustment or the exclusion of illegal aliens, but the exclusion of imputed persons from a state's population count could affect the apportionment of Congressional seats.

The imputation procedure accounted for approximately 761,000 persons out of a final count of 226.5 million in the 1980 Census (0.34 percent of the total). Not surprisingly, some states have more of these "imputed" persons than others. Following the 1980 Census, the State of Indiana sued the Census Bureau in an effort to have the people that were imputed by the Census Bureau removed from the count of each state's population that is used for reapportionment. The Census Bureau prevailed in court.

The Result

While adjustment has become an issue in which significant political power is presumed to be at stake, what would actually happen to Congressional apportionment if the 1990 Census were adjusted? Analysis of the impact of possible adjustments to census data requires a projection of state populations to 1990 and some underlying assumptions about the undercount in the 1990 Census. The three major assumptions used in this Demographer's Page are outlined below:

1. Blacks and Hispanics will be undercounted by 5 percent in every state.
2. The number of illegal aliens counted in each state in the 1990 Census will be the same as those estimated in the 1980 Census.
3. The number of people "imputed" in each state's total population in 1990 will be the same as in 1980.

Table 2. The Impact of Various Adjustment Scenarios on Reapportionment Following 1990 Census

State	Projected Congressional seats based on no adjustment	Scenarios. Difference from column 1 based on results of scenario						
	1	2	3	4	5	6	7	8
Ala	7	0	0	0	0	0	0	0
Alaska	1	0	0	0	0	0	0	0
Ariz	7	0	0	0	0	0	0	0
Ark	4	0	0	0	0	0	0	0
Calif	50	0	−1	0	0	+1	−1	−1
Colo	6	0	0	0	0	0	0	0
Conn	6	0	0	0	0	0	0	0
Del	1	0	0	0	0	0	0	0
Fla	22	0	0	0	0	0	0	0
Ga.	12	0	0	0	0	0	0	0
Hawaii	2	0	0	0	0	0	0	0
Idaho	2	0	0	0	0	0	0	0
Ill	20	0	0	0	0	0	0	0
Ind	10	0	0	0	0	0	0	0
Iowa	5	0	0	0	0	0	0	0
Kan	4	0	0	0	0	0	0	0
Ky	7	0	0	−1	0	−1	0	0
La	8	0	0	0	0	0	0	0
Maine	2	0	0	0	0	0	0	0
Md	8	0	0	0	0	0	0	0
Mass	10	0	0	0	0	0	0	0
Mich	16	0	0	0	0	0	0	0
Minn	8	0	0	0	−1	−1	0	0
Miss	5	0	0	0	0	0	0	0
Mo	9	0	0	0	0	0	0	0
Mont	1	0	0	0	0	0	0	0
Neb	3	0	0	0	0	0	0	0
Nev	2	0	0	0	0	0	0	0
N H	2	0	0	0	0	0	0	0
N J	14	0	0	0	0	0	0	0
N M	3	0	0	0	0	0	0	0
N Y	31	0	0	0	0	0	0	0
N C	12	0	0	0	0	0	0	0
N Dak	1	0	0	0	0	0	0	0
Ohio	19	0	0	0	0	0	0	0
Okla	6	0	0	0	0	0	0	0
Oreg	5	0	0	0	0	0	0	0
Pa.	20	0	+1	+1	+1	+1	+1	+1
R.I.	2	0	0	0	0	0	0	0
S C	6	0	0	0	0	0	0	0
S D	1	0	0	0	0	0	0	0
Tenn	9	0	0	0	0	0	0	0
Tex	31	0	0	0	0	0	0	0
Utah	3	0	0	0	0	0	0	0
Vt	1	0	0	0	0	0	0	0
Va.	11	0	0	0	0	0	0	0
Wash.	8	0	0	0	0	0	0	0
W.Va	3	0	0	0	0	0	0	0
Wis.	8	0	0	0	0	0	0	0
Wyom	1	0	0	0	0	0	0	0

Source: PRB projections

Eight different scenarios for adjusting the census are listed in Table 1. The results of these scenarios for reapportionment of Congress following the 1990 Census are presented in Table 2.

It is clear that for most states, none of the adjustments described here will make any difference. For 46 out of the 50 states, none of the adjustments

Continued on next page

alone or in combination would cause the state to gain or lose a seat. The four states that might be affected by an adjustment of the types discussed here are California, Kentucky, Minnesota, and Pennsylvania.

Interestingly, Scenario 2, minority undercount adjustment, the modification of the census figures which has received the most attention, would not result in movement of any seats between states. Some past studies which indicate a big impact from this type of adjustment failed to adjust both the black and Hispanic populations, or made an adjustment in one state or locality but not others.

However, each of the other seven adjustment scenarios would result in a different apportionment of Congress.

The elimination of undocumented aliens from the numbers used for apportionment, Scenario 3, would move a seat from California to Pennsylvania. This is not surprising when you realize that the 1980 Census figures for California included an estimated 1 million illegal aliens. When these million people are taken out of California's projected population, that state loses a seat. Since Pennsylvania is projected to have gained almost enough population for an additional seat anyway, it is not surprising that the seat lost by California goes to Pennsylvania.

Scenario 4 (subtracting imputed persons from the census counts for each state) would result in one seat moving from Kentucky to Pennsylvania.

Scenarios 5 through 8 involve various combinations of adjustments. Scenario 5, which includes minority undercount adjustment and subtraction of undocumented aliens, would result in a seat moving from Minnesota to Pennsylvania.

Scenario 6, minority undercount adjustment and subtraction of imputed persons, would result in the movement of two seats. Under this scenario, Kentucky and Minnesota would each lose a seat, and California and Pennsylvania would gain one seat each.

Scenario 7, which includes subtracting undocumented aliens and imputed persons from each state's figure, would result in one seat moving from California to Pennsylvania.

Finally, if all three adjustments were made simultaneously (Scenario 8), one seat would move from California to Pennsylvania.

Conclusions

It appears that none of the adjustments to census figures that are being urged upon the Bureau by various interest groups are likely to make a big difference in the overall apportionment of Congress following the 1990 Census. Nonetheless, for those few states that may be in jeopardy of losing a seat, and for those states that see the prospect of gaining an additional seat, the adjustment question is significant.

Under most adjustment scenarios, Pennsylvania is likely to gain an additional seat. Minnesota and Kentucky are not likely to gain an additional seat under any of the adjustment scenarios, but they may lose a seat depending on which adjustments are selected. California is unique in that it may gain a seat under some adjustment scenarios, or lose a seat under different adjustment scenarios.

It is important to note that even though adjustment of census figures is not likely to have a major impact on Congressional reapportionment in 1990, adjustment of census figures would have a significant impact on the distribution of federal and state funding based on population figures. Furthermore, adjustment would have an impact on the redrawing of election district boundaries for federal, state, and local offices that follows the census. Consequently, it should be clear that adjustment issues are important for many reasons.

References

1. U.S. Bureau of the Census, "The Coverage of the Population in the 1980 Census," PHC80-E4 (Washington, DC: GPO, February 1988), Table A-80.4.
2. *Ibid.*, Table 7-1.
3. Black, Hugo L., Remarks in the Senate in *Congressional Record*, Vol. 71, May 1929, p. 2078.
4. Williams, Jennifer and David C. Huckabee, *Proposed Exclusion of Illegal Aliens from the Population Used to Apportion the House of Representatives: A Methodological and Policy Analysis*, Congressional Research Service Report 88-418 Gov., June 8, 1988.

THE UNITED STATES COMMISSION ON CIVIL RIGHTS

The United States Commission on Civil Rights, first created by the Civil Rights Act of 1957 and reestablished by the United States Commission on Civil Rights Act of 1983, is an independent, bipartisan agency of the Federal Government. By the terms of the Act, as amended, the Commission is charged with the following duties pertaining to discrimination or denials of equal protection based on race, color, religion, sex, age, handicap, or national origin, or in the administration of justice: the investigation of discriminatory denials of the right to vote; the study of legal developments with respect to discrimination or denials of equal protection; the appraisal of the laws and policies of the United States with respect to discrimination or denials of equal protection; the maintenance of a national clearinghouse for information respecting discrimination or denials of equal protection; and the investigation of patterns or practices of fraud or discrimination in the conduct of Federal elections. The Commission is also required to submit reports to the President and the Congress at such times as the Commission, the Congress, or the President shall deem desirable.

THE STATE ADVISORY COMMITTEES

An Advisory Committee to the United States Commission on Civil Rights has been established in each of the 50 States and the District of Columbia pursuant to section 105(c) of the Civil Rights Act of 1957 and section 6(c) of the United State Commission on Civil Rights Act of 1983. The Advisory Committees are made up of responsible persons who serve without compensation. Their functions under their mandate from the Commission are to: advise the Commission of all relevant information concerning their respective States on matters within the jurisdiction of the Commission; advise the Commission on matters of mutual concern in the preparation of reports of the Commission to the President and the Congress; receive reports, suggestions, and recommendations from individuals, public and private organizations, and public officials upon matters pertinent to inquiries conducted by the State Advisory Committee; initiate and forward advice and recommendations to the Commission upon matters in which the Commission shall request the assistance of the State Advisory Committee; and attend, as observers, any open hearing or conference which the Commission may hold within the State.