OVERSIGHT OF THE 2000 CENSUS: SERIOUS PROBLEMS WITH STATISTICAL ADJUSTMENT REMAIN

HEARING

BEFORE THE

SUBCOMMITTEE ON THE CENSUS

OF THE

COMMITTEE ON GOVERNMENT REFORM AND OVERSIGHT HOUSE OF REPRESENTATIVES

ONE HUNDRED FIFTH CONGRESS

SECOND SESSION

SEPTEMBER 17, 1998

Serial No. 105–201

Printed for the use of the Committee on Government Reform and Oversight

U.S. GOVERNMENT PRINTING OFFICE

54–930 CC WASHINGTON : 1999

For sale by the U.S. Government Printing Office
Superintendent of Documents, Congressional Sales Office, Washington, DC 20402
ISBN 0-16-058357-8

COMMITTEE ON GOVERNMENT REFORM AND OVERSIGHT

DAN BURTON, Indiana, *Chairman*

BENJAMIN A. GILMAN, New York
J. DENNIS HASTERT, Illinois
CONSTANCE A. MORELLA, Maryland
CHRISTOPHER SHAYS, Connecticut
CHRISTOPHER COX, California
ILEANA ROS-LEHTINEN, Florida
JOHN M. McHUGH, New York
STEPHEN HORN, California
JOHN L. MICA, Florida
THOMAS M. DAVIS, Virginia
DAVID M. McINTOSH, Indiana
MARK E. SOUDER, Indiana
JOE SCARBOROUGH, Florida
JOHN B. SHADEGG, Arizona
STEVEN C. LaTOURETTE, Ohio
MARSHALL "MARK" SANFORD, South Carolina
JOHN E. SUNUNU, New Hampshire
PETE SESSIONS, Texas
MICHAEL PAPPAS, New Jersey
VINCE SNOWBARGER, Kansas
BOB BARR, Georgia
DAN MILLER, Florida
RON LEWIS, Kentucky

HENRY A. WAXMAN, California
TOM LANTOS, California
ROBERT E. WISE, JR., West Virginia
MAJOR R. OWENS, New York
EDOLPHUS TOWNS, New York
PAUL E. KANJORSKI, Pennsylvania
GARY A. CONDIT, California
CAROLYN B. MALONEY, New York
THOMAS M. BARRETT, Wisconsin
ELEANOR HOLMES NORTON, Washington, DC
CHAKA FATTAH, Pennsylvania
ELIJAH E. CUMMINGS, Maryland
DENNIS J. KUCINICH, Ohio
ROD R. BLAGOJEVICH, Illinois
DANNY K. DAVIS, Illinois
JOHN F. TIERNEY, Massachusetts
JIM TURNER, Texas
THOMAS H. ALLEN, Maine
HAROLD E. FORD, JR., Tennessee

BERNARD SANDERS, Vermont (Independent)

KEVIN BINGER, *Staff Director*
DANIEL R. MOLL, *Deputy Staff Director*
DAVID A. KASS, *Deputy Counsel and Parliamentarian*
LISA SMITH ARAFUNE, *Deputy Chief Clerk*
PHIL SCHILIRO, *Minority Staff Director*

SUBCOMMITTEE ON THE CENSUS

DAN MILLER, Florida, *Chairman*

THOMAS M. DAVIS, Virginia
JOHN B. SHADEGG, Arizona
VINCE SNOWBARGER, Kansas
RON LEWIS, Kentucky

CAROLYN B. MALONEY, New York
ROD R. BLAGOJEVICH, Illinois
DANNY K. DAVIS, Illinois

EX OFFICIO

DAN BURTON, Indiana

HENRY A. WAXMAN, California

THOMAS W. BRIERTON, *Deputy Staff Director*
KELLY DUQUIN, *Professional Staff Member*
DAVID McMILLEN, *Minority Counsel*

CONTENTS

OVERSIGHT OF THE 2000 CENSUS: SERIOUS PROBLEMS WITH STATISTICAL ADJUSTMENT REMAIN

THURSDAY, SEPTEMBER 17, 1998

HOUSE OF REPRESENTATIVES,
SUBCOMMITTEE ON THE CENSUS,
COMMITTEE ON GOVERNMENT REFORM AND OVERSIGHT,
Washington, DC.

The subcommittee met, pursuant to notice, at 10:03 a.m., in room 2154, Rayburn House Office Building, Hon. Dan Miller (chairman of the subcommittee) presiding.

Present: Representatives Miller, Lewis, Davis of Virginia, Shadegg, Snowbarger, Maloney, Blagojevich, and Davis of Illinois.

Staff present: Thomas W. Brierton, deputy staff director; David Flaherty, senior data analyst; Kelly Duquin, professional staff member; Michelle Ash, minority counsel; David McMillen and Mark Stephenson, minority professional staff members; and Ellen Rayner, minority chief clerk.

Mr. MILLER. Good morning. There is a quorum present and the Subcommittee on the Census will now come to order. I want to welcome our guests here this morning. I think we'll have a very worthwhile and beneficial hearing this morning. I think it is worthwhile for those of us here and hopefully the media is covering because this is a very critical issue. All eight of you who are here today will be able to contribute to the decision process.

I will have an opening statement. The ranking member, Mrs. Maloney, will have an opening statement. Then we'll proceed with a series of questions for each of the three panels. We have 5 minutes of questions per Member. Then we'll proceed to the second panel and then the third panel. Hopefully we don't have many votes today. Someone is going to check on the vote schedule for today, but let's hope we don't have many interruptions.

Today we are here to receive testimony from some very distinguished statisticians. Contrary to popular belief, there are many in the statistical community who oppose sampling for use in the 2000 census, and fear a disaster should this risky experiment be allowed to continue.

Before we go any further, I want to draw everyone's attention to an Associated Press story from Vermont. The story reports that the unemployment rate has dropped to the lowest level in the State since 1988. But more relevant to this hearing, the story says that the reported unemployment rate is likely wrong because of statistical sampling. "But the department warned that the sharp decline

may also have been caused by a statistical sampling error by the U.S. Census Bureau. The numbers will be revised by the end of the year when more accurate data will be available."

A statistical sampling error by the U.S. Census Bureau. This is why we cannot, under any circumstances, have sampling in the 2000 census. Any error whatsoever in the Clinton sampling plan, which according to testimony has 3,600 parts to it, will be unrecoverable. There will be no time for adjustment before the report is due to the President. No time for adjustment; before States will need to redraw district lines.

I submit to the minority and those supportive of sampling, if the Census Bureau made a statistical sampling error in calculating the unemployment rate of one of the least populated States in the Nation, with a population of 589,000, how can we trust that they won't make a sampling error estimating 10 percent of the U.S. population, some 26 million people? It is time for the minority to end the charade. It is time for a full enumeration.

At our last hearing, we heard from the Under Secretary of Commerce. He confirmed one of the biggest fallacies put forth by my colleagues on the other side and the supporters of sampling. The National Academy of Sciences, or for that matter any other scientific group, has not endorsed the Clinton plan, to use sampling in the 2000 census. What the National Academy of Sciences and other groups have endorsed is the concept of sampling. They have not endorsed this specific plan. In fact, Under Secretary Shapiro said that not only have they not endorsed this plan, it would be impossible for them to do so. The plan has 3,600 parts, and has yet to be completed; 3,600 parts, that sounds like an awfully complicated plan to me, with thousands of chances for making a mistake.

The significance of this lack of an endorsement cannot be overemphasized. Endorsing a theory is a far cry from endorsing a specific plan. For example, I can tell you that it's possible for you to build a rocket to go to the moon. However, that doesn't mean that your rocket won't blow up on the launching pad. That is our point. This sampling scheme, the largest statistical experiment in history, with 3,600 parts, has not been empirically proven to work. We cannot afford for the 2000 census to blow up on the launching pad or in flight.

While it is true that in 1990 the census was 98.4 percent correct, an A+ in anybody's book, the potential for error with sampling could be much greater. With sampling, one small error in the 3,600 parts compounded through 26 million people would be devastating. When sampling was attempted in 1990 to correct for the perceived undercount, there were numerous errors that would, if not discovered, have wrongly sent a congressional seat from Pennsylvania to Arizona. It took the Census Bureau almost 2 years to correct for all the sampling errors. How can a plan that is much more complex find and correct any errors in a shorter period of time, 5 months, before the numbers are due to the President? The answer is, they can't. What's worse, because the administration is insisting on a one number census, there is no hard data to fall back on. The only word for this plan is irresponsible.

The American Statistical Association convened a blue ribbon panel to review the issue of sampling in the census. In the most recent issue of Amstat News, August/September 1998, the president of the American Statistical Association, David Moore, writes in his personal column, "The American Statistical Association takes no position on the details of any proposed use of sampling in the 2000 Census." Let me repeat. "The American Statistical Association takes no position on the details of any use of statistical sampling in the 2000 Census." Clearly, neither the Academy of Sciences or the American Statistical Association knows if a rocket will fly or blow up on the launching pad.

Now I have heard the argument made by the other side that somehow we are being unadventurous by not wanting to experiment with the 2000 census. That is ridiculous. We test drugs before we make them available to the public. We test new designs before we put them in airplanes. It's simply not prudent public policy to conduct the largest statistical experiment in history at the same time as the 2000 census. If a statistical sampling error was made in Vermont, one of our smallest States, imagine what the damage would be if a statistical sampling error was made in the 2000 census. I would hope that despite Mrs. Maloney's avid support for sampling, she is as troubled as I am by this AP story. Will the rocket fly or blow up on the launching pad?

At our last hearing, the ranking member of the subcommittee, Mrs. Maloney, took a good deal of time during her opening statement and in questioning to dismiss any ideas that had been put forth to use enhanced enumeration methods to improve the 2000 census. Despite the naysayers, the majority on this committee and in Congress will find ways to improve the census. We do believe that there are community outreach programs that will improve the accuracy of the 2000 census as well as practical considerations, like improving the address list, which according to the National Academy of Sciences accounted for 50 percent of the undercount. Mrs. Maloney also issued a challenge to the majority to present an alterative plan for the 2000 census. Mrs. Maloney, it seems, is missing a fundamental point. We have a plan. It's called full enumeration. It's legal and it's constitutional.

Your plan, known as sampling or as the President calls it, political polling, has been found illegal. I can assure everyone here that with or without the help of the minority, we are going to prepare for a full enumeration. Obviously this process would be easier with the help of the pro-sampling groups and the minority. But make no mistake about it, if the minority wastes 5 or 6 months ignoring the handwriting on the wall and not helping us plan for a full enumeration, there is a real danger that the 2000 census will be at risk.

Mrs. Maloney.

[The prepared statement of Hon. Dan Miller and the articles referred to follow:]

STATEMENT OF CHAIRMAN DAN MILLER (FL-13)
SUBCOMMITTEE ON THE CENSUS
STATISTICIANS OPPOSED TO SAMPLING HEARING
SEPTEMBER 17, 1998

Today we are here to receive testimony from some very distinguished statisticians.

Contrary to popular belief, there are many in the statistical community who oppose sampling for use in the 2000 Census, and fear a disaster should this risky experiment be allowed to continue.

But before we go any further, I want to draw everyone's attention to an Associated Press story from Vermont. The story reports that the unemployment rate has dropped to the lowest level in the state since 1988.

But more relevant to this hearing// the story says that the reported unemployment rate is likely wrong because of statistical sampling. "But the department warned that the sharp decline may also have been caused by a **statistical sampling error** by the U.S. Census Bureau!

The numbers will be revised by the end of the year when more accurate data will be available."

A "statistical sampling error by the U.S. Census Bureau." This is why we cannot under any circumstances// have sampling in the 2000 Census.

Any error whatsoever in the Clinton sampling plan which according to testimony has **3600** parts to it, will be unrecoverable. There will be no time for adjustment before the report is due to the President. No time for adjustment before states will need to redraw district lines.

I submit to the minority and the supporters of sampling: If the Census Bureau made a "statistical sampling error" in calculating the unemployment rate of one of the least populated states in the nation, with a population of 589,000, how can we trust that they won't make a sampling error estimating 10% of the U.S. population -- some 26 million people?

It's time for the minority to end this charade. It's time for a full enumeration!

At our last hearing we heard from the Under Secretary of Commerce, he confirmed one of the biggest fallacies put forth by my colleagues on the other side and the supporters of sampling.

The National Academy of Sciences, or for that matter any other scientific group, has **not** endorsed the Clinton Administration's plan to use sampling in the 2000 Census.

What the National Academy of Sciences and other groups have done, is endorse the **concept** of sampling. They have not endorsed this **specific** plan.

In fact, Under Secretary Shapiro said, that not only have they **not** endorsed this plan, it would be impossible for them to do so, because the plan, which has over **3600** parts to it, has yet to be completed. **3600** parts, that sounds like an awfully complicated plan to me with thousands of chances to make a mistake.

The significance of this lack of an endorsement by any group cannot be overemphasized.

Endorsing in **theory** is a far cry from endorsing a **specific** plan. For example, I can tell you that it's possible for you to build a rocket to go to the moon.

However, that doesn't mean that your rocket won't blow up on the launching pad.

And that's our point!

This sampling scheme, the largest statistical experiment in history, with over **3600** parts has not been empirically proven to work. We can't afford for the 2000 Census to blow up on the launching pad, or in flight!

While it is true that in 1990 the Census was 98.4 percent accurate, an A + in anyone's book, the potential for error with "sampling" could be much greater. With sampling, one small error in the 3600 parts compounded through 26 million people would be devastating.

When sampling was attempted in 1990 to correct for the perceived undercount, there were numerous errors that would, if not discovered, have wrongly sent a congressional seat from Pennsylvania to Arizona.

It took the Census Bureau almost two years to correct for all of the sampling errors.

How can a plan that is much more complex, find and correct any errors in a shorter period of time, five months, before the numbers are due to the President? The answer is, they can't!.

What's worse because the administration is insisting upon a one number Census, there is no hard data to fall back on.

The only word for this plan is **irresponsible.**

The American Statistical Association convened a Blue Ribbon panel to review the issue of sampling in the Census. In the most recent issue of Amstat News August/September 1998, the President of American Statistical Association David Moore writes in his personal column,

"The American Statistical Association also takes no position . . . on the **details** of any proposed use of statistical sampling in the 2000 Census."

Let me repeat that in case you missed it.

"The American Statistical Association takes no position on the **details** of any use of statistical sampling in the 2000 Census."

Clearly, neither the National Academy of Sciences or the American Statistical Association knows if the rocket will fly or blow up on the launching pad?

Now, I've heard the argument made by the other side that somehow we are being unadventurous by not wanting to experiment with the 2000 Census.

That's ridiculous.

We test drugs before we make them available to the public. We test new designs before we put them in airplanes.

It's simply not prudent public policy to conduct the largest statistical experiment in history at the same time as the 2000 Census. If a statistical sampling error was made in Vermont, one of our smallest states, imagine what the damage would be if statistical sampling error was made in the 2000 Census.

I would hope, that despite Ms. Maloney's avid support for sampling, she is as troubled as I am by this AP story.

Will the rocket fly or blow up on the launching pad?

At our last hearing, the ranking member of the subcommittee took a good deal of time during her opening statement and in questioning to dismiss any ideas that have been put forth to use enhanced enumeration methods to improve the 2000 Census.

Despite the naysayers, the majority on this subcommittee and in Congress will find ways to improve the Census. We do believe that there are community outreach programs that will improve the accuracy of the 2000 Census, as well as practical considerations like improving the address list which according to the National Academy of Sciences accounted for 50 percent of the undercount in 1990.

Ms Maloney also issued a challenge to the majority to present an alternative plan for the 2000 Census. Ms. Maloney, it seems, is missing a fundamental point.

We have a plan, it's called a full enumeration.

It's legal and constitutional.

Your plan known as "sampling," or as the President calls it, political polling// has been found illegal.

I can assure everyone here, that with or without the help of the minority, we are going to prepare for a full enumeration. Obviously, this process would be easier with the help of the prosampling groups and the minority.

But make no mistake about it, if the minority wastes five or six months with the minority ignoring the handwriting on the wall, and not helping us plan for a full enumeration there is a real danger that the 2000 Census will be at risk.

The Associated Press
September 15, 1998, Tuesday, AM cycle

SECTION: State and Regional

LENGTH: 221 words

HEADLINE: August unemployment down to lowest level since 1988

DATELINE: MONTPELIER, Vt.

BODY:
Vermont's unemployment dropped to its lowest level since 1988 in August with a rate of 2.9 percent.

The August rate was down six tenths of a percent from July, and 1.6 percent from the national rate.

Susan D. Auld, the commissioner of the employment and training department attributes the drop to a rise in jobs in the service industry.

"The service industry received a boost in August from the temporary summer youth employment program, which contributed to the improved job statistics,"
said Auld. "The strong demand for labor pushed the state's unemployment rate to its lowest point since 1988."

But the department warned that the sharp decline may also have been caused by a statistical sampling error by the U.S. Census Bureau. The numbers will be revised by the end of the year and more accurate data will be available. [emphasis added]

The greatest gains in employment occurred in government and services, which added 1,000 jobs each. Manufacturing also continued to improve, with the largest rise in wood and food products. Tourism employment rose in the retail and lodging industry, while health services employment also continued to improve.

Losses were felt in construction, which dropped 200 jobs, and retail employment which also fell by 200.

The annual job growth rate was up from past months to 1.8 percent.

PRESIDENT'S CORNER

Statistics, Policy, and ASA

THE THEME OF THE 1998 JOINT STATISTICAL Meetings was "Statistics: A Guide to Policy." This choice of theme was inspired by a comment made by Fred Mosteller when I interviewed him for the first issue of the *Journal of Statistics Education* in 1993. You can read the entire interview by visiting *JSE* at *www.stat.ncsu.edu/info/jse/homepage.html*. Here is the comment I have in mind:

> "Usually people think (and many statisticians tend to think) that once good data are available, then the answer to the policy question is at hand. But that usually is not true, because policy implies politics, and politics implies controversy, and the same data that some people use to support a policy are used by others to oppose it. So it's very difficult to handle policy questions, but nevertheless data does help the debate."

We have all been reminded by the long-running debate over plans for the 2000 Census that "policy implies politics, and politics implies controversy." ASA has some involvement in this debate, as well as in less-publicized issues such as proposed legislation on intellectual property that might hinder statistical analysis of databases. Let me reflect a bit on ASA's role.

Because of their technical expertise, individual statisticians are often deeply involved in policy issues. Think, for example, of the role of statisticians on panels that recommended policy on mammograms for younger women. Attempting to follow the data does not shield statisticians from considerable heat when the issues are heated, nor should it. Statistical expertise is not exercised in a clean room. Yet statisticians have an obligation to follow the data. We are not lawyers, whose duty is to the client rather than to the truth. ASA can support individual statisticians under pressure through a public statement of professional standards. Visit ASA's Web site to read and comment on the draft *Ethical Guidelines for Statistical Practice* prepared by our Committee on Professional Ethics.

A professional society should be more hesitant than individuals to enter policy debates. ASA filed an amicus brief in the census court cases, but the brief responds only to attacks on statistical sampling as a general method. From the brief:

> "ASA takes no position on the appropriate disposition of this case or on the legality or constitutionality of any aspect of the 2000 census. ASA also takes no position in this brief on the details of any proposed use of statistical sampling in the 2000 census.
>
> "ASA is, however, concerned to defend statistically designed sampling as a valid, important, and generally accepted scientific method for gaining accurate knowledge about widely dispersed human populations."

From one point of view (endorsed unanimously by the ASA Board) this is a minimalist defense of statistical science. From another, any intervention in our adversarial legal system, however strictly constrained, is de facto an intervention on one side or the other. How far should ASA go? Members have told me in strong terms both that we have gone too far and that we have not gone far enough.

The Census controversy helps frame the wider question. There is a hierarchy of issues concerning the census:

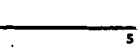
- The status of sampling and statistical estimation based on sampling as a scientific method in general;
- The non-partisan status of government statistical offices;
- The potential usefulness of sampling and statistical estimation in fixing the census undercount;
- The specific proposals for use of sampling and estimation under consideration by the Census Bureau;
- The Bureau's ability to carry out the details of the proposals in the time remaining before the 2000 Census.

ASA's amicus brief addresses only the first. Speaking individually in my February column, I responded to attacks on the first two, with passing comments on the third. ASA's Blue Ribbon Panel addressed the first and third while avoiding judgment on the fourth and fifth. This level of ASA involvement seems appropriate to me. The last two areas in the list above require detailed knowledge for informed judgment. They are in the realm of specific advice (about which statisticians may not be unanimous) rather than broad principle. The first three are sufficiently broad to be of concern to the profession as a whole and therefore to ASA.

DAVID MOORE

My own opinion is that ASA (and other groups concerned that our government produce data both honestly and competently) have given too little attention to the second issue: the non-partisan status of government statistical offices. William Safire (*New York Times*, December 7, 1997) implied that government statisticians are both politicized and lazy. He characterized Census Bureau plans as "having a statistician put a thumb on the scale." George Will (*Indianapolis Star*, June 15, 1998) took a more subtle approach: this administration, with its "remarkable record of lawlessness," will manipulate the sample selection. Government statisticians are not explicitly accused, but are presumably too weak to resist blatant interference. These writers echo Republican charges. Democrats reply that congressional Republicans want to specify what statistical procedures the Census Bureau may and may not use. The politicians agree that (other) politicians are trying to manipulate the 2000 Census for their own vile purposes.

The likely result is two intertwined disasters for the statistical profession. Constant accusations of political motivations in Census Bureau planning, combined with the impressive level of distrust of government that characterizes U.S. public opinion, will lead to distrust of government statistics in general. If the census is politicized, why not the unemployment rate or the CPI? A conspiracy lurks behind every weighted mean. The second disaster will feed the first: the 2000 Census may well fail to be credible. Charles Schultze (*Washington Post*, June 7, 1998) pointed out the logistical complexity of the decennial census with or without sampling. He called it "perhaps the biggest and most complex task the civilian arm of the federal government has to undertake." While the political parties maneuver for uncertain advantage, time is running out to do the planning, hiring, training, and rehearsal that depend on larger decisions not yet made. The parties may or may not gain a few House seats by not leaving statistical issues to statisticians. Both the statistical profession and the nation will lose.



Mrs. MALONEY. Thank you, Mr. Chairman. First of all, I would like to sincerely thank you for agreeing to invite three witnesses at the request of the minority. Among them are two noted statisticians and a former Director of the Census Bureau, Dr. Barbara Bryant. I encourage our audience to stay until the last panel to hear Dr. Bryant. She is the only witness here today who clearly has hands on experience in conducting a census.

Mr. Chairman, last week I expressed my disappointment that you have not scheduled any subcommittee hearings to review the alternative methods for counting that you and other opponents of scientific sampling methods have proposed. I suggest that if you are serious about using administrative records to identify uncounted people; or hiring mail carriers to knock on doors of unresponsive households; or planning more promotion and outreach to reduce the differential undercount, that you would exercise your authority, as chairman, to hold hearings on these alternatives. I certainly would welcome the opportunity to explore these alternatives in depth since they may well be very useful, if not in 2000, then perhaps in 2010.

That is what this oversight panel should do, bring in experts on all sides of the issues, explore prior evaluations, and determine whether the methods stand a reasonable chance of reducing the persistent disproportionate undercount of minorities and the poor that traditional counting methods can't seem to fix. Instead, I suspect you will choose once again to throw darts at the scientific methods, almost all knowledgeable experts agree hold the best promise of eliminating that undercount. This hearing is déjà vu all over again. We are accomplishing nothing new or useful today that will help the Census Bureau do the best job possible, in the year 2000.

At a similar hearing this past spring, you invited three scientists who had no prior experience with census operations and planning to testify against the Bureau's 2000 census plan. You did not invite anyone who had knowledge of the process and the Bureau's specific plan. With all due respect, Mr. Chairman, the hearings you have held so far this year have not been designed to elicit useful discussion of the complex issues involved in preparing for and taking a census. They have merely been exercises in cross examination designed to poke holes wherever possible in the Bureau's census plan, avoiding the difficult task of thoroughly reviewing census operations in an effort to improve the process.

My Republican colleagues on the subcommittee have done little more than ask leading questions in order to get the answers they want to hear, instead of soliciting thoughtful thorough answers that might help us probe genuine difficulties in the census process and seek bipartisan solutions.

In short, this subcommittee has squandered precious time and resources that could have been spent on constructive oversight. Instead, the majority has been on a crusade to destroy credible scientific work. Tragically the consequence, intended or not, may be a failed census of any design. Equally troubling are the irresponsible attacks on well accepted scientific methods and principles that you, Mr. Chairman, with a doctorate in statistics, should know are unwarranted and untrue. What a sad legacy for this panel to leave.

It is not a legacy that any one of us should be proud of. I certainly am not because I know the American people expect and deserve more from their elected officials.

Mr. Chairman, last week you, in your own words, "extended an olive branch" to outside groups supporting a fair and accurate census, to those who support the use of modern statistical methods in achieving that result. Well, I encourage you, Mr. Chairman, to reach out to La Raza, to the Leadership Council on Civil Rights, to the United States Conference of Mayors, to the National Governors Association, to the National League of Cities, and to the NAACP. Set up meetings with these organizations. Have your staff meet with them. I suspect you haven't contacted them yet, but you should. If you do, you will hear the true grassroots support to which you have so far been deaf, the voice of those who deserve to be counted, who want to be counted, but who were not counted in 1990.

Mr. Chairman, we know that the 1990 census had an undercount of 8.4 million people, and that if we conducted the 2000 census using 1990 methods, we will miss at least that many people again. We have yet to hear credible Republican solutions which would solve that problem. It is getting late. We better start hearing them soon or we won't be able to use them anyway.

In conclusion, I would like to submit to the record a collection of editorials from newspapers across the country, supporting the use of statistical methods to achieve a fair and accurate census. Included in these editorials is the New York Times, Impeding an Accurate Census. They talk about efforts to block scientific methods. The Boston Globe, the Atlantic Constitution, and San Francisco Chronicle, the Los Angeles Times, and I have a series of them from across the country. I would like to put them into the record accompanying my opening comments. Thank you, Mr. Chairman. I yield back the balance of my time.

Mr. MILLER. Thank you, Mrs. Maloney.

Mrs. MALONEY. May I put these editorials in the record?

Mr. MILLER. Yes. Without objection, they will be put in the record.

[The prepared statement of Hon. Carolyn B. Maloney follows:]

Opening Statement -- Rep Carolyn B Maloney
September 17, 1998

First, Mr Chairman, I want to thank you for agreeing to invite three witnesses at the request of the minority Among them are two noted statisticians and former Director of the Census Bureau, Dr Barbara Bryant I encourage our audience to stay until the last panel to hear Dr Bryant She is the only witness here today who clearly has hands on experience in conducting a census

Mr Chairman, last week I expressed my disappointment that you have not scheduled any subcommittee hearings to review the alternative methods for counting that you and other opponents of scientific sampling methods have proposed.

I suggested that if you are serious about using administrative records to identify uncounted people, or hiring mail carriers to knock on doors of unresponsive households, or planning more promotion and outreach to reduce the differential undercount, that you would exercise your authority as chairman to hold hearings on these issues.

I certainly would welcome the opportunity to explore these alternatives in depth, since they may well be useful -- if not in 2000 then perhaps in 2010.

That is what this oversight panel should do: bring in experts on all sides of the issues, explore prior evaluations, and determine whether the methods stand a reasonable chance of reducing the persistent, disproportionate undercount of minorities and the poor that traditional counting methods can't seem to fix

Instead, I suspect you will choose once again to throw darts at the scientific methods almost all knowledgable experts agree hold the best promise for eliminating that undercount. This hearing is deja vu all over again. We are accomplishing nothing new or useful today that will help the Census Bureau do the best job possible in 2000.

At a similar hearing this past spring, you invited three scientists who had no prior experience with census operations and planning to testify against the Bureau's 2000 census plan. You did not invite anyone who had a knowledge of the process and the Bureau's specific plans.

With all due respect, Mr Chairman, the hearings you have held so far this year have not been designed to elicit useful discussion of the complex issues involved in preparing for and taking the census

They have merely been exercises in cross-examination, designed to poke holes wherever possible in the Bureau's census plan -- avoiding the difficult task of thoroughly reviewing census operations in an effort to improve the process.

My Republican colleagues on this subcommittee have done little more than ask leading questions in order to get the answers they want to hear, instead of soliciting thoughtful, thorough answers that might help us probe genuine difficulties in the census process and seek bipartisan solutions

In short, this subcommittee has squandered precious time and resources that could have been spent on constructive oversight. Instead, the Majority has been on a crusade to destroy credible scientific work. Tragically, the consequence, intended or not, may be a failed census of any design.

Equally troubling are the irresponsible attacks on well-accepted scientific methods and principles that you, Mr. Chairman, with a doctorate in statistics, should know are unwarranted and untrue.

What a sad legacy for this panel to leave. It is not a legacy that any one of us should be proud of. I certainly am not, because I know the American people expect and deserve more from their elected representatives.

Mr. Chairman, last week you "extended an olive branch" to outside groups supporting a fair and accurate census -- to those who support the use of modern statistical methods in achieving that result.

Well I encourage you Mr. Chairman, to reach out to La Raza, to the Leadership Council on Civil Rights, to the US Conference of Mayors, to the National Governors Association, to the National League of Cities and to the NAACP.

Set up meetings with these organizations -- have your staff meet with them. I suspect you haven't contacted them yet, but you should. If you do you will hear the true grassroots support to which you have so far been deaf. The voice of those who deserve to be counted, who want to be counted, but who were <u>not</u> counted in 1990.

Mr. Chairman, we know that the 1990 census had an undercount of 8.4 million, and that if we conduct the 2000 census using 1990 methods, we will miss at least that many people. We have yet to hear credible, Republican solutions which would solve that problem. It's getting late, we better start hearing them soon, or we won't be able to use them anyway.

In conclusion, I'd like to submit for the record this collection of editorials from newspapers around the country, supporting the use of statistical methods to achieve a fair, and accurate census 2000.

Thank you Mr. Chairman.

Houston Chronicle

RICHARD J. V. JOHNSON, Chairman and Publisher

G. E. McDAVID, President
JOHN W. SWEENEY, Vice President, General Manager
JACK H. STANLEY, Vice President, Operations
DALE E. NISSEN, Vice President, Manufacturing
ROBERT E. CARLQUIST, Vice President, Administration
GARY W. RANDAZZO, Vice President, Sales
JOYCELYN MAREK, Vice President, Marketing & Electronic Products
DWIGHT M. BROWN, Vice President, Advertising
HOLLIS G. PRICE, Vice President, Circulation
ANN TURNBACH, Vice President, Human Resources
RICHARD A. WINEGARDEN, Vice President, Chief Financial Officer

JACK LOFTIS, Executive Vice President, Editor
TONY PEDERSON, Vice President, Managing Editor

FRANK MICHEL, Associate Editor
TOMMY MILLER, Deputy Managing Editor

SUSAN BISCHOFF,
DAN CUNNINGHAM,
FERNANDO DOVALINA,
WALTER JOHNS,
Assistant Managing Editors

Houston Chronicle is a division of The Hearst Corporation

NUMBERS GAME

Politicizing census count harmful to Houston

The shameless game of political football over the upcoming U.S. census continues, and places like Houston risk becoming big losers.

The Republican-controlled U.S. House has passed a measure to withhold some funding next year to the Census Bureau in an effort to force the bureau to abandon plans to use a process called "statistical sampling," which could give a more accurate count of the nation's population in the year 2000.

Because of a serious undercount here in the 1990 census, Houston has lost many millions of dollars in federal funding that is tied to population. If a similar undercount occurs again in 2000, Houston and Texas stand to lose billions over the next decade.

Many Republicans are concerned that a more accurate count will skew redrawn voting districts to favor Democrats. That's a complicated political concern for which a fair solution could be worked out. But the census process should not be held hostage over it.

If allowed to go forward, the Census Bureau would, as constitutionally mandated, enumerate the population with the time-tested methods of questionnaires and personal interviews by census-takers. Only then, would the inevitable undercount be corrected with more highly accurate statistical sampling.

Statistical sampling would not, as some have mislead the public to believe, be used *instead* of proper enumeration of the citizenry. It would use recounts of some neighborhoods to make the proper enumeration more accurate and thus fairer to fast-growing and sprawling regions like Houston.

Would Republicans put their own party interests ahead of the citizens' right to a fair and accurate return on the tax dollars we all send to Washington? It appears that they would.

That may be good politics. But count it as bad – very bad – public policy.

THE BUFFALO NEWS

82 PAGES

HEAD-COUNT GAMES

Does census silliness mean no more polls?

So, maybe we'll get half a census in 2000.

Maybe we'll count just the people east of the Mason-Dixon line, since the House only approved half the money to prepare for the decennial head count.

That would be no sillier than what the Republican majority proposes for political reasons: a census that covers the whole country, but that is guaranteed to be far more inaccurate than necessary because it shuns the use of the best available methods.

The $34 million House bill passed the other day to fund the departments of Commerce, Justice and State included the unusual proviso that Commerce — which conducts the population survey — could use only half the $952 million designated for census preparations next year. The rest would be held up pending approval by Congress.

That huge string was attached because Commerce — backed by the National Academy of Sciences, among others — wants to use statistical sampling in hard-to-count areas to make the enumeration more accurate.

But Republicans object, which leaves them in the position of arguing for a shoddy census.

It's conceded that the last census undercounted the U.S. population by a net 4 million people, with most of the undercount centered in minority neighborhoods. That undercount is critical — particularly for states like New York — because the census figures are used to determine the number of congressional seats a state gets, its legislative districts and various categories of federal aid.

But Republicans — no doubt fearful that counting more minorities will help Democrats — complain with a straight face that sampling is too unreliable.

Of course, they'd have a lot more credibility making that argument if they'd pledge to run their re-election campaigns without polls — which are nothing more than rough forms of the precise statistical samples census officials would use.

The huge commerce-justice-state budget bill is one of 13 that must be passed to prevent another government shutdown. Nevertheless, President Clinton has threatened to veto it if the final version contains the silly prohibition against greater census accuracy. He certainly should.

The New York Times

Founded in 1851

ADOLPH S. OCHS, Publisher 1896-1935
ARTHUR HAYS SULZBERGER, Publisher 1935-1961
ORVILLE E. DRYFOOS, Publisher 1961-1963
ARTHUR OCHS SULZBERGER, Publisher 1963-1992

ARTHUR OCHS SULZBERGER JR., Publisher

SOLOMON B. WATSON IV, Executive Editor
JOSEPH LELYVELD, Executive Editor
JOHN M. GEDDES, Deputy Managing Editor

Assistant Managing Editors
SOMAN GOLDEN BEHR, JACK ROSENTHAL
ALLAN M. SIEGAL

GERALD M. BOYD, Senior Page Editor
PHILIP M. BOFFEY, Deputy Editorial Page Editor

JANET L. ROBINSON, President, General Manager
DANIEL H. COHEN, Senior VP, Advertising
RICHARD H. GILMAN, Senior VP, Operations
SCOTT HEEKIN-CANEDY, VP, Planning
MARC Z. KRAMER, VP, Labor Relations
DENNIS L. STERN, VP, Human Resources
JAMES L. TERRILL, VP, Chief Financial Officer
DAVID A. THURM, VP, Production

PENELOPE MUSE ABERNATHY, President, News Services
MARTIN A. NISENHOLTZ, President, Electronic Media

Impeding an Accurate Census

The ruling by a special three-judge panel in Washington that statistical sampling cannot be used in the 2000 census will be a setback for the democratic process if allowed to stand. It would prohibit the use of sampling to supplement traditional methods that have failed to produce an accurate count for purposes of apportioning representatives in Congress.

The decision will almost certainly be appealed to the Supreme Court, as it should be. A panel of Federal judges in the Fourth Circuit is hearing another lawsuit on the same sampling question, and their judgment is also likely to reach the Court.

The ruling found that sampling is not allowed under the Federal Census Act, a judgment that turns on a reading of some ambiguous wording in the act. The court avoided the larger question of whether sampling is unconstitutional, as Republicans in Congress have long argued. The Republicans fear that a more precise count will turn up more immigrants, minorities, the poor and the young, who are likely to swell population numbers in urban centers that lean toward the Democrats. But because an overt campaign against a fair count would be unseemly and politically risky, the Republicans have seized upon the abstract argument that sampling violates the constitutional requirement of "actual enumeration." But it is bizarre to say that the Constitution's drafters were wedded to a single, head-by-head methodology that, as the country grew, became inadequate to its purposes. The whole point is to produce an accurate count.

A strict head count depends entirely on locating people at specific addresses. With more people having no permanent address, living with relatives or becoming homeless, accuracy declines. The 1990 census, the most costly in history, left more than eight million people uncounted, and more than four million were counted twice or in the wrong place. Blacks and Hispanics were also more likely to be uncounted than whites.

These problems are likely to become more pronounced in the upcoming census. Professional statisticians, including experts convened by the National Academy of Sciences, have said that this gap can be scientifically corrected by adding sampling techniques. The Census Bureau plans to estimate the undercounted and make adjustments based on a highly detailed survey of 750,000 households around the nation.

Even though the ruling did not touch the Republicans' favorite constitutional argument, House Speaker Newt Gingrich has hailed it as a victory. But a ruling where accuracy loses can hardly be called a triumph for anyone but entrenched politicians.

The Boston Globe

WILLIAM O. TAYLOR, *Chairman of the Board* — BENJAMIN B. TAYLOR, *Publisher*
MATTHEW V. STORIN, *Editor* — HUGH GREENWAY, *Editor, Editorial Page*

WILLIAM S. HUFF, *President* — HELEN W. DONOVAN, *Executive Editor*
STEPHEN E. TAYLOR, *Executive Vice President* — GREGORY L. MOORE, *Managing Editor*

Founded 1872

CHARLES H. TAYLOR, *Publisher 1873-1921* WILLIAM O. TAYLOR, *Publisher 1921-1951* WM. DAVIS TAYLOR, *Publisher 1953-1977*
JOHN I. TAYLOR, *President 1963-1977* LAURENCE L. WINSHIP, *Editor 1955-1965* THOMAS WINSHIP, *Editor 1965-1984*

Senseless vote on the census

When something is broken, sensible people try to fix it. The US census was badly flawed in 1990, but Congress wants to pretend there is no problem. Without action, the 2000 count will be worse.

In 1990 the census undercounted America by 4 million people – 1.6 percent of the population. That's a net figure. In all, 8.5 million people – mostly poor, minorities, and young children – were overlooked, while 4.5 million – many of them students and people who moved – were counted twice. Census officials conceded it was the first time the count had been less accurate than the previous census.

To remedy the situation, the Census Bureau wants to use a system under which more than 90 percent of the population would be counted individually, and statistical sampling would be used to count those hardest to find. The sampling would not be guesswork. Intensive personal interviewing in a target area would produce information that would be used to estimate the undercount in similar areas.

Congress is insisting that the Census Bureau conduct an actual enumeration and count only real people individually. In an ideal world this method would be best, but it would be far more expensive than reason or Congress could allow. As it is, the primary mechanism is a mailed survey, but only two-thirds of the forms are returned, and that number has been declining, further blurring the census numbers.

In Congress, substantive arguments have been overtaken by raw partisanship. In a vote this week to block the sampling, 222 Republicans and 5 Democrats outvoted 198 Democrats and 2 Republicans. The Republicans feel that most of the uncounted would lean Democratic and would produce more Democratic districts, so they should remain uncounted.

The problem is not that the census, using statistical sampling, would create millions of fictional residents – "virtual people," as Speaker Newt Gingrich called them. The problem is that millions of real, flesh-and-blood people will be left out – turned into fictional characters – if sampling isn't allowed. It is a problem that needs to be fixed.

The Atlanta Constitution

RON MARTIN, *Editor*

CYNTHIA TUCKER
Editorial Page Editor

JOHN WALTER
Managing Editor

Editorials

In census, accuracy counts

In a review of the 1990 national census, statisticians at the U.S. Census Bureau discovered that they had overlooked 4.4 percent of the nation's African-American population, 5 percent of Hispanics and 12.2 percent of American Indians living on reservations.

Appalled by those numbers, bureau officials began to look for ways to improve the count for the 2000 census. They discovered that even if they spent a lot more money to hire a lot more workers to canvass neighborhoods much more intensely, the effort would produce only a slight improvement in accuracy. If they wanted to make the census truly accurate, as the U.S. Constitution requires, they would have to augment the traditional head-count method used as far back as King Herod's biblical times.

Republicans suggest that statistical sampling is part of a conspiracy by the Clinton administration to distort the census so that areas with a higher Democratic turnout get more seats in Congress.

■

The method they chose, statistical sampling, is something akin to political polling, only much more scientific and sophisticated. It would be used in addition to, not as a replacement for, the usual head-count method. The Census Bureau field-tested its new approach in 1995, found it accurate, and then asked the National Academy of Sciences to review its work.

In two separate peer-reviewed reports, an NAS panel enthusiastically endorsed statistical sampling. "A census of acceptable accuracy and cost is not possible without the use of sampling procedures," the outside scientists found.

That's a pretty straightforward statement. Unfortunately, the prospect of a truly accurate census is threatening to some people. House Speaker Newt Gingrich (R-Ga.), a man who usually celebrates the progress that science and technology can bring, is fighting to ensure that only the archaic method of counting heads is used by the Census Bureau.

That's particularly ironic because in other contexts, Gingrich is an avid enthusiast of statistical sampling. For example, he poll-tested the wording of every item in his famous Contract With America until it got an approval rating of at least 60 percent. Yet in this case, he and other Republicans profess to distrust the technique. They suggest that statistical sampling is part of a conspiracy by the Clinton administration to distort the census so that areas with a higher Democratic turnout get more seats in Congress.

There are several problems with the accuracy of that scenario.

First, the folks at the Census Bureau are highly professional career people with a deep respect for the importance of their work. They would blow the whistle immediately on any attempt by politicians in either party to jiggle the numbers.

Second, if an administration wanted to cook the numbers, and if it had the support of Census Bureau professionals in doing so, it could do so just as easily under the head-count method as with statistical sampling.

And third, the proposal to use statistical sampling came from within the bureau, not from Democratic politicians. And the necessity and accuracy of statistical sampling have been confirmed by outside experts and scientists of unquestioned credentials.

The Republicans have from time to time made halting, awkward attempts to reach out to Hispanic, black and other minority and immigrant voters, without much success. Their stand on statistical sampling indicates that they do not really expect to improve on that performance.

Rather than compete for support from those groups, the GOP intention is to pretend that they don't exist and thus deny them the representation in Congress that is their right.

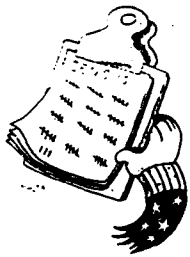
A22

San Francisco Chronicle

THE VOICE OF THE WEST

Making the Census Count

COUNTING NOSES doesn't sound like a political act, but the next census is turning into just that. Republican forces in Congress are only willing to give the Census Bureau half a year's budget because the proposed head-counting effort will use statistical projections, not person-by-person totals, to round out the final numbers in the national tally in 2000. GOP objections threaten to undercut a reasonable effort by census takers to produce an accurate number.



The sampling method is controversial because it goes beyond counting people who answer surveys and doorbell interviews to include a best-guess estimate of others who cannot be located. The 1990 census missed an estimated 8.4 million and double-counted 4.2 million others.

There is a reasoned debate among statisticians about the best methods to use, but a majority view argues that census takers going door-to-door will never produce the most accurate population number. Congress has produced its own overlay of argument. Republicans fear statistical sampling will be widely used in minority-heavy urban areas, typically Democratic strongholds. Democrats, in turn, push the sampling methods as the soundest way to reach a final number that favors them when congressional boundaries are redrawn following the decadal census.

Yes, politics intrudes. But the Republican alternative is a strict interpretation of "actual enumeration" as stated in the Constitution. Relying on mailed-out surveys and household visits isn't practical in a country with an estimated population of 270 million.

It will take a different counting method to produce an approximate number. Current plans call for counting 90 percent of the households in a census tract and then use calculations to nail down the number of people living in the final 10 percent of the dwellings. This method saves time and money. It should also produce an accurate number, one that fulfills the duty of the Census Bureau and stands the test of good reason.

LOS ANGELES TIMES EDITORIALS

MARK H. WILLES, *Publisher*
KATHRYN M. DOWNING, *President and Chief Executive Officer*
MICHAEL PARKS, *Editor and Executive Vice President*
JANET CLAYTON, *Editor of the Editorial Pages and Vice President*

Showdown in Census Feud

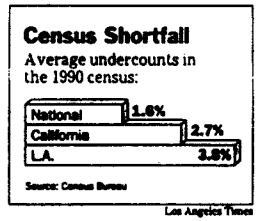
House Republican leaders left Washington last week after inserting wording in an appropriations bill that essentially prohibits the Census Bureau from using "statistical sampling" to improve the accuracy of its traditional head count in the 2000 census.

A showdown seems inevitable when Congress returns next month, for President Clinton has threatened to veto the bill unless the language is removed. The bill funds not only the census but the departments of Commerce, State and Justice as well.

The census has been a popular target since 1792, when George Washington issued the first presidential veto in disagreement with Congress' interpretation of the 1790 census figures. The numbers are critical, for they determine how congressional districts are drawn and federal dollars are distributed.

Demographers and statisticians agree that sampling is the most cost-effective way of preventing a recurrence of the embarrassing errors that afflicted the 1990 census, which missed more than 4 million Americans and counted another 4 million twice. Those errors have hurt California disproportionately, for its undercount of 2.7% was far higher than the average national undercount of 1.6%. Los Angeles had the second-highest undercount of all major U.S. cities, at 3.83% or 138,808 people.

Politics enters the picture because the kind of people whom sampling should catch—those owning no homes—lean toward Democratic candidates, while the folks overrepresented in a traditional head count, those owning multiple homes, tend to favor Republicans.



House Speaker Newt Gingrich is leading the Republican opposition against sampling, calling the census "an issue of great importance to our party." But while the use of sampling in the year 2000 could cost the GOP some votes, Gingrich's opposition could cost it more by alienating the two constituencies traditionally slighted when sampling is not used, Latinos and African Americans.

In 1991, Gingrich defended one of those very constituencies in a census undercount, writing a letter urging the Bush administration to use sampling to correct for a low enumeration of African Americans in Georgia in 1990. "If the undercount is not corrected, it would have a serious negative impact on Georgia," Gingrich wrote, for "minority voting strength would be greatly diluted."

Gingrich should reread that letter and support the Census Bureau's current attempt to do what he wished it had done in 1990.

A18 WEDNESDAY, AUGUST 26, 1998

The Washington Post

AN INDEPENDENT NEWSPAPER

The Census Decision

THOUGH FRAMED in constitutional and scientific terms, the objections to the use of statistical sampling techniques in the census always have been chiefly political. That sampling tools can produce a more accurate census is a matter on which there is wide scientific agreement. The claims that such techniques would violate the Constitution's requirement of an "actual enumeration" of the population are, likewise, a thin cover for the fact that past censuses have undercounted urban dwellers and minorities. Since the census is used in part to assign congressional districts, the undercount is believed by many to help Republicans, and whether that situation will be preserved is what this fight is really about.

It is worth noting, therefore, that the decision Monday by a federal three-judge panel to block the use of sampling was based on neither constitutional nor scientific concerns. It was based, rather, on the language of the law. And in this area, the census issue is actually a tough one. The law says that "except for the determination of population for purposes of apportionment of Representatives in Congress among the several States, the Secretary [of Commerce] shall, if he considers it feasible, authorize the use of the statistical method known as 'sampling' in carrying out" the census. The sticky question is whether the word "except" in this sentence means sampling cannot be used in apportionment enumeration or whether it simply means it does not have to be.

The Justice Department argued that the provision makes sampling mandatory in all situations except apportionment, and makes it discretionary in that context. The House of Representatives, which brought the suit as an institution, argued otherwise. And Judge Royce Lamberth—writing for a unanimous panel—rejected the department's view, analogizing the law's construction to the sentence "Except for Mary, all children at the party shall be served cake." Wrote Lamberth, "One would expect that the person who issued [this] directive . . . would be quite surprised to learn that Mary had been served cake." The criticism of the opinion from defenders of sampling was loud and immediate, but Judge Lamberth's is hardly a crazy reading of the law.

Nor is the Justice Department's argument obviously wrong. The department cites several examples of laws constructed similarly to this one that do imply discretion, and cites also another provision of law in which the authority to use sampling seems clearer. Judge Lamberth notes that other courts have gone the other way. Because the administration's policy objective—an accurate census—is an eminently reasonable one, this decision should certainly be appealed. Judge Lamberth's opinion raises the possibility that the law requires a census less accurate than the best one science could devise. If he is right about this and Congress exults in that fact rather than changing the law, an apolitical census will have been dealt a significant blow.

THE MONITOR'S VIEW

"First the blade, then the ear, then the full grain in the ear."

Counting Controversy

COUNTING people can be a pretty simple process, but not when the numbers reach into hundreds of millions. At that point, you rely on the best possible estimate, using the best statistical tools.

That's what the US Census Bureau has been gearing up to do. The bureau's experts planned to use sampling methods to correct the serious undercount in the 1990 census, when upwards of 8 million people were missed and more than 4 million were counted twice. But their plans ran smack into partisan politics.

Republicans in Congress attacked sampling, asserting its use was unconstitutional and worrying, doubtless, that it would increase the count of people likely to vote Democratic – poor, often minority, residents – and accordingly alter congressional districts. Democrats rallied behind sampling for obvious reasons.

As part of a compromise worked out in Congress, GOP lawmakers brought suit against the sampling plan, and this week they won. A panel of federal court judges found unanimously that the Census Act of 1976 does indeed forbid the use of sampling for purposes of congressional reapportionment. But the critical question (sidestepped by the judges) is whether the Constitution itself forbids sampling when it calls for an "actual enumeration." That question now moves on to the Supreme Court.

What tends to get obscured by this partisan battle is the Census Bureau's desire to do the best job possible. A head count of every person in the US is a practical impossibility. Sampling would be used in addition to an actual count – by mailed forms and follow-up visits – of 90 percent of households. It offers a statistically sound way of getting closer to the truth, given the bureau's limitations of staff and money. It would better approximate an "actual enumeration," and perhaps the high court will see it that way.

But that ruling won't come until next spring. Meanwhile, the Census Bureau has to get moving. It may have to do the best it can with old (more labor intensive and costly) methods. Republicans have been holding its budget hostage to the sampling controversy. That has to stop.

Chicago Tribune

Section 1 Wednesday, August 26, 1998

A setback for census sampling

Caught in the middle of a political dogfight over whether statistical sampling should be used in the year 2000 census, a three-judge federal panel Monday ruled in favor of the Republicans—but required some fancy legal footwork to do so.

Left unresolved by the court was the constitutionality of statistical sampling, in which scientific methods are used to account for populations that are underrepresented in census counts. The Clinton administration ought to take the matter up to the U.S. Supreme Court, which could well reverse the panel's decision.

At issue is whether the Census Bureau should use statistical sampling to get a more accurate count of the population. According to the bureau's own calculations, the 1990 census missed about 8.4 million people, while 4.4 million other folks were counted *twice*.

Most of the errors occurred in large cities and involved African-Americans and Hispanics, two groups notoriously reticent about filling out census forms and difficult to track down in all the urban nooks and crannies.

The bureau wants to do an actual count of 90 percent of the population and then do sampling-based projections on the rest of the people. There is ample evidence that sampling would achieve a more accurate final count.

That's where partisan politics barge in: Most of the undercounted are likely to lean toward the Democratic Party, and the GOP—nervous about its majorities in Congress—adamantly opposes anything that would give the Democrats an edge, either in reapportionment battles or distribution of federal funds.

House Speaker Newt Gingrich filed suit in February to stop the Clinton administration's plan to use sampling on grounds that it would violate the constitutional mandate for an "actual enumeration" of the population, a literal head count. That is a thin thread because the Constitution also provides for the census to be conducted "in such manner as [Congress] shall by law direct."

The federal panel ducked the constitutional issue, basing its decision on a reading of federal laws that seem to offer contradictory directives on sampling. The panel's reasoning isn't outlandish, but it's also quite possible the high court would reach a different interpretation of the federal laws.

The bottom line is that sampling would produce the fairest, most accurate count of the population. The Supreme Court should take the opportunity to examine if that, indeed, is what federal law will allow.

Newsday
EDITORIALS
"Where there is no vision, the people perish."

Accuracy, Not Politics, Must Drive Census

House Republican leaders should stop playing political games with the 2000 census.

Months ago, they agreed that census officials would be allowed to prepare to use the kind of statistical adjustments that professionals say would make the count more accurate: the 1990 census severely undercounted minorities, immigrants and the poor. In exchange, House Speaker Newt Gingrich (R-Ga.) got public money for a lawsuit challenging the constitutionality of that kind of statistical sampling.

Now Republicans want to take the Commerce Department appropriations bill and fund the Census Bureau for just six months. That would make it easy for them to derail preparation for the 2000 census unless they get the counting method they want. They are willing to sacrifice accuracy to prevent the counting of people likely to be Democrats. That's outrageous, and it will hurt New York, with its large minority population.

The Constitution calls for an "actual enumeration" of the nation's people. The courts will decide if sampling fits within that definition. What's important is accuracy, not adhering to historical methods that don't work well with today's huge, mobile population.

Three separate panels of the National Academy of Sciences have endorsed the use of modern sampling methods. About 90 percent of the nation's people would be counted in the traditional direct fashion. Sampling would be used only to fine-tune those numbers. The House should fund the Census Bureau for the full year and let it get on with its work.

et Top Court Decide How Census Counts Americans

ederal appeals court's ruling that scien- sampling methods cannot be used in cting the 2000 census left a key ques- nanswered: Is sampling constitutional? ead of grappling with that core ques- he three-judge panel parsed conflicting ns of the Census Act, one that says ling can be used at the discretion of the nerce secretary and one that allows ling except for the purpose of appor- nent. The judges gave more weight to rohibition regarding apportionment be- , they said, it is more specific.

The administration is wisely appealing this ruling. The issue is important to states, like New York, where undercounting can deprive citizens of federal aid and representation in Congress. What's needed is a definitive decision from the Supreme Court.

The Constitution calls for an "actual enumeration" of the nation's people. The top court needs to decide if that language prohibits the use of modern statistical methods to fine-tune a flawed national headcount. The answer *should* be no. What matters is accuracy, not adherence to outdated methods.

But this is, at bottom, a political fight. Democrats want the census to use sampling to provide a more accurate count of minorities, immigrants and renters who have historically been undercounted; Republicans want to rely solely on an actual head count. Both assume the undercounted are likely to be in poor areas represented by Democrats. At stake are billions of federal aid dollars and possibly the balance of power in the Republican-run House. Partisan concerns are serious stuff in Washington, but outside the Beltway it's accuracy that should count.

LOS ANGELES TIMES EDITORIALS

Make the Census Count

Every decade since 1790, the U.S. government has taken a population count. The process of compiling the census has evolved with a changing nation, from gathering information by horseback to door-to-door questioners to mail-in household surveys. The plan has been to include statistical sampling in the 2000 census in order to better count minorities, immigrants, renters and others who for a variety of reasons are often not picked up by traditional census methods.

Each nose matters, and California benefits from each resident's presence on the rolls. That's reason for displeasure with Monday's ruling by a three-judge federal panel that rejected the use of sampling, agreeing with House Republicans who sued to block use of the valuable statistical tool. Census data is used to draw or realign congressional districts and shape allocations of billions of federal dollars to states. If the decennial head count fails to find elusive citizens, they are inadequately represented politically and in terms of government programs. The solution lies in sampling, say Democrats, who traditionally attract the votes of many of those who tend to be uncounted.

This issue of sampling and representation is front and center in the GOP lawsuit and demands a prompt Supreme Court review. The GOP, led by House Speaker Newt Gingrich, specifically challenged the Clinton administration's contention that the laws governing the census allow sampling for political apportionment. Meanwhile, the Census Bureau is appropriately proceeding with a two-track preparation, using sampling in one version and excluding it from the other.

Getting the numbers right by using sampling is critical. In the 1970 census, a small undercount was identified through sampling and the census was adjusted accordingly without dispute. There was no sampling in the 1980 census, but sampling conducted after the 1990 census identified nearly 9 million people as missed in the traditional head count—including 1 million in California—and 4.4 million who were counted twice.

The 1990 undercount cost California an estimated $500 million in federal funds and shorted it one House seat. An undercount in the 2000 census could cost California $1 billion in federal funds and two additional House seats.

From day one the dispute over sampling has been about power, not accuracy. A neutral judgment from the Supreme Court is warranted, promptly.

The New York Times

Founded in 1851

ADOLPH S. OCHS, Publisher 1896-1935
ARTHUR HAYS SULZBERGER, Publisher 1935-1961
ORVIL E. DRYFOOS, Publisher 1961-1963
ARTHUR OCHS SULZBERGER, Publisher 1963-1992

[illegible] Publisher

JOSEPH LELYVELD, Executive Editor
BILL KELLER, Managing Editor
JERALD M. ROSE, Deputy Managing Editor
JOHN M. GEDDES, Deputy Managing Editor

Associate Managing Editors
SOMA GOLDEN BEHR, JACK ROSENTHAL
CAROLYN LEE, ALLAN M. SIEGAL

[illegible], Associate Page Editor
PHILIP M. BOFFEY, Deputy Editorial Page Editor

JANET L. ROBINSON, President, General Manager
DANIEL H. COHEN, Senior VP, Advertising
RICHARD H. GILMAN, Senior VP, Operations
SCOTT H. HEEKIN-CANEDY, VP, Planning
MARC Z. KRAMER, VP, Labor Relations
DENNIS L. STERN, VP, Human Resources
JAMES L. TERRILL, VP, Chief Financial Officer
DAVID A. THURM, VP, Production

PENELOPE MUSE ABERNATHY, President, News Services
MARTIN A. NISENHOLTZ, President, Electronic Media

Keeping the Census Fair

In 1990, for the first time in decades, the Bureau of the Census produced a count that was less accurate than the one that preceded it. Under Congressional direction, the agency consulted the National Academy of Sciences and came up with a new, more statistically sophisticated procedure. But Republican leaders have declared war on the plan. In effect, they are fighting for a census in which some Americans count for more than others.

Changes in American life style, in which families double up, children shuttle among relatives and adults move into and out of relationships make the population increasingly hard to pin down. The Census Bureau estimates it missed about 8.4 million people in 1990, most of them residents of rental housing. About 4.4 million others — many of them college students and people who owned two homes — were counted twice. The results were disturbingly skewed along racial and ethnic lines. One in 10 black males were missed by the census, and 1 in 20 Hispanics. About half the Americans who never got counted were children.

To keep the same thing from happening in the year 2000 census, officials decided to use statistical sampling techniques to check the accuracy of the count. Field workers will be dispatched to 750,000 randomly selected houses to perform a second mini-census. The two results will be compared, and the bureau will try to reconcile the differences.

Sampling techniques are the state of the art in statistics, but Republicans contend that they violate the constitutional mandate for "actual enumeration" of the population. The party actually fears that since poor and minority Americans tend to be Democrats, correcting the census will help the opposition.

The Republicans are attempting to starve the Census Bureau into submission. The proposed appropriations bill for the Commerce Department contains funding for census preparations for only the next six months. If the bill becomes law, next March President Clinton will have a choice of accepting the Republicans' demands for a less accurate count or allowing preparations for the census to come to a halt.

The President may feel obliged to veto a budget bill that endorses such an unfair procedure. To avoid that, the House should adopt a Democratic amendment to the Commerce budget that would finance the census for the entire year, and direct the bureau to be prepared to use either counting method. The Republicans have challenged use of sampling in court, and in the unlikely event the Supreme Court decides to take up the issue, the judiciary can decide the matter. Otherwise, Congress must stop trying to politicize the census.

Pittsburgh Post-Gazette

Founded 1786

Paul Block, publisher, 1927-1941 Paul Block Jr., co-publisher, 1942-1997
William Block, co-publisher, 1942-1989, chairman, 1990-

Jon Robinson Block
chairman and editor-in-chief

William Block Jr.
chief executive officer and publisher

Jon G. Craig Jr., editor and vice-president
managing editor

Robert B. Hagan vice-president and general manager
Robert J. McCue vice-president of sales and marketing

Counting and the court

Judges reject a sensible system for the census

A special federal court panel has ruled against the Clinton administration's plan to adjust the 2000 census by using scientific sampling. If the Census Bureau does not prevail at the U.S. Supreme Court — which seems certain to take up the case — an untrustworthy count is guaranteed. Let us count the ways that will hurt the public interest.

The census, of course, provides important demographic information, but its figures are also the basis for the apportionment of seats in the U.S. House of Representatives. It is that fact, above all, that has made sampling controversial.

Few people seriously argue that sampling will not bring a more accurate count. While the popular mind may consider it an attempt to substitute something like a public opinion poll for the job of counting heads, the reality is far different. The forms will still go out to households. Census workers will still follow up. Indeed, their goal is to reach a minimum of 90 percent of households in each census tract.

But because it is impossible to reach everyone, sampling would be done at the margins to adjust the figures and better reflect reality. This is not a bogus exercise: the Census Bureau already uses sampling widely in gathering data. It is a method endorsed by a panel of the National Academy of Sciences. You would have to be a Luddite to be against it.

Or a Republican. The Republican-controlled House brought the suit ruled upon this week. One of its main arguments is that Article 1, Section 2, of the Constitution requires that an "actual enumeration" be made. Since an actual count will be done, and sampling is intended only to supplement it, the Republicans' interpretation seems to us overly narrow.

Actually, the three federal judges ruled on narrow grounds of their own — they found for the plaintiffs on the basis of wording in an amended section of the Census Act. This legislation allows sampling but not, the Republicans successfully argued, for the purpose of apportionment. As the decision acknowledges, other courts have ruled differently on this issue. It is another reason why the Supreme Court must have the last word.

Behind the Republicans' resistance to sampling lies the fear that the Clinton administration will manipulate the census data for its own political gain. Given the abuses reported in the naturalization process for new citizens in the recent past, that fear is — sadly — not entirely preposterous, although the Census Bureau does enjoy a solid reputation for professionalism. But if the Republican leadership really was serious about that objection, it could devise some mechanism to monitor and guarantee the process to its satisfaction.

The truth is that the Republicans fear not so much a manipulated census but an accurate one. In the 1990 census, which happened to be the first in America's history that was less accurate than the one before, 4 million people were undercounted. As GOP strategists well know, the missing Americans included groups not usually known as Republican supporters. It is estimated that 4.4 percent of African Americans were missed, 5 percent of Hispanics and 12.2 percent of American Indians on reservations (by contrast, only 0.7 percent of non-Hispanic whites were not counted).

Legislative apportionment based upon a faulty count of minorities and poor people is an exercise in disenfranchisement, — one that threatens serious consequences for the long-term health of [illegible] democracy. We can only hope that the Supreme Court reads the law in that light.

San Francisco Chronicle
THE VOICE OF THE WEST

EDITORIALS

Blocking the Census For Political Gain

COUNTING NOSES in the United States is a difficult — and politically loaded — task. A decision by a three-judge panel blocking the sensible practice of statistical sampling to project correct population numbers can only worsen the chances of an accurate census in 2000.

This is no idle academic debate. Big money changes hands with a federally-posted census total. California may have lost $120 million in federal aid because of missed head counts in the 1990 census. Medicare, welfare and transit funds go where the population resides, a factor that should help a growing state.

A legal ruling threatens to hinder the effort to count America's population.

There are also clear political stakes. Mailed census forms followed up by doorstep interviews by census takers sound like a good way to count the population. But customarily left out are children, minorities and renters who avoid being tallied for many reasons.

By measuring a small area for these missed numbers, the Census Bureau projects a final number it believes is more accurate for a wider zone. Skeptics think this technique is a guesstimate, but careful testing of such a current experiment under way in Sacramento has taught statisticians to use the device properly.

Republicans in Congress, who filed the successful suit, believe this method is flawed. It evades constitutional language calling for "actual enumeration" and, more specifically to the lawsuit, goes against language barring sampling for all-important political redistricting. It's important to note the judges did not pass on constitutional questions or the worth of sampling, only whether fuzzy wording in federal law allows its use.

Democrats claim the suit is an exercise in political avoidance. Undercounted minorities, chiefly in California and the Southwest, are not likely to comprise safe Republican districts. Thus, a higher population count reached by sampling won't help the GOP, now with a thin majority in Congress.

The next step in the dispute is a likely appeal to the Supreme Court by the Clinton administration. In the meantime, it is making plans for a conventional census count, adding $1 billion to the $4 billion budget.

Legal arguments to the side, the country is poorly served by a partial census. Three panels of statistical experts and the head of the Census Bureau believe sampling is sound. In a diverse, mobile society, head counting is a complex task. Sampling is the proper tool to learn who we are.

FOUNDED BY JOSEPH PULITZER, DECEMBER 12, 1878

ST. LOUIS POST-DISPATCH

JOSEPH PULITZER EDITOR & PUBLISHER 1878-1911
JOSEPH PULITZER EDITOR & PUBLISHER 1912-1955
JOSEPH PULITZER EDITOR & PUBLISHER 1955-1986
CHAIRMAN 1970-1993
MICHAEL E. PULITZER CHAIRMAN & PRESIDENT 1993-

NICHOLAS G. PENNIMAN IV PUBLISHER
TERRANCE C.Z. EGGER GENERAL MANAGER

COLE C. CAMPBELL EDITOR
RICHARD K. WEIL JR. MANAGING EDITOR
CHRISTINE A. BERTELSON EDITORIAL PAGE EDITOR
ARNIE ROBBINS DEPUTY EDITOR

CENSUS

Let no person be invisible

Forget the law and the Constitution for a moment. The goal of the census is to achieve the most accurate possible count. The best way to do that — the scientists at the National Academy of Sciences say — is to use the modern techniques of statistical sampling.

Unfortunately, the law, the Constitution and the GOP pose difficult hurdles.

On Monday, a federal court in Washington ruled that federal law does not allow statistical sampling for apportionment.

The three-judge court accepted the evidence showing that the census is increasingly inaccurate. In 1990 it missed 8 million people — 12 percent of Native Americans, 5 percent of Hispanics, 4.4 percent of African-Americans and less than 1 percent of whites.

The results of that undercount may shortchange minorities in Congress and mean less federal money for cities like St. Louis.

At issue in the court case was a 1976 law requiring the Census Bureau to use sampling "except" for apportionment. The Clinton administration said this meant the Census Bureau *must* use sampling for other census purposes, but still *could* use sampling for apportionment. The federal court disagreed. Congress wouldn't have made such a "momentous" change in the census in such an "oblique" way, it said. The court has a point.

Congress could fix the problem by passing a new law, but probably won't because Republicans fear a loss of power if more minorities are counted. So the Clinton administration will appeal to the Supreme Court. Even if the court agrees with the administration on the 1976 law, there would still be the constitutional hurdle.

The Constitution requires an "actual enumeration." Webster's has two definitions of enumerate. The first — "to determine the number of" — lends support to sampling. The second — "name one by one" — does not. There's no telling which definition would move the court.

Congress should not wait to find out. If partisanship stands in the way of changing the 1976 law, Congress should at least increase the number of counters for poor areas in 2000.

In a government of the people, no person should be invisible.

The Philadelphia Inquirer

ROBERT J. HALL, Publisher and Chairman
ROBERT J. ROSENTHAL, Editor and Executive Vice President

WILLIAM K. GOARD, Managing Editor
PHILIP DIXON, Associate Managing Editor
SANDRA L. WOOD, Associate Managing Editor
RONALD PATEL, Sunday Editor

JANE R. EISNER, Editor of the Editorial Page
CHRIS SATULLO, Deputy Editorial Page Editor
ACEL MOORE, Associate Editor

A14 Wednesday, August 26, 1998

EDITORIALS

Counting on Congress

Despite a ruling against statistical projections, lawmakers should help assure an accurate census.

The 1990 Census missed more than 8 million people. The 2000 Census may be worse.

What's maddening is that experts know how to make the count more accurate, but may not be allowed to.

The problem isn't scientific, but political. The undercount is concentrated among blacks, Hispanics and other minorities who tend to vote Democratic. So Republican leaders like it just fine.

Now, there's an unfortunate twist in the partisan war over how to conduct the 2000 Census. A panel of federal judges has just barred the Census Bureau from supplementing the traditional "head count" with statistical projections to make the numbers more accurate.

The science behind such projections has been endorsed by the American Statistical Association and by the National Academy of Sciences. These expert groups and others have made it clear that the traditional mailings and door-knocking aren't sufficient to measure a society in which families have fragmented, government is widely mistrusted, and minorities feel marginalized.

Faced with this expert consensus, Republicans on the Hill have argued that a Census Bureau that is part of the Clinton administration can't be trusted to do the projections objectively. In any event, they say, such estimations would violate the Constitution's requirement for an "actual enumeration" of the population.

The constitutional issue may turn out to be a close call, but the three-judge panel didn't get to it. Instead, they concluded that a 1976 federal law prohibits the use of statistical projections in divvying up congressional districts among states.

It would make sense for Congress to give the Census Bureau clear legal authority to employ such projections — especially in light of advances made in such techniques since 1976 — but the Republicans aren't shooting for fairness. And the unanimity of the ruling suggested to surprised Democrats that their view will be a tough sell in the Supreme Court.

The biggest potential losers, though, are not the political partisans wrestling for control of the House, but the undercounted minorities and the cities where they are concentrated. From Philadelphia to New Orleans to Los Angeles, such regions are being cheated out of full political representation — and some of the federal aid that's divided according to Census numbers.

The Republicans know full well that the traditional Census is increasingly inaccurate. Yet they self-righteously huff and puff that it would be wrong to stray from the traditional, antiquated methods. They deserve a two-faced Oscar for cynicism.

Mr. MILLER. I look forward to having the opportunity to meet with groups. Unfortunately they are sending the message they don't want to talk to us, but we are hoping that during the break after October, and some field hearings, we can go into communities and find out how we can improve our methods of collecting the full enumeration.

I hope in your statement you didn't mean that this is a waste of time. We have eight very distinguished statisticians here. We have talked about statistics. I think you would have to agree, I would agree that all eight are highly respected and eminent statisticians, knowledgeable of the census issue. You did make a statement that it wasn't worth the trouble to even have this. I don't think you really mean that, do you?

Mrs. MALONEY. Sometimes I do. These hearings have been——

Mr. MILLER. Well I will assure the eight people that I will listen.

Mrs. MALONEY. Unfortunately, very partisan.

Mr. MILLER. I will be listening.

Mrs. MALONEY. Just as a point of information, the groups have indicated to me that they would like to meet with you. If you are telling me they are refusing to meet with you, I certainly would like to really personally intercede to set up these meetings.

Mr. MILLER. Well we will be——

Mrs. MALONEY. There is a cross or a conflict in communications, because they are indicating to me they would like to meet with you.

Mr. MILLER. Well, great. We are looking forward to that. I hope you will look forward to this hearing too, because these people have something to contribute. I am disappointed that you don't think it's worth the effort. I'm sorry that you feel that way.

Mrs. MALONEY. I always look forward to our discussions, Mr. Miller.

Mr. MILLER. Let's proceed with the first panel. Dr. Breiman and Dr. Ylvisaker. If you would remain standing, we need to swear you in here.

[Witnesses sworn.]

Mr. MILLER. Let the record state that they both answered in the affirmative.

Let me proceed with the opening statements. Dr. Breiman, if you would like to proceed first. Dr. Ylvisaker will go next. Then we'll go to questioning. Welcome.

STATEMENTS OF LEO BREIMAN, PROFESSOR OF STATISTICS, U.C. BERKELEY; AND DONALD YLVISAKER, PROFESSOR OF STATISTICS, UCLA

Mr. BREIMAN. I thank Chairman Miller and all the members of the subcommittee for inviting me to testify today regarding census adjustment. Thirteen of my years as a career statistician were spent as a freelance statistical consultant——

Mr. MILLER. Excuse me 1 minute, if I may. I meant to ask one thing. I would like to ask all the panelists if they would hold the microphone a little bit closer. But would you also briefly tell us where you are now, and what qualifies you with respect to census? So that we know on all eight of you what your qualifications are, briefly.

Mr. BREIMAN. Yes. I am professor emeritus of statistics, University of California, Berkeley. I think I have spent almost 50 years as a statistician. Thirteen of my years as a career statistician were spent as a freelance statistical consultant, participating in dozens of projects involving large amounts of data. I learned, sometimes painfully, that the first and often most important thing a statistician needs to do is check the data quality, look at the details.

In 1991, the Census Bureau carried out extensive evaluation procedures aimed at estimating errors in the adjustment estimates. These were summarized in numerous reports which I carefully studied along with other related documents. My study is summarized in a paper, "The 1991 Adjustment: Undercount or Bad Data," published in 1994 in the well-regarded journal of Statistical Science, and which I submit for the record.

Based on my close examination of thousands of pages of reports, I concluded that at least 70 percent of the initial 1991 undercount estimate came from data errors. The paper is filled with quantitative descriptions of numerous error sources. It has been reviewed by statisticians in favor of adjustment and by the Census Bureau. No errors have been found.

The initial undercount estimate by the Bureau in 1991 was 5.3 million people. The numbers in a later census report and also on page 15 of the Bureau's CAPE report, giving known errors and a bias due to data errors, results in an undercount estimate of only 2.3 million people. This is 50 percent less, 57 percent less than the initial 5.3 million estimate. This means that the Bureau is admitting that 57 percent of its initial undercount estimate came from data errors. So the Bureau and I agree that the major part of the proposed 1991 adjustments came from errors, although 70 percent is more realistic than 57 percent.

This raises the question of how the Census Bureau, known for the accuracy of its surveys, produced adjustments so error-filled. The key reason is that in the undercount estimation procedure, the effect of small errors is magnified. Here is a simplified picture. The 1990 procedure and the year 2000 ICM share the same approach. The census is followed several months later by a mini-census in many randomly selected areas, each containing a few hundred people. An effort is made to match the people found by the mini-census to the people found by the census in the same areas. The percent of non-matches is an estimate of the undercount in the area. Then these estimates are used to estimate undercounts for all other locales in the country. A calculation shows that an error in failing to match 1 person in 100 can often lead to a doubling of the undercount estimate. A failure to match 2 people in 100 can often lead to a tripling of the undercount estimate.

Knowing this magnification effect, the question becomes whether it is possible to get match accuracy of 98 or 99 percent, particularly in those areas we think are most susceptible to undercount. The experience of the 1990 adjustment shows that it is not possible. There are too many people moving in and out of places. There are too many survey forms with garbled or missing or falsified data. There are too many errors in addresses. There are simply too many sources of error. My article documents many of these. The problems

are inherent in the nature of the mobile and diverse U.S. population.

After reviewing the census 2000 operational plan, my opinion is that the effective data errors on the ICM will be just as significant and probably more so than in the 1990 adjustment. These errors will be spread unevenly around the country in an unpredictable pattern. The ICM will reflect the extent of errors and not the undercount. The bottom line is that one cannot correct the errors in the census by adding more errors to it.

Thank you for the opportunity to state the conclusion of many hours, days, and months of looking at the fine print.

[The prepared statement of Mr. Breiman follows:]

Testimony on Adjusting the Census for Undercount
to the House Subcommittee on the Census
September 17, 1998

Leo Breiman
Professor Emeritus of Statistics
University of California at Berkeley

I thank Chairman Miller and all the members of the Subcommittee for inviting me to testify today regarding Census adjustment. Thirteen of my years as a career statistician were spent as a free-lance statistical consultant participating in dozens of projects involving large amounts of data. I learned, sometimes painfully, that the first and often most important thing an statistician needs to do is check the data quality--look at the details.

In 1991 the Census Bureau carried out extensive evaluation procedures aimed at estimating errors in the adjustment estimates. These were summarized in numerous reports which I carefully studied along with other related documents. My study is summarized in the paper "The 1991 Adjustment--Undercount or Bad Data" published in 1994 in the well-regarded journal Statistical Science, and which I submit for the record.

Based on my close examination of thousands of pages of reports, I concluded that at least 70% of the initial 1991 undercount estimate came from data errors. The paper is filled with quantitative descriptions of the numerous error sources. It has been reviewed by statisticians in favor of adjustment and by the Census Bureau. No errors have been found.

The initial undercount estimate by the Bureau in 1991 was 5.3 million people. The numbers in a later report, and also on page 15 of the Bureau's CAPE Report giving known errors and the bias due to data errors, results in an undercount estimate of 2.3 million people. This is 57% less than the initial 5.3 million estimate. This means that the Bureau is admitting that 57% of its initial undercount estimate came from data errors. So we are not far apart. Both of us, the Bureau and I, agree that the major part of the proposed 1991 adjustments came from errors.

This raises the question of how the Census Bureau, known for the accuracy of its surveys, produced adjustments so error-filled. The key reason is that in the undercount estimation procedure the effects of small errors is magnified. Here is a simplified picture.

The 1990 procedure and the year 2000 ICM share the same approach. The Census is followed several months later by a mini-census in many randomly selected areas, each containing a few hundred people. An effort is made

to match the people found by the mini-census in each area to people found by the Census in the same area. The percent of non-matches is an estimate of the undercount in the area. Then these estimates are used to estimate undercounts for all other locales in the country.

A calculation shows that an error in failing to match one person in 100 can often lead to a doubling of the undercount estimate in the area--an error in failing to match two people in 100 can often lead to a tripling of the undercount estimate.

Knowing this magnification effect, the question becomes whether it is possible to get match accuracy of 98 or 99%, particularly in those areas we think are most susceptible to undercount. The experience of the 1990 adjustment is that it is not possible. There are simply too many sources of error. There are too many people moving in and out of places, there are too many survey forms with garbled or missing data, and so on. My article documents many of these error sources. The problems are inherent in the nature of the mobile and diverse United States population.

After reviewing the Census 2000 Operational Plan, my opinion is that the effects of data errors on the ICM will be just as significant as in the 1990 adjustment and these errors will be spread unevenly around the country in an unpredictable pattern. The bottom line is that one cannot correct the errors in the Census by adding more errors to it.

Thank you for this opportunity to address the committee. I will be happy to answer any questions.

Mr. MILLER. Thank you very much.

Dr. Ylvisaker. Am I pronouncing it close enough?

Mr. YLVISAKER. That's pretty good.

Mr. MILLER. Thank you.

Mr. YLVISAKER. I thank Chairman Miller and the other members of the subcommittee for this invitation to speak at today's hearing. You have my written testimony. Since time is short, I will simply touch on the points that deserve some emphasis.

I would like to start with a personal view of the statistical community since it's brought up quite a bit. Not a lot of statisticians have actually thought a whole lot about the census. By a whole lot, I mean investing an enormous amount of time to get current with details. Statisticians who have thought seriously about the proposed adjustment methods are divided over the possibility that they can correct counts or shares. In my opinion, the line that divides them on this issue is quite a familiar one. It has to do really with the degree to which, or limitations, one might put on statistical methods. Some of us are being more cautious perhaps than others.

It is particularly unfortunate to me that the proposed methodology is so closely identified with statistical sampling. This skews the census debate in the larger statistical community because statisticians who are not familiar with details will vigorously and quite properly defend sampling.

Turning to methodology, the ICM proposed for 2000 is a descendent of the PES. So one might look back at the 1990 outcome. The 1990 adjustment procedure was not simple to begin with. It saw considerable complication before it was finished. The sample of 167,000 housing units first produced raw undercounts for 1,392 poststrata. That's about 140 housing units per poststrata. It was known from the start that the undercount estimate for an individual poststrata would have to be based in part on what happened in other poststrata. And the smoothing of rates, based in part on what happened in other poststrata, accounted for a good deal of the complexity of the process.

In 1991, the Undercount Steering Committee found itself unable to evaluate the aspects of the full process in the time allotted to them. In 1992, the Committee on Adjustment of Post Censal Estimates later in 1992 endorsed the simplification that came when the number of poststrata was reduced from 1,392 to 357. Looking back, it is pretty clear that the plan that was put in place in 1990 was no longer with us, and certainly no longer with us now.

By 1992, it was also clear that a good portion of what had been supposed at the time of the secretary's decision to be measured undercount was in fact measured non-sampling error. A portion of this error could be removed, not all of it.

A third problem, and probably the most serious of all, is the problem of heterogeneity of poststrata. This has been observed in a variety of studies. It should come as a surprise to no one, the impact of such error on the undercount estimation is quite uncertain. The problem in the future is that it is a continuing problem. No one can say anything definitive about heterogeneity of poststrata or correlation bias.

Going on to the methodology in census 2000, there are new features that set it apart from the 1990 plan. Poststrata descriptions are thought more efficient and will stay within State boundaries. A larger survey will be taken with the prospect of gaining more information. Sampling for non-response follow-up has brought in to better allocate resources and reduce costs.

At the same time the ICM, on 750,000 households as opposed to 167,000, is still going to be too small to keep sampling error within required bounds. If the number of poststrata goes up considerably, as it must having 51 States and the District of Columbia to deal with, we are going to have the property that estimates of undercount rates in certain poststrata are going to have to depend on undercount rates in other poststrata. This is going to require some form of smoothing again. A simpler form of smoothing has been mentioned. The latest information that I have is not totally clear on exactly how this is going to take place. The difficulty is going to be that we are going to use dual system estimation. This is going to contain the same kind of errors that have been there previously. There figures to be more opportunity for such errors to occur, given the size and the tight time schedule of the operation.

Overriding the issue is the question of correlation bias. This remains a barrier to concluding that improvement has really been made.

There still seems to be much that is unsettled. Poststrata in each of the 50 States and the District of Columbia will have to be decided. These are individual decisions. The way in which undercount estimates will be finally obtained is not evident in the latest census documents.

One cannot but be impressed by the number of decisions that accompany the task of planning and implementation of such a large scale survey, and the analysis that's going to have to follow it. The added element of sampling for non-response follow-up seems destined to add its own complexity to this mix.

Again, I find it particularly unfortunate that the public is assaulted with a great deal of misinformation at the present time. We are not talking about sampling, scientific method, and so on. We are talking about a very specific set of techniques that happen to employ sampling. When this is understood clearly, we'll be somewhat advanced. The problem I see with it now is that the proposed adjustment method is not a simple method. It's a time when accessibility of census methodology is of utmost importance. People can't agree and may not cooperate in what they do not understand. Thank you.

[The prepared statement of Mr. Ylvisaker follows:]

Census Undercount Issues

Testimony prepared for the House Subcommittee on the Census
September 17, 1998

Donald Ylvisaker
Department of Statistics
UCLA

I thank Chairman Miller and the other members of the Subcommittee for the invitation to speak at today's hearing. Thanks are also due the staff for providing me with much helpful information.

My direct involvement in census matters began with a Joint Statistical Agreement with the Bureau in 1990-1991, this to assess the test census carried out in East Los Angeles in 1986. Some other formal involvement followed, including work for the Department of Justice during the preparation phase of the district court case in the spring of 1992, and membership on the Panel of Experts to the C.A.P.E. Committee in the summer of 1992.

Less direct contact with the census debate has been within the statistics community. The issue here is not whether sampling is generally useful, there is overwhelming evidence that it is. The pertinent question is whether techniques of the type under consideration, and featuring sampling, are capable of improving the census with respect to the undercounting of minorities. Statisticians who have thought seriously about the question are divided over the answer. This division is roughly along lines that divide statisticians on other large issues, indicative of a deep-rooted philosophical split in their ranks.1

Can adjustment of the traditional headcount be accomplished through statistical means so that the differential undercounting problem is ameliorated? I have great respect for the people at the Bureau who have been put in the position of having to bring forward a solution to this problem while under intense scrutiny. At the same time, I do not feel that methods now proposed are capable of improving matters so that one could, with any confidence, call for their use. This conclusion follows from the arguments outlined below.

What is required? One needs a procedure for adjustment that is simple enough to be understood, and hence to be viewed as acceptable, if grudgingly, by the principals on whom it falls.2 Moreover, prespecification of what is intended should be complete enough that such acceptance can be reached in a timely way. All this takes place in the face of a problem that is enormously difficult - the provision of corrected counts down to the census block level.

What is proposed? In the PES of 1990, and in the ICM planned for 2000, the population is segmented into poststrata, an example of which might be the collection of male Hispanic owners, aged 50+, living in rural California. When it comes to the question of being counted in the census, it is expected that differences between people in distinct poststrata are generally larger than differences between people in the same poststrata. The intention is to invoke the homogeneity of like persons to bring about rather precise estimates of the undercounting of the particular group.

Given homogeneous and stable poststrata of responsive and well-identified individuals, the method of capture (census) - recapture (PES/ICM) allows one to estimate true counts, translate them to adjustment factors for poststrata, and apply these factors to preliminary counts at various levels of aggregation. In theory, such estimates of adjustment factors are subject to a sampling error that is controlled by appropriate choice of sample size. In census practice, people are not well-identified and there are significant problems in matching them; populations are not stable so one must deal with movers; there is nonresponse and imputations might follow. Difficulties of this sort bring in nonsampling error, or bias. Moreover, poststrata are not homogeneous and this results in correlation bias. As it happens, good estimates of sampling error

are relatively easy to make compared to estimates of nonsampling error. Correlation bias, by its nature, does not allow direct estimation.

1990 Results. The PES used a sample of 167,000 housing units to estimate adjustment factors for 1392 regionally based poststrata. The size of the sampling error in the poststrata estimates, and suspect differentials when estimates were viewed in combination, meant estimates had to be revised. One aspect of this revision had been prespecified, the smoothing step, while another had not, a presmoothing step was added during the estimation phase. The technical discussion in the Undercount Steering Committee Report of 1991 found the full process "too complex to evaluate directly within the available time" and allowed that the true sampling error of estimates was likely understated by a substantial amount.3 Direct evaluation was completed at a later date and the understatement of sampling errors was substantiated.4 In words, the process that had evolved to account for all circumstances, foreseen and not, was too complex to be fully understood in real time, and the optimistic view of its basic properties was found to be unwarranted.

Postcensal estimation had to be considered in 1992 and this allowed a period of time during which the PES and its evaluation studies could be re-evaluated. C.A.P.E. reduced the number of poststrata from 1392 to 357 to avoid quantifying sampling error as it played out through the smoothing process that had been used. This simplified the estimation method at the possible expense of added heterogeneity within poststrata. Serious attention was paid to nonsampling error and, in particular, a significant computer error and the re-examination of 104 PES blocks reduced the estimated national undercount from 2.1% to 1.6%. C.A.P.E. deemed revised estimates to be an improvement at the national and state levels, but found them short of this standard for smaller areas.5 The Committee further estimated that 45% of the smaller figure was due to measured nonsampling error as opposed to measured undercount. In his study, Breiman found that about 80% of the measured undercount was attributable to nonsampling error.6 Neither finding is made more palatable by the fact that correcting for correlation bias is generally thought to mean there are additional uncounted people - correlation bias cannot be directly estimated so additions cannot be sited with distributive accuracy. By 1992, if not earlier, the 1991 decision to not adjust could be said to have considerable technical support.

Plans for 2000.7 Census 2000 poststrata are closer in spirit to those thought efficient in 1992 than those used in 1990. At the same time, they will not cross state boundaries. The ICM will be based on a sample of 750,000 housing units, a near five-fold increase over the PES. A new ingredient is the use of sampling for nonresponse follow-up, a program that sets a goal of 90 percent response, or more, in each eligible census tract.

From Waite and Hogan regarding poststrata and sample size:

"If, for example, we define six race/origin groups, seven age/sex groups, two tenure categories and three geographic groups in a typical state, a total of 252 poststrata would be required, more than the sample will support. "

This brings one back to the question of smoothing estimates of poststrata adjustment factors. To do this via the proposed raking is to use a methodology closer to the surface than the one employed in 1990. Still, such smoothing must rely on initial (raw) dual system estimates and these are subject to the same bias problems previously met. Moreover, one finds:

"Research on the best characteristics to use to define the marginal constraints, and on the number of dimensions to use for the raking matrix will continue. We have not decided whether raking will be used in 2000."

In plain terms, the fixing of poststrata and the estimation of adjustment factors for them is not yet settled. Evidently a large number of judgments remain to be made, in light of the fact that individualized decisions are required in each of 50 states and in the District of Columbia.

The increase to a sample size of 750,000 housing units is both a blessing and a curse. One hopes to find smaller sampling error in estimates of undercount rates, but size is a dangerous element in the realm of nonsampling error. No doubt much useful information emerged from the PES about matching, mover and imputation policies to help control nonsampling errors of the type previously found. At the same time, it is safe to say that these 1990 lessons will fall short of being able to deal with unforeseen problems arising in 2000, for the following reasons: the ICM is a much larger scale operation than the PES; the ICM has a tighter time schedule than did the PES; the introduction of sampling for nonresponse follow-up will interact with the ICM in ways that are not likely to be fully understood beforehand.

Does the Census 2000 Plan provide what is required? By the standards mentioned earlier, evidently not. It is not a simple adjustment process. That it is not well-understood follows from the fact that acceptance of it is still commonly confused with the appropriateness of statistical sampling. Prespecification appears to be far from a reality. Finally, the question of whether one is correcting counts will not be answered before (or after) the year 2000, as the experience of the early 1990s demonstrates.

1Such a view of statisticians is expressed well in Paul Meier's testimony to be found in the transcript of the United States District Court, Eastern District of New York, trial of City of New York et al. v. United States Department of Commerce et al., at page 1972. It is also given in more colorful language in a 1986 paper concerning the seminal discrimination case Hazelwood School District v. U. S. - see the Alternative Outlooks section of "What happened in Hazelwood" by Meier, Sacks and Zabell, the paper appears in Statistics and the Law (DeGroot, Fienberg and Kadane, eds.), Wiley, New York.

2A specific call for simplicity can be found as Point 4 in the Future section of the C.A.P.E. Committee Report, at page 34.

3This language is found in Appendix 6 of the 1991 Report of the Undercount Steering Committee.

4Fay, Robert E. (1992). "Inferences for small domain estimates from the 1990 Post-Enumeration Survey." Technical Report, Bureau of the Census, Washington, D. C.

5Recommendation 4 of the C.A.P.E. Committee Report.

6Breiman, Leo (1994). "The 1991 Census Adjustment: Undercount or Bad Data?" (with discussion). Statistical Science 9, 458-475, 521-527.

7 I rely in this section on selected portions of the most recent Bureau document at hand – "Statistical methodologies for Census 2000- -Decisions, Issues, and Preliminary Results", by Preston Jay Waite and Howard Hogan. It was presented to the Joint Statistical Meetings in Dallas on August 13th. Quotations that appear are from that paper.

Mr. MILLER. Thank you all very much. The majority staff has prepared a glossary of terms that are used, if anybody wants it. I think it is on the table back here to make sure everybody is comfortable with that.

I made a statement that this is the largest statistical experiment in history. I say that because based on empirical evidence, we don't have a lot to go on as to what they are doing and have proposed in the year 2000. What do you say about that statement? What else is there other than the 1990? Is that the only large experiment with sampling that we have used or is there any other large experiment like this? We've had a dress rehearsal which is not that large.

Mr. BREIMAN. Well, Congressman, I don't understand why the Bureau is going in this direction, given the experience and the errors in the 1990 PES. It just baffles me. I mean it's like buying a pig in a poke. Whereas in 1990, it was known to be a bad pig. So why are we trying to do it again with a bigger pig in the poke?

Mr. MILLER. Doctor?

Mr. YLVISAKER. I don't know about the largest. I would hesitate to say it's the largest. Without thinking, one tends to think of the polio vaccination program as a very large experiment. I think it is cited as one. A very careful experiment had to be run in that circumstance. The absolute size is not so important, but I think when one begins to think about the detail that must go into a very few months of decisionmaking, I am certainly boggled by size.

Mr. MILLER. But based on empirical evidence, the 1990 is the only real thing we can really go back on.

Mr. YLVISAKER. I think we must depend on the 1990 census, at least to learn lessons. I am not sure these lessons are going to be sufficient for 2000, because one is changing two things. We changed the methodology, and we changed the size at the same time. Who is to say what is going to happen.

Mr. MILLER. Both of you agree that the 1990 attempt of using sampling was a failure. Is that right?

Mr. BREIMAN. Definitely.

Mr. YLVISAKER. I think there is no way to disagree with the fact that this was not a success.

Mr. MILLER. I know you can't generalize for the statistical community. They spent 2 years analyzing it and never really felt comfortable.

Will the increasing sample size make a difference?

Mr. YLVISAKER. It is going to have two effects. It is going to complicate matters a good deal. It is supposed to have some positive effect as well. I am more nervous about the possible negative effect of size.

Mr. MILLER. The negative effect of size?

Mr. YLVISAKER. It requires more people. It requires more time. Many, many decisions have to be made on a variety of issues. Size is not necessarily good at a certain level.

Mr. MILLER. They are going to allow half the amount of time for their sampling, with 5 times the larger sample.

Mr. BREIMAN. If I may weigh in on this? The size of the sample does not affect the critical problem, which is how well can you match people. To what percentage can you match people? Can you achieve say a 99 percent match rate? Size has nothing to do with

that. I mean you can sample a million households. You are still faced with the problem of how accurate can your matching be, because as I noted, just 1 percent error in matching can double the undercount estimate. That matching remains a property of the U.S. population in terms of how well you can match.

Mr. MILLER. Dr. Breiman, in your statement you said that 70 percent of the initial undercount came from data errors. Would you explain a data error?

Mr. BREIMAN. Yes. OK. Part of it was a million of that undercount was due to a computer error in classifying people. OK? Which was later found and corrected. The remainder of it was that in their extensive evaluations and followup evaluations, the Bureau did things such as checking, did a double check on the matching by doing an independent matching later on. So they had some idea of what the error and the matching was. They factored that in. Then there were followup interviews done after the PES, in which people went out and they gathered more information in these followup interviews. For instance, they found there were more matches than had originally been found by the PES.

Another very interesting thing is this, that 250,000 arrived this way. The Census Bureau decided to more carefully rematch 104 block clusters, these little areas, out of over 5,000. Now these were low match-rate areas. But just rematching 100 of them out of the 5,000 they found enough matches, additional matches, to lower the undercount estimate by 250,000. Now one of the effects of the larger——

Mr. MILLER. Excuse me. The yellow light is on, and we are going to the 5 minute rule. The red light just came on. The bottom line is that the error rate was so great in the sampling experiment in 1990 and basically those same errors are going to continue in 2000. Is that what you are saying?

Mr. BREIMAN. It could be even worse because they are going to have to send and train more matching teams into the field.

Mr. MILLER. And they have less time to do it in.

Mr. BREIMAN. Right.

Mr. MILLER. OK. Thank you. We may have time for a second round.

Mrs. Maloney.

Mrs. MALONEY. Dr. Breiman.

Mr. BREIMAN. Yes.

Mrs. MALONEY. And Dr. Ylvisaker, thank you both for your testimony.

Dr. Breiman, during the litigation over the 1990 census, you were called upon to testify in the May 1992 proceedings.

Mr. BREIMAN. Yes.

Mrs. MALONEY. At that time, did the court accept your testimony as an expert witness?

Mr. BREIMAN. No. It did not.

Mrs. MALONEY. OK. I would like to request the chairman if I could put in the record two pages from the transcript of that case where they ruled that he was not an expert witness. Is that admissible?

Mr. MILLER. What is the purpose?

Mrs. MALONEY. To show that in the 1992 proceedings, the court ruled he was not an expert witness on statistical sampling.

Mr. MILLER. Is the purpose just to try to smear somebody?

Mrs. MALONEY. Not to smear someone, but I think that's important because it reflects on the credibility of his testimony, not only during the court then, but today. I think it is fair to have part of the legal proceeding where this is so stated entered into the record.

Mr. MILLER. Without objection.

Mr. SNOWBARGER. Mr. Chairman, I would object to that. I don't think it's a policy here to impeach our witnesses. She already has made it a part of the record, the point that she was trying to make. We have not seen copies of the transcript and I would object.

Mr. MILLER. Objection heard.

Mrs. MALONEY. Dr. Breiman, during your testimony you asked, it wasn't your spoken testimony today but in your written testimony you asked that a copy of your 1994 article from Statistical Science be made part of the record. Is it not the case that there were a series of criticisms of that article published by that journal?

Mr. BREIMAN. My article was a discussion article, which means that other statisticians were invited to comment on this article. I would very much like to also include the discussion for the record. In particular, Dr. Fienberg, Dr. Ericksen, the other names escape me. Numbers of others commented. I think the comments would be very interesting.

Mrs. MALONEY. Well thank you because I do believe that those criticisms, along with your response in this dialog, should be included likewise in the record along with the article.

Mr. BREIMAN. I agree.

Mrs. MALONEY. Without objection.

Mr. MILLER. Is there objection? If not, all of the articles will be included in the record.

[The information referred to follows:]

Statistical Science
1994, Vol. 9, No. 4, 458–537

The 1991 Census Adjustment: Undercount or Bad Data?

Leo Breiman

Abstract. The question of whether to adjust the 1990 census using a capture–recapture model has been hotly argued in statistical journals and courtrooms. Most of the arguments to date concern methodological issues rather than data quality. Following the Post Enumeration Survey, which was designed to provide the basic data for adjustment, the Census Bureau carried out various evaluation studies to try to determine the accuracy of the adjusted counts as compared to the census counts. This resulted in the P-project reports, which totaled over a thousand pages of evaluation descriptions and tables. Careful scrutiny of these studies together with auxiliary sources of information provided by the Census Bureau is used to examine the issue of whether the data gathered in the Post Enumeration Survey can provide reliable undercount estimates.

Key words and phrases: Census, Post Enumeration Survey, nonsampling error, undercount.

1. INTRODUCTION AND SUMMARY

To give the setting for this paper, we begin with a simple example. Consider a project undertaken to find the total fish population of a large pond. Efforts are made to catch all of the fish and paint a red X on their backs. In total 10,000 fish are caught and marked. To see if this effort gave a complete count, 100 fish were later caught and examined. Of these, 98 had X's on their backs, and two did not. If the recapture (second catch) is done at random with each fish in the pond having the same recapture probability, and if the population of the pond stays the same between the initial catch and the recapture, then an approximately unbiased estimate for the total pond population is 10,204 and there is an estimated undercount of 2.0%. Such estimates are called capture–recapture estimates.

However, suppose that subsequent study revealed that there may have been X's on the backs of the two fishes, but perhaps the X's had not been well painted on to begin with or that the examination had not been well carried out. Instead of there being a 2% undercount, what may be true is that there was 2% bad data, or perhaps there was a 1% undercount and 1% bad data. The question of how much of the data is bad is fundamental to knowing how accurate the undercount estimate is.

Leo Breiman is Professor, Department of Statistics, University of California, Berkeley, California 94720.

The effort to adjust the 1990 census for undercount was arduous. It consisted of following the census with the Post Enumeration Survey (PES), covering 380,000 persons; matching the census and PES records; and then computing capture–recapture estimates of the undercount. We examine two related questions. First, what are the sources of errors in the PES and matching, and how big are these errors? Second, what is the effect of the errors on the undercount estimates? In the end, our objective is to see whether, in retrospect, the census adjustment proposed in 1991 is statistically justifiable.

Along the way, we will try to give full references to the relevant published literature. Many of the important documents are unpublished internal Bureau of the Census reports or reports generated by other government agencies. We will cite these, noting that they are available on request from the proper agency. Some references are included that provide background information but do not have an immediate connection to the issues raised in the text. These are described in Section 7.

1.1 Background

The issue of whether to adjust the 1990 census using a capture–recapture model has been one of the most highly publicized and important statistical issues of the past decade. It has serious political and economic consequences. Census counts are used to apportion congressional and legislative seats and to

distribute tens of billions of dollars that flow from the national government to states, counties and cities.

Part of the long-term Census Bureau planning for 1990 was to follow the census with a Post Enumeration Survey (PES) covering over 150,000 households, to match census persons to PES persons and, using capture–recapture assumptions, to compute adjusted estimates of the census counts at the national, state and local levels.

The Department of Commerce decided in 1987 not to issue official 1990 census counts that were statistically adjusted. The estimates based on the PES data and matching would be used only to "provide a careful evaluation of the coverage of the 1990 Census" (Hogan and Wolter, 1988). However, under legal pressure from a group of cities and states who favored adjustment, the Department of Commerce in July 1989 agreed to initiate a new decision making process (see Department of Commerce, 1991b, Section 4, Appendix 1).

In March 1990, prior to the census and the PES, the Department of Commerce published guidelines to follow in deciding whether the adjusted census counts would be officially adopted (Department of Commerce, 1989). The most important, from a statistical point of view, was Guideline 1: "The Census shall be considered the most accurate count of the population of the United States, at the national, state, and local level, unless an adjusted count is shown to be more accurate."

The census interviews of households not returning their questionnaires by mail took place in May–July 1990. The PES was carried out in July–August 1990. The adjusted counts for the nation, states and larger counties and cities were released in June 1991 and showed an estimated national undercount of about 2%—5 million people (Department of Commerce, 1991a). The largest estimated undercounts were, as expected, among minorities in central cities.

On July 15, 1991, the Secretary of Commerce made the decision not to adjust (Department of Commerce, 1991b) and a group of cities and states (including, e.g., California, Los Angeles, Atlanta, New York, Florida and Texas) sued to force adoption of the adjusted numbers. Statisticians testifying for the plaintiffs were Barbara Bailar, Eugene Ericksen, Stephen Fienberg, John Rolph, John Tukey and Kirk Wolter. Testifying for the government were Peter Bounpane, Robert Fay, David Freedman, Paul Meier and Kenneth Wachter. Leo Breiman assisted the defendants. The testimony ended on May 28, 1992.

On April 13, 1993, the court issued its holding that "the decision against adjustment shall not be disturbed...." (U.S. District Court, 1991). The holding was based on the grounds that, since reasonable statisticians could differ on the merits of adjustment, "the Secretary's decision not to adjust the 1990 Census count was neither arbitrary nor capricious." The court did not base its ruling on the relative accuracy of the adjusted versus nonadjusted counts, and controversy over this latter issue will probably continue in the pages of statistical journals for years to come (with reasonable statisticians on both sides).

1.2 The Adjustment Method

The adjustment undertaking was extraordinarily complex. Over the decade preceding the 1990 census, much research had been concentrated on this issue by statisticians inside and outside of the Census Bureau. Four rehearsals aimed at uncovering deficiencies in the methodology were carried out. The first was in Mississippi in 1986, the second in Los Angeles in 1986, the third in North Dakota in 1987 and the last in Missouri and Washington in 1988. Numerous published papers, committee reports and thousands of pages of internal Bureau of the Census studies dealt with various issues involved. The reference section lists many of these.

The method used in 1990 to estimate adjusted counts at all levels down to the census block consisted of defining 1,392 poststrata on the basis of the following: age; sex; race or ethnicity; renter or owner; place type (i.e., central city, nonincorporated area, etc.); and geographic location. For instance, one poststratum is male, ages 10–19, black renter, central city, New England. All persons in the population, with some insignificant exceptions (Undercount Steering Committee, 1991), are in one of the poststrata.

Then the capture–recapture idea is applied, that is, there is an original survey (the census), followed by another survey (the PES). A matching is done to see how many of the persons found in the second survey were also found in the first. The census count is corrected for erroneous enumerations. In each poststratum, the corrected census count, the PES count and the number of matches are used to compute a capture–recapture estimate of the poststratum population. The population estimate divided by the original census count is the raw adjustment factor for the poststratum.

The population estimate is called the Dual System Estimate (DSE) (Wolter, 1986b). We stretch terminology by referring to the adjustment factors and the undercounts also as DSE estimates. Now we can give an explanation for the poststrata definition. The a priori belief was that the stratification used created population pools (each poststratum) having approximately equal recapture probabilities, thus validating the capture–recapture assumptions.

Statistical techniques were then used to produce a set of smoothed adjustment factors having smaller variances than the raw factors. The estimated variance–covariance matrix of the raw adjustment factors was "presmoothed," that is, regressed against some explanatory variables including, for example, indicators for gender, age and minority. Then the raw adjustment factors were smoothed using a Bayesian version of linear regression and the presmoothed variance–covariance matrix, with the variables in the regression selected using a "best subsets" method and Mallows' C_p. See Freedman et al. (1993) for a more detailed description.

Adjusted counts for each local unit are computed by slicing it into poststrata, adjusting each poststratum for its undercount and then recombining. For instance, suppose that a certain city has a census count of 10,000 in a poststratum and that this poststratum has an estimated adjustment factor of 1.035. Then the adjusted city count in this poststratum is $10{,}000 \times 1.035 = 10{,}350$. The overall adjusted count for the city is obtained by adjusting the counts in all poststrata intersecting the city population and adding these up.

The extent of the error introduced into the DSE estimates by using common adjustment factors over potentially diverse populations is a subject of controversy (Freedman and Wachter, 1994). For instance, the adjustment factor for the poststratum male, ages 10–19, black renter, central city, New England is used to adjust the counts for all central cities in New England. There is also controversy about the statistical validity of the smoothed adjustment factors and their estimated variances and covariances (Freedman et al., 1993). This paper, however, deals only with the issue of the errors in the DSE estimates attributable to errors in the PES data and the matching process.

1.3 Summary

This paper is laid out in two parts. First, Sections 2 and 3 describe the steps leading to the DSE estimates and the evaluation procedures and documents. Second, Section 4 discusses sensitivity of the undercount estimates to errors, and Sections 5 and 6 give an analysis of the errors and potential errors. Here is a brief summary and conclusion as an introduction and a road map.

The PES covered 380,000 people in 169,000 households. These households were in 5,290 block clusters selected using a stratified random design. Then efforts were made to match the PES data to the census data from the same blocks. This was a multistep process involving the reinterview of some households where better information was needed. Cases that could not be resolved into a match category were sent to an imputation process. (See Section 2.)

After the PES was completed, evaluation data was gathered through a rematching study, some field reinterviews and examination of quality assurance records. The evaluation results were summarized in documents called P-project reports. These studies form the main source of information for the error discussions in this paper. Due to the smaller sample sizes in the evaluation data, results were presented only at highly aggregated levels, that is, either as weighted to national figures or to the nation divided into 13 large evaluation strata defined by minority versus nonminority, by central city versus non–central city and by geographic region, Northeast, South, Midwest, West (see Section 3.5, Table 4). (See Section 3.)

To understand the subsequent error analysis, it is first necessary to understand that small errors in the PES can result in large errors in the undercount estimates. The DSE undercount estimates are computed from estimates of the census counts corrected for erroneous enumerations, the PES counts and the number of matches. Then, for instance, errors of 0.5% in the match count and in the PES count can together lead to a 50% error in the undercount estimate. Also, because of the uneven weighting due to the sampling design, mistakes in a handful of people may have large effects on undercount estimates. (See Section 4.)

There are many potential sources of errors. For instance, the number of matches can be erroneous because of mistakes made by the matching teams or because some of the data is missing, unreliable or fabricated. Persons moving in or out of the sample blocks after census day (April 1, 1990) can be an error source. Deciding just which housing units are in the designated blocks is not always simple, and mistakes can result in errors in the DSE estimates. The final imputation process is another potential error source. The P-studies are examined to see what evidence there is concerning the magnitudes of the errors and data quality. (See Section 5.)

Some error estimates can be obtained from the evaluation data aggregated to the national level and to the level of the 13 evaluation strata. The DSE national undercount estimate was 2.1%. The Census Bureau Total Error Report [P16] estimated that various errors led to an upward bias of 0.7% in the DSE estimate. Correcting the initial estimate for these errors gives an estimate of 1.4%.

After the Secretary's decision a coding error in processing the PES data was discovered that lowered the undercount estimate by 1,000,000 persons. The rematching of some suspect blocks led to a 250,000-person decrease. As a result, the Census Bureau low-

ered the DSE estimate to 1.6% with an estimated upward bias of 0.7% (Mulry, 1992b). Correcting for this bias drops the estimate to 0.9%. Our scrutiny of the P-studies showed some errors not included in the total error report that lower the national estimate to 0.4%. Other problems could lower the figure even more. (See Section 6.)

Thus, using the Census Bureau estimates, 55% of the DSE national undercount estimate is due to bad data or processing errors; our estimate is at least 80%. Also, there is an uneven distribution of errors. The corrections downward in the five minority evaluation strata are considerably larger than in the nonminority strata (see Section 6.3, Table 16). Some individual evaluation strata were affected more than others. For instance, the minority, central-city stratum in the South has a DSE undercount estimate of 5.7%. The corrected estimate is 1.5%.

Our error analysis does not always agree with the Census Bureau analysis, and the differences are pointed out in Section 6 and the Appendix. The most recent account of the Census Bureau analysis is Mulry and Spencer (1993). This is one of a collection of articles about the 1990 undercount published in the September 1993 issue of the *Journal of the American Statistical Association* (*JASA*). Note, however, that this article does not correct for the coding error or other errors found by the Census Bureau in late 1991 and early 1992 (see Hogan, 1993, in the same collection). Therefore the error estimates, the loss function analysis and conclusions in the article do not reflect current knowledge.

1.4 Conclusions

The PES data are not reliable enough to give accurate undercount estimates. To a substantial extent the 1991 DSE undercount estimates are artifacts reflecting data quality. The largest part of the original undercount estimate is due to bad data and processing error—80% on the national level. It is difficult to deduce how much more of the undercount estimate is similarly affected.

Because of the relatively small sample sizes in the evaluation data, results are broken out by the 13 evaluation strata rather than the 1,392 poststrata. Thus, it is not known what effect the error rates and decreases in the undercount estimates indicated by the evaluation data would have on the undercount estimates at the county, city and state level. However, the fact that 80% of the undercount estimate aggregated to the national level is due to poor data quality indicates that the DSE estimate state and local adjustments are largely reflections of bad data in an unpredictable pattern.

TABLE 1
PES household interview outcomes

Interview with household member	93.7%
Proxy interview	4.2%
Noninterview	1.6%
Out-of-scope	0.5%

The results of this study should not be taken to mean that I believe that the true 1990 undercount is as low as 0.4% or even 0.9%. My focus was on whether the 1990–1991 DSE estimate process produced reasonable estimates of the true undercount. To that, my answer is no; there were simply too many sources of error. The accuracy necessary in dozens of diverse areas to keep the total error within the requisite bounds was simply not attainable. It is not attainable now and probably not in the future.

2. PRODUCTION OF THE UNDERCOUNT ESTIMATES

The procedure that produced the raw DSE undercount estimates for the 1,392 poststrata can be put into four phases. The first, which we call the initial PES, is similar to many other sample surveys. Addresses are listed and interviewers go from one household to the next. The PES was carried out in July–August 1990.

The next three phases are unique: they consist of the matching, the follow-up interviews to obtain additional information on hard-to-match cases (November–December 1990) and more matching followed by imputation. Planning for the PES and the subsequent phases is discussed in Anolik (1990), Biemer and Stokes (1989), Childers et al. (1987), Hogan (1989), Hogan and Wolter (1988) and Wolter (1987a). Descriptions of operations and results are given in Department of Commerce (1991b, Section 4), in Hogan (1993), and in Undercount Steering Committee (1991).

2.1 The Initial PES

The Post Enumeration Survey covered 380,000 people in 169,000 households. These households were in 5,290 block clusters selected using a stratified random design. Woltman, Alberti and Moriarty (1988) summarize the sample design of the PES, and Hogan (1990) gives an overall description.

The first phase of the PES consisted of listing addresses in the designated blocks. Then interviewers covered the address lists. The questionnaire covered name, age, sex, race or ethnicity, owner or

renter, marital status, relation to head of household and address as of April 1, 1990 (census day). A summary of the outcomes of household interviews is given in Table 1 (Diffendal and Belin, 1991, Table 3.1).

"Proxy interviews" are interviews with nonhousehold members such as neighbors, apartment managers, landlords or other knowledgeable respondents. "Out-of-scopes" are persons that do not belong in the PES sample for a variety of reasons (i.e., home on a military leave, visiting for the weekend, duplicate record, etc.).

2.2 Matching

The persons found by the PES in the designated block clusters are referred to as the P-sample. The persons found by the census in the same block clusters constitute the E-sample. The next phase of the PES undercount estimation procedure consisted of attempting to match E-sample census records to P-sample PES records. The Census Bureau has done research on improving matching techniques for over a decade, and the methods used on the 1990 data were comprehensive and well rehearsed.

The matching was done in three phases. The first phase consisted of computer matching (Jaro, 1989). This matched 75% of the records. In the next phase, the records unmatched by the computer went through each of two independent tracks. One track consisted of two matching teams; the other, of one. In the third phase, an adjudication team assigned a match status code to all records on which the two previous tracks had disagreed. There was also a special team assigned to double-check match status codes in suspect data.

The match codes are complex and differ for the P- and E-samples. They fall into three basic categories. For the P-sample, these are match, nonmatch and unresolved. For the E-sample, these are correctly enumerated, erroneously enumerated and unresolved. Persons could be erroneously enumerated in the census for a variety of reasons; for example, if they were born after April 1, 1990, or died before this date, or if they were enumerated at the wrong address or were enumerated more than once.

We have not been able to find a single document giving a detailed but readable description of the matching rules, codes and procedures. What exists are lists of match codes and lengthy, detailed manuals of instructions for the matching teams. The best brief summaries are in Department of Commerce (1991b, Section 4) and in West, Corby and Van Nest (1989). A more extensive description of the similar matching procedures used in the 1988 rehearsal is in Childers and Hogan (1989a).

TABLE 2
PES follow-up household interview outcomes

	P-sample	E-sample
Interview with household member	81.6%	79.2%
Proxy interview	17.0%	19.6%
Noninterview or out-of-scope	1.4%	1.3%

2.3 The Production Follow-up

After the initial matching procedure, many persons were unresolved or unmatched. To get additional information about these cases, a follow-up survey was done. This included the following:

1. all people in the E-sample not matched to the P-sample;
2. all P-sample whole household nonmatches;
3. proxy interview P-sample partial household nonmatches.

A total of 47,000 households were sent to follow up [P2, Table 3.2]. The outcomes of household interviews [P2, Tables 6 and 7] are given in Table 2.

2.4 Imputing the Unresolved

After the new information gathered in the production follow-up was used in the matching procedure, 12,500 persons in the E- and P-samples remained unresolved [P1]. Statistical models were used to impute how many of the unresolved P-sample persons should be assigned as matches in each poststratum and how many of the unresolved E-sample persons should be assigned as correctly enumerated in each poststratum. These models are described in Belin et al. (1993) and Diffendal and Belin (1991).

Following the imputation procedures, the numbers in each poststratum needed to estimate the poststratum adjustment factor are on hand: the census count; the P-sample count; the number correctly enumerated by the census; and the number of matches. The output consists of the raw adjustment numbers for each poststratum (see Section 4.1).

3. THE EVALUATION DATA AND STUDIES

The Census Bureau planned and carried out an evaluation of the process leading to the DSE estimates. Sources of potential error were categorized, and projects were designed to gather and/or analyze data in order to estimate the magnitudes of the errors. The evaluation data gathering and analysis were mainly done following the completion of the

PES. The results are detailed in 22 evaluation project reports, referred to as the P-project reports.

Data for the evaluation projects (except [P13] and [P18]) come from three sources. The first consists of records from the quality control and quality assurance procedures that were ongoing during the PES and matching. The second is a rematching study carried on using a subsample of the P- and E-records but different matching teams. The third, and most important, is an evaluation follow-up (EFU) which reinterviewed a subsample of the P-sample and E-sample persons.

The narratives and tables of the P-project reports form the basis for most of the error analysis in this report. For this reason we give a more detailed overview of the evaluation and the P-project reports. This consists first of a listing of the reports and other associated references. Second, a description of the evaluation data sources is given. The results in the P-project reports are usually given by aggregation into 13 evaluation strata (the definitions of these strata are given in Section 3.4).

3.1 The Evaluation Studies

All project reports were obtained from the census. Various types of error are connected with evaluation project reports as follows:

- matching errors, P-sample—[P7], [P8];
- matching errors, E-sample—[P9a], [P10];
- interviewer fabrications—[P5], [P5a], [P6];
- census day address errors—[P4];
- missing data and imputation error—[P1], [P2], [P3];
- incorrect address coding—[P11];
- correlation bias—[P13];
- errors due to late census data—[P18];
- total error summary—[P16].

The full references for the P-project reports are listed at the end of the reference section. Recently, extracts from some of these reports have appeared in proceedings volumes or journals: [P2] in Gbur (1991a); [P4] in West, Mulry, Palmer and Petrik (1991); [P5] in Tremblay, Stokes and Greenberg (1991); [P7] and [P10] in Davis, Mulry, Palmer and Biemer (1991); and [P16] in Mulry and Spencer (1991, 1993). The published articles omit much of the detail given in the parent P-project reports. An overall view of the purposes and methods of evaluation is given in Hogan (1989). Section 4 of Department of Commerce (1991b) contains executive summaries of all of the evaluation project reports, but there is no single document that gives full descriptions of all of the evaluation projects.

The Census Bureau's intention was to make the P-project reports the definitive collection of information regarding the evaluation of the PES and the subsequent undercount estimates. Almost all of the material in this report is drawn from these evaluation studies. Supplementary data about the P-sample reinterviews in the EFU and the effects of the coding error discovered in late 1991 were supplied by the Census Bureau.

3.2 Quality Control and Assurance Data

During the interviewing phase of the PES, an ongoing quality control operation was in place to confirm that the PES interviewers visited the correct households and conducted the interviews according to survey procedures, and to conduct reinterviews of questionable work. Overall, about 35% of the P-sample was reinterviewed by phone or personal visits in this quality control operation [P5]. This data was used in [P5] to estimate the extent of undetected fabricated interviews.

In production matching, at least three and up to five different teams were involved in the final decision on each record not matched by computer. Logs were kept on the intermediate decisions. These were used in [P8] to quantify the ranges of disagreement between the various teams. This data source is referred to as the quality assurance results.

3.3 The Matching Error Study

In the Matching Error Study, at each processing center a sample of P- and E-records were selected and rematched. According to the bureau, the rematching was done more carefully than the production matching. Report [P10] notes that in the rematching "all match codes... were reviewed by MRS, the most highly trained matching personnel." In production matching the matching review specialists (MRS's) reviewed only a fraction of the codes assigned. Rematching was undertaken for 71,000 P-sample cases.

The rematching was not independent of the original matching. The rematch teams had available to them all of the match codes assigned in the production matching. The Matching Error Study estimates what happens when the matching is repeated using the same matching rules, the same computer algorithm and with the previous matching information available to the rematch team. Brief descriptions are given in reports [P7] and [P10].

3.4 Evaluation Follow-up (EFU)

In this phase, carried out in February 1991, a subsample of households were reinterviewed to get additional information about Census Day address errors.

TABLE 3
EFU household interview outcomes

	P-sample	E-sample
Interview with household member	87.7%	85.2%
Proxy interview	10.4%	13.4%
Noninterview	1.9%	1.4%

noninterviews, fabricated data, imputations and so on. The interviewers had available the records of the past interviews. About 11,000 households were reinterviewed, and data collected for 27,000 persons ([P4], [P9a]). The design of the EFU is described in Reports [P3], [P4], [P5a] and [P9a]. The interview results are shown in Table 3 [P2, Tables 8, 9 and 11].

The interviewers used were more experienced than those used in the PES. To quote report [P3], "A staff composed of only current survey interviewers was used for the EFU interviewing. The interviewers hired and trained for the PES and the Census were primarily temporary employees. . . ." Report [P4] states that the team who matched the data from the EFU "consisted of Matching Technicians (Techs) and Matching Review Specialists (MRS), the highest level and most trained of the matchers from the PES."

3.5 The Evaluation Strata

Because the sample sizes in the evaluations, particularly in the EFU, are small when allocated down to the 1,392 poststrata, most of the information in the evaluation reports is aggregated to the 13 evaluation strata defined in Table 4. In particular, these 13 strata are used in [P16], the total error report. These strata are defined by geographic region, central city versus non–central city and minority versus nonminority. We will categorize information either using these same 13 strata or by aggregating to the national level. Table 4 lists the raw DSE undercount estimates for each evaluation stratum [P16]. They have considerable variation, being large in the five minority strata (1, 3, 5, 8 and 11) and smaller in the nonminority strata. Secondary variations appear by region and place type.

4. EFFECTS OF SMALL PES ERRORS

The two foregoing sections have described the production of the PES undercount estimates and described the evaluation data and reports, and they form the backdrop for our error analysis. However, to assess the effects of errors in the PES we need to cross another bridge. It turns out that small errors in the PES can lead to large errors in the undercount estimates. There are two reasons for this.

4.1 The Equation for Computing Undercount

Suppose that in a certain population pool, the census count was N_C. If N_{EE} persons were erroneously enumerated, then $N_{CE} = N_C - N_{EE}$ is the number of persons correctly enumerated in the pool. The persons enumerated by the census form the E-sample.

Later, another survey of the same pool enumerates N_P persons (the persons in the P-sample) and, of these, N_M can be matched to the E-sample. Then the dual system estimate [except for a technical adjustment in how N_{CE} is computed (Mulry and Spencer (1991)] for the total number of persons in the pool is defined by

$$N_{DSE} = N_{CE} \times (N_P/N_M).$$

The undercount estimate is $N_{DSE} - N_C$ and the percent undercount is this difference expressed as a percent of N_{DSE}. Suppose, for example, that $N_C = 1.020$, but it is determined that $N_{EE} = 10$ so N_{CE} is 1,010. Suppose also that the number of people found in the second survey is $N_P = 1{,}000$, and that there are $N_M = 970$ matches. Then $N_{DSE} = 1.041$ and the undercount estimate is 2.0%.

All three numbers N_{CE}, N_P and N_M are estimates subject to error. In particular, N_P is subject to some of the same errors as N_C. The undercount estimates are sensitive to errors in N_P and N_M, less so to errors in N_{EE}. For instance, if N_P decreases by 0.5% to 995, and N_M increases by 0.5% to 975, then N_{DSE} decreases to 1.031 and the undercount estimate to 1.0%. On the other hand, if the two changes go 0.5% in the other direction, then the undercount estimate increases to 3.1%. Thus we have the following:

> Two 0.5% errors in estimating N_M and N_P can result in a 50% error in the undercount estimate.

4.2 The Weighting Effect

The effect of errors is further complicated by disparities in weighting. To get the DSE population estimate in a poststratum, the numbers used in the formula above are weighted up from numbers in the sample blocks. The 5,290 block clusters in the PES were not randomly selected from all U.S. block clusters. Instead they were randomly selected from predefined sampling strata.

On the average, one person in the PES corresponds to 650 in the U.S. population, but because of the stratification and nonresponse this can be uneven. There are some block clusters where one person weights up

TABLE 4
Evaluation strata and their estimated undercount

Location	Place type	Race–ethnicity	DSE undercount (%)
1 Northeast	Central city	Minority	6.83
2 Northeast	Central city	Nonminority	−0.75
3 United States	Non–central city	Minority	5.43
4 Northeast	Non–central city	Nonminority	0.01
5 South	Central city	Minority	5.68
6 South	Central city	Nonminority	1.94
7 South	Non–central city	Nonminority	1.82
8 Midwest	Central city	Minority	3.97
9 Midwest	Central city	Nonminority	1.28
10 Midwest	Non–central city	Nonminority	0.39
11 West	Central city	Minority	6.14
12 West	Central city	Nonminority	2.13
13 West	Non–central city	Nonminority + Indian	1.84

to over 10,000 in the U.S. population. Mistakes in a handful of people in such block clusters would be highly magnified. For example, in a certain block cluster a single unmatched family consisting of five persons contributed 45,000 to the undercount estimate.

Another illustration consists of two PES block clusters in which a low match rate increased the national undercount estimate by almost one million people. There were 648 persons in the two block clusters. The problem is discussed in a 1991 Census Bureau memorandum (Hogan, 1991). The cause was investigated and determined as faulty census geocoding. The two block clusters had their influence downweighted so they contribute only 150,000 to the estimated undercount.

The question is not whether the PES was accurate compared to other sample surveys or whether the matching was accurate compared to other matching projects, but whether they were accurate in terms of the resulting undercount estimates. Small errors in estimating N_P and N_M lead to large errors in undercount estimates. Errors in a small number of persons can be disproportionately magnified by the weighting.

5. ERROR SOURCES AND DATA QUALITY

Estimates for some types of errors in the national and in the 13 evaluation strata undercount estimates can be based on the evaluation data. These are covered in the next section and show that well over half of the national DSE undercount estimate is due to bad data.

However, there is a considerable amount of other information in the P-studies that is relevant to data quality issues and potential error sources in the DSE estimation process. This information is diverse and comes from many different reports. Putting these different pieces together gives insight on the difficulty of the DSE estimation process. The information available covers the following:

- matching errors;
- fabrications;
- census day address errors;
- geocoding errors;
- unreliable interviews;
- missing data;
- imputation.

Because many different facets of the data quality issue are covered, this section contains a fair amount of detail. At the end, we summarize and look at the implications.

5.1 Matching

Matching records from two different files of human data with differing names, ages, missing sex or race identifiers and different addresses can involve difficult decisions. In the present situation, one file consisted of the PES records. The other file consisted of the census records. The evaluation material concerning matching comes from a rematching study, from ongoing quality assurance records and from the reinterviews. Error rates for matching weighted to the nation and the 13 evaluation strata were derived from the rematching study and are given in [P16].

In this section we examine disagreement rates as totaled over individual cases. The marginal disagreement rates can be much smaller. For instance, weighted to the total population, the number of P-sample matches in the rematching study differs from the number in production matching by only 0.18%, while the disagreement rate is 1.8%.

TABLE 5
P-sample match–rematch disagreement rates

Whole sample	1.8%
Unresolved group	23.8%

TABLE 6
E-sample match-rematch disagreement rates

Whole sample	2.1%
Unresolved group	35.4%

However, similarities in highly aggregated marginal totals cannot be used to infer that match–rematch differences are small at the level of the 1,392 poststrata. Substantial poststrata differences can "average out" to small differences in marginal totals at aggregated levels. For instance, in the 13 evaluation strata, on average the number of P-sample matches in the rematching differs from the number in the production matching by 0.41%. This is over double the 0.18% difference found using the aggregation to the total population. We can expect larger differences at the state and city level.

5.1.1 *Rematching study data.* The primary sources of information about matching errors are the rematching study reports [P7] (P-sample) and [P10] (E-sample). In the study, over 70,000 E- and P-sample persons were rematched by more experienced matching teams and compared with the original matching. One indication of matching accuracy is in the disagreement rates between the category assigned by the production matching and that assigned in the rematch.

The disagreement rates for the P-sample persons are obtained as follows. In the production matching, each person in the P-sample was categorized as a match, nonmatch, unresolved or out-of-scope. The production out-of-scope persons were assigned zero weight in the sample and their numbers do not appear in the tables. The rematching did a similar categorization. The total disagreement rate is the number of people categorized differently in the two matching procedures as a percentage of the total number categorized.

Tables 16–28 of report [P7] give cross-tabulations of counts in the production versus rematch categories in each evaluation stratum with numbers weighted to the total population. The disagreement rate was computed for each evaluation stratum and averaged over the strata to give the first row of Table 5.

The second row of Table 5 gives the disagreement rate between the match and rematch teams on the membership of the unresolved category. This number is computed as the percentage of all persons categorized as unresolved by the rematch team that are categorized differently in the production matching. These numbers are also weighted to the total population and averaged across evaluation strata. Cases in the E-sample get put into three categories: correctly enumerated; erroneously enumerated; and unresolved. The primary evidence concerning errors in this categorization comes from the rematching study report [P10]. Table 6 summarizes the extent of the disagreement calculated the same way as for Table 5 and based on [P7, Tables 42–54].

TABLE 7
SMG1–SMG2 *disagreement rates*

Matched	10.7%
Not matched	6.6%
Unresolved	31.2%

5.1.2 *Quality assurance results.* Another source of information concerning disagreement in production matching is given in report [P8]. In the production matching process, the first phase was computer matching. This matched 75% of the cases. After the computer matching and some clerical matching, two teams (SMG1 and SMG2) worked, independently of each other, on the cases not matched by the computer.

Report [P8, Table 3.1] gives data concerning the disagreement rate between these two teams. We give the results since they comprise the only data available where two teams worked independently on matching the same cases. The breakdown in Table 7 gives the percentage disagreement on the major match categories. Overall, the disagreement rate was about 10% on the cases handled by the two teams, that is, the cases not matched by the computer. The disagreement rates were computed taking the SMG1 results as the base, that is, 10.7% of the cases categorized as "Match" by the SMG1 team were put into other categories by the SMG2 team.

5.2 Fabrications in the P-Sample

From [P5a], "Interviewers may fabricate people in the P-sample housing units. The creation of fictitious individuals has the effect of decreasing the PES match rate causing the estimate of coverage error [undercount] to be too large." Also, the effect is differential. The general belief is that the more difficult the area is to survey, the higher the fabrication rate

(Stokes and Jones, 1989). Thus, one expects high fabrication rates in minority, central-city areas—exactly those areas which have the highest estimated undercounts.

There are three studies estimating the extent of fabrications in the PES. The first is the evaluation field study; the second uses the data from quality control; and the third is the [P6] project, which attempted to quantify fabrication rates by looking at interviewers with unusually high nonmatch rates. The estimates of error due to fabrications used in [P16] were based on the P-sample data in the EFU.

Out of 14,444 cases in the P-sample EFU data, 13 were identified as fabrications, including 12 blacks [P5a]. These 13 weight to the national total as 0.03% of the cases. Based on the [P16] estimates, this 0.03% rate inflated the DSE national undercount estimate by 50,000 persons (see Section 6.1, Table 15). Thus, at a similar scaling, an undetected fabrication rate of 1% would have inflated the DSE undercount estimate by 1,650,000.

The EFU fabrication estimate may be low. To quote from report [P5a], "The data for the study were collected in the EFU which was not designed specifically to detect fabrication.... Thus, it is possible that the EFU did not identify more cases as fictitious because there was not enough new and additional interviewer information to establish that the cases were fictitious in the PES."

Report [P5] gives estimates of the fabrication rates based on quality control data gathered during the PES. During operations, quality control (QC) found that 0.26% of the household interviews were fabricated. Their estimate is that 0.06% of the remaining cases on a national level are fabrications. It is difficult to know how firm the basis is for this estimate. In particular, report [P5, page 4] states: "A limitation for this project (estimation of the fabrication rates) is the incompleteness and inconsistency of the QC data sources across RCC/ROs" (where RO is regional offices, and RCC is census centers).

Report [P6] tries to estimate the fabrication rate by identifying interviewers with an unusually high nonmatch rates. Of these interviewers, only 38% were identified as problem interviewers by quality control procedures. The report [P6, page 12] states: "It has been the speculation that in data collections such as the Census Bureau's current surveys between $\frac{1}{2}$ and $1\frac{1}{2}$% of the interviews are fabricated (Biemer and Stokes, 1989). The results from this study indicate that in an undertaking such as the PES the percentage is higher. Here, with the exception of two regions, the range was from 2.1 to 5.97%. In two regions, the percentage went as high as 7.79 and 8.79."

The report concludes [P6, page 15] that "Overall, between .9 and 6.5% of the interviewers were found to have high nonmatch rates. This compares favorably with the expectation that between 2 to 5% of interviewers are dishonest in their data collection." The estimates in [P6] are based on fitting mathematical models and on assuming that high nonmatch rates for an interviewer as compared to interviewers in neighboring blocks is a strong indication of fabrication. There is no data available to verify these assumptions.

At present, the effects of P-sample fabrication on the DSE estimates of undercount are difficult to quantify. The QC and EFU estimates seem low for reasons given above. The [P6] estimates seem high. The potential range is large, going from 0.03% to 8.79%. On the E-sample side, fabrication is treated as a component of erroneous enumeration. Report [P9a], which uses the E-sample EFU data to estimate errors in the erroneously enumerated counts, does not break out E-sample fabrication separately.

5.3 Census Day Address Errors

One of the most difficult sources of error to pin down is errors in the location of residence on census day (April 1, 1990). People who moved into the sample blocks after census day would not have been enumerated by the census as living in the sample blocks. If they appear in the PES sample, the nonmatch rate is erroneously inflated.

The EFU study found 334 P-sample respondents who were originally classified as nonmovers but new information revealed as after census day inmovers. Weighted to a national level, this represents 1,410,000 persons [P4]. These cases were then rematched using the new information. The computations in report [P16] show that this resulted in a decrease of 811,000 people in the estimated national undercount.

In total (nationally weighted) 41% of the newly discovered movers were originally matched in the PES. The implication is that the matching process incorrectly matched people not resident in the area on census day to people counted in the area by the census on the census day. Although the absolute number of these people is small, they weight up to 580,313 nationally.

5.4 Geocoding Errors

Geocoding, in this context, is the assignment of housing unit addresses to the selected sample blocks. The census makes one such assignment and the PES another. Errors in geocoding affect undercount es-

imates. Suppose the PES erroneously assigned a housing unit to one of the sample blocks. Then the P-sample persons in the unit could probably not be matched against any E-sample persons, and the nonmatch rate would be inflated. These errors are the subject of [P11], and the following discussion is based on that report.

In an effort to minimize the effects of geocoding errors, search areas consisting of one or two rings of blocks surrounding a sample block were defined. If an E-sample case cannot be matched to a P-sample case in its block, then it goes to follow-up and the interviewer is instructed to draw a sketch of the location of the housing unit. This sketch is then used to get a geocoding. If the location is in the sample block or search area and the enumeration is correct in other respects, the case is classified as correctly enumerated (CE); if not, as erroneously enumerated (EE). For a P-sample case in the block with no match in the block, a match in the search area is sought. If one exists, the case is put into match status.

Overall, 4.08% of the P-sample was matched to the Census through geocoding to the surrounding blocks. However, only 2.29% of the E-sample got CE status in surrounding blocks. The difference, weighted to the national level, is "an approximate excess of 4,296,000 in the P-sample population" [P11, Attachment]. The implication of this result is that if the surrounding blocks search had not been done, then geocoding errors would have caused a doubling of the DSE national estimated undercount, to over 4%. On the other hand, using a larger search area might well have produced a much lower undercount estimate.

The EFU included some of the E-sample households that had sketches made of their locations and geocoded in the production follow-up. For these households, the EFU interviewers made a second location sketch and a second geocoding was performed. Putting the geocodings into three categories (located inside a sample block, located in a search area, located outside of both) there is a 20% disagreement rate between the two geocodings [P11, Table 3.1].

5.5 Reliability of Interview Data

Some information reflecting PES interview reliability can be obtained from the EFU. Of the EFU P-sample interviews 4% were rejected as being unreliable [P4]. The EFU interviewers were regular census employees and more experienced than the PES interviewers. Thus, the 4% rejection rate seems surprisingly high. Of the rejected interviews 58% were with family members, 13% with neighbors, 13% with the apartment manager or landlord and 16% other.

The rejection rate is higher for minorities and central cities. Report [P4, Table 5.1.4] gives the ratio of the number of rejected interviews to the total number of EFU interviews broken down by evaluation strata. It is generally large where the estimated undercount is large. An analysis of the implications of this 4% rejection rate on the accuracy of the DSE estimates has not been carried out.

TABLE 8
Percentage change in match status using new EFU information

Correctly enumerated	7.2%
Erroneously enumerated	32.8%

The EFU collected data for 11,992 E-sample persons [P9a]. The new interview information was given to the matching team along with the PES production matching information. Match status could be changed from the PES production match status only if new, relevant and reliable information regarding a case was present in the EFU interview. The cross-tabulated data, weighted to the nation, comparing status before and after use of the EFU information is given in [P9a, Table 35]. Changes are summarized in Table 8.

Over 2,000,000 persons classified as "correctly enumerated" in the PES became classified either as "erroneously enumerated" or "unresolved" after use of EFU data. Over 1,600,000 persons originally in the "erroneously enumerated" category moved to "correctly enumerated" or "unresolved." The implication is that a substantial fraction of the interview data did not give reliable results in the original PES matching. The analogous data for P-sample persons does not appear in any of the P-reports.

5.6 Missing Data

Interviews can result in missing information for some of the people in the household. Missing data can affect the PES estimates in two ways. First, it can make matching more difficult and error-prone. Second, assignment of persons to a poststratum depends on some questionnaire characteristics. If these are missing, the person may be assigned to the wrong poststratum.

Report [P2] contains relevant data, weighted to the nation. Table 9, based on [P2, Table 3.3] gives the percentage missing for some PES and census questionnaire characteristics. The percentage of missing data is highest in those strata where the estimated undercount is highest. This property is true not only for race, but for all other characteristics. Table 10 gives the correlations between percentage of missing characteristics and percentage undercount over the 13 evaluation strata [P2, Table 3.4].

TABLE 9
Percentage of missing data

Characteristic	P-sample	E-sample
Race*	2.5	11.8
Age	0.7	2.4
Sex	0.5	1.0
Tenure	2.3	2.5

*Report [P2] states "the race variable . . . is a combination of race and Hispanic origin."

TABLE 10
Correlations of undercount estimates with percentage of missing data

Characteristic	P-sample	E-sample
Tenure	0.5	0.8
Sex	0.5	0.8
Age	0.6	0.7
Race	0.7	0.8

After matching, the missing data is filled in by a hot-deck imputation algorithm (Diffendal and Belin, 1991, Appendix 2). This serves two purposes. One is to allocate the persons to poststrata. The second is that complete information is necessary for the imputation of the unresolved persons into match categories. This latter procedure is discussed in the next section.

5.7 Imputation of the Unresolved Cases

At the end of production matching, there were 5,359 unresolved persons in the E-sample, and 7,156 unresolved persons in the P-sample [P1]. These 12,515 are among the persons having the most incomplete and least reliable data in the PES and the census. Over two-thirds of the unresolved people in the PES sample are after census day movers (52%) or possible movers (16%), and 45% are minority (Diffendal and Belin, 1991, Table 3.6). In the E-sample 32% have unresolved geocoding [P1].

Although the unresolved account for only 1.6% of the total combined PES and census samples, the estimates of the undercount strongly depend on what category they are finally assigned. If all unresolved PES sample cases are assumed to be matches and all census sample unresolved assumed to be erroneously counted by the census, then the DSE national estimate is 1,000,000 less than the census. At the other extreme, the DSE estimate is 9,000,000 more than the census.

The Census Bureau handles the unresolved P-sample by using a complex hierarchical logistic regression model that depends on estimating coefficients for dozens of variables (for details see Belin et al., 1993, and Diffendal and Belin, 1991). For each unresolved person in the P-sample, the P-sample model is used to compute a match probability. In each poststratum, the number of matches is increased by the sum of the match probabilities of the P-sample unresolved persons in the stratum.

The E-sample unresolved are treated using a different hierarchical logistic regression model (see Belin et al., 1993, and Diffendal and Belin, 1991). For each unresolved person in the E-sample, the E-sample model is used to compute a correct enumeration probability. Then the number of correctly enumerated people in each poststratum is increased by the sum of the correct enumeration probabilities of the E-sample unresolved people in the stratum. Neither model has been tested prior to their use on the 1990 PES data.

The outputs of these models have a significant and differential effect on the undercount estimates. In the five minority evaluation strata, imputation adds, on the average, 1.2% to the undercount estimates. For instance, in evaluation stratum 8, the imputation increases the undercount estimate from 2.8% to 4.0%, and in stratum 11, from 4.2% to 6.1%. In the nonminority evaluation strata, imputation adds an average of 0.6% to the undercount estimates. Thus, the imputation modeling is a significant contributor to the larger estimated undercounts in the minority strata. (These numbers were computed from Ericksen, Estrada, Tukey and Wolter, 1991, Table 11, Appendix C).

Significant proportions of key variables used in the models (such as age, sex, race, etc.) have been previously imputed. Of the P-sample unresolveds, 28% have at least one characteristic missing in their data. In the E-sample the percentage is 38% (Diffendal and Belin, 1991, pages 13 and 26). The coefficients of the variables are estimated using the data from the PES and the PES follow-up. This involves the further assumption that the final unresolved group is similar in nature to the persons resolved in the PES follow-up.

The only data available for assessment of the models comes from the EFU. After the imputation models were used to assign match and correct enumeration probabilities to unresolved PES persons, a subsample of these people were reinterviewed in the EFU. The new information was sent back to the matching teams and rematching was carried out. The results, weighted to the nation, are given in Table 11 (from [P3, Tables 3.1 and 3.2]).

The imputation models predicted that 42% of all of the P-sample persons in Table 11 would be matches and that 78% of all of the E-sample persons would be correctly enumerated [P3, Tables 3.3 and 3.4, nationally weighted]. The large proportion of the cases

TABLE 11
Rematch results for unresolved groups

P-sample unresolved		E-sample unresolved	
Match	12%	Correct enumeration	62%
Nonmatch	27%	Erroneous enumeration	17%
Unresolved	59%	Unresolved	21%
Out-of-scope	3%		

left unresolved, particularly in the P-sample, makes conclusions uncertain.

In computing and using the match probability, the assumption is that a high computed probability of a match implies that the person is very likely a true match. Thus, one would expect that, for a P-sample unresolved person with a high computed match probability, additional information would show that the person is indeed a match. This can be examined by looking at the new match status (Table 12) resulting from the EFU reinterview information for different ranges of computed match probabilities (from [P3, Table 3.1]).

As the match probabilities increase, the proportion of resolved cases that result in matches increases, but so does the proportion of unresolved cases. Report [P3], in summary of Table 11, states: "Thus, for P-sample persons, the imputation process is consistent with EFU results. However, the high percentage of unresolved persons in the EFU (58.55 percent) may limit the utility of this result."

Table 13 for the new match status for the E-sample unresolved versus the imputed correct enumeration probability is taken from [P3, Table 3.2]. There is no evidence here of an association between the probabilities computed by the imputation model and the enumeration status as determined by the EFU reinterview information.

The most important thing about estimating the undercount is not its total magnitude, but its differential effect on the 50 states and on thousands of counties and cities. These differential effects are estimated using the allocation of each PES-surveyed person into one of 1,392 poststratum. On the average there are 270 persons and 9 unresolved cases per poststratum. How these nine cases are resolved is an important determinant of whether there will be a high or low estimated undercount for that poststratum.

Evidence for the accuracy of the imputation models at more disaggregated levels is not available. Report [P3] does not give a table of the imputation results by evaluation strata. In the total error report [P16], the imputation estimates in each of the 13 evaluation strata are assigned zero bias. No explanation is given.

Other ways of looking at the imputation results are presented in Belin et al. (1993). If attention is confined to the 316 P-sample persons resolved in the EFU (first two numbers of first column, Table 11), the model gives an accurate prediction of the proportion of matches (Belin et al., 1993, Table 3, page 1157). The predicted proportion of matches categorized by imputed match probabilities (as in Table 12, first two rows) is not as accurate. The results in Belin et al. (1993), are not comparable to Tables 11 and 12 because weighting to the nation is not used and the categorization is different.

There is another evaluation report that deals with the imputation models ([P1], "Analysis of reasonable alternatives"). This work is intended as a sensitivity analysis and not as an assessment of accuracy.

5.8 Summary of Data Quality Evidence

It was noted in Section 4 that several errors of the size of $\frac{1}{2}$% could have a large effect in the national undercount estimates. The analysis of the evaluation data not only shows sources of error potentially larger than $\frac{1}{2}$%, but also many such sources, the following in particular:

1. In the PES, 6.3% of the interviews were with other than household members. In the PES follow-up, 19.2% were with other than family members.
2. The percentage disagreements in the P-sample rematching study are all well above 1%, and the average over the evaluation strata is 1.8%. In the E-sample, the match code disagreement averaged over the evaluation strata is 2.1%. In the E- and P-sample, the disagreement on the makeup of the unresolved groups averaged 27.6%.
3. The fact that 4% of the EFU interviews were rejected as being unreliable is disturbing, since the EFU interviewers were more experienced than the PES interviewers. The implication of 4% unreliable information in the EFU needs to be considered in judging the reliability of the PES data.
4. Substantial changes (7 and 33%) in enumeration status assigned in production matching resulted when a rematching was done using the EFU E-sample reinterview data. This is a reflection of the reliability of the interview data used in production matching.
5. The imputation models used to assign unresolved persons into match or nonmatch, correctly enumerated or not, are previously untested, and the EFU evidence concerning their performance is inconclusive. There is no evidence concerning

TABLE 12
EFU rematch results versus imputed match probabilities

New match status	Imputed match probability			
	0–25%	25–50%	50–75%	75–100%
Match (%)	6	13	17	15
Nonmatch (%)	44	26	15	7
Unresolved (%)	47	57	66	77
Out-of-scope (%)	3	4	2	1

TABLE 13
EFU rematch results versus imputed correct enumeration probabilities

New match status	Imputed correct enumeration probability		
	0–50%	50–75%	75–100%
Correct enumeration (%)	67	50	65
Erroneous enumeration (%)	10	26	15
Unresolved (%)	23	24	20

TABLE 14
Correlations with DSE undercount estimates

Percent unresolved P-sample	0.7
Percent unresolved E-sample	0.8
P plus E match–rematch disagreements	0.6
Missing data (average)	0.7
Rejected EFU interviews (P plus E)	0.5

TABLE 15
Decreases in the DSE undercount estimates due to the evaluation data

Number	Reason
553,000	P-sample rematching*
811,000	Census day address errors*
50,000	Fabrications*
624,000	E-sample rematching*
−473,000	E-sample reinterview*
290,000	Ratio estimator bias*
183,000	Late late census data
164,000	New out-of-scopes in rematch
358,000	New out-of-scopes in reinterview
537,000	P-sample reinterview
128,000	Reinterview of noninterviews
1,018,000	Computer coding error
Total 4,243,000	

accuracy at the poststratum level, or even at the level of the 13 evaluation strata.

There are other indications of serious errors: the EFU found 334 persons that moved into the area after census day but were not identified as inmovers by the PES. This number weights up nationally to 1,410,000 movers not correctly identified by the PES. The EFU had a 20% geocoding disagreement rate with the PES follow-up. Reported P-sample fabrication rates may be significant underestimates.

Correlations between the undercount estimates and data quality indicators computed over the 13 evaluation strata are shown in Table 14.

The correlations of the undercount estimates with measures of bad data indicate that it is difficult to gauge what is being measured by the undercount estimates. The data quality is worst where the undercount estimates are the highest—in the minority strata. One conjecture may be that the correlations of estimated undercounts with bad-data measures show that where there are large amounts of bad data there are also large real undercounts. However, the difficulty is that we do not know how much of the estimate is due to bad data. Section 6 gives evidence on this issue.

6. CHANGES IN THE DSE ESTIMATES INDICATED BY EVALUATION DATA

6.1 Summary of Decreases in Undercount

The original national DSE undercount estimate is 2.1%, or 5,275,000 persons. All through the evaluations, as more experienced personnel were used to collect data or to rematch, it was seen that the original DSE undercount estimates were too high. Report [P16] attempts to give an overview and summary of

all DSE errors as estimated by the evaluation studies. In the last two years, studies of the DSE estimate errors have been published (Hogan, 1990; Mulry and Spencer, 1991, 1993). These are based, essentially, on the [P16] analysis.

There are omissions in [P16], and Table 15 gives a more comprehensive summary using both the applicable parts of the [P16] report and data gathered from the other P-projects as part of this report. The numbers given are the *decreases* in the DSE national undercount estimate indicated by the evaluation data, and their sources will be discussed below.

As a result of these decreases, the corrected undercount estimate is 1,032,000, or 0.4%. The corrections reduce the estimated undercount to about one-fifth of the original DSE value.

The first six entries in Table 15, marked with an asterisk, are listed in report [P16] and are computed from the data in Tables 1–13 of [P16]. Each of these tables lists, for each evaluation stratum, the numbers (in the column labeled "mean") that get substituted into the formula at the bottom of page 5 of [P16] to correct the DSE estimate. This was done, one error at a time, in the order in which they were listed in the [P16] tables. The results were then added across evaluation strata to give the tabulated results.

Report [P16] gives a second set of undercount estimates that includes an additional error source called model bias, more commonly known as correlation bias. This is the bias ascribed to the existence of persons unreachable by any survey. However, because these bias estimates are (and must be) based on highly speculative assumptions and have only a tenuous connection with any data, they are not included in our discussion.

The results of this paper were circulated to the Census Bureau in March, 1992. In the reviews there was agreement with the numbers given in Table 15 except, perhaps, in two areas. The first is the question of what is included in "census day address error." The second is the possibility that there are compensating factors to the "new out-of-scope" errors. The issues not resolved with the Census Bureau are discussed in the Appendix.

Using the bias estimate in the original version of [P16] drops the undercount estimate from 2.1% down to 1.4%. A later (June 1992) total error report (Mulry, 1992b) using different evaluation strata states the DSE as 1.6% with an upward bias of 0.7%, leading to a bias corrected estimate of 0.9%. Thus, the Census Bureau has come about halfway toward the estimate given in this paper. Using their current estimates, 55% of the original DSE undercount is due to bad data. Our estimate is 80%.

6.2 Error Sources Not Included in [P16]

There are six sources of error listed above which are not included in the [P16] report.

6.2.1 *Late late census data.* Some census data came in after the DSE estimates were computed. Because of time constraints, a compromise procedure was worked out which used only part of the late census data. In the blocks most affected, the late census data was matched to the P-sample data and the DSE estimates revised accordingly. Report [P18] estimates that if all of the late census data had been used, the DSE undercount estimate would be reduced by 183,000 people. The relevant descriptions and tables are contained in report [P18].

6.2.2 *New out-of-scopes.* P-sample out-of-scopes are cases that do not belong in the P-sample. They are out-of-scope for reasons such as being duplicate records, fictitious records, wrong addresses and so on, and they should be subtracted out of the number of persons in the P-sample.

Both in the P-sample rematch study and in the evaluation P-sample reinterviews, many cases that were originally classified as nonmatches were reclassified as out-of-scope. The estimated decrease in the undercount was obtained by decreasing the size of the weighted P-sample by the weighted number of persons switched from nonmatch to out-of-scope, and then recomputing the DSE estimate.

The out-of-scope P-sample corrections come from two sources. The data from the rematch study is taken from report [P7]. The reinterview data does not appear in the evaluation reports and was supplied by the Census Bureau upon request. To avoid overlap with census day address error analysis, the reinterview data used does not include the inmovers reported on in [P4].

6.2.3 *P-sample reinterview.* In the evaluation, the EFU P-sample reinterview persons were rematched using the new information. The changes in match status were given in tables similar to the tables in the P-sample rematch study. The number of new matches in each evaluation stratum was computed using the same method as the Census Bureau used to compute the number of new matches from the tables in the P-sample rematch report [P7]. The data used was supplied by the Census Bureau and does not include inmovers.

6.2.4 *Noninterview error.* Some households that were treated by the PES as noninterviews (refusals, no one home, etc.) and adjusted for by being

"weighted out" were revisited in the EFU. The EFU was successful in getting interviews from 75% of these households. This new data was then supplied to the matching teams. As a result, report [P3] states: "At the national level, an estimated 102,403 more matches than are indirectly added by the noninterview adjustment would be added to the PES total." The relevant data is in report [P3], which gives, by evaluation stratum, the increased number of matches.

6.2.5 *Computer coding error.* In late 1991 the Census Bureau found an error in its computer code which resulted in sometimes classifying E-sample persons as correctly enumerated when they should have been classified as erroneously enumerated (Hogan, 1993). Correcting this error gave a revised undercount estimate slightly over 1,000,000 lower than the original estimate. The data on the effects of the coding error by evaluation stratum were transmitted to me in February 1992 by the Census Bureau. The numbers given in the June 1992 error report (Mulry, 1992b) indicate that only a minor portion of the error overlaps with other sources of error treated in [P16].

6.3 Decreases by Evaluation Stratum

Table 16 gives the distribution of the undercount changes by evaluation strata. The first column is the original DSE estimate undercount estimate. The second column gives the undercount estimates corrected as shown in [P16]. The third column gives the estimates corrected for the additional evaluation data noted above, and the fourth column gives the total change downward.

One trend seen in Table 16 is that the largest decreases occur in those strata having the highest original estimated undercount. In the five minority strata, the average decrease is 2.7%. In the nonminority strata, the average decrease is 1.3%.

6.4 Potential for Further Decreases and Changes

Even though the evaluation evidence cuts the undercount estimate by 80%, the available figures are probably conservative. The dependence in the rematch study tends to minimize discrepancies. The real fabrication rate may be larger than that used [P16]. The effect of the imputation models is largely unknown. Besides these, there are other potential changes not previously discussed and not included in Table 15.

(1) In the P-sample rematching and reinterview, 413,000 persons switched from an original classification of unresolved to out-of-scope. Many of these persons were imputed as nonmatches in the DSE estimates and should be treated as switching from nonmatch to out-of-scope. Using the best available estimates of the numbers of imputed nonmatches among these switchers lowers the undercount estimate by an additional 210,000. Most of the decrease affects the five minority evaluation strata, with their average undercount estimate going from 2.9% to 2.7%. The average undercount estimate in the nonminority strata stays about the same.

(2) The final undercount estimate for evaluation stratum 12 seems odd. Even though it is a nonminority stratum, it winds up having the third highest estimated undercount (3.1%). This is higher than three of the minority strata and is over twice as large as any other nonminority stratum. This may be due to a mistake in the census computation (see the Appendix). If so, the undercount estimate is further reduced by 236,000 persons, and the final estimate in stratum 12 drops to 1.1%. This latter figure is consistent with the other nonminority strata.

(3) Another problematic area is the fact that with a search area of 1–2 rings of blocks around a selected block, 4.1% of the P-sample were matched in the surrounding blocks, while only 2.3% of the E-sample got correct enumeration status in surrounding blocks. Suppose the Census Bureau had decided to use a search area of 6–8 rings of blocks. With this much larger search area more matches and more correct enumerations would be found in the surrounding blocks.

To see what the magnitude of the change might be, say that both rates increased by 20%, that is, the number matched in the surrounding blocks went to 4.9% of the total number of matches instead of 4.1%, and the correct enumerations to 2.8% instead of 2.3%. These assumptions seem conservative, but the undercount estimate would decrease by another 1,000,000.

There is some evidence that shows that the estimated undercount would significantly decrease with a larger search area. Concerning the two block clusters mentioned in the introduction that contributed almost one million to the estimated undercount, the June 1991 Census Bureau memorandum states: "the matching . . had been done correctly. However, approximately 75 percent of the non-matching people could have been converted to a match if the search area had been expanded."

Another piece of evidence is that in the rehearsal for the PES in Los Angeles in 1986, although relatively few households were matched outside their

TABLE 16
Undercount changes by evaluation stratum

Stratum	Original DSE	[P16] corrected	Total corrected	Change
1	6.8	5.4	3.7	3.2
2	−0.8	−1.3	−2.5	1.8
3	5.4	3.9	2.6	2.9
4	0.0	−1.1	−1.6	1.6
5	5.7	3.2	1.5	4.1
6	1.9	1.1	0.3	1.6
7	1.8	1.6	0.6	1.2
8	4.0	4.2	3.0	1.0
9	1.3	0.5	−0.5	1.8
10	0.4	−0.1	−0.7	1.1
11	6.1	5.7	3.9	2.3
12	2.1	3.7*	3.1*	−1.0*
13	1.8	0.7	−0.2	2.1

* These numbers may be erroneous due to a possible mistake in report [P16]. See the discussion in the Appendix. Correcting the mistake gives the numbers 1.7, 1.1 and 1.0.

blocks, 38% of those that were matched outside were matched more than five blocks away (Wolter, 1987a).

(4) The Census Bureau carefully rematched 104 block clusters having large numbers of nonmatches and erroneously enumerated persons (Hogan, 1993). The result was a further decrease of 250,000 in the estimated undercount. This has not been included in Table 15 because the possible overlap with other error sources listed. Getting a decrease of 250,000 in the estimated undercount by rematching 104 out of 5,290 block clusters raises the question of what additional changes might result if the Census Bureau had the resources for similarly rematching the rest.

7. COMMENTS ON REFERENCES

References that describe the methods used in the adjustment rehearsals and the evaluations of the outcomes are as follows: Anolik (1988) for Mississippi 1986; Hogan and Wolter (1988), Schenker (1988) and Stokes and Jones (1989) for Los Angeles 1986; Anolik (1989) for North Dakota 1987; Childers and Hogan (1989a, b; 1990), Diffendal (1988) and Mulry and Dajani (1989) for Missouri and Washington 1988. These publications are interesting in that they describe the outcomes of early efforts that led to the methodology used in 1990. Because the rehearsals were smaller in scale than the 1990 effort, their evaluations were sometimes more detailed.

Lessons learned in the rehearsals are summarized in Hogan (1989). Discussion and planning for the 1990 adjustment are given in Anolik (1990), Biemer and Stokes (1989), Childers et al. (1987) and West, Corby and Van Nest (1989). The most comprehensive view of the problems to be faced in 1990 is in the Bureau of the Census document Wolter (1987a).

The Secretary of Commerce's decision (Department of Commerce, 1991b) came with a six-inch-high stack of back-up material and contains good summary descriptions of the census, the adjustment procedure and the evaluations, as well as other useful material. Differing views toward adjustment are contained. In particular, two committee reports have informative views of the outcome of the adjustment project. The Undercount Steering Committee consisted of Census Bureau statisticians, with a majority favoring adjustment (Undercount Steering Committee, 1991). The Special Advisory Panel consisted of statisticians from outside the Census Bureau and split evenly on adjustment. Because of this split, the Special Advisory Committee submitted a number of reports, with Ericksen, Estrada, Tukey and Wolter (1991) and Wachter (1991) being the most substantive. Another interesting and informed view appears in Freedman (1991) (also contained in Department of Commerce, 1991b).

APPENDIX: ISSUES UNRESOLVED WITH THE CENSUS BUREAU

After review of this paper by the Census Bureau, two issues remained unresolved. First, the possibility was raised that some of the persons originally classified as out-of-scope by the PES, with a subsequent reduction in P-sample size, might truly be in-scope. If so, then these should be added back into the P-sample size, therefore increasing the undercount estimate and canceling out some of the effect of the new out-of-scope errors.

The only data available in the P-studies concerning the original out-of-scopes is in a sample of 193

PES out-of-scope persons sent to the EFU. The results are given in Report [P3, Table 3.6]. Many of these did come back into scope, but then were usually classified as matches. The Census Bureau provided a weighting to the national level of Table 3.6. Based on these numbers and assuming a match rate of 90% and that 50% of the unresolved are imputed as matches, the overall effect would be to increase the undercount estimate by less than 12,000 persons.

The second problem came up when investigating the estimated undercount in evaluation stratum 12. The estimate is increased by an upward adjustment of 1.27% attributed to census day address error [P16, Table 12]. However, the formula and example initially given to us by the Census Bureau for computation of the census day address error shows that the adjustment can only be negative, and it is negative in the other 12 evaluation strata.

In December 1991, I was informed that the formula was incorrect, that the tables in [P4] were wrong and that the census day address error computation in [P4] and [P16] included all of the errors found in the P-sample reinterviews regardless of whether they were census day address errors or not. This was surprising, since this error is consistently referred to both in [P4] and [P16] as census day address error. Furthermore, some of the tables in [P4] are reproduced in the published paper titled "Address reporting error in the 1990 post-enumeration study" (West, Mulry, Parmer and Petrik, 1991).

I have been unable to obtain from the bureau any more specific information regarding their method for computing census day address error. If the P-sample interview error is actually included in the census day address error, 537,000 persons would be subtracted from the total decrease of 4,243,000 persons detailed in Table 15. The result would be to lift the estimated undercount from 0.4% to 0.6%.

ACKNOWLEDGMENTS

The learning and study process leading to this paper was helped by the cooperation of many persons at the Census Bureau who patiently went over complex details of the PES, matching and evaluations with me. John Thompson, Chief, Statistical Support Division, Bureau of the Census, provided invaluable assistance by pointing me toward the right people and setting up meetings with them.

David Freedman provided many helpful comments and discussion which resulted in a much improved paper and I am glad (more or less) that he got me involved in these census issues. Ken Wachter also contributed insightful comments and discussion. Mark Hansen carried out spotlessly clean computer runs and Robert Koyak provided important information. Terry Speed has been untiring in his insistence on clear prose and good organization; Robert Kass made suggestions which improved readability; and numbers of anonymous referees contributed to a better paper.

P-Project Reports

All of the P-project reports referenced are in one of the 1990 Coverage Studies and Evaluation Memorandum Series issued by the Statistical Support Division, Bureau of the Census, and are all dated in July 1991. They have the main title "1990 Post-Enumeration Survey Evaluation Project P(#)" followed by title and author. We list project number, memorandum series identification, title and author(s) for each of the reports referred to in the text.

P1 Series #A-9, Analysis of reasonable alternatives, Stephen Mack, Eric Schindler and Joe Schafer.

P2 Series #B-4, Distribution of missing data rates, Philip M. Gbur.

P3 Series #C-2, Evaluation of imputation methodology for unresolved match status cases, Philip M. Gbur.

P4 Series #D-2, Quality of reported census day address, Kirsten K. West.

P5 Series #E-4, Analysis of PES P-sample fabrications from PES quality control data, Antoinette Tremblay.

P5a Series #F-1, Analysis of P-sample fabrication from evaluation follow or -up data, Kirsten K. West.

P6 Series #G-2, Fabrication in the P-sample: interviewer effect, Kirsten K. West.

P7 Series #H-2, Estimates of P-sample clerical matching error from a rematching evaluation, Mary C. Davis and Paul Biemer.

P8 Series #I-2, Matching error—estimates of clerical error from quality assurance results, Michael Ringwelski.

P9a Series #K-2, Accurate measurement of census erroneous enumerations, Kirsten K. West.

P10 Series #L-2, Measurement of the census erroneous enumerations—clerical error made in the assignment of enumeration status, Mary C. Davis and Paul Biemer.

P11 Series #M-2, Balancing error evaluation, Randall Parmer.

P13 Series #O-3, Use of alternative dual system estimators to measure correlation bias, William Bell.

P16 Series #R-6, Total error in PES estimates by evaluation post strata, Mary H. Mulry.

P18 Series #T-1, The evaluation of late late census data in the 1990 post enumeration study, Nicholas Alberti.

Heterogeneity and Census Adjustment for the Intercensal Base

D. Freedman and K. Wachter

Abstract. Current techniques for census adjustment involve the "synthetic assumption" that undercount rates are constant within "poststrata" across geographical areas. A poststratum is a subgroup of people with given demographic characteristics; poststrata are chosen to minimize heterogeneity in undercount rates. This paper will use 1990 census data to assess the synthetic assumption. We find that heterogeneity within poststrata is quite large, with a corresponding impact on local undercount rates estimated by the synthetic method. Thus, any comparison of error rates between the census and adjusted counts should take heterogeneity into account.

Key words and phrases: Census adjustment, synthetic method, loss function analysis, small area estimation.

1. INTRODUCTION

The decennial census is used to apportion Congress and to allocate tax moneys. Over the course of a decade, intercensal (more correctly, "postcensal") population estimates are developed for states and other geographical areas. These estimates are used to allocate tax moneys in noncensus years. In effect, census figures are rolled forward using data on births, deaths and migration patterns.

The census is generally thought to suffer from a small undercount, which is differential by race, ethnicity and area. The Census Bureau proposed to adjust the 1990 census to correct the undercount. They divided the population into "poststrata," which are relatively broad demographic subgroups thought to be more or less homogeneous with respect to undercount rates. Adjustment factors for the various poststrata were estimated from a special sample survey done after the census, called the Post Enumeration Survey, or PES. For more information and references, see Fay (1992a), Hogan (1993), Freedman (1991), Freedman et al. (1993) or Freedman, Wachter, Cutler and Klein (1994).

The Census Bureau's proposal to adjust the census by these methods was rejected by the parent agency (Department of Commerce, 1991b). Subsequently, the bureau considered a somewhat different procedure for adjusting the census as a base for the intercensal estimates. Among other things, the poststratification was changed (Bureau of the Census, 1992a, b).

To adjust the census and to compare the errors associated with adjusted and unadjusted counts, the Census Bureau makes the "synthetic assumption" that undercount rates are constant within poststrata across geographical areas. This paper will use 1990 census data to assess the synthetic assumption, focusing on the bureau's revised poststratification for the intercensal base. We find that heterogeneity within poststrata is quite large.

Any comparison of error rates in the raw and adjusted census counts should take into account the error due to heterogeneity. Otherwise, the comparison may be quite biased against the census. The bureau's "loss function analysis," which attempts to provide unbiased estimates of risk, appears to suffer from this problem and is unconvincing for that reason among others (Bureau of the Census, 1992b; Freedman, Wachter, Cutler and Klein, 1994; Wachter and Speed, 1991). The impact of heterogeneity on risk estimates is discussed in section 5.

Before proceeding, we outline the "synthetic method" used to adjust state counts, indicating where the synthetic assumption comes into play. California is the lead example. On census day (April 1, 1990), members of 147 different poststrata were resident in that state.

D. Freedman is Professor of Statistics, and K. Wachter is Professor of Demography and Statistics, Department of Statistics, University of California, Berkeley, California 94720. *Freedman and Wachter testified for the Department of Justice in City of New York et al. v. Department of Commerce et al., a lawsuit in which New York sought to compel census adjustment.*

One poststratum, code #301, consisted of blacks age 0–17 living in owner-occupied housing in rural areas anywhere in the United States. The census counted a nationwide total of 764,400 such persons. The "adjustment factor" for poststratum #301 was estimated from PES data as 1.058. In other words, the nationwide population in poststratum #301 as of census day was estimated as

$$1.058 \times 764.400 = 808.735.$$

rather than the census count of 764,400.

The adjustment factor of 1.058 was estimated using PES data from all across the country, California to Maine. However, to adjust California by the synthetic method, the multiplier of 1.058 is applied to the 4,260 members of poststratum #301 who were resident in California on April 1, 1990, according to the census. In other words, the number of blacks age 0–17 living in owner-occupied housing in rural areas in California on April 1, 1990, is estimated as

$$1.058 \times 4260 = 4,507.$$

rather than the census count of 4,260.

A similar procedure is applied to each of the 147 poststrata with members resident in California, and the adjusted population of California is obtained by summing the 147 products. (The PES sampling frame covered about 98% of the population; the "residual population" is not covered by the counts above and has to be added in separately.)

Each poststratum has its own adjustment factor, estimated from PES data covering many states. However, this factor is applied to the members of the poststratum resident in California. That is where the synthetic assumption enters: undercount rates are assumed to be constant within poststrata across states. Our object is to assess the degree to which the synthetic assumption holds.

For previous work on synthetic adjustment, with a review of the literature, see Wachter and Freedman (1994). For discussion by Census Bureau staff, see Fay and Thompson (1993) or Kim, Blodgett and Zaslavsky (1993). For a discussion of adjustment from other perspectives, see Breiman (1994), Choldin (1994), Citro and Cohen (1985), Ericksen and Kadane (1985), Ericksen, Kadane and Tukey (1989), Ericksen, Estrada, Tukey and Wolter (1991), Fay (1992a), Freedman (1991), Freedman and Navidi (1986, 1992), Hengartner and Speed (1993), Hogan (1993), Hogan and Wolter (1988), Mitroff, Mason and Barabba (1983), Schenker (1993), Schirm (1991), Schirm and Preston (1987), Singh (1992), Steffey (1993), Steffey and Bradburn (1994), Wolter (1986a, 1991), Wolter and Causey (1991) and Ylvisaker (1991). Choldin (1994) reviews work on the 1990 census, and is generally sympathetic to adjustment. On the year 2000, see Steffey (1993) or Steffey and Bradburn (1994), who support current plans for integrating the census, coverage measurement, and adjustment; components would not be made available for separate analysis: hence the description as a "one-number census."

TABLE 1
Proxy variables considered by the Bureau of the Census (1992b)

Substitution rate: the percentage of persons in each group whose entire census records were imputed. (For about 1% of the population, the census determines the number of persons in a household, e.g., by interviewing a neighbor, but has no personal information about the occupants. Then occupants are chosen at random from nearby houses, and their personal characteristics are imputed to the occupants of the target household.)

Allocation rate: the percentage of persons in each group with at least one out of six critical items in the census record "allocated," that is, imputed.

Multiunit housing rate: the percentage of persons in each group who were living in multiunit housing (such as apartment buildings).

Nonmailback rate: the percentage of persons in each group who failed to mail back their census questionnaire. (The denominator is the number to whom forms were mailed.)

Mobility rate: the percentage of persons in each group who were living at a different address five years ago.

Poverty rate: the percentage of persons in each group whose incomes were below the poverty line.

2. RESULTS

Proxy Variables

Of course, direct information on variability of undercount rates within poststrata is hard to come by. Although the PES sampled roughly 400,000 persons, there are 357 poststrata in the scheme for adjusting the postcensal base; the average sample size is about 1,000 persons per poststratum. Crossing the sample with the 50 states produces nugatory sample sizes. To investigate the synthetic assumption, the Census Bureau turned to proxy variables, that is, variables thought to resemble undercount rates in terms of heterogeneity. A "variable" is a numerical characteristic of population groups, with values determined from census data. The relevant groups are the poststrata, the states and the poststratum × state cells. We will consider the six proxies shown in Table 1. The first four are based on census short-form data. The last two come from the census long-form sample. (The denominator for long-form variables consists of population sizes estimated from the long-form sample.)

There is no sampling variability in the first four proxies, and essentially none in the last two, since the census long-form sample is so large. Of course, there is sampling variability in estimated under-

TABLE 2
Age–sex groups defined by the Bureau of the Census (1992b) for adjusting the 1990 census base; intercensal estimates

	Male		Female
0–17		1	
18–29	2		4
30–49	3		5
50+	6		7

count rates, since these are estimated from a sample survey.

The undercount rates are what everybody cares about, but the PES is too thin on the ground to make any very strong case about homogeneity in such rates. That is why the Bureau turned to the proxies. The great advantage of the proxies is complete (or virtually complete) data. The drawback is that the proxies may not behave like undercount rates, in terms of heterogeneity.

Poststrata

The poststrata considered in Bureau of the Census (1992b) for adjusting the intercensals are shown in Tables 2 and 3. Basically, there are 7 age-sex groups shown in Table 2, crossed with the 51 "poststrata groups" (PSG's) shown in Table 3. Thus, there are $51 \times 7 = 357$ poststrata. These are identified by two- or three-digit code numbers. Poststratum #11 is poststratum group 1 crossed with age-group 1; poststratum #517 is PSG 51 crossed with age-group 7. These poststrata are also referred to by sequence number: code #11 is the first in sequence, and code #517 is 357th in sequence.

Age-sex group 1 (top of Table 2) consists of males and females ages 0–17. Age-sex group 7 (bottom right) consists of females age 50 and over.

Poststratum group 51 (at the bottom of Table 3) consists of American Indians on reservations anywhere in the United States; PSG 30 (in the middle of the table) consists of blacks living in owner-occupied housing in rural areas anywhere in the United States; PSG 1 (at the top left) consists of non-Hispanic whites living in owner-occupied housing in large urban areas in the Northeast region of the United States. (The Northeast census region consists of Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey and Pennsylvania.)

Basically, the population is split by race and ethnicity (non-Hispanic white, Hispanic, black, Asian and Pacific Islander, American Indian; the first category contains any residual). There is another split on home ownership, and another on place type (big cities, other urban, rural); there is a final split on census region. However, some of the cells in the resulting cross-tabulation are collapsed, as shown in Table 3. The geographic scope of the poststrata is worth attention.The poststrata for whites are regional. However, most of the minority poststrata are national; for instance, small urban areas are combined from coast to coast.

The Data Set

To describe the data, consider a 357×51 matrix, with rows for poststrata and columns for the states (and Washington, D.C.). There is one matrix showing population counts; the (1, 1) entry gives the number of persons in the first poststratum (code #11) living in Alabama; the (357, 51) entry gives the number of persons in the 357th poststratum (code #517) living in Wyoming. (It is a numerical coincidence that there are 51 areas and 51 PSG's.)

Likewise, there is a matrix for each proxy variable, showing counts by poststratum and state. The (1, 1) entry in the "substitution" matrix gives the number of substituted persons in the first poststratum (code #11) in Alabama, and so forth. There is a matrix showing counts of persons in the mail universe (to whom census forms were mailed), used as denominators for nonmailback rates. Finally, there is a matrix showing the population counts as estimated from the census long-form sample, which are used as denominators for poverty and mobility rates. Counts are based on the PES target population (e.g., the institutional population is excluded, as are persons in the wilds of Alaska). These data were kindly provided by the Census Bureau.

Descriptive Statistics

We begin with some descriptive statistics. Table 4 summarizes data on the undercount, as estimated from the PES corrected for certain biases (Bureau of the Census, 1992b). The overall undercount rate is 1.580%. The root mean square standard error (SE) of state undercount rates is 0.466%; this measures sampling error. The SD of undercount rates across states (corrected for sampling error) is about 0.653%; this measures state-to-state differences in undercount rates.

Table 5 gives descriptive statistics for the proxies. Column 1 shows levels, that is, national rates. Column 2 shows the SD of the rates for the 51 areas (states and Washington D.C.). Thus, 0.899% of census forms were substitutions; the SD of the state rates was 0.408%. The third column shows the SD of the rates for the 357 poststrata. The last column shows the root mean square (r.m.s.) over the 357 poststrata of the within-poststratum-across-state SD. For example, take the first poststra-

TABLE 3
Poststrata groups (PSG's) defined by the Bureau of the Census (1992b) for adjusting the 1990 census base; intercensal estimates

	Northeast	South	Midwest	West
Non-Hispanic white				
Owner				
Urbanized areas 250 K+	1	2	3	4
Other urban	5	6	7	8
Nonurban	9	10	11	12
Nonowner				
Urbanized areas 250 K+	13	14	15	16
Other urban	17	18	19	20
Nonurban	21	22	23	24
Black				
Owner				
Urbanized areas 250 K+	25	26	27	28
Other urban		29		
Nonurban		30		
NonOwner				
Urbanized areas 250 K+	31	32	33	34
Other urban		35		
Nonurban		36		
Nonblack hispanic				
Owner				
Urbanized areas 250 K+	37	38	39	40
Other urban		41		
Nonurban		42		
Nonowner				
Urbanized areas 250 K+	43	44	45	46
Other urban		47		
Nonurban		48		
Asian and Pacific Islander				
Owner		49		
Nonowner		50		
American Indians on reservations		51		

TABLE 4
Summary statistics on undercount rates, in percent, corrected for certain biases in the PES

National rate	1.580%
Root Mean Square SE of rates for 50 states and Washington D.C.	0.466%
SD of rates across 50 states and Washington, D.C.	0.653%

tum, code #11. That poststratum had members resident in six states. In those states, the substitution rates were

0.0056 0.0043 0.0059 0.0056 0.0038 0.0087.

The SD of the six rates is 0.0017. Similarly for the remaining 356 poststrata. The r.m.s. of the 357 SD's is 0.00857 = 0.857%.

Table 5 does contain some good news for the proadjustment side: poststrata absorb more variability than do states (column 3 is bigger than column 2). There is also bad news: within-poststratum variability is larger than across-state variability (column 4 is bigger than column 2, except for multiunit housing). Differences between state substitution rates, for example, cannot be explained by differences in demographics; indeed, controlling for the poststrata makes the differences larger rather than smaller. This may seem a bit paradoxical, but see Section 4, which makes the connection with analysis of variance.

If the synthetic assumption held true, the SD's in the last column of Table 5 would be negligible. Instead, they are larger than the state-to-state differences summarized in column 2. Thus, Table 5 confirms what is evident a priori: there is quite a lot of heterogeneity within poststrata across geographical areas. Moreover, Table 5 strongly suggests that the synthetic method is counterproductive at the state level; it might be wiser to adjust each state on the basis of its own PES data.

Error in Adjustment Due to Heterogeneity

We consider the root mean squared error in synthetic estimates of state-level rates for the proxies.

TABLE 5

Levels and SD's for proxies, in percent: the first SD is over the 50 states and Washington, D.C.; the second SD is over the 357 poststrata; the last SD is within poststrata across state, r.m.s. over state

Proxy	Level (%)	Standard Deviations (%) Across states	Across poststrata	Within poststrata across states
Substitution	0.899	0.408	0.701	0.857
Allocation	16.034	3.573	7.209	3.922
Multiunit housing	22.145	8.926	29.318	8.755
Nonmailback	25.613	4.716	12.324	6.491
Mobility	42.738	5.270	48.918	7.949
Poverty	12.918	4.132	13.586	7.616

assuming that poststrata-level rates are given. Substitutions are the lead example. Suppose the substitution rate for each poststratum (the row total of substitutions divided by the row total of population) is known, but the entries in the matrix (the number of substitutions in each poststratum × state cell) are unknown. These entries could be estimated synthetically: multiply the poststratum substitution rate by the population count in the poststratum × state cell. Then add the entries in each column of the matrix to get an estimated number of substitutions in the corresponding state. Finally, divide by the state population to get each state's substitution rate. The objects of interest are these estimated state substitution rates.

In effect, each state's substitution rate has been estimated by the synthetic method, just as the Census Bureau estimates undercount rates. With substitutions, the exact answers are available from census data. Therefore the r.m.s. error of the synthetic method, across the 50 states and Washington, D.C., can be computed. This error is due solely to failures in the synthetic assumption, that is, variations in the substitution rate within poststrata across states. The r.m.s. errors are shown in column 1 of Table 6, in percent. For example, with substitutions, the r.m.s. error is about 0.31%. With allocations, the r.m.s. error is about 1.79%.

At first glance, these errors may seem rather small, supporting the synthetic assumption. However, much depends on the standard of comparison. The sampling error in estimated state undercount rates (r.m.s. SE across the 50 states and Washington, D.C.) is about 0.5%; see Table 4. Thus, error due to heterogeneity in the proxies is of the same order as, or even larger than, sampling error in estimated undercount rates. If heterogeneity in undercount rates is comparable to heterogeneity in the proxies, then heterogeneity is a major source of error in census adjustment by the synthetic method. On the other hand, if the proxies are not comparable to the undercount rate with respect to heterogeneity, they seem to provide little evidence about the degree to which the synthetic assumption holds.

Of course, scale matters, and the variables in Table 6 have different overall levels. For example, the Census Bureau's estimate for the undercount rate (from the Post Enumeration Survey corrected for certain biases) is 1.580% (Table 4). The overall substitution rate is 0.899% (Table 5). Practice at the Bureau suggests scaling the substitution rate by the factor $1.580/0.899 = 1.758$, before doing the calculations. The second column of Table 6 gives results for this "level-scaling." (The arithmetic for substitutions is $1.758 \times 0.31 = 0.54$.) As Table 6 shows, substitutions now have the worst r.m.s. error of the proxies, somewhat larger than the r.m.s. SE of state undercount rates.

There are other ways to scale besides levels. One possibility is to equalize state-to-state variability. The SD of undercount rates across states (corrected for sampling variance) is about 0.653% (Table 4). One could scale substitutions to have the same SD, multiplying by $0.653/0.408 = 1.600$ before doing the calculations. Results for this SD-scaling are shown in the third column of Table 6. (The arithmetic for substitutions is $1.600 \times 0.31 \approx 0.50$.)

Apparently, there are at least two scales on which the proxies differ from the undercounts: level and SD. Generally, it is not possible to match on both scales; although with substitutions or multiunit housing, you come very close. As Table 6 shows, conclusions depend on scaling. Levels matter, so do variances and so would covariation (common patterns in residuals across states). No scaling is perfect, which implies there may be no good proxies for the undercount. Furthermore, the proxies are all positive, while undercounts may be positive or negative, the latter possibility corresponding to overcounts. Thus, caution is in order. Still, Table 6 suggests that

TABLE 6
Root mean square errors of synthetic estimates, in percent; six proxies for undercount

Proxy	Root Mean Square Error (%)		
	Raw	Level-scaled	SD-scaled
Substitution	0.31	0.54	0.50
Allocation	1.79	0.18	0.33
Multiunit housing	3.41	0.24	0.25
Nonmailback	2.37	0.14	0.33
Mobility	3.53	0.13	0.44
Poverty	3.01	0.37	0.48

error in census adjustments due to heterogeneity is comparable in magnitude to the error caused by sample variability, and perhaps even larger.

One additional comparison may be useful. The interest in adjusting state populations is due to differential undercounts, which vary from state to state. The size of these differentials can be measured by the SD of the state rates. Take substitutions as a proxy. The state-to-state differences are on the order of 0.4% (column 2 in Table 5). The error due to heterogeneity is about 0.3% (column 1 in Table 6). This error is about 75% of the effect of interest. Results for other proxies are similar. Heterogeneity is not trivial.

The Sign Test

Come back to the 357 × 51 matrix of population counts for poststrata by states, and the corresponding matrix of substitutions. In each cell, we have the actual number of substitutions from census records, as well as the number estimated from the synthetic method. For each cell, let

$$\text{residual} = \text{estimate} - \text{actual}.$$

In cells with no population, the residual is 0. Otherwise, the residual will be positive or negative.

A strict interpretation of the synthetic assumption would require the residual to be identically 0. Clearly, the data contradict this strict form of the assumption. A weaker interpretation is that the residuals vary randomly (in some sense) from cell to cell and are more or less symmetric. If that is so, residuals would tend to cancel when adjusting states.

A sign test of the weaker interpretation can be made as follows. For each state, there are cells with nonzero population counts, corresponding to poststrata with members resident in that state. A typical state has 147 nonzero cells, although Alaska has 98; no state has more than 147. (There are 7 age groups in Table 2 and 21 rows in Table 3: $7 \times 21 = 147$.) In each nonzero cell, the weak interpretation of the synthetic assumption suggests that the sign of the residual is random. Thus, for each state, the number of positive residuals should be binomial with success probability 1/2, the number of trials being that state's number of nonzero cells. As a result, nearly half the nonzero residuals should be positive in each state. This prediction can now be compared to the data.

TABLE 7
Empirical SD's, in percent: Six Proxies for undercount; the theoretical SD, based on the synthetic assumption, is about 4.4%

Proxy	Empirical SD (%)
Substitution	24.5
Allocation	24.0
Multiunit housing	26.5
Nonmailback	22.3
Mobility	16.8
Poverty	23.7

For each state, count the number of positive residuals and divide by the number of cells with nonzero population. Then compute the SD of the resulting empirical distribution on [0, 1]. The empirical SD for substitutions is 24.5%; the SD suggested by the binomial model is 4.4%; details are in Section 4. The empirical distribution is fairly uniform, the model distribution is rather concentrated. In other words, the distribution of residuals is quite different from the predictions of the synthetic assumption. Cancellation of residuals is an unlikely assumption.

Results for the other proxies are rather similar. The six empirical SD's are shown in Table 7. The binomial SD's are all about 4.4%, since the number of cells with nonzero population is virtually the same for all populations considered (census, mail universe and long-form sample estimated population).

3. SUMMARY AND CONCLUSIONS

Data on proxy variables suggest substantial failures in the synthetic assumption. The r.m.s. error in

estimated state undercount rates arising from failures in the synthetic assumption seems comparable in magnitude to the r.m.s. error arising from sampling variability. Investigators who compare errors in adjusted and unadjusted census data should take heterogeneity into account; otherwise, the analysis may be quite biased against the census. The Census Bureau's loss function analysis appears to suffer from this problem (Bureau of the Census, 1992b) and is unconvincing for that reason among others.

Postscript

At a briefing on 23 December 1992, the Bureau announced its decision not to adjust the census. The Bureau argued that its "loss function analysis" was robust against heterogeneity; that improvements could be made at the state level; but that, for smaller areas like cities and countries, adjustment was of doubtful value (Bureau of the Census, 1992c, d; 1993).

4. MATHEMATICAL DETAILS

Index the 357 poststrata by i and the 51 areas (states and Washington, D.C.) by j. Let c_{ij} be the census count of members of poststratum i resident in state j on census day. Write subscript "+" for addition over a subscript: thus

$$c_{i+} = \sum_{j=1}^{51} c_{ij}$$

is the ith row sum in the c-matrix, corresponding to the census count of members of poststratum i in the 50 states and Washington, D.C.

Substitutions are the lead example. Let s_{ij} be the number of substitutions in the cell corresponding to poststratum i and area j. The first column in Table 5 reports levels, that is, s_{++}/c_{++}. The second column reports the SD of the state rates, that is, the SD of the 51-vector whose jth entry is s_{+j}/c_{+j}. The third column reports the SD of the poststratum rates, that is, the SD of the 357-vector whose ith entry is s_{i+}/c_{i+}. These statistics depend only on the marginal distributions.

The fourth column in Table 5 depends on interior cells. It reports the root mean square of the 357-vector whose ith entry is the standard deviation of

$$\{s_{ij}/c_{ij}: j = 1, \ldots, 51 \text{ and } c_{ij} > 0\}.$$

Allocations and multiunit housing are similar. For nonmailback rates, the denominator is the number of persons in the mail universe, rather than the census count. For mobility and poverty rates, the denominator is the estimated census count from the long-form sample.

The following example illustrates how controlling for poststratum increases cross-state variability; contact will be made with analysis of variance. Let x_{ij} be some numerical characteristic of the ijth cell. Let U be a random row index, and let V be a random column index. Consider the random variable

$$X = x_{UV}.$$

Choose U and V with weights proportional to the census counts, so

$$P\{U = i \text{ and } V = j\} = c_{ij}/c_{++},$$

where c_{ij} is the census count in the cell and subscript "+" denotes summation over an index. Take $x_{ij} = s_{ij}$, the number of substitutions in cell (i, j). Now $E\{X \mid V\}$ gives the state substitution rates, and $E(\text{var}\{X \mid U\})$ is the mean across poststrata of the within-poststratum variance; means and variances are weighted by population size. Thus, $E(\text{var}\{X \mid U\})$ is the analog of within-poststratum-across-state variance. However, there is no necessary inequality between $E(\text{var}\{X \mid U\})$ and $\text{var}(E\{X \mid V\})$, although both are less than $\text{var}(X)$:

$$E\Big(\text{var}\{X \mid U\}\Big) + \text{var}\Big(E\{X \mid U\}\Big) = \text{var}\, X.$$

The weighted results may be of some interest (Table 8). For two out of the six proxies, conditioning on poststratum adds variability to the state rates.

It may be useful to put the calculation for Table 6 in algebraic terms. As before, let s_{ij} be the number of substitutions in cell (i, j), and let c_{ij} be the census count. Let $\lambda_i = s_{i+}/c_{i+}$ be the substitution rate for poststratum i. Then the estimated number of substitutions in cell (i, j) is $c_{ij} \times \lambda_i$, the estimated number of substitutions in state j is $\sum_i c_{ij} \times \lambda_i$, and the estimated substitution rate in state j is

$$(\textstyle\sum_i c_{ij} \times \lambda_i)/(\sum_i c_{ij}).$$

For the sign test, let n_j be the number of positive cells in state j, that is, the number of poststrata i such that the census count $c_{ij} > 0$. Let ξ_j be binomial, with n_j trials and success probability 1/2, independent in j. Let $\eta_j = \xi_j/n_j$. The number of positive residuals is modeled as ξ_j, according to the weak form of the synthetic assumption. So the empirical distribution for the fraction of positive residuals is modeled as

$$\{\eta_j: j = 1, \ldots, 51\}.$$

TABLE 8

Standard deviations for proxies, in percent: the first SD is over the 50 states and Washington, D.C.; the second SD is over the 357 poststrata; the third SD is within poststrata across states, r.m.s. over states; the last SD is taken over all poststrata × state cells; all SD's are weighted by population

	Standard Deviations (%)			
Proxy	Across states	Across poststrata	Within poststrata across states	Across all cells
Substitution	0.26	0.65	0.35	0.74
Allocation	2.37	6.24	1.92	6.53
Multiunit housing	9.44	27.28	5.56	27.84
Nonmailback	4.40	12.89	2.97	13.22
Mobility	5.70	21.67	4.34	22.10
Poverty	3.40	12.31	3.98	12.93

This distribution has empirical variance

$$s^2 = \frac{1}{50}\sum_{j=1}^{51}(\eta_j - \bar{\eta})^2 \quad \text{where } \bar{\eta} = \frac{1}{51}\sum_{j=1}^{51}\eta_j.$$

As usual,

$$E\{s^2\} = 0.25 \times \frac{1}{51}\sum_{j=1}^{51}\frac{1}{n_j} = 0.044 = 4.4\%.$$

The model involves binomials with different numbers of trials. Thus, the empirical distribution is a slightly nonstandard object. However, it seems to be informative.

In principle, skewness in the residuals may arise because some cells are relatively small, and the substitution rate (for instance) is close to 0. Consider, then, a "superpopulation" version of the synthetic assumption. The null hypothesis is that, for each poststratum i, the number of substitutions s_{ij} is binomial, with success probability p_i and number of trials c_{ij}; the numbers are independent across states j; the success probability depends only on the poststratum i. The alternative hypothesis allows the success probability to depend on i and j.

Treating poststrata as independent, the usual likelihood test statistic is $\chi^2 = 2.2 \times 10^5$ on 6,510 degrees of freedom. Taking the 357 poststrata one at a time, the mean number of degrees of freedom is 18 and the mean χ^2 value is 614. The most favorable poststratum (sequence number 258) offers five degrees of freedom and $\chi^2 = 12$, so $P = 3.6\%$. The next most favorable poststratum (256) also has five degrees of freedom but $\chi^2 = 20$ so $P = 0.13\%$. The remaining 355 poststrata all have $P < 0.1\%$. About 20% of the positive cells have fewer than five expected substitutions but censoring such cells does not change the results. The superpopulation idea, whatever its conceptual merits or demerits, is not viable.

5. LOSS FUNCTION ANALYSIS

Heterogeneity is appreciable, adding substantial uncertainty to estimated undercounts for states and smaller areas. That is our message so far. Now there is another topic: the impact of heterogeneity on what the Census Bureau calls "loss function analysis", that is, estimation of risks for the census and adjustment.

The Bureau's loss function analysis is done on population shares rather than rates. To keep the focus on heterogeneity, we ignore problems created by bias and sampling error in the Post Enumeration Survey. Then risk is just total squared error in population shares. (For background, see Freedman, Wachter, Cutler and Klein, 1994.)

In the presence of heterogeneity, estimated risks for the census and for adjustment are severely biased. Hence, the risk difference may be biased. The sign of the bias in the risk difference can go either way; cancellation is also a possibility. Bias in estimated risks is the topic of this section.

To address heterogeneity, the Bureau uses "artificial population analysis," creating hypothetical true and census populations from data on proxies. Artificial population counts have to be generated for each poststratum × state cell. There are many ways to do this, but our approach is quite straightforward. The census goes in as itself. Next, each proxy variable corresponds to a set of poststratum × state cell counts, for instance, the number of substitutions in each cell. For the "true population," we use

$$\text{census} + \lambda \times \text{proxy}.$$

The scale factor λ is chosen to make the overall undercount rate in the artificial population match 1.6%, the rate estimated for the census of 1990:

$$\lambda \times \frac{\text{total of proxy cell counts}}{\text{total of census cell counts}} \approx 1.6\%.$$

TABLE 9

Loss function analysis using the proxies: risks have been multiplied by 10^8; scale factors are chosen to match the level of undercount. The "true risks" are computed using data on proxies in each poststratum × state cell. The "estimated risks" are computed using the synthetic assumption. The estimated risk for adjustment is 0, so the estimated risk for the census coincides with the estimated risk difference; DIFF = 8 × SUB − 0.2 × NMB

Proxy	Scale factor	True risks		True risk difference	Estimated risk difference
		Census	Adjustment		
SUB	1.8	51	26	25	33
ALL	0.10	14	3	11.1	10.6
MUH	0.072	223	27	196	138
NMB	0.063	35	6	29	24
MOB	0.038	26	3	23	22
POV	0.12	40	19	21	24
DIFF	**	459	381	77	224

The scale factors are shown in column 1 of Table 9; SUB stands for substitutions, ALL for allocations and so forth.

Of course, the proxies are part of the census population, while omitted persons are disjoint from that population. For present purposes, that may not matter. On the other hand, all methods for generating artificial populations have a basic problem: the hypothetical "true population" may not be in close correspondence to the actual population.

In effect, Table 6 reports the results of artificial population analysis for estimates of adjustment factors, while the present section deals more directly with shares. Wolter and Causey (1991) review Bureau research on artificial population analysis for the 1980 census; also see Bureau of the Census (1992d, 1993) on the 1990 census.

Turn now to the algebra. Recall that c_{ij} is the census count in poststratum i and state j, so the census count in state j is

$$c_{+j} = \sum_{i=1}^{357} c_{ij}.$$

The total census population is

$$c_{++} = \sum_{i=1}^{357}\sum_{j=1}^{51} c_{ij}.$$

The census share for state j is $\text{csh}_j = c_{+j}/c_{++}$.

Substitutions are the lead example; s_{ij} is the number of substitutions in poststratum i and state j; and the "true undercount" t_{ij} in that cell is taken as $t_{ij} = 1.8 \times s_{ij}$. (As noted above, scale factors like the 1.8 are chosen to make the level of the proxy correspond to the level of undercount: $1.8 \times s_{++}/c_{++} \approx 0.016$.)

The "true population share" of state j is

$$\text{tsh}_j = (c_{+j} + t_{+j})/(c_{++} + t_{++}).$$

The total squared error for the census is then

$$\sum_{j=1}^{51}(\text{csh}_j - \text{tsh}_j)^2.$$

This is $51/10^8$, as shown in column 2 of Table 9.

The synthetic method estimates t_{ij} as $c_{ij} \times t_{i+}/c_{i+}$, where t_{i+} is assumed to be known. So the adjusted population for state j is

$$a_{+j} = \sum_{i=1}^{357} c_{ij} \times t_{i+}/c_{i+},$$

and the adjusted share for state j is

$$\text{ash}_j = a_{+j}/a_{++}.$$

The total squared error for adjustment is

$$\sum_{j=1}^{51}(\text{ash}_j - \text{tsh}_j)^2.$$

This is $26/10^8$, as shown in column 3 of Table 9.

The risk difference is the total squared error for census shares minus the total squared error for adjusted shares. This works out to $51 - 26 = 25/10^8$, as shown in column 4 of Table 9. Positive numbers favor adjustment: the census makes larger errors. Of course, in application the true shares tsh_j would not be known. However, we are assuming no bias in the Post Enumeration Survey and no sampling error; so the "adjustment factors" t_{i+}/c_{i+} are known with certainty.

Invoking the synthetic assumption, the error due to adjustment is estimated as 0. On the same basis, the squared error for the census is estimated

as

$$\sum_{j=1}^{51}(\text{csh}_j - \text{ash}_j)^2.$$

This is $33/10^8$, as shown in the last column of Table 9. (The estimated risk for the census coincides with the estimated risk difference.) In this example, the estimated risk difference based on the synthetic assumption is about one-third larger than the true risk difference. In other words, the Bureau's loss function analysis, which rides on the synthetic assumption, overstates the advantages of an adjustment based on the synthetic method. In the presence of heterogeneity, loss function analysis is biased toward adjustment.

Results for other proxies in Table 9 are computed in a similar way. (Here, the denominators for nonmailback rates are the census counts; the denominators for mobility rates and poverty rates are census counts estimated from the long-form sample.) The poverty rate (POV) goes like the substitution rate. For allocations (ALL) and mobility (MOB), however, heterogeneity causes almost no bias in estimated risk differences. For multiunit housing (MUH) and nonmailback rates (NMB), heterogeneity creates a bias against adjustment.

A further complication should be mentioned: the proxies are all positive, but undercounts can be negative—when the corresponding cell has been overcounted. To indicate the possibilities, we use DIFF = 8 × SUB − 0.2 × NMB as a proxy. The bias against the census is remarkable, a factor of 3. ($\text{DIFF}_{++}/c_{++} \approx 2.1\%$, the estimated level of undercount on July 15, 1991; 8 × SUB more or less matches the gross omissions while 0.2 × NMB matches the erroneous enumerations; see Department of Commerce, 1991b.) In short, Table 9 shows that almost anything can happen.

A bit of algebra may make the situation clearer. The bias in the loss function analysis is the difference between the true risk difference and the estimated risk difference. For state j, this is

$$(\text{csh}_j - \text{tsh}_j)^2 - (\text{ash}_j - \text{tsh}_j)^2 - (\text{csh}_j - \text{ash}_j)^2$$
$$= 2 \times (\text{ash}_j - \text{csh}_j) \times (\text{tsh}_j - \text{ash}_j).$$

Thus, loss function analysis is "conservative," that is, biased against adjustment, when adjustment of shares is conservative:

$$\text{ash}_j > \text{csh}_j \quad \text{and} \quad \text{tsh}_j > \text{ash}_j$$

or

$$\text{ash}_j < \text{csh}_j \quad \text{and} \quad \text{tsh}_j < \text{ash}_j.$$

TABLE 10

Correlations of proxy rates with undercount rates; the unit of analysis is either the poststratum or the state

Proxy	Unit of Analysis	
	Poststratum	State
Substitution	0.40	0.28
Allocation	0.0080	0.10
Multiunit housing	0.24	0.0021
Nonmailback	0.43	0.37
Mobility	0.33	0.53
Poverty	0.42	0.44

Otherwise, loss function analysis is biased against the census.

It may be noted that none of the proxies are well correlated with undercounts (Table 10). Allocations and multiunit housing are particularly weak in this respect. Furthermore, the biases (like the risks themselves) are driven by data for only a few of the 51 areas.

A minor inconsistency in our scaling should be noted too: as the denominator for its undercount rate, the bureau uses the estimated true population; our denominator for the proxy rates is the analog of the census population. In principle, these rates should be level-scaled to $1/(1 - 0.0158) \approx 1.605\%$ rather than 1.580%.

Table 9 quantifies bias in estimated risks due to heterogeneity—for the proxies. For undercount rates, however, the bias is unknown. The Census Bureau seems to argue that, for a majority of its proxies, the bias is negligible or runs against adjustment (Bureau of the Census, 1993). However, we think the basic question is still open: are undercounts more like substitutions, or allocations, or multiunit housing, let alone DIFF? The data do not answer this question.

To sum up, in proxies for the undercount rate, heterogeneity is appreciable. This creates substantial extra uncertainty in estimates for states and smaller areas. The same conclusions are likely to hold for undercounts themselves. With respect to loss function analysis, the bias due to heterogeneity may be substantial; the data do not decide the issue.

ACKNOWLEDGMENTS

We thank Richard Cutler (Utah), Terry Speed (U.C. Berkeley) and Amos Tversky (Stanford) for helpful comments. Research partially supported by NSF Grant DMS-92-08677 (to D. Freedman).

Can We Reach Consensus on Census Adjustment?

Thomas R. Belin and John E. Rolph

Abstract. Attempting a complete headcount is an imperfect method for carrying out a census, as is modifying an attempted headcount with sample-based adjustments. It is a mistake to assume that one approach enjoys a scientific presumption over the other. There are important details available from evaluation studies of the 1990 decennial census that reflect upon the accuracy of adjusted and unadjusted census figures. Decisions about adjustment might therefore be based on comparing the accuracy of alternative census-taking strategies at some level of aggregation of the population. In any such comparison, the choices of an appropriate level of aggregation, the factors defining the aggregation, and appropriate loss criteria are important issues to decide in advance. After providing context for decisions about census-taking strategy, we comment on the recent literature on census adjustment, including the papers by Freedman and Wachter and by Breiman contained in this issue; we also discuss the Census Bureau's plans for the year 2000. We conclude that the 1990 approach to summarizing the accuracy of an adjusted census can be improved upon, but that many of the criticisms of census adjustment do not reflect a balanced decision-making perspective. We also conclude that the Census Bureau is pursuing constructive research in evaluating a "one-number census," and we suggest that statisticians have a role to play in avoiding the costly legal battles that have plagued recent censuses by assisting in the process of deciding on a design for the 2000 census.

Key words and phrases: Decision, loss function, census adjustment, one-number census.

For an effective adjustment procedure to be widely accepted, given that not all localities will benefit, it is important that there be as widespread understanding and agreement as possible within the professional community of statisticians that a general reduction in differential coverage error is sufficiently desirable to accept adverse impacts on some individual localities. —Panel on Decennial Census Methodology of the Committee on National Statistics, National Academy of Sciences, 1985.

1. THE PROSPECT OF CONSENSUS IN A CONTENTIOUS DEBATE

That the widespread agreement among statisticians envisioned by the National Academy panel does not exist is clear from the preceding papers, by Freedman and Wachter (1994; hereafter FW) and by Breiman (1994). We are far from sure that consensus is attainable on census adjustment, but we think it should be possible to settle issues that in the past have stood in the way of consensus.

Our primary goal in this article is to set forth the key substantive issues in the debate over statistical adjustment to address differential undercount in the decennial census. To sharpen our focus, we have tried to separate statistical principles from the many relevant technical details that would attend any proposed adjustment. Our hope is that by highlighting

Thomas R. Belin is Assistant Professor in Residence, Department of Biostatistics, UCLA School of Public Health, Los Angeles, California 90024-1772. John E. Rolph is Professor of Statistics, University of Southern California Business School, Los Angeles, California 90089-1421.

points of contention and contributing to a sensible framework for discussion, we might help in putting our profession's collective energies to more constructive use than having us argue about philosophical differences in a courtroom every 10 years from here on in.

Several themes permeate the adjustment debate: the interplay of personal and scientific principles, the idea that census adjustment should be viewed as a decision problem, the role of formal loss functions in such a decision, and the question of what level of aggregation is appropriate for evaluating alternative census numbers. The paper by FW discusses the impact of a particular source of uncertainty on the kinds of evaluations that might be used to decide about adjustment, and the paper by Breiman raises several provocative points in connection with the proposed adjustment of the 1990 census. We discuss the papers by FW and by Breiman with the goal of relating them to the broader context of the adjustment controversy.

Although we wish to promote consensus, not all of our comments, particularly about the latter paper, are conciliatory. Consensus-building sometimes requires criticizing ideas around which consensus will never emerge, and this we have endeavored to do.

The two of us have come at the census undercount problem from different angles. Belin was a graduate student in the Harvard Statistics Department whose thesis work on computer matching of census and post-enumeration survey (PES) data brought him to the Census Bureau, and he worked at the Bureau from mid-1990 to mid-1991 developing programs that were used for the imputation of cases with unresolved enumeration status in the 1990 PES. Rolph is presently an American Statistical Association representative to the Census Advisory Committee of Professional Associations and served as a member of a National Academy of Sciences committee in the mid-1980's that produced the report from which the opening quotation is taken; he also testified in the *New York et al. v. Commerce Department et al.* lawsuit for the plaintiffs on issues pertaining to regression smoothing techniques used in the proposed adjustment procedure.

The rest of the article is organized as follows. Section 2 reviews central elements of the adjustment controversy, with special attention to issues that involve human judgment. In Section 3, we discuss the 1990 adjustment operation, adding details not contained in the FW and Breiman papers, in order to provide the groundwork for our later comments on those papers in Sections 4 and 5. The growing support for a "one-number census" in the year 2000 is the subject of Section 6, and we offer some concluding remarks in Section 7.

2. THE ADJUSTMENT DEBATE

Scientists, politicians, government officials and members of the judiciary, among others, have articulated viewpoints on strategies for census-taking that frequently conflict. There is now a voluminous statistical literature on the subject: for example, Alho, Mulry, Wurdeman and Kim (1993); Belin et al. (1993); Bell (1993); Bryant (1991, 1992, 1993); Citro and Cohen (1985); Cohen (1990, 1991); Cressie (1992); Darroch, Fienberg, Glonek and Junker (1993); Datta et al. (1992); Ericksen and DeFonso (1993); Ericksen and Kadane (1985); Ericksen, Kadane and Tukey (1989); Ericksen, Estrada, Tukey and Wolter (1991); Fay, Passel, Robinson and Cowan (1988); Fay and Thompson (1993); Fienberg (1989–1992, 1992a,b, 1993, 1994a); Freedman (1991, 1994); Freedman and Navidi (1986, 1992); Freedman et al. (1993); Hogan (1992a, 1993); Hogan and Wolter (1988); Isaki, Schultz, Diffendal and Huang (1988); Kruskal (1991); Martin, Brownrigg and Fay (1990); Meyer and Kadane (1992); Mulry and Spencer (1991, 1993); Robinson, Ahmed, Das Gupta and Woodrow (1993); Rolph (1993); Royce (1992); Schirm (1991); Schirm and Preston (1987); Wachter (1991, 1993a, 1994); Wachter and Speed (1991); Wolter (1991); Wolter and Causey (1991); Zaslavsky (1993a); along with the documentation of the 1991 adjustment decision (Mosbacher 1991a); the transcript and exhibits from the *New York et al. v. Commerce Department et al.* lawsuit over the proposed 1990 adjustment (U.S. District Court 1992); the special sections of the *Journal of the American Statistical Association* in September 1993, of *Jurimetrics* in Fall 1993 and of *Survey Methodology* in June 1988, December 1988 and June 1992; and the recent National Research Council reports by the Panel on Census Requirements in the Year 2000 and Beyond and by the Panel to Evaluate Alternative Census Methods (Committee on National Statistics, 1993a,b). Our goal here is to provide some context for the adjustment debate and to do so in a way that makes the "consensus" in our title a plausible objective.

2.1 Accuracy and the Census

Thomas Jefferson's suspicion of an undercount in the 1790 census (Ericksen, Kadane and Tukey, 1989) has often been cited in the adjustment debate as an early sign in our history that the census needs fixing. There was a differential undercount between blacks and whites after the Civil War (Anderson, 1988), and the Census Bureau has published black–white differentials based on demographic analysis from the 1940 census onward. To obtain estimates of differential undercount broken down into subpopulations defined by geography and other characteristics, the

Census Bureau has used "dual system estimates," which are described by Breiman and to which we allude later. The "adjustment debate" surrounding the 1990 decennial census has centered on the comparison of census numbers that include a correction based on estimates of differential undercount and census numbers that contain no such correction. We now describe some of the issues at the heart of the debate.

2.2 "Personal and Political" Considerations in Loss Functions

Formal yardsticks to compare the adjusted to the unadjusted counts in the form of loss functions were used by the Census Bureau in their evaluations of 1990 census results. There has been criticism of the idea of reducing decisions about adjustment to a comparison of "loss estimates" that synthesize large quantities of data with fairly crude statistical measures. Kenneth Wachter, in his report to the Secretary of Commerce as a member of the Secretary's Special Advisory Panel on the 1990 Census, wrote the following (Wachter, 1991, page 5, emphasis in original):

> *I do not believe that any highly aggregated index or loss function is appropriate for summing up overall accuracy.* It is informative to understand how much the outcomes of calculations with different versions of such aggregated indices differ. But the choice among them is not a scientific choice. Each such index involves implicit value judgments about different sorts of error. For example, each index determines whether a few large errors are more serious than a great many smaller errors. *Whether we agree with a particular tradeoff is a matter of personal and political values. It should not be disguised as science.*

An example that illustrates what Wachter has in mind as a relevant "personal and political" consideration is his concern for future funding of censuses (Wachter, 1991, page 42, emphasis in original):

> *I believe that Census adjustment would reduce the stake that individuals, civic leaders, and Congressional representatives would have in coverage improvement efforts. If coverage improvement efforts suffer, our knowledge of the characteristics of the hard-to-count population will suffer too. Adjustment would also increase the political leverage of technical decisions. Extra efforts to guarantee the Census Bureau's independence and objectivity would be required.*

Wachter (1991, pages 42–43) further argues that the "most serious consequence" of adjustment on future census efforts is that politicians would be likely to reduce funding for data collection, notably the expensive efforts to collect information on especially hard-to-count individuals, which he describes as a "tragedy" because the information lost at the margin is "especially precious as background for policy." Although information on hard-to-count people may be precious, data from the "coverage improvement" efforts Wachter defends do not seem to have an especially prominent role in current policymaking; also, their cost-effectiveness is questionable (Ericksen and DeFonso, 1993). In any event, we do not endorse Wachter's speculative predictions about the consequences of using estimation techniques in census-taking, but we do agree that there are a variety of relevant nonstatistical issues that should be aired and discussed by all parties, statisticians and nonstatisticians alike.

One reason it has been so difficult to build consensus on census adjustment is that the "rules of the game" do not seem to be fixed. For example, Wachter (1991, page 4) commented (emphasis in original):

> I have heard versions of an *a priori* argument contending that we should always be able to do better by using all our data, combining all the information we have, than by restricting ourselves to one part or another. This is meant to be an argument that we can do better by combining the Census enumeration with the PES than by using the Census alone.... The assumptions of this argument....do not hold for the Census and the PES. In ordinary life, we all often make wrong decisions figuring in new information, when we would have been better off to stay with what we thought at first. *The question of whether counts can be improved is an empirical question, not an* a priori *one.*

When combined with his comment that "I do not believe that any highly aggregated index or loss function is appropriate..." one page later in his report, this passage presents us with a philosophical quandary. On the one hand, Wachter argues that adjustment is an empirical question and that it is only from evidence in the data that we can reliably infer that adjusted counts are more accurate. On the other, however, Wachter argues that no index for summarizing the evidence from data is an unambiguous measure of whether one estimate is better than

another. How he would have us proceed is unclear. Our view is that it would be helpful, at the very least, to agree upon what type of evidence from the data would be viewed as favorable for adjustment.

Some may accept Wachter's position on both fronts and may then reason that adjustment is philosophically untenable. We reject this argument, as it fails to heed the warning cited by several commentators on the adjustment debate (e.g., Tukey, 1991; Fienberg, 1992a), namely, "Don't let the perfect be the enemy of the good." Substantial gains in accuracy have been shown to be possible through adjustment by research such as Schirm and Preston (1987), Isaki, Schultz, Diffendal and Huang (1988) and Zaslavsky (1993a), and several investigators have found that comparisons between unadjusted and adjusted counts are apt to agree across reasonable loss function choices (see Citro and Cohen, 1985; Schirm and Preston, 1987; Mulry, 1992a; Mulry and Spencer, 1993; Zaslavsky, 1993a). The issue of how to balance evidence from the data with *a priori* considerations is admittedly difficult, but we see no reason to ignore the evidence from research on loss function analysis.

2.3 "Validation" of Adjustment Methods

In responding to the argument "Nothing is perfect, and don't let the best be the enemy of the good," Freedman and Navidi (1992) write the following:

> We think the census is imperfect, but good. We think the smoothing models are quite questionable, and the arguments to defend them are bad. Proponents of adjustment have an obligation to state their assumptions and produce data to validate them. The models don't have to hold perfectly, but departures from assumptions and their impacts need to be studied. Otherwise, the algorithms have no justification except familiarity.

Of course, the adjustment process should be fully documented with its assumptions spelled out, and indeed that happened. But did proponents of the change that was made early in 1994 in the way unemployment figures are produced (Plewes, 1994) "validate" their assumptions? What would it take to "validate" such assumptions, and who gets to decide whether the assumptions are "validated?"

Freedman offered a hint of what kind of validation he had in mind in some personal correspondence with Belin (Freedman, 1993a). The discussion focused on results presented in Belin et al. (1993) comparing predictions from the P-sample imputation model to results from evaluation follow-up (EFU) interviewing, an operation that in part investigated individuals whose enumeration status in the census was unresolved. We later describe the EFU operation in more detail, and we also examine the accuracy of imputation further in the context of Breiman's paper. For present purposes, it is relevant that fewer than half of the individuals whose census enumeration status was unresolved in the PES were able to be resolved through the additional review. However, among those EFU cases that were resolved, 31.6% were determined to have been enumerated in the census, and the imputation model had predicted that 32.2% of these cases had been enumerated with a predictive standard error of 2.2%. Belin (1993) asked Freedman:

> I am curious, since you seem to think that the EFU data constitute evidence against the approach we used, can you offer a hypothetical scenario in which the EFU data would come out in such a way that you would not criticize our imputation approach? I really do not mean to be facetious here; I simply do not know what might silence some of the critics on this point.

Freedman (1993a) wrote the following:

> [W]hat would have to happen in EFU to persuade me? If all or nearly all of the imputed cases in production had been in the sampling universe for EFU, and a very substantial fraction of the sampled cases had been resolved in EFU, and the match status as determined in EFU had on average matched the imputed probabilities (conditional on some covariates of interest), and EFU fieldwork had been done more or less properly, then I would be quite persuaded, indeed, this would be a very serious argument. Since none of these conditions are met, I find the assertion that EFU validates the model to be somewhat frivolous.

Freedman's position here is very extreme. With the designation of "unresolved" status coming much later than the original interview, roughly six months later in the 1990 PES, one could never design a PES–EFU program to guarantee the high response rate that Freedman demands as a precondition to "validation." There is uncertainty inherent in any adjustment operation due to the inevitable presence of unresolved cases and the inevitable difficulty of following them up. We think it makes sense to weigh this uncertainty against the foregone conclusion that the census headcount will never be perfectly calibrated in all parts of the country for all population subgroups.

It strikes us as an extreme position to be so intolerant of uncertainty from a source such as unresolved enumeration status that one would reject adjustment out of hand without considering the degree of bias evidence in the census headcount. Meanwhile, the assertion that 31.6% does not constitute close agreement with 32.2 ± 2.2% raises questions about what standard Freedman is using in his evaluation of the data.

Ultimately, the Freedman–Navidi position, "We think the census is imperfect, but good," begs the question, because the determination that the census is "good" is in the eye of the beholder. Is there no room in statistical science for those who say, "We think the census could be better"?

2.4 Producing Census Counts as a Decision Problem

The 1990 attempted headcount resulted in completely implausible census counts. So did the proposed 1990 adjustment. By "completely implausible" we mean that there is evidence in available data that suggests the presence of systematic errors in the allocation of population counts to areas. The attempted headcount was completely implausible because at aggregated levels there was a clear bias in the coverage of minority populations. The proposed adjusted counts were completely implausible because there is demonstrable heterogeneity in undercount rates within poststrata assumed to have constant within-stratum undercount rates (FW; Hengartner and Speed, 1993). That a choice has to be made between imperfect alternatives is obvious. Population counts have to be produced, so we have to choose from among "implausible sets" of counts.

Zaslavsky (1993a) has proposed what appear to be less implausible models than either of these other two, a development that we applaud. Surely Zaslavsky's models are not perfect, but violations of the assumptions of Zaslavsky's models appear to be harder to find than violations of the assumptions underlying either an unadjusted census or an adjustment based purely on dual-system estimation. In our opinion, the census would be better if it were harder than it currently is to find systematic errors in it.

2.5 Guidelines for a Decision on Adjustment

Important constraints attend this decision problem, notably time, available methodology, and the need to have the solution be broadly acceptable to a skeptical public and its elected officials. In such a decision setting, there are many valid nonstatistical considerations that touch on issues of fairness, precedent for future censuses, and so on. However, there is a need for clarification of standards or rules to guide the decision-making process.

The 1990 adjustment decision differed from that in 1980 in that the Commerce Department was required as part of settlement of a lawsuit to produce "guidelines" for its decision. These guidelines, which themselves were the target of a subsequent legal challenge, put forward some general presumptions and highlighted several issues as relevant for decision-making purposes. It bears repeating that there is an important distinction to be drawn between statistical and nonstatistical considerations. "Guidelines" as we conceive of them are primarily useful as a way of prioritizing nonstatistical issues so that statistical results can usefully inform the decision-making process. In his paper, Breiman comments as follows:

> The most important [guideline], from a statistical point of view, was Guideline 1: "The Census shall be considered the most accurate count of the population of the United States, at the national, State, and local level, unless an adjusted count is shown to be more accurate."

One might also say that Guideline 1, which we do not necessarily endorse, was the most important guideline from a political point of view, in the sense of that term used by Wachter (1991).

Freedman and Navidi (1992) advance the notion that there is a "burden of proof" that any proposed adjustment must surmount in order to be accepted, apparently viewing such a burden as something of a basic canon of science since adjustment is based on statistical models. Meanwhile, in distinguishing themselves from "modelers," Freedman and Navidi appear not to acknowledge the assumption that the attempted headcount leads to constant undercount rates across poststrata. However, this assumption is part of the implicit model that underlines the unadjusted counts. In arguing that the burden of proof is a statistical issue, Freedman and Navidi shed more light on their own biases than they do on the broader adjustment debate:

> As the exchange with Fienberg indicates, modelers are reluctant to accept the burden of proof. Once they make an assumption, it is taken as truth unless it can be disproved. Even then, they may view the assumption as useful until it can be replaced by some other assumption.
>
> Language is used in a specialized way. An assumption is "reasonable" if the modelers think it is reasonable. If questioned, they introspect again. The introspection

> confirms the original conclusion; after all, the assumptions are by now familiar parts of the technical literature. The modelers become indignant at those who do not share the faith.

As for the psychological profile of the "modeler," we think it applies equally well to those whose "faith" has it that the attempted headcount leads to constant undercount rates across poststrata. In fact, the assumption that the undercount rate is constant everywhere is stronger than the assumption that undercount rates are constant within poststrata, so perhaps the criticisms apply to a greater extent to those whose faith favors unadjusted counts. Meanwhile, it is naive to suggest that "modelers" have a monopoly on using language in a specialized way for rhetorical purposes, as our discussion below of Breiman's paper amply demonstrates. However, Freedman and Navidi are certainly correct that rhetoric has created a gulf across which it is hard to move toward consensus.

To summarize, "personal and political" opinions deserve to be considered in the adjustment debate, and developing guidelines may be a useful way to summarize nonstatistical issues, but there is not a sound basis for rejecting adjustment out of hand on scientific grounds.

2.6 What Is the Right Loss Function?

Crucial to any decision about statistical estimation in the census are the "measures of performance" (Isaki, Schultz, Diffendal and Huang, 1988) chosen to assess the accuracy of census numbers. Such measures have been variously referred to as measures of improvement (Wolter, 1987b), measures of adjustment success (Schirm and Preston, 1987) and measures of misproportionality (Tukey, 1983), but commonly are referred to using the generic term "loss functions" (e.g., Spencer, 1980a,b; Citro and Cohen, 1985; Woltman, 1991; Woltman et al., 1991; Freedman and Navidi, 1992; Mulry and Spencer, 1993; Zaslavsky, 1993a). We see a prospect for genuine consensus in this area.

We begin by recounting some statements made in the report of the Panel on Decennial Census Methodology in 1985 that we believe apply equally well today as they did back then:

1. "[I]t is impossible to determine a single loss function that is appropriate for evaluating every effect of an error in census numbers: each use of the census numbers has a different effect resulting in different components of loss" (Citro and Cohen, 1985, page 279).
2. "It must be accepted that no adjustment procedure can be expected to simultaneously reduce the error of all census information for every location in the United States" (Citro and Cohen, 1985, page 281).
3. "[A]djustment should be undertaken when there is reasonable certainty that appreciable reduction in the general differential coverage error will be achieved. A relatively trivial reduction would not be worthwhile, since adjustment will surely cost time and resources to implement, and doubt about whether the adjustment did or did not reduce the differential coverage error would impair public confidence in census figures. Furthermore, knowledge of a subsequent adjustment might reduce public cooperation, thus lowering the completeness of the census count" (Citro and Cohen, 1985, page 281).
4. "The panel believes that it is substantially more important to reduce the general error per person than the general error per place. Hence, the panel does not recommend the use of loss functions for measuring the total error that weight each political jurisdiction equally, e.g., that determine the proportion of the 39,000 revenue sharing jurisdictions that gained or loss through adjustment, regardless of the number of people in each jurisdiction. Rather, the panel believes that the contribution to total loss attributable to an area should reflect the size of its population" (Citro and Cohen, 1985, page 282).

There is widespread agreement that distributional accuracy, or focusing on population shares rather than population totals, is the appropriate goal (Schirm and Preston, 1987; Undercount Steering Committee, 1991; Mosbacher, 1991b; Mulry and Spencer, 1993; Zaslavsky, 1993a; Rolph, 1993). However, the fourth point, that it is important to consider the size of jurisdictions in assessments of accuracy as opposed to counting the number of domains made more accurate, is underappreciated. Much of politics is place-oriented, so political interest is often in error per place, not error per person. This is an issue where getting statisticians together with students of politics might be helpful in sharpening the debate. The rationale for the panel's recommendation that the focus be on reducing error per person was based on the idea that each citizen's interest should be weighted equally for fairness reasons.

This point evidently did not sink in among key decisionmakers on adjustment of the 1990 census. Then-Secretary of Commerce Robert Mosbacher, the day after he announced his decision not to adjust the 1990 census, commented in congressional testimony (Mosbacher, 1991b):

> Based on the measurements so far completed, the Census Bureau estimated that the proportional share of about 29 States would be made more accurate and about 21 States would be made less accurate by adjustment.... When the Census Bureau made allowances for plausible estimates of factors not yet measured, these comparisons shifted toward favoring the accuracy of the census enumeration. Using this test, 28 or 29 States were estimated to be made less accurate if the adjustment were to be used.

This testimony was given only a few weeks after the Census Bureau's Undercount Steering Committee, in its report, took great pains to point out that it is a fallacy to suppose that the break-even point between adjustment and nonadjustment occurs where the shares are more accurate in exactly half of the jurisdictions (Undercount Steering Committee, 1991):

> Suppose [conditions hold that imply] the census distribution [of proportionate shares for states] is a close approximation to the truth but differs randomly by about 1 percent, and the PES is unbiased but has a sampling variance of the same magnitude. Thus, the sets of two proportions are equally accurate.... Under assumptions of normality, the negative sign [for the difference between the loss for a given state under the unadjusted census versus under the adjusted census, implying an advantage to the unadjusted census] should appear in an expected 68 percent of the states, or about 34 out of 50. Intuition that the break-even point is when half of the states have negative losses and half have positive is *not* correct. [emphasis in original] Further, when the ratio of [the variance of the biases in unadjusted state population shares] to [the variance of PES estimates of state population shares] is 2 to 1, strongly favoring an adjustment, then a similar calculation gives that 59 percent would have negative signs, or about 29.5 states. The expected number of negative signs is about 21 at a ratio of [the variance of biases in unadjusted state shares to the variance of PES estimates of state shares] of 5 to 1. Under this simple model, observing 21 negative signs is consistent with a strong positive effect of adjustment on the measurement of the true population proportions.

The secretary's reliance on a summary of the results that probably had the effect of misleading the assembled legislators is troubling. It may reflect the place-oriented outlook of a political appointee who was unaware of the apparent paradox or a conscious choice to present results in a way that favored a particular interpretation. Either way, this portion of Mosbacher's rationale for not adjusting the census is weak, sounding like an argument that the presidency of the United States should be awarded to the person who wins the most number of states.

The National Academy panel in the mid-1980's (Citro and Cohen, 1985) cites earlier work by Kadane (1984) that provides a rationale for the choice of a loss function for congressional seat allocation and by Spencer (1980b) for choosing a loss function appropriate for revenue-sharing purposes. Tukey (1983) and Wolter (1987b), along with the National Academy panel, provide strong arguments about the importance of taking into account the size of units in assessing their contribution to aggregated loss. Zaslavsky (1993a) adds to these arguments a seemingly reasonable condition that reduces the field of contenders of candidate loss functions, namely, that the aggregated loss should be insensitive to division of domains of consideration if undercount is uniform within that domain. Zaslavsky accordingly restricts attention to absolute error $\Sigma|X_k - T_k|$, where X_k is the estimated population share for domain k and T_k is the true population share for domain k, and size-weighted squared relative error $\Sigma(X_k - T_k)^2/T_k$. We endorse the loss function choices in Zaslavsky (1993a) for the reasons outlined there and by earlier authors (e.g., Hartigan, 1992).

Finally, we note that the methods of evaluation in Zaslavsky (1993a) represent a substantial advance over the loss function analysis strategy implemented in recent Census Bureau decisions on adjustment. Rather than assuming that adjusted counts corrected by estimated biases can be taken as true population counts, Zaslavsky reflects the fact that the estimated biases have variances of their own that need to be included in the analysis for a proper accounting of uncertainty. This is a technical advance that could remove a major source of controversy over adjustment evaluation.

2.7 What Is the Right Level of Analysis?

Beyond the issue of choosing among loss functions, another concern is the level of geographical or other aggregation at which to apply loss functions. Freedman (1991) discusses small-sample issues that attend the problem of making adjustments at disagregated levels. Freedman uses as an example the city of Stockton, California, which has a population of about 200,000, roughly a quarter of whom are Hispanic. Stockton cannot have very

many individuals included in a PES sample of 5,000 blocks around the country, and no stable estimates of a Hispanic undercount rate would be available from data in Stockton alone. The Census Bureau's 1990 stratification pooled Stockton's Hispanics with other Hispanics in Southern California and pooled its non-Hispanics with other non-Hispanics throughout the Western states in order to estimate undercount rates. Freedman argues that "Stockton is the rule, not the exception," since there are far more local jurisdictions around the country than there were blocks in the 1990 PES; implicitly, the argument seems to be that adjustment is untenable on its face.

On the other hand, as noted by Zaslavsky (1993b), direct sample-based estimates of undercount rates for demographic subgroups at the most aggregated level, that is, for the country as a whole, are accurate enough to detect significant differences between groups. For example, no one disputes that blacks and whites are differentially undercounted. At a less aggregated level, for example, at the level of 50 states, sophisticated "composite" estimates, or weighted averages of unadjusted and adjusted counts with weights derived from estimates of the precision of the unadjusted and adjusted counts, may be shown to be more accurate than unadjusted counts (Zaslavsky, 1993a). However, there is some level of aggregation, which Zaslavsky (1993b) terms "the gray zone," where composite estimates improve accuracy, but the case is difficult to prove. Also, there is what Zaslavsky (1993b) terms "the red zone," a level of aggregation where adjustment may make estimates less accurate, although again it may be hard to prove that these small-area adjustments give less precise estimates of population shares.

Once again, "personal and political" considerations can also enter the picture, as loss functions may not reflect fairness considerations that many deem important. Cain (1992) reports a systematic effect at the level of congressional districts which raises fairness questions that would be difficult to capture in an explicit loss function analysis. In the trial on the 1990 adjustment, Cain presented a regression of estimated percent undercount on percent Democratic registration among registered voters in the 52 newly apportioned California congressional districts, which were created based on unadjusted data. He found that a difference of 10% in Democratic registration was associated with a difference of 0.8% in estimated undercount, with worse undercounts in districts that were more Democratic.

If the choice of a level of aggregation at which to evaluate census numbers is not decided in advance, it will be extremely difficult not only to build consensus around a particular set of census numbers but even to design an undercount evaluation program like the 1990 PES that would be broadly acceptable.

3. A SUMMARY OF THE ATTEMPTED HEADCOUNT AND ADJUSTMENT-RELATED OPERATIONS

Although the descriptions of the census-taking process in the papers by Freedman and Wachter and by Breiman are not incorrect, they are incomplete because they do not mention some key features of the applied setting that result in a decision-making framework where the alternatives are imperfect and there are crucial constraints, such as time. We give our own synopsis of the census-taking process to provide a context for our later comments. See Hogan (1992a, 1993), Ericksen, Estrada, Tukey and Wolter (1991) and Freedman et al. (1993) for more detailed accounts. The process that generated the official 1990 population counts we term an "attempted headcount" for reasons that the descriptions below will make clear.

3.1 Address Lists and Initial Enumeration Attempt

In 1990, in most areas the census assembled address lists from a variety of public and private sources and sent a census form to each household on the address list. In other areas, notably rural areas where addresses such as "Rural Route 1" might be common to a number of households, census enumerators attempted to visit each housing unit in person. Four different enumeration strategies were used by the Census Bureau, with all areas of the country being classified into one of these four "types of enumeration areas." Most households received one census form, but some received none and some received more than one. The official Census Day was April 1, 1990, with respondents to the census asked to list residents of the housing unit as of that day. In addition to trying to include all housing units in its address list, the Bureau attempted to enumerate homeless individuals with teams of enumerators visiting places thought "likely" to have homeless people spending the night. The 1990 PES was focused solely on the population in housing units and was not designed to measure undercount of the homeless population; in the Census Bureau jargon that Breiman invokes, the homeless population was "out-of-scope" for the PES.

3.2 Follow-up

When a household did not respond to the first attempt of the Census Bureau to count individuals living there, census enumerators sought an interview with a household member during May or June of

1990 in "nonresponse follow-up." When no household member was available, a "last-resort" interview was sought with a neighbor, landlord, or some other person who claimed knowledge about the residents of a household. When a census enumerator suspected that a housing unit was occupied but did not have direct evidence, individuals were added to households through "imputation" or "substitution," whereby the number of individuals from a nearby household is taken as the number for the household in question.

In addition, there were several "coverage improvement" programs. These include the following: counting people who might be vacationing or traveling; double-checking households that were originally classified as vacant; running newspaper advertisements asking "Were You Counted?" that people could clip and send to the Bureau; and reviewing administrative lists in an attempt to count parolees and probationers. One of these operations, the "Parolee/Probationer Follow-up Program," was not "prespecified," or planned in advance of its mid-1990 inception, but was undertaken when an originally planned "Parolee/Probationer Check" resulted in much lower than expected response. About 2.9 million people were added to the population between August and December of 1990 (Ericksen and DeFonso, 1993); these later enumerations were considerably more likely to be erroneous (duplicate, fictitious, etc.) than enumerations obtained closer to the official census day (Ericksen, Estrada, Tukey and Wolter, 1991; Belin and Diffendal, 1991; Griffin and Wajer, 1993; Ericksen and DeFonso, 1993).

3.3 Post Enumeration Survey (PES)

Before headcounting efforts began, PES workers (also employed by the Census Bureau but not involved in the attempted headcount) canvassed a sample of blocks, recording addresses and sketching maps, that were to be included in the PES based on a stratified sampling scheme considering region of the country, "place type" (i.e., central city or one of a few types of non–central city locations), 1980 percentage of minority population in the area, 1980 percentage of renters in the area, and "type of enumeration area" (i.e., census-taking methodology used in the area, reflecting the source of the household address in the master address list and the strategy for contacting the household, such as by mail or by personal interview) (Woltman, Alberti and Moriarity, 1988). The PES interviewing was done in person, mostly in July and August, or three to four months after census day. Some areas with high levels of nonresponse in the initial stages of PES interviewing were recanvassed by experienced interviewers who work regularly for the Census Bureau on its monthly household surveys.

The PES design was a stratified cluster sample, with all housing units in sampled blocks targeted for inclusion in the PES sample, except in some large blocks where subsampling of the block took place. The stratification in the design was intended to improve the precision of estimates of differential undercount between certain population subgroups, with some groups anticipated to have high undercounts sampled at a greater than proportional rate. Sampling weights based on inverse probabilities of selection of the block were spread among the individuals in the block, with an adjustment made to individual sampling weights when subsampling was done and when interviews for particular households were not obtained.

3.4 Matching the E-Sample and P-Sample

Breiman uses the Census Bureau jargon "E-sample" and "P-sample," the former consisting of the individuals living in PES blocks on April 1, 1990, and the latter consisting of the individuals living in PES blocks at the time of the PES. These samples overlap, but one reason they can differ is that people move between census day and the time of the PES.

Names of individuals from census forms, which are not routinely entered into census databases, were keypunched into an E-sample data file to supplement demographic information from the census form, and names and associated characteristics from the PES were entered into a separate P-sample data file. Name, address, and personal characteristics provide the basis for matching records from these two data files; when names are not available, matching is not possible, so undercount estimation needs to accommodate such cases. A computer matching operation identified obvious matches, after which information was printed out onto forms that were reviewed by clerical matching teams, one form for each housing unit with a line for each individual in that housing unit. Sometimes, for reasons that will become clear, this process is called before-follow-up matching; Breiman refers to these matching efforts by their Census Bureau acronyms, SMG1 and SMG2. Separate matching operations were undertaken for individuals who moved between the time of the census and the PES; for these individuals, queries in the PES interview about the address where the individual resided on April 1 were used to search census data files to see whether the person had been enumerated.

3.5 PES Follow-up

The "PES follow-up," which took place in October and November of 1990, attempted to resolve discrepancies between the census and PES rosters by send-

ing interviewers to the households in question. Both census and PES enumerators can fabricate the existence of people, and both sets of interviewers miss residents. Thus, for both E-sample cases that did not match a PES case and P-sample cases that did not match a census case, interviewers attempted to find out whether the individual lived at the sampled housing unit on census day. The rule of thumb is that two "sightings" in the field establish the existence of individuals. So, for example, individuals observed in both the census and the PES need not be followed up; in addition, a second "sighting" in PES follow-up is taken as establishing the existence of individuals who had been included in the PES but not the census.

Information from PES follow-up interviewing was returned to processing offices for "after-follow-up matching" by teams of matching clerks, who attempted to resolve individuals as having been enumerated or not having been enumerated based on the additional information from the field. Some cases inevitably remain unresolved in this process, perhaps because a follow-up interview was not obtained. Since these cases were selected for follow-up interviewing on account of a discrepancy between the census and the PES, they are clearly not a random subsample of the PES sample. Thus, ignoring unresolved cases in undercount estimation would lead to predictable biases.

3.6 Estimation for Unresolved Cases

For these remaining unresolved cases, rather than attempting further follow-up interviewing, logistic regression models based on characteristics observed on each individual were used to impute a probability of having been correctly enumerated. These probabilities were used to estimate undercount rates and to characterize uncertainty about those estimated rates (Schafer and Schenker, 1991; Belin et al., 1993). The logistic regressions used not only personal characteristics such as age, race and sex, but also processing characteristics such as "before-follow-up match status," which reflects information such as that there is a person in the census who appears to be a possible match for the given unresolved P-sample individual or that the unresolved person was part of a whole household of individuals who did not match anyone in the census even though the address where they resided was included in the census address list. As shown in Belin et al. (1993), such distinctions clearly have great predictive value.

3.7 Adjustment Poststrata

Estimation of undercount rates was done for various poststrata defined by the following: geography, place type, race or ethnic origin, owner or renter status, age, and sex. For the 1991 proposed adjustment discussed by Breiman, there were 1,392 poststrata; for the 1992 proposed adjustment of the base for intercensal estimates discussed by FW, there was a coarser stratification into 357 poststrata. Sampling variances were estimated by a jackknife method, where the units being jackknifed were the primary sampling units, namely, the sampled blocks.

3.8 Smoothing Models

For the 1991 proposed adjustment, a hierarchical "smoothing" regression model was used that related poststratum-level undercount rates to covariates. Because large estimated undercounts are typically associated with large variance estimates or, more specifically, because the error in estimating the mean for a Poisson-like undercount rate is apt to be correlated with the estimated variance, "raw" sample variance estimates were modeled to produce "presmoothed" variance estimates for input into the smoothing of undercount estimates (Hogan, 1993).

There were a few poststrata where the combination of presmoothing and smoothing led to undercount estimates that did not appear to pass a test of face-validity. For example, Asian–Pacific Islander males aged 20–29 in New York City had a direct estimate of a 34% undercount, and presmoothing approximately halved its estimated standard deviation from 14.4% to 6.2%. That poststratum became an influential point in the regression; for 12 poststrata involving Asians and Pacific Islanders in New York City, the mean $\pm$ SD of their estimated undercount rates went from $4.3 \pm 20.4\%$ before smoothing to $13.4 \pm 2.0\%$ after smoothing. The considerable shrinkage of the undercount estimates for these poststrata is a reflection of limited sample sizes, and this degree of sensitivity was unusual. To lessen the impact of influential and outlying poststrata, which were identified through analysis of regression residuals (Bell, 1991; Hogan, 1993), the variances of the undercount estimates for these poststrata were left "un-presmoothed." This yielded final undercount estimates for the 12 Asian–Pacific Islander poststrata in New York City of $9.8 \pm 1.9\%$, these being regarded as superior to the estimates with all variances presmoothed as a matter of face-validity.

Much controversy attended the use of such models in the *New York v. Commerce Department* lawsuit (Freedman, 1992; Freedman et al., 1993); the issue of smoothing was perhaps the most prominent technical issue argued there. In that context, the second author expressed strong support for such models being used as they were by the Census Bureau in their 1991 proposed adjustment (Rolph, 1992, 1993).

While far from a perfect fix to a challenging technical problem, the decision to presmooth was based on the technical issue of avoiding bias in estimating Poisson-like means. We mention this to refute the recent suggestion by a group of authors (Freedman et al., 1993) that a "circular" argument was the basis for presmoothing, or, as those authors put it, that presmoothing is "defended because it delivers answers that agree with a priori expectations [about minority undercounts]." These comments ignore the rationale for the decision four years earlier at the Census Bureau to carry out presmoothing for bias-reduction reasons.

3.9 Evaluation Studies

To inform officials better in the decision-making process, the Census Bureau attempted to synthesize components of error in undercount estimates into a "total error model" (Mulry and Spencer, 1991, 1993). Such errors were estimated by a series of projects such as a rematch study, in which additional matching clerks reviewed a sample of cases, with results later compared to those of the original PES matching clerks, and "evaluation follow-up" (EFU) interviewing, where a final attempt was made to reach individuals in the PES sample to assess undetected PES fabrication, the actual enumeration rate of individuals who were unresolved in PES processing and so on. Breiman (1994) focuses on the accuracy of these studies.

In addition, a "loss function analysis" was used to inform the adjustment decision by estimating the costs of errors associated with the attempted headcount and the adjusted counts. Here, simple functions are used to approximate the costs of errors, or losses. Comparisons of the unadjusted and adjusted counts took place at the level of a small number (13 in the 1991 loss function analysis and 10 in the 1992 loss function analysis) of groupings of PES poststrata into "evaluation poststrata" that reflect broad geographical and racial–ethnic characteristics, and comparisons also took place at more disaggregated levels such as the state level. In both the 1991 and 1992 analyses, the latter of which corrected some of the errors cited by Breiman, notably the computer processing error, the Census Bureau calculations favored the adjusted counts (Mulry and Spencer, 1993; Fay and Thompson, 1993). Not all observers have found these loss function analyses to be decisive, and some commentators have criticized these approaches as being based on arguable approximations in the calculations of total error and for relying on undercount estimates that have been corrected for estimated bias (e.g., Wachter and Speed, 1991).

In this broad context, Freedman and Wachter have focused on one issue that was ignored in the Census Bureau's loss function analysis, namely, heterogeneity of undercount rates within poststrata, and Breiman's article targets many quantities that figure into these loss calculations and some other quantities that were omitted. We now turn to the interesting and informative article by Freedman and Wachter.

4. FREEDMAN AND WACHTER'S STUDY OF HETEROGENEITY

Heterogeneity of undercount rates within strata or poststrata is but one of many sources of error underlying sample-based population estimation. Presumably, the amount of heterogeneity bias in estimates depends heavily on the level of aggregation at which analysis takes place. Hengartner and Speed (1993) and Schafer (1993) engage in a lively debate over heterogeneity in undercount rates in the context of the proposed 1990 adjustment, shedding further light on the issue of an appropriate level of aggregation for analysis; Wolter and Causey (1991) discuss heterogeneity in the context of the proposed 1980 adjustment. Freedman and Wachter attempt to add to our understanding the heterogeneity bias, and their efforts are most welcome.

We commend FW for adopting the approach of analyzing proxies to gain insight into the plausible effects of adjustment. This methodology was used in Isaki, Schultz, Diffendal and Huang (1988), which has been invoked by some as a rationale for adjustment. By lending legitimacy to the general approach, FW make it easier to discuss the merits of the technical issues. Being able to agree upon methods of analysis is an essential step toward broader consensus on these issues.

A major point of agreement we have with FW is their conclusion about standard errors for small areas. Without focusing on details, we agree with their assessment at the outset of Section 5: "Heterogeneity is appreciable, adding substantial uncertainty to estimated undercounts for states and smaller areas." Omitting heterogeneity from small-area standard errors can clearly be a substantial error. Perhaps this work by FW could be used to estimate analogs to survey design effects to inflate standard errors by a fixed factor or by a set of factors that might depend on characteristics of the area. Would FW be willing to embrace the idea of taking this next step, or do they have an alternative solution in mind to the problem of estimating small-area standard errors?

Freedman and Wachter also amply demonstrate that the synthetic assumption for carrying adjustments down to small areas is false. Schafer's (1993) comment on the similar conclusion of Hengartner

and Speed (1993) was, "The Census Bureau is not surprised," and neither are we. This is an empirical conclusion about which there is broad consensus.

Our main criticism of FW is their interpretation of results from their loss function analysis. Their Table 9 considers seven proxy variables, four of which show the estimated risk difference to be lower than the true risk difference and three of which show the estimated risk difference to be higher. This does not surprise us. Their own algebra shows that the bias in the estimated risk difference for domain j is given by $2 \times (\text{ash}_j - \text{csh}_j) \times (\text{tsh}_j - \text{ash}_j)$, where tsh_j is the true population share, csh_j is the census population share and ash_j is the adjusted (synthetically estimated) population share. Contributions to bias are positive if the true undercounts are higher than the synthetic undercount estimates in the same domains where the adjusted population shares are estimated by the synthetic method to be higher than the census population shares, that is, in those domains where the PES identified large undercounts. As noted by Fay and Thompson (1993), it does not seem obvious why such a correlation would occur. Without such a correlation, one would expect the risk difference to be fairly symmetrically distributed around zero, as it is in FW's Table 9. A reviewer pointed out the following to us (emphasis in the original):

> A priori I would expect that at the local level, areas with high proportions of hard-to-count groups (highly segregated minority areas, areas in which everybody is a renter) would tend to have worse undercount FOR A GIVEN POSTSTRATUM than those with more of a mix (e.g., comparing black renters in a segregated renter-only area to black renters in an integrated area with many homeowners); this makes the covariance positive, favoring adjustment. It is less obvious to me how this would work out at the state level.

We think FW do a commendable job pointing up the limitations of proxy variable analysis while also illustrating its utility. So when they conclude, "In short, Table 9 shows that almost anything can happen," we think another reasonable voice might have phrased it, "These results support the theoretical argument that heterogeneity bias should, on average, favor neither adjusted nor unadjusted figures." It appears to us that FW are arguing a point as if they were in an adversarial proceeding. We favor the more balanced interpretation in Fay and Thompson (1993).

The possible presence of heterogeneity bias might not be irrelevant to a decision on adjustment, however, if there is some presumption in advance favoring either adjustment or nonadjustment. Fay (1993) pointed to the risk of heterogeneity bias as one of several factors that led him to change his mind about the 1991 adjustment decision. Fay had originally supported adjustment, but a few months after the decision he concluded from a subsequent analysis that the sampling variances may have been underestimated by a substantial factor and reconsidered his original support for adjustment. When it was pointed out to him, in a meeting where he was presenting his findings, that the amount by which he claimed the sampling variances were underestimated still would not have overturned the loss function advantage of the adjusted counts, Fay (1992b) replied, "Yes, but our margin of comfort is gone." When asked what he had in mind when he made that comment, Fay (1993) explained:

> The loss function analysis was incomplete. We didn't know whether we had all the biases in it fully measured. We can make quite a list including that we didn't know what to do at the time about heterogeneity bias.... I was saying if there are pieces left out of the puzzle, then...we have to allow for more uncertainty.

5. BREIMAN'S PAPER

Breiman's paper poses a substantial challenge. He focuses on a number of details of the adjustment procedure about which neither of us had more than general knowledge prior to writing this article, along with a few topics about which we have specialized knowledge. A primary reason that it is hard to respond to Breiman is that some of his depictions are very misleading. We wish he had not included material in his article that we believe deserves the harshest kind of criticism, but he did, and we feel obliged to point out what we view as unprofessional practices.

5.1 Breiman's Planned Testimony and the Current Paper

Although Breiman states that he assisted the defendants in the *New York v. Commerce Department* lawsuit, he did not note that this paper has the same title and numerous similarities to a manuscript (Breiman, 1992b) that he prepared to offer in support of his testimony (which it turned out he did not give). Indeed, many aspects of the current text are identical to the version prepared for the trial. This earlier version is relevant to material we discuss later.

5.2 The Computer Coding Error and a Contradiction in Breiman's Analysis

Breiman is correct to consider the potential effect of factors not included in the Total Error Model anal-

ysis on decisions about census adjustment, but inconsistencies in the standards he applies make it appear that his primary aim is to discredit the whole idea of adjustment. One of Breiman's main points is that if every area were undercounted by the same amount, then adjustment would be superfluous and would only add error: "The most important thing about estimating the undercount is not its total magnitude, but its differential effect on the 50 states and on thousands of counties and cities." Why then, in an apparently contradictory approach, does he summarize evidence on data quality, and on the computer coding error in particular, in terms of effects on the net national undercount rate?

That the computer coding error substantially affected the overall undercount estimate and looms as a warning signal for any future effort to undertake a complicated adjustment is widely acknowledged. The error was that E-sample cases that match at the PES sample address with a PES mover should be counted as erroneous enumerations rather than correct enumerations since the PES person reported that they should have been counted elsewhere on census day. This would affect differential undercounts between jurisdictions to the extent that people move to some types of areas more than they move to other types of areas, but that type of analysis is absent from Breiman's paper. The issue was addressed by Mulry and Spencer (1993), who wrote:

> The accuracy of the adjustment is affected by many kinds of errors. Some of the errors cancel, others do not.... To decide whether adjustment improves accuracy, it is necessary to consider net error.... There is no guarantee that all kinds of error can be identified, or even that the identified errors' moments are accurately estimated. For example, the original total error analysis failed to detect a large processing error that occurred when the wrong computer program was used to edit clerical match codes. Underestimating errors in the adjustment does not necessarily imply overestimation of the accuracy of the adjustment relative to the census, however.

Breiman's attempt to stretch the left endpoint of an interval estimate for the national undercount rate is at odds with his stated view that the issue of differential undercount is paramount.

When Breiman does focus on differential undercount in his Table 16, his commentary reflects the widespread view that the issue of differential undercount of minorities is important. Whey, then, does he not label evaluation poststrata as minority or nonminority? To allow readers to see the differential minority undercount, we reproduce Breiman's Table 16 with information on minority status added (see Table 1). Other authors (e.g., Mulry and Spencer, 1993; Zaslavsky, 1993a) add interval estimates in addition to labeling minority status.

TABLE 1

Minority evaluation strata	Total corrected undercount	Nonminority evaluation strata	Total corrected undercount
1	3.7	2	−2.5
3	2.6	4	−1.6
5	1.5	6	0.3
8	3.0	7	0.6
11	3.9	9	−0.5
		10	−0.7
		12	3.1*
		13	−0.2

*Might be 1.1.
Source: Breiman (1994, Table 16).

Even after accounting for the computer error, the Census Bureau's loss function calculations favor the adjusted over the unadjusted counts at the state level (Mulry, 1992a,b; Fay and Thompson, 1993) and at other less aggregated levels as well (Mulry, 1992a,b; Mulry and Spencer, 1993). In fact, the reanalysis of corrected data in Zaslavsky (1993a, b) is generally more favorable to adjustment than the analysis with uncorrected data. Zaslavsky considers several strategies for combining census counts and dual system estimates (DSE) of poststratum population totals (i.e., adjusted counts based on both census and PES data) to minimize the estimated risk of the estimator for a given loss function; he evaluates the performance of these strategies under various loss functions, such as the absolute-error and size-weighted squared relative error (SWSRE) criteria mentioned in our Section 2.6. Using the data set developed for the proposed 1991 adjustment, Zaslavsky finds that the linear combination between census and DSE minimizing the SWSRE estimated risk attaches 12.3% weight to the census and 87.7% weight to the DSE; using a revised data set that corrected the computer coding error and included other modifications (Mulry, 1992b), a similar analysis suggests that the SWSRE estimated risk is minimized when negative weight is attached to the census.

5.3 Matching Error and Another Contradiction

Breiman states that net undercount should be measured roughly as a difference between omissions and erroneous enumerations, allowing some errors in the census to be canceled by other errors in the census. However, in the PES, particularly in his

discussion of matching error, Breiman does not take into account that errors can cancel. Why, for example, does Breiman report matching error results primarily in terms of "disagreement rates"? If that is the standard approach, then the disagreement rate between the census and the PES is not his 2.1% net undercount figure, but more like 5.6% (General Accounting Office, 1991).

False matches and false nonmatches have biasing effects in opposite directions, and the loss function analyses carried out by the Census Bureau attempt to account for such effects. Curiously, after criticizing PES matching for its high disagreement rates, Breiman then uses the Census Bureau's calculations of the effects of matching error in his Table 15.

5.4 Fabrication and Obfuscation

On the issue of fabrication, Breiman states:

> The report concludes [P6, page 15] that "Overall, between .9 and 6.5% of the interviewers were found to have high nonmatch rates. This compares favorably with the expectation that between 2 to 5% of interviewers are dishonest in their data collection." If these latter figures are anywhere near truth, then the DSE numbers overestimate the undercount by millions of persons.

If Breiman's "millions of persons" had any basis in reality, then PES follow-up results would have been replete with fictitious PES enumerations. In fact, PES follow-up revealed 86 individuals who were fictitious enumerations in households where nonfictitious enumerations were recorded and 138 whole households that were fabricated out of 377,381 individuals in the P-sample in 171,390 housing units. Needless to say, fictitious enumerations detected in PES follow-up are not included in dual system estimates of the population. They do not inflate counts; they prevent the counts from being inflated.

In addition, quality-control checks suggesting 2–5% interviewer fabrication does not translate to 2–5% of cases entered in the P-sample database being fabrications. Indeed, the evidence suggests that undetected fictitious enumerations in the PES were a small fraction of detected fictitious enumerations (West, 1991), and rough calculations assuming average weights suggest that detected fictitious enumerations prevented an inflation of the population total by some tens of thousands of individuals. Thus, using Breiman's calculations, the quality control operations in the district offices helped prevent the dual system estimates from being inflated by "millions of people." However, despite his rating of the PES in his conclusion as "excellent judged only as a sample survey and matching operation," Breiman discounts the possibility that quality control at the original interviewing stage, PES follow-up interviewing and evaluation follow-up interviewing would have been able to catch fictitious enumerations.

5.5 Correlation Bias and the Weight to Attach to Imprecise Estimates of Nonsampling Errors

Although Breiman criticizes the [P16] total error analysis for ignoring sources of nonsampling error, he then states:

> Report [P16] gives a second set of undercount estimates that includes an additional error source called model bias, more commonly known as correlation bias. This is the bias ascribed to the existence of persons unreachable by any survey. However, because these bias estimates are (and must be) based on highly speculative assumptions and have only a tenuous connection with any data, they are not included in our discussion.

Correlation-bias estimates have more than a "tenuous connection with any data"; on the contrary, estimates are based on combining PES data with evidence from vital records about the ratio of the size of the male population to that of the female population. These data are used to estimate a parameter that characterize dependence in omission rates between the census and PES (Bell, 1993).

Bell (1993) does point out that assumptions underlying these estimates are speculative, so that Breiman's treatment of attaching no weight to the results, or, equivalently from a subjectivist point of view, attaching an infinite standard error to correlation bias estimates, may not be completely unreasonable. We would caution, however, that Bell (1993) and other empirical studies of correlation bias such as Zaslavsky and Wolfgang (1993) suggest that correlation bias is appreciable; in Bell's words, "assuming independence (no correlation bias) is even more restrictive [than alternative estimators of correlation bias] and appears to be refuted for adult males by the data (subject to the limitations of data quality, including those discussed in this section)." We also read Bell's Table 3 to imply that the primary evidence of differential correlation bias appears to be in the 20–29-year-old age group, with the differential bias going in the direction that suggests even higher black undercount.

In sum, less accurate bias estimates should be given less weight in loss calculations, whether formal or informal. Using this standard, Zaslavsky (1993a)

provides a more reasoned approach to the reflection of evidence about biases than does Breiman's paper, which appears to treat some bias estimates as having infinite standard errors and other bias estimates as having zero standard errors.

5.6 Breiman's Table 13 as "Bad Data"

Breiman's Table 13, which originally appeared in a Census Bureau report, contains a flaw that makes the table's column headings not directly comparable with the entries in the interior of the table. This flaw was pointed out in Census Bureau memoranda written by Belin to senior Census Bureau officials (Belin, 1991a,b). Prior to receiving these memoranda, there had been plans to raise in the Undercount Steering Committee report the same concern Breiman has raised with his Table 13; however, based on the information in Belin's memoranda, a decision was made not to mention the issue. Because the details are cumbersome, we postpone a full discussion to the Appendix.

Breiman may not have known about the relevant internal Census Bureau documents when he was writing earlier drafts of the article, but Belin was contacted by one of the referees of Breiman's paper on its initial review and gave the reviewer copies of these memoranda. Our understanding is that the reviewer mentioned the flaw in Table 13 in a referee's report to Breiman. Breiman appears to have made no effort to address this criticism, as the table itself is unchanged from the earlier version of his paper, and the text, which is quoted in the Appendix, is identical in substance.

5.7 The Prism through which Breiman Views Tables 11 and 12

A *Statistical Science* article is not a place to play games with the base of percentages. An informative comparison in Breiman's Tables 11 and 12 is between the imputed probabilities and the fraction of cases resolved in EFU that were found to have been included in the census. This comparison yields an average predicted probability of 32.2% for PES cases that were originally unresolved but later resolved in the EFU as compared to a proportion of 31.6% for the enumeration rate among resolved EFU cases (Belin et al., 1993). Some of the decimal places of accuracy are lost in Breiman's rounding, but it is plain enough from Table 11 that $12/(12 + 27) = 30.8\%$, so rounding error does not appear to be the problem here. A similar perspective emerges from our Table 2.

Belin et al. (1993) provide a more detailed analysis of the same EFU data that includes appropriate predictive standard errors, demonstrating good agreement between imputed probabilities and enumeration rates from the EFU not only in the aggregate but also across important subgroups. Wachter (1993b) criticized the interpretation of Belin et al. (1993) because the EFU-resolved cases do not amount to a random subsample of the EFU sample, which generated discussion of the value of the EFU data in the presence of nonresponse, but Wachter did not deny that the agreement between what was observed and what was predicted was extremely good. Breiman calls the interpretation of Belin et al. (1993) "optimistic."

TABLE 2
Data from Breiman's Table 12 for resolved evaluation follow-up (EFU) cases

Imputed probability of having been in census	0–25%	25–50%	50–75%	75–100%
Fraction of resolved EFU cases determined to have been in census	12%	33%	53%	68%

Regarding Table 12, the 1992 version of Breiman's paper reads: "Note that the percent in the match category is virtually constant from an imputed probability of 25% to 100%. One would do just as well by assigning each person to match or other than match by tossing a coin with probability 15% of heads." The current version reads: "As the match probabilities increase, the proportion of resolved cases that result in matches increases. Another, more curious, trend also seems to be present—the higher the model match probabilities, the higher the proportion of unresolved persons the matching team finds." The latter is an improvement, but it seems to us that Breiman is grasping for something to criticize.

For us, the important question raised by this discussion is not about imputation methods. The real question is whether the statistics profession will require Census Bureau statisticians to respond to every criticism raised by a professional outside the Census Bureau, no matter how questionable, before sanctioning some kind of adjustment methodology.

5.8 Breiman's "Unresolved Issues" with the Census Bureau

Breiman attempts to use the fact that he circulated a draft of his paper at the Census Bureau to corroborate his conclusions: "The results of this paper were circulated to the Census Bureau in March 1992. In the reviews there was agreement with the numbers given in Table 15 except, perhaps, in two areas...[that] are discussed in the Appendix." In the Appendix, he discusses "census day address error," which refers to cases where the PES incorrectly recorded an individual's residence as of April 1, 1990, implying that the PES matching operation may have been searching in the wrong location for a census

enumeration. Such PES cases may then be incorrectly designated as nonmatches to the census, which would inflate undercount estimates. In his interaction with Bureau staff, Breiman questioned the formula for computing census day address error, and he asserts that the matter was never resolved:

> In December 1991, I was informed that the formula was incorrect, that the tables in [P4] were wrong and that the census day address error computation in [P4] and [P16] included all of the errors found in P-sample reinterviews regardless of whether they were census day address error or not. This was surprising, since this error is consistently referred to both in [P4] and [P16] as census day address error. Furthermore, some of the tables in [P4] are reproduced in the published paper titled "Address reporting error in the 1990 post-enumeration study" (West, Mulry, Parmer and Petrick, 1991).

Breiman then states that he was "unable to obtain from the Bureau any more specific information regarding their method for computing census day address error."

Our conversations with Census Bureau officials lead us to conclude that some of Breiman's assertions are misleading. First, Breiman's inability to obtain the information he desired should not be interpreted as suggesting that the Census Bureau was uncooperative. John Thompson (1994) of the Census Bureau explained to us:

> Leading up to the lawsuit, we transmitted a lot of information to Leo and others. To my knowledge, we had provided the answers to all of his requests, and I was unaware that he was waiting for anything. I thought we had agreed with Leo that there was a mistake in the earlier documentation that had been corrected in the documentation of the 1992 adjustment decision.

Mary Mulry (1994) of the Census Bureau echoed this sentiment, pointing out that the original calculation was correct and that it was just the documentation that was wrong. The effect of census day address error on undercount estimates goes in one direction, but other PES data collection errors could have effects in either direction. As Breiman notes, the Bureau had included all sources of data collection error in its original calculations but had summarized them under the label "census day address error." Responding to input from Breiman, Bureau staff changed their documentation by substituting "error in address reporting and other P-sample data collection errors" (Mulry, 1992b) for what had previously been labeled "census day address error." Thus, Breiman's inability to obtain "more specific information regarding their method for computing census day address error" appears to have resulted from Bureau officials deciding that separate calculations for each type of error were inessential so long as the accounting of cases was complete and the documentation of methods was accurate. Breiman cites the report by Mulry (1992b) in his paper, but he ignores the change in the documentation.

Clearly, having the ability to reproduce results is crucial, and questions arise as to appropriate standards for documentation. Is it enough that a Census Bureau employee familiar with detailed codes and jargon can reproduce results, or must the documentation be understandable to outside professional statisticians? What is the significance of a document that is in error if the actual calculation that was carried out is correct?

Many aspects of the 1990 adjustment procedure were rushed, so for the future we need to be realistic in assessing how long various operations take. Meanwhile, Breiman's criticism notwithstanding, our view is that the 1990 adjustment procedure was acceptably documented. We also think that it would have been better for Breiman to resolve his dispute somewhere other than in these pages.

Also, the Census Bureau's corroboration of Breiman's numbers should not be interpreted as Census Bureau endorsement of his conclusions. For example, according to the July 20, 1992, minutes of the Committee on Adjustment of Postcensal Estimates (CAPE) at the Census Bureau, that meeting was devoted to discussing a 13-page critique of an earlier version of Breiman's work by Hogan (1992b). Thompson (1994) also said: "His numbers appear to be correct, but the inferences are Leo's interpretation of the data. The main point I'd like to make is that if journals publish criticisms of the Census Bureau, then they should offer the Census Bureau an opportunity to comment." The Census Bureau is a professional agency. Ultimately, we do not believe that Breiman does justice to the Census Bureau in his paper.

6. YEAR 2000: WILL CONSENSUS EMERGE AROUND A "ONE-NUMBER CENSUS"?

The Census Bureau is planning for the year 2000 census and beyond (Holmes, 1994). Tortora, Miskura and Dillman (1993) summarize the design options from the perspective of senior Bureau officials. Current plans call for selecting the design of the next census by December 1995. A primary goal is to de-

sign a "one-number census," which Tortora, Miskura and Dillman (1993) define in the following way (emphasis in original):

> The term "One-Number Census" names the concept that *the decennial census is designed to produce the best possible single set of results by legal deadlines, and that those results are based on an appropriate combination of counting, assignment, and statistical estimation techniques.*

The evolution of the one-number census flows in many ways from the adjustment experience in 1990. We briefly describe some of the connections.

Apart from the technical issues we have discussed, the political reality is that there cannot be two population numbers, no matter how good either one of them is, without politicizing the process. In 1991, we saw the Secretary of Commerce being required to choose one set of numbers or the other after the consequences of his decision were apparent, a scenario that would open any decision to the charge of being politically motivated. As Barbara Everitt Bryant, the director of the Census Bureau during the adjustment controversy over the 1990 census, recently wrote (Bryant, 1993):

> It becomes clear from the experience of 1990 that a two-number census did not work. With a 200-year history of never adjusting the census, the production of two numbers made one number the adversary of the other, with the consequent demand for statistically impossible proof that one number is better than the other at all levels at which data are produced.

The statistical community may think of adjustment as a technical issue, and policymakers may endorse this view so long as the technical decisions are made in advance of knowing their effects. The concept of a one-number census recognizes this imperative.

Although Tortora, Miskura and Dillman (1993) do not require that "statistical estimation techniques" be used in order for a census to qualify as a one-number census, current plans for the 1995 test census suggest that statistical methods are high on the Census Bureau's research agenda. The debate over adjustment is thus imbedded in the consideration of a one-number census, with one distinction being that "adjustment" in a one-number census is built into the census design and is viewed as an integral part of the census. One reflection of how the thinking, or at least the nomenclature, has evolved at the Bureau is that the paper by Tortora. Miskura and Dillman does not contain the word "adjustment."

Tortora, Miskura and Dillman (1993) are aware that there are some major hurdles to clear before a one-number census could become reality, which these Bureau officials pose as an opportunity and a challenge to statisticians outside the Bureau:

> We have proposed that the 1995 Census Test be a one-number census.... Determining how to do this combines some intriguing technical, perceptual, and operational issues. One set of issues involves the estimation-based coverage improvement methods to be used.... Another set of issues has to do with determining if the new one-number approach, which will rely less on counting techniques and more on estimation than the 1990 census, yields the desired improvement.... What measures are needed to make a convincing case? Is it sufficient to show that we have improved the relative coverage of population groups; or is it important to also show that the "proportionate shares" of the total population across all areas are improved? In either case, what is the geographic level at which such improvement must be shown?... We are currently discussing these issues within the Census Bureau and invite the input of the technical community on them.

Along with the political imperative mentioned above, the push for a one-number census appears also to be based on other nonstatistical considerations, such as operational feasibility. Time constraints imposed by legal deadlines would seem to eliminate the possibility of carrying out a 1990-style PES and EFU program and then allowing enough time to deliberate over a comparison of an "unadjusted" and an "adjusted" census. Bob Fay, the Senior Mathematical Statistician at the Bureau, described recently in the following exchange with Tom Belin (Fay, 1993):

> Fay: The state of science will not be such in the year 2000 that we can achieve consensus by looking at a lot of evaluation data over a four-week period and deciding how to adjust the census. It now seems that if we're going to make this change we'll have to just decide ahead of time to live with something, and then, go back and look and see what we've done over a two-year period. But in fact, even the notion of guidelines or the two-track approach may not be the way the society decides to adjust its data.

> Belin: In other words, the type of guidelines, if you will, that you're imagining would say, "Just do it" and then, review it.
>
> Fay: Yes. And in some ways I think Barbara Bailar and Kirk Wolter were ahead of me on this by ten years.... The only subtle difference is to design the census around the idea that it has to be ready within the usual scheduled timetables for the census rather than delay the census eight months or longer.

Another compelling consideration, mentioned by Bryant (1993), is the desire to avoid tying up the top statisticians at the Census Bureau in litigation.

A number of important voices have registered support for the idea of a one-number census. The report of the Panel to Evaluate Alternative Census Methods (Committee on National Statistics, 1993b, page 2) states:

> One key message is that the dual objectives of reducing the differential undercount and controlling costs will require expanded use of sampling and statistical estimation.... We endorse the Census Bureau's stated goal of achieving a one-number census in 2000 that incorporates the results from coverage measurement programs, including programs involving sampling and statistical estimation, into the official census population totals. We recommend that research on alternative methodologies continue in pursuit of this goal.

Bryant (1993) also endorses the idea of a one-number census in forceful terms:

> The 1990 experience also demonstrated that differential undercount cannot be reduced by more intensive enumeration efforts. It is not possible to count 100% of the population, although the public and many elected officials have an unrealistic expectation that this can be done. Combining statistical estimation of those missed with the best reasonable effort to enumerate all in a one-number census should improve overall accuracy.... The census has changed evolutionarily with the times. Now is the time to make a larger than evolutionary leap.

To facilitate discourse, terms like "overall accuracy" should be defined to distinguish, for example, between "state-level" and "block-level" accuracy, as we discuss in Section 2.7. However, there appears to be a growing consensus around the idea of integrating statistical estimation into the census-taking process instead of generating two sets of counts and deciding between them. Hopefully, this confluence of views will create momentum for statisticians with widely varying perspectives to reach the kind of consensus envisioned by the National Academy panel in the mid-1980's.

7. CONCLUSION

It is true that there are a great many sources of error in adjustment and that the details matter greatly. Not all sources of uncertainty in the proposed adjustment of the 1990 census were quantified, and there was a major processing error discovered after senior Census Bureau officials had already recommended adjustment. Freedman and Wachter zero in on one of the fundamental issues about which there is genuine philosophical disagreement, and despite their own leanings against adjustment they have produced analyses that help clarify a technical issue of great interest. This is constructive. On the other hand, Breiman is not careful with his facts, and one of the most harmful aspects of Breiman's paper, we believe, is that Breiman has cultivated a sense of distrust of government statisticians that we regard as unjustified and irresponsible.

Ultimately, we would press our colleagues to embrace the idea of a "negotiated settlement" to the adjustment controversy. Crucial design decisions need to be made in the next few years. As a profession, we should press policymakers in Congress to exercise their prerogative and responsibility to, as the Constitution puts it, "direct" the census. By this we mean that Congress should state what it needs from the Census Bureau and should support the best professional judgment of the Census Bureau on how to get there. If present trends in designing the next census continue and get widespread support from the statistics community, we may be able to avoid full-blown court cases after the year 2000 census that would waste time, money and the energies of outstanding members of the profession. To statisticians who pride themselves on improving society by avoiding wasteful public and private spending, this looms as a calling.

APPENDIX: BREIMAN'S TABLE 13 AS "BAD DATA"

Breiman's Table 13 on the accuracy of E-sample imputation originated in the preliminary report for project P3 at the end of May 1991 (Gbur, 1991b). The preliminary report also contains a table summarizing evidence about P-sample imputation, which Breiman has reproduced as his Table 12. Belin, who

played a prominent role in producing the imputed probabilities, was still working at the Census Bureau at the time the preliminary P3 report was circulated. At the time, the P-sample results were generally received by senior Census Bureau officials as being supportive of the imputation methodology, contrary to Breiman's interpretation of the results. However, the E-sample results were a source of concern among Bureau officials, much as they are in Breiman's account, largely due to the much greater than expected fraction of individuals classified as correctly enumerated from the EFU who received imputed probabilities of 50% or less of having been correctly enumerated. Belin was told that the Undercount Steering Committee, which consisted of senior Bureau officials, was planning to include a comment about the discrepancy in the E-sample results in their final report. This is important because the absence of any mention of this discrepancy in Undercount Steering Committee (1991) confirms that senior Bureau officials agreed with Belin's analysis of the situation at the time.

Belin and his colleagues quickly identified a flaw in the construction of the E-sample table. PES matching rules for E-sample cases that were unmatched in before-follow-up matching declare that the E-sample case is a correct enumeration only if three conditions hold: a follow-up interview must establish the person to exist, the E-sample address has to be correctly assigned to the sample block or to an adjacent block, and the individual cannot be duplicated in a search area defined as a ring of three "surrounding blocks." The latter condition is required because otherwise it would remain unclear whether the correctly enumerated individual was the one in the sample block or the one in the surrounding block. To reflect this uncertainty in the case of a duplicate in a surrounding block, the standard Census Bureau procedure, as described in Belin et al. (1991), was to assign an imputed probability of correct enumeration of 50% when an E-sample case had one surrounding block duplicate, 33.3% for an E-sample case that matched two other census records in surrounding blocks, and so on.

The preliminary P3 report, according to its author, placed all E-sample cases having surrounding block duplicates in the 0–50% category in the table. However, EFU operations were only concerned with the first of the three conditions mentioned above; there had been no review of whether the E-sample case in the sample block was "the" correct enumeration as opposed to one of the surrounding block duplicates being "the" correct enumeration. Another contribution to the imputed probability of correct enumeration, namely, the possibility of incorrect assignment of residences to geographical locations, known as geocoding error, was also not part of the EFU review. Almost 90% of the cases that appear in the 0–50% category of what is now Breiman's Table 13 were attached to surrounding block duplicates.

Belin (1991a, b) wrote two memoranda to Census Bureau superiors on this matter. Using the same data set as was used to generate the preliminary P3 report, he found that the average imputed probability for individuals in the 0–50% column, excluding the contributions of surrounding block duplicates and possible geocoding error, was 88.7%. The correct enumeration rate among resolved cases from the EFU was $67/(67 + 10) = 87.0\%$.

As noted earlier, there is no mention of the discrepancy in the E-sample table in Undercount Steering Committee (1991). However, the author of the P3 preliminary report, despite receiving copies of both of Belin's memoranda, unfortunately left the table unedited in the final P3 report (Gbur, 1991c). It is understandable both that Breiman would have taken the table to support the position he was set to argue in court and that he would not have been aware of Belin's memoranda, which, although included in the administrative record of the adjustment decision, amounted to only five pages out of thousands.

During the review of Breiman's article for *Statistical Science*, an anonymous referee who knew that Belin had worked on imputation in the PES contacted Belin to discuss Table 13. Belin provided copies of the memoranda, and we understand that the referee conveyed an explanation of this issue to Breiman.

Between Breiman (1992) and Breiman (1994), Table 13 did not change, and the discussion surrounding it changed from, "There is no sign of a correlation between the probabilities computed by the imputation model and the enumeration status as determined by the EFU reinterview information" to "There is no evidence here of an association between the probabilities computed by the imputation model and the enumeration status as determined by the EFU reinterview information." Belin wrote the Executive Editor of *Statistical Science* to make him aware of the situation and to request an opportunity to respond, which grew into our current collaboration.

ACKNOWLEDGMENTS

We are grateful to Bob Fay, Mary Mulry and John Thompson for helpful discussion and for responding to requests for information. We also thank Norman Bradburn, Louis Gordon, Rob Kass, Don Ylvisaker and two referees, whose comments greatly improved the text.

REFERENCES

ALHO, J. M., MULRY, M. H., WURDEMAN, K. and KIM, J. (1993). Estimating heterogeneity in the probabilities of enumeration for dual-system estimation. *J. Amer. Statist. Assoc.* **88** 1130–1136.

ANDERSON, M. J. (1988). *The American Census: A Social History*. Yale Univ. Press.

ANOLIK, I. (1988). The rural post-enumeration survey in east central Mississippi, Statistical Research Division Report, Series RR 88/10, Bureau of the Census, Washington, D.C.

ANOLIK, I. (1989). The 1987 Post-Enumeration Survey. In *Proceedings of the Survey Research Methods Section* 710–715. Amer. Statist. Assoc., Alexandria, VA.

ANOLIK, I. (1990). Matching movers in a post-enumeration survey: results from 1988 and plans for 1990. In *Proceedings of the Survey Research Methods Section* 536–540. Amer. Statist. Assoc., Alexandria, VA.

ANOLIK, I. and HOGAN, H. (1991). Non-response conversion: the experience of the 1990 Post-Enumeration Survey. In *Proceedings of the Survey Research Methods Section* 658–663 (P3). Amer. Statist. Assoc., Alexandria, VA.

BATEMAN, D., CLARK, J., MULRY, M. and THOMPSON, J. (1991). 1990 Post-Enumeration Survey evaluation results. In *Proceedings of the Social Statistics Section* 21–30. Amer. Statist. Assoc., Alexandria, VA.

BELIN, T. R. (1991a). Interpretation of results from evaluation project P3. Statistical Support Division, Bureau of the Census, Washington, D.C. (June 5, 1991.)

BELIN, T. R. (1991b). Comparison of E-sample results from P3 evaluation of production imputation results. Statistical Support Division, Bureau of the Census, Washington, D.C. (June 26, 1991.)

BELIN, T. R. (1993). E-mail correspondence to D. Freedman, August 31, 1993.

BELIN, T. R. and DIFFENDAL, G. J. (1991). Results from the handling of unresolved enumeration status, missing characteristic data, and noninterviews in the 1990 Post-Enumeration Survey. STSD Decennial Census Memorandum Series V-112. Bureau of the Census, Washington, D.C.

BELIN, T. R., DIFFENDAL, G. J., MACK, S., RUBIN, D. B., SCHAFER, J. L. and ZASLAVSKY, A. M. (1993). Hierarchical logistic-regression models for imputation of unresolved enumeration status in undercount estimation (with discussion). *J. Amer. Statist. Assoc.* **88** 1149–1166.

BELIN, T. R., DIFFENDAL, G. J., RUBIN, D. B., SCHAFER, J. L. and ZASLAVSKY, A. M. (1991). Documentation of handling of unresolved enumeration status in 1990 census/Post-Enumeration Survey. STSD Decennial Census Memorandum Series V-98. Bureau of the Census, Washington, D.C.

BELL, W. R. (1991). Standardized residuals from regression models for variances of adjustment factors. Statistical Research Division, Bureau of the Census, Washington, D.C. (May 16, 1991.)

BELL, W. R. (1993), Using information from demographic analysis in Post-Enumeration Survey estimation. *J. Amer. Statist. Assoc.* **88** 1106–1118.

BIEMER, P. and STOKES, S. (1989). The optimal design of quality control samples to detect interviewer cheating. *Journal of Official Statistics* **5** 23–39.

BREIMAN, L. (1992). The 1991 census adjustment: Undercount or bad data? Preliminary version, technical report, Dept. Statistics, Univ. California, Berkeley.

BREIMAN, L. (1994). The 1991 census adjustment: Undercount or bad data? *Statist. Sci.* **9** 458–475.

BRYANT, B. E. (1991). Recommendation to Secretary of Commerce Robert A. Mosbacher on whether or not to adjust the 1990 census. Bureau of the Census, Washington, D.C. (June 1991.)

BRYANT, B.E. (1992). Decision of the director of the Bureau of the Census on whether to use information from the 1990 Post-Enumeration Survey (PES) to adjust the base for the Intercensal population estimates produced by the Bureau of the Census. *Federal Register* Docket 920895–2347, December 29, 1992.

BRYANT, B. E. (1993). Guest commentary. Census-taking for a litigious, data-driven society. *Chance* **6** (Summer) 44–49.

BUREAU OF THE CENSUS (1992a). Request for comments on the proposed options for incorporating information from the Post-Enumeration Survey into the intercensal population estimates produced by the Bureau of the Census. *Federal Register* **57** 35,562–35,564 (August 10).

BUREAU OF THE CENSUS (1992b). Assessment of accuracy of adjusted versus unadjusted 1990 census base for use in intercensal estimates. Report of the Committee on Adjustment of Post Censal Estimates, (August 7) Suitland, MD.

BUREAU OF THE CENSUS (1992c). Decision on whether to incorporate information from the Post-Enumeration Survey into the base for intercensal population estimates. *Federal Register* **57** 42,939–42,940 (September 17).

BUREAU OF THE CENSUS (1992d). Additional research on accuracy of adjusted versus unadjusted 1990 census base for use in intercensal estimates. Addendum to the Report of the Committee on Adjustment of Post Censal Estimates, Suitland, MD (November 28).

BUREAU OF THE CENSUS (1993). Decision of the director of the Bureau of the Census on whether to use information from the 1990 Post-Enumeration Survey (PES) to adjust the base for the intercensal population estimates produced by the Bureau of the Census. *Federal Register* **58** 69–77. (January 4).

CAIN, B. (1992). Exhibit for *New York et al. v. Commerce Department et al.* U.S. District Court, Eastern District of New York.

CHILDERS, D., DIFFENDAL, G., HOGAN, H., SCHENKER, N. and WOLTER, K. (1987). The technical feasibility of correcting the 1990 census In *Proceedings of the Social Statistics Section* 36–45. Amer. Statist. Assoc., Alexandria, VA.

CHILDERS, D. R. and HOGAN, H. (1989a). The 1988 Post Enumeration Survey, methods and preliminary results. Statistical Research Division, Bureau of the Census, Washington, D.C. (October 11, 1989.)

CHILDERS, D. R. and HOGAN, H. (1989b). The 1988 Post Enumeration Survey, methods and preliminary results. In *Proceedings of the Survey Research Methods Section* 685–689. Amer. Statist. Assoc., Alexandria, VA.

CHILDERS, D. R. and HOGAN, H. (1990). Results of the 1988 dress rehearsal post enumeration survey. In *Proceedings of the Survey Research Methods Section* 547–552. Amer. Statist. Assoc., Alexandria, VA.

CHOLDIN, H. M. (1994). *Looking for the Last Percent: The Controversy over Census Undercounts*. Rutgers Univ. Press.

CITRO, C. F. and COHEN, M. L., eds. (1985). *The Bicentennial Census: New Directions for Methodology in 1990*. National Academy Press, Washington, D.C.

COHEN, M. L. (1990). Adjustment and reapportionment—analyzing the 1980 decision. *Journal of Official Statistics* **6** 241–250.

COHEN, M. L. (1991). Choice and conflict about census data: adjusting the American census count. *Journal of Public Policy* **11** 357–398.

COMMITTEE ON NATIONAL STATISTICS (1993a). Planning the decennial census: interim report. Report of the Panel on Census Requirements in the Year 2000 and Beyond, National Research Council, Washington, D.C.

COMMITTEE ON NATIONAL STATISTICS (1993b). A census that mirrors America: interim report. Report of the panel to Evaluate Alternative Census Methods. National Research Council, Wahington, D.C.

CRESSIE, N. (1992). REML estimation in empirical Bayes smoothing of census undercount, *Survey Methodology* **18** 75–94.

DARROCH, J. N., FIENBERG, S. E., GLONEK, G. F. V. and JUNKER, B. W. (1993). A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. *J. Amer. Statist. Assoc.* **88** 1137–1148.

DATTA, G. S., GHOSH M., HUANG, E. T., ISAKI, C. T., SCHULTZ, L. K. and TSAY, J. H. (1992). Hierarchical and empirical Bayes methods for adjustment of census undercount: the 1988 Missouri dress rehersal data. *Survey Methodology* **18** 95–108.

DAVIS, M. C., MULRY, M., PARMER, R. and BIEMER, P. (1991). The matching error study for the 1990 post enumeration study. In *Proceedings of the Survey Research Methods Section* 248–253. Amer. Statist. Assoc., Alexandria, VA.

DEPARTMENT OF COMMERCE (1989). *Federal Register* **54**(238) 51,002–51,004.

DEPARTMENT OF COMMERCE (1991a). *Department of Commerce News* June 13.

DEPARTMENT OF COMMERCE (1991b). Decision on whether or not a statistical adjustment of the 1990 decennial census of population should be made for coverage deficiencies resulting in an overcount or undercount of the population. Docket No. 91281–1181, July 15, 1991. Office of the Secretary of Commerce, Washington, D.C. *Federal Register* **56** 33,582–33,642 (July 22nd).

DIFFENDAL, G. (1988). The 1988 test of adjustment related operations in Central Los Angeles County. *Survey Methodology* **14** 71–86.

DIFFENDAL, G. and BELIN, T. (1991). Results of procedures for handling noninterviews, missing characteristic data, and unresolved enumeration status in 1990 census/Post-Enumeration Survey. STSD Decennial Census Memorandum Series #V-112, Bureau of the Census, Washington, D.C.

ERICKSEN, E. P. and DEFONSO, T. K. (1993). Beyond the net undercount: how to measure census error, *Chance* **6** (Fall) 38–43.

ERICKSEN, E., ESTRADA, L., TUKEY, J. and WOLTER, K. (members, Special Advisory Panel) (1991). Report on the 1990 census and the Post Enumeration Survey. Report submitted to the secretary of the Department of Commerce, June 22, 1991.

ERICKSEN, E. P. and KADANE, J. B. (1985). Estimating the population in a census year (with discussion). *J. Amer. Statist. Assoc.* **80** 98–131.

ERICKSEN, E. P., KADANE, J. B. and TUKEY, J. W. (1989). Adjusting the 1980 census of population and housing. *J. Amer. Statist. Assoc.* **84** 927–943.

FAY, R. E. (1992a). Inferences for small domain estimates from the 1990 Post-Enumeration Survey. Technical report, Bureau of the Census, Washington, D.C.

FAY, R. E. (1992b). Personal communication (presentation in a meeting during Census Bureau Annual Research Conference).

FAY, R. E. (1993). Personal communication (interview with T. Belin).

FAY, R. E., PASSEL, J. S., ROBINSON, J. G. and COWAN, C. D. (1988). *The Coverage of the Population in the 1980 Census*. U.S. Department of Commerce, Washington, D.C.

FAY, R. E. and THOMPSON, J. H. (1993). The 1990 Post Enumeration Survey: statistical lessons in hindsight. In *Proceedings of the 1993 Annual Research Conference* 71–91. Bureau of the Census, Washington, D.C.

FIENBERG, S. E. (1989–1992). An adjusted census in 1990? Assorted articles. *Chance* **2–5**.

FIENBERG, S. E. (1992a). Comment on "Should we have adjusted the U.S. census of 1980?" by D. A. Freedman and W. C. Navidi. *Survey Methodology* **18** 25–28.

FIENBERG, S. E. (1992b). Bibliography on capture–recapture modelling with application to census undercount adjustment. *Survey Methodology* **18** 143–154.

FIENBERG, S. E. (1993). The New York City adjustment trial: witness for the Plaintiffs. *Jurimetrics* **34** 65–84.

FIENBERG, S. E. (1994a). Ethical and modelling considerations in correcting the results of the 1990 decennial census. In *Ethics Modelling* (W. A. Wallace, ed.) 103–114. Pergamon Press, New York.

FREEDMAN, D. (1991). Adjusting the 1990 census. *Science* **252** 1233–1236.

FREEDMAN, D. A. (1992). Testimony in *New York et al. v. Commerce Department et al.* U.S. District Court, Eastern District of New York, 89 CV 3474.

FREEDMAN, D. A. (1993a). E-mail correspondence with T. Belin, September 5, 1993.

FREEDMAN, D. A. (1993b). Adjusting the census of 1990. *Jurimetrics* **34** 99–106.

FREEDMAN, D. and NAVIDI, W. (1986). Regression models for adjusting the 1980 census (with discussion). *Statist. Sci.* **1** 1–39.

FREEDMAN, D. and NAVIDI, W. (1992). Should we have adjusted the U.S. census of 1980? (with discussion). *Survey Methodology* **18** 3–74.

FREEDMAN, D. and WACHTER, K. (1994). Heterogeneity and census adjustment for the intercensal base. *Statist. Sci.* **9** 476–485.

FREEDMAN, D., WACHTER, K., COSTER, D., CUTLER, R. and KLEIN, S. (1993). Adjusting the census of 1990: the smoothing model. *Evaluation Review* **17** 371–443.

FREEDMAN, D., WACHTER, K., CUTLER, R. and KLEIN, S. (1994). Adjusting the census of 1990: loss functions. *Evaluation Review* **18** 243–280.

GBUR, P. M. (1991a). Missing data in the 1990 Post Enumeration Survey. In *Proceedings of the Survey Research Methods Section* 646–651. Amer. Statist. Assoc., Alexandria, VA.

GBUR, P. M. (1991b). PES Evaluation Project P3: evaluation of imputation methodology for unresolved match status cases. Preliminary version, Statistical Support Division, Bureau of the Census, Washington, D.C. (May 31, 1991.)

GBUR, P. M. (1991c). PES Evaluation Project P3: evaluation of imputation methodology for unresolved match status cases. Final version, 1990 Coverage Studies and Evaluation Memorandum Series #C-2, Statistical Support Division, Bureau of the Census, Washington, D.C. (July 12, 1991.)

GENERAL ACCOUNTING OFFICE (1991). The 1990 census: final preparations for a possible adjustment. Report GAO/T-GGD-91-26, June 19, 1991.

GRIFFIN, D. H. and WAJER, S. C. (1993). Results from the 1990 search match operation. In *Proceedings of the Survey Research Methods Section*. Amer. Statist. Assoc., Alexandria, VA. To appear.

HARTIGAN, J. A. (1992). Comment on "Should we have adjusted the U.S. census of 1980?" by D. A. Freedman and W. C. Navidi. *Survey Methodology* **18** 44–50.

HENGARTNER, N. and SPEED, T. P. (1993). Assessing between-block heterogeneity within the post-strata of the 1990 Post-Enumeration Survey (with discussion). *J. Amer. Statist. Assoc.* **88** 1119–1129.

HOGAN, H. (1989). Nine years of coverage evaluation research: What have we learned? In *Proceedings of the Survey Research Methods Section* 663–668. Amer. Statist. Assoc., Alexandria, VA.

HOGAN, H. (1990). The 1990 Post-Enumeration Survey: an overview. In *Proceedings of the Survey Research Methods Section* 518–523. Amer. Statist. Assoc., Alexandria, VA.

HOGAN, H. (1991). Downweighting outlier small blocks. STSD 1990 Decennial Census Memorandum Series #V-109. addressed to John Thompson. Chief, Statistical Support Division, Bureau of the Census. Washington, D.C. June 18, 1991.

HOGAN, H. (1992a). The 1990 Post-Enumeration Survey: an overview. *Amer. Statist.* **46** 261–269.

HOGAN, H. (1992b). Discussion of "The 1991 census adjustment: undercount or bad data" (June 1992 version) by L. Breiman. Minutes of the Committee on Adjustment of Postcensal Estimates, Bureau of the Census, Washington, D.C. (July 20, 1992.)

HOGAN, H. (1993). The 1990 Post-Enumeration Survey: operations and results. *J. Amer. Statist. Assoc.* **88** 1047–1060.

HOGAN, H. and WOLTER, K. (1988). Measuring accuracy in a post-enumeration survey. *Survey Methodology* **14** 99–116.

HOLMES, S. A. (1994). Census officials plan big changes in gathering data. *The New York Times*, May 16, 1994, page A1.

ISAKI, C. T., SCHULTZ, L. K., DIFFENDAL, G. J. and HUANG, E. T. (1988). On estimating census undercount in small areas. *Journal of Official Statistics* **4** 95–112.

JARO, M. (1989). Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. *J. Amer. Statist. Assoc.* **84** 414–420.

KADANE, J. B. (1984). Allocating congressional seats among the states when state populations are uncertain. Technical Report 309, Dept. Statistics, Carnegie Mellon Univ.

KIM, J. J., BLODGETT, R. and ZASLAVSKY, A. (1993). Evaluation of the synthetic assumption in the 1990 Post-Enumeration Survey. Technical report, Bureau of the Census, Washington, D.C.

KRUSKAL, W. (1991). Recommendation to Secretary of Commerce Robert A. Mosbacher on adjustment of the 1990 census, June 1991.

MARTIN, E., BROWNRIGG, L. A. and FAY, R. E. (1990). Results of ethnographic studies of 1988 dress rehearsal census coverage. In *Proceedings of the Survey Research Section*, 530–535. Amer. Statist. Assoc., Alexandria, VA.

MEYER, M. M. and KADANE, J. B. (1992). Reconstructing the adjusted census for Florida: A case study in data examination. *Journal of Computational and Graphical Statistics* **1** 287–300.

MITROFF, I., MASON, R. and BARABBA, V. (1983). *The 1980 Census: Policymaking Amid Turbulence*. Heath, Lexington, MA.

MOSBACHER, R. A. (1991a). The decision of the secretary of commerce on whether a statistical adjustment of the 1990 Census of Population and Housing should be made for coverage deficiencies resulting in an overcount or undercount of the population. *Federal Register* **56** 140, 33,582–33,688, (July 22, 1991.)

MOSBACHER, R. A. (1991b). Testimony before the House Subcommittee on Census and Population, Committee on Post Office and Civil Service, Serial No. 102–21, July 16, 1991. U.S. Government Printing Office, Washington, D.C.

MULRY, M. H. (1992a). Loss function analysis for the Post Census Review (PCR) estimates. Decennial Statistical Studies Division, Bureau of the Census, Washington, D.C.

MULRY, M. H. (1992b). Total error of Post Census Review estimates of population. Decennial Statistical Studies Division. Bureau of the Census, Washington, D.C. (June 1992.)

MULRY, M. H. (1994). Personal communication, January 25, 1994.

MULRY, M. and DAJANI, A. (1989). The matching error study for the 1988 Dress Rehearsal. In *Proceedings of the Social Statistics Section* 704–709. Amer. Statist. Assoc., Alexandria, VA.

MULRY, M. and SPENCER, B. (1991). Total error in PES estimates of population (with discussion). *J. Amer. Statist. Assoc.* **86** 839–863.

MULRY, M. H. and SPENCER, B. D. (1993). Accuracy of the 1990 census and undercount adjustments. *J. Amer. Statist. Assoc.* **88** 1080–1091.

PLEWES, T. (1994). The redesigned current population survey. *Chance* **7** 35–41.

ROBINSON, J. G., AHMED, B., DAS GUPTA, P. and WOODROW, K. A. (1993). Estimation of population coverage in the 1990 United States census based on demographic analysis (with discussion). *J. Amer. Statist. Assoc.* **88** 1061–1079.

ROLPH, J. E. (1992). Testimony in *New York et al. v. Commerce Department et al.* U.S. District Court, Eastern District of New York, 89 CV 3474.

ROLPH, J. E. (1994). The census adjustment trial: reflections of a witness for the plaintiffs. *Jurimetrics* **34** 85–98.

ROYCE, D. (1992). A comparison of some estimators of a set of population totals. *Survey Methodology* **18** 109–126.

SCHAFER, J. L. (1993). Comment on "Assessing between-block heterogeneity within the post-strata of the 1990 Post-Enumeration Survey" by N. Hengartner and T. P. Speed. *J. Amer. Statist. Assoc.* **88** 1125–1127.

SCHAFER, J. L. and SCHENKER, N. (1991). Variance estimation with imputed means. In *Proceedings of the Survey Research Methods Section* 696–701. Amer. Statist. Assoc., Alexandria, VA.

SCHENKER, N. (1988). Handling missing data in coverage estimation, with application to the 1986 test of adjustment related operations. *Survey Methodology* **14** 87–97.

SCHENKER, N., ed. (1993). Special section on the 1990 undercount. *J. Amer. Statist. Assoc.* **88** 1044–1166.

SCHIRM, A. L. (1991). The effects of census undercount adjustment on congressional apportionment. *J. Amer. Statist. Assoc.* **86** 526–541.

SCHIRM, A. L. and PRESTON, J. (1987). Census undercount adjustment and the quality of geographic population distributions (with discussion). *J. Amer. Statist. Assoc.* **82** 965–990.

SINGH, M. P., ed. (1992). Special section on census undercount measurement methods and issues. *Survey Methodology* **18** 1–154.

SPENCER, B. D. (1980a). *Benefit–Cost Analysis of Data Used to Allocate Funds. Lecture Notes in Statist.* **3** Springer, New York.

SPENCER B. D. (1980b). Implications of equity and accuracy for undercount adjustment: a decision-theoretic approach. In *Proceedings of the Conference on Census Undercount* 204–216. Bureau of the Census, Washington, D.C.

STEFFEY, D. L., ed. (1993). *A Census That Mirrors America*. National Research Council, Washington, D.C.

STEFFEY, D. L. and BRADBURN, N. M., eds. (1994). *Counting People in the Information Age*. National Research Council, Washington, D.C.

STOKES, L. and JONES, P. (1989). Evaluation of the interviewer quality control procedure for the Post-Enumeration Survey. In *Proceedings of the Survey Research Methods Section* 696–698. Amer. Statist. Assoc., Alexandria, VA.

THOMPSON, J. H. (1994). Personal communication, February 1, 1994.

TORTORA, R. D., MISKURA, S. M. and DILLMAN, D. A. (1993). Onward towards a 2000 census design: research results. In *Proceedings of the Survey Research Methods Section* 120–128. Amer. Statist. Assoc., Alexandria, VA.

TREMBLAY, A., STOKES, S. L. and GREENBERG, B. S. (1991). Estimation of PES fabrications from quality control data. In *Proceedings of the Survey Research Methods Section* 242–247. Amer. Statist. Assoc., Alexandria, VA.

TUKEY, J. W. (1983). Affidavit in *Cuomo v. Baldridge*, Civil Action 80-4550. U.S. District Court, Southern District of New York.

TUKEY, J. W. (1991). Testimony before the House Subcommittee on Census and Population, Committee on Post Office and Civil Service, Serial No. 102–19, June 27, 1991. U.S. Government Printing Office, Washington, D.C.

UNDERCOUNT STEERING COMMITTEE (1991). Technical assessment of the accuracy of unadjusted versus adjusted 1990 census counts, June 21, 1991. Bureau of the Census, Washington, D.C.

U.S. DISTRICT COURT (1991). Eastern District of New York (Judge Joseph M. McLaughlin), Memorandum and Order 88 CV 3474, 92 CV 1566, 92 CV 2037, April 13, 1993.

U.S. DISTRICT COURT (1992). *New York et al. v. Commerce Department et al.* U.S. District Court, Eastern District of New York (Judge Joseph M. McLaughlin), 89 CV 3474, 92 CV 1566, 92 CV 2037.

WACHTER, K. W. (member, Special Advisory Panel) (1991). Recommendations on 1990 census adjustment. Report submitted to the Secretary of Commerce, June 17, 1991.

WACHTER, K. W. (1993a). The census adjustment trial: an exchange. *Jurimetrics* **34** 107–115.

WACHTER, K. W. (1993b). Ignoring nonignorable effects. Comment on "Hierarchical logistic regression models for imputation of unresolved enumeration status in undercount estimation" by T. R. Belin, G. J. Diffendal, S. Mack, D. B. Rubin, J. L. Schafer and A. M. Zaslavsky. *J. Amer. Statist. Assoc.* **88** 1161–1163.

WACHTER, K. and FREEDMAN, D. (1994). Local heterogeneity in 1990 census coverage. Technical report, Dept. Statistics, Univ. California, Berkeley.

WACHTER, K. W. and SPEED, T. P. (1991). Comment on "Total error in PES estimates of population: the dress rehearsal of 1988" by M. H. Mulry and B. D. Spencer. *J. Amer. Statist. Assoc.* **86** 858–861.

WEST, K. K. (1991). PES Evaluation Project #P5a: analysis of P-sample fabrication from evaluation followup data, 1990 Coverage Studies and Evaluation Series Memorandum #G-2, Statistical Support Division, Bureau of the Census, Washington, D.C. (July 12, 1991.)

WEST, K., CORBY, C. and VAN NEST, G. (1989). Quality assurance for the clerical matching operation in the 1990 Post-Enumeration Survey. In *Proceedings of the Survey Research Methods Section* 699–703. Amer. Statist. Assoc., Alexandria, VA.

WEST, K. K., MULRY, M. H., PARMER, R. and PETRIK, J. (1991). Address reporting error in the 1990 Post-Enumeration Survey. In *Proceedings on Survey Research Methods Section* 236–241. Amer. Statist. Assoc., Alexandria, VA.

WOLTER, K. (1986a). Comment on "Regression models for adjusting the 1980 census," by D. A. Freedman and W. C. Navidi. *Statist. Sci.* **1** 24–28.

WOLTER, K. (1986b). Some coverage error models for census data. *J. Amer. Statist. Assoc.* **81** 338–346.

WOLTER, K. (1987a). Technical feasibility of adjustment. Memorandum to the Undercount Steering Committee, Bureau of the Census, Washington, D.C.

WOLTER, K. M. (1987b). Comment on "Census undercount adjustment and the quality of geographic population distributions" by A. L. Schirm and S. H. Preston. *J. Amer. Statist. Assoc.* **82** 978–980.

WOLTER, K. M. (1991). Accounting for America's uncounted and miscounted. *Science* **253** 12–15.

WOLTER, K. M. and CAUSEY, B. (1991). Evaluation of procedures for improving population estimates for small areas. *J. Amer. Statist. Assoc.* **86** 278–284.

WOLTMAN, H. F. (1991). Final report for loss function analysis for places. Statistical Support Division, Bureau of the Census, Washington, D.C.

WOLTMAN, H., ALBERTI, N. and MORIARITY, C. (1988). Sample design for the 1990 census Post-Enumeration Survey. In *Proceedings of the Survey Research Methods Section* 529–534. Amer. Statist. Assoc., Alexandria, VA.

WOLTMAN, H. F., GRIFFIN, R. A., NAVARRO, A., BATES, L. and MULRY, M. H. (1991). Post-Enumeration Survey Evaluation Project P16: total error model—loss function evaluation. Statistical Support Division, Bureau of the Census. Washington, D.C.

YLVISAKER, D. (1991). A look back at TARO. Technical report, Dept. Mathematics, UCLA.

ZASLAVSKY, A. M. (1993a). Combining census, dual-system, and evaluation study data to estimate population shares. *J. Amer. Statist. Assoc.* **88** 1092–1105.

ZASLAVSKY, A. M. (1993b). Composite estimators for multiple domains and information sources via hierarchical Bayes linear models and Gibbs sampling. Presentation at UCLA Workshop on Model Selection, Bayes Factors, and Sensitivity Analysis. (February 19, 1993.)

ZASLAVSKY, A. M. and WOLFGANG, G. S. (1993). Triple system modeling of census, Post-Enumeration Survey, and administrative list data. *Journal of Business and Economic Statistics* **11** 279–288.

Comment

Ian Diamond and Chris Skinner

From this side of the Atlantic, the amount of attention devoted to the adjustment of the U.S. census can sometimes seem bewildering. Censuses are conducted for many purposes and raise many methodological problems. For perhaps most purposes, the effects of coverage errors in the census seem likely to be minor compared, for example, to the effects of nonresponse in most sample surveys. Nevertheless, there are important uses of census data in Great Britain, as in the United States, where coverage errors do matter, most notably in the preparation of local area population estimates, which are widely used by central government for the allocation of resources. Furthermore, the adjustment issue has particular significance for the 1991 censuses of England and Wales and of Scotland since, unlike in 1981 and before, the estimates of the national under-

Ian Diamond and Chris Skinner are Professors, Department of Social Statistics, University of Southampton, Southampton S017 1BJ, United Kingdom.

count based on postenumeration data differ considerably from the estimates based on rolling forward the 1981 census to 1991 by adjusting for intercensal births, deaths and net migration.

In our comments, we respond to the Editor's invitation to broaden the discussion by describing some work that has recently been undertaken using data from the 1991 census of England and Wales to obtain population estimates. Following this discussion, we refer specifically to the three papers.

An important aim of the census in England and Wales is to provide a base from which to calculate midyear estimates of the population. For many reasons censuses in England and Wales need to be taken in early spring, typically in April, and so from the outset adjustments have to be built into the estimation process so as to yield midyear estimates. Thus, the issue for the 1991 census was never one of whether to adjust but rather of how to adjust. In addition to the adjustment to the midyear date, there were also initial adjustments made to account for infant undernumeration, and for students and members of the armed forces who were not at their usual address on census night. After these adjustments, the overall undercount in England and Wales has been estimated by two principal strategies. The first has been via a postenumeration survey, the Census Validation Survey (CVS). This gave an estimated undercount of 240,000 residents. The second strategy has been to use demographic methods to roll forward the population count from 1981. This gave an estimated undercount of 1,113,000 residents. After considerable work it was decided that the rolled forward estimate provided the most reliable figure on which to base the overall population estimate. This meant that after adjusting for the CVS there remained a difference of 873,000 residents between the estimates based on the two alternative strategies.

Much work has been undertaken to attempt to explain this discrepancy (e.g., Office of Population Censuses and Surveys, 1993). For ages 45–79, the two sets of estimates are in good agreement. For the under-45's, the decision to believe the estimates from the rolled-forward approach was supported by a comparison of the demographic structure of the population under the two approaches. This showed that the age/sex structure of the census count adjusted by the CVS was markedly short of young people, predominantly male, particularly those in their 20's—groups which are universally acknowledged as hard to enumerate. As a result, the final national count and age/sex distribution was adjusted at ages 0–44 to match the rolled-forward estimates. At ages above 80, auxiliary data from the Department of Social Security on pensions were used to adjust to the final midyear estimate.

The shortfall of the CVS undercount estimate has led to scrutiny of the methodology of coverage error estimation based on postenumeration surveys. The British approach is quite distinct from the dual system approach in the United States. Instead of attempting an independent replication of the census count, "dependent reenumeration" is used to follow up the census in sampled areas with the aim of assessing the "true" count. Experienced interviewers are given data from household census forms and conduct reinterviews to assess the coverage and quality of the responses. In addition, buildings are relisted, nonresidential buildings are checked, addresses recorded in the census as vacant or absent are followed up and the occupation of buildings with multiple households is checked.

The undercount rate can then simply be estimated by comparing the census count with the revised "true" counts in the sample areas. The accuracy of estimates based on such "dependent reenumeration" as compared to dual system estimates clearly depends on the validity of the assumptions underlying each procedure, as discussed by Breiman for the U.S. case. In principle, the dependent reenumeration approach avoids the problems of correlation bias in the dual system estimate and is less critically dependent on the matching of census respondents to postenumeration survey respondents, which as Breiman emphasises may be prone to error. On the other hand, the assumption that the "true count" in sampled areas can be determined exactly is clearly heroic and is strongly refuted by the evidence referred to earlier. A considerable number of persons seem to have been missed on both occasions either because of deliberate avoidance or because of problems in making contact. There is a clear need to develop a new strategy for the CVS for the next census in England and Wales, perhaps adopting a form of dual system estimation. This will require more comprehensive as well as bigger samples.

Although methods of demographic analysis such as those described above have strong advantages for the estimation of coverage error at a national level, problems of estimating internal migration make them difficult to use to distribute the estimated undercount geographically. In England and Wales, infants, armed forces personnel and students can be allocated to local areas using vital registration and administrative data. However, for the majority of the population, local area population estimates were obtained by using a demographic estimation procedure to allocate weights to local area counts.

The local authorities were divided into 10 broad types of areas (based on a classification of 1981 census data). The sex ratios (males per 100 females) obtained by adjusting each five-year age group in each

local area type by the national undercount for that age group were then compared with sex ratios calculated from the 1971 and 1981 estimates. Since several of the 1991 estimates seemed implausible, a procedure for adjusting the estimates to attain a "target" sex ratio in each of the 10 groups of areas was developed (Diamond, 1993). After much exploratory analysis it was decided to use this adjusted approach only in the three five-year age groups 20–24, 25–29 and 30–34. This was for two main reasons. First, there was little evidence of implausible sex ratios as a result of using a uniform adjustment in other age groups. Second, the majority of the estimated national undercount occurred in this age range. Adjustments using the demographic estimation approach were incorporated into the final midyear estimates for local authorities and have been generally accepted although work to estimate populations for smaller areas continues. Whereas synthetic estimation has so far not been used explicitly, some methodological investigation of these techniques has taken place (e.g., Skinner, 1991).

We now turn specifically to the three papers under discussion. Breiman draws our attention to a range of sources of nonsampling error which may affect the dual system estimates. His discussion draws on the extensive evaluation work of the U.S. Census Bureau, and Breiman's attempt to summarise the combined effects of these errors on dual system estimates might be viewed as offering an alternative to the total error analysis reported, for example, in Mulry and Spencer (1991, 1993). We are not qualified to comment on the detailed differences between these alternatives but, like Belin and Rolph, find it hard to view the point estimates of Breiman as a methodological advance over the interval estimates of Mulry and Spencer.

We have described how, in the British case, demographic analysis has been used to validate postenumeration survey methodology at a national level. For the U.S. case, we note that the 95% interval (1.00%, 2.25%) of Mulry and Spencer (1993) for the national undercount rate agrees well with the estimate from demographic analysis of 1.85% (Robinson, Ahmed, Das Gupta and Woodrow, 1993). We would be interested to hear how Breiman views differences between results in his Table 15 (and perhaps similar results disaggregated by age and gender) and estimates from demographic analysis.

Freedman and Wachter examine the impact of the heterogeneity of undercount rates between states within post strata. Belin and Rolph refer to this impact as heterogeneity bias. We feel this risks confusion with the biasing effect on the dual system estimator of heterogeneity of capture probabilities between different individuals (correlation bias). Freedman and Wachter's discussion focuses not on properties of the dual system estimator *per se* but rather on the use of synthetic estimation across states. Like Belin and Rolph and others, we do not find the heterogeneity between states surprising. It seems more natural to interpret the impact of the heterogeneity in terms of variance rather than bias. Thus, the important questions that we feel flow from Freedman and Wachter's work are (a) how well can the additional uncertainty induced by the heterogeneity be quantified in appropriate interval estimators and (b) are there ways that this additional uncertainty can be reduced (for example, the procedures for England and Wales, described earlier, implicitly assume that sex ratios can be used as auxiliary variables at the local area level to reduce error). We encourage Freedman and Wachter to develop work in these directions.

As to the future of adjustment methodology, the experience of England and Wales suggests great scope for demographic estimation, and we are encouraged to see developments in that direction in the United States (Robinson, 1994). This point has been well made in an excellent review by Himes and Clogg (1992). However, if the method used is to be based on an analysis of intercensal birth records (Robinson, 1994), then it will be necessary to assess the accuracy of birth records as suggested by Wachter (1994).

Belin and Rolph provide wise comments on the nature of the adjustment debate. The importance that this debate take place in an open and scientific manner cannot, we feel, be overemphasized. Not only does it allow census agencies in other countries to share experiences, but it is to be hoped that it also enables our understanding of methodology to progress in a cumulative manner. To come to a consensus on adjustment does, of course, require more than just scientific progress. Our experience is that for there to be a common will to adjust there must first be an agreement that certain groups will be harder to enumerate than others and so a disproportionate adjustment for some groups will be seen to be fairer than either a uniform adjustment nationally or even no adjustment at all. In England and Wales the general agreement that young people, particularly men, in inner city areas were more likely to be underenumerated led to an acceptance of the demographic method used to estimate the mid-1991 population. It is also important for there to be a common recognition that the final census figure is an estimate subject to error. Clearly such errors must be minimized and standard errors estimated, but to expect that a census estimate must be perfect is unrealistic.

ACKNOWLEDGMENT

The research for this discussion was supported by ESRC Grant H507255122.

Comment

Eugene P. Ericksen, Stephen E. Fienberg and Joseph B. Kadane

Our discussion of these three papers focuses primarily on those by Breiman and by Freedman and Wachter. Our observations are consistent with many of those made by Belin and Rolph in their paper, but this should surprise few people who are aware of both our previously stated positions on undercount adjustment and the role that two of us (Ericksen and Fienberg) had as expert witnesses in the recent New York City census adjustment litigation, described in part by Belin and Rolph.

1. BREIMAN ON BAD DATA

Until recently, many Census Bureau and other experts on census coverage equated the net undercount rate with the omission rate. For example, in a report describing the coverage of the 1970 census, the Bureau of the Census (1975) analyzed the consistency between the demographic estimate of undercount and the omission rate given by a postenumeration survey. They made no mention of the possible existence of erroneous enumerations, even though the bureau measured such errors as part of its first such survey in 1950.

In 1980, Bureau Director Vincent Barabba, in explaining his decision not to adjust the census gave as one of two reasons the fact that the net undercount was close to zero. Later analysis exposed the problem with his conclusion. First, the bureau reckoned that about 3 million undocumented aliens had been left out of the demographic estimate of the national population which was the basis of the conclusion that there was no undercount. Second, survey data collected in the 1980 Census Postenumeration Program indicated best estimates of 13 million omissions and 10 million counting errors, which are the sum of substitutions and erroneous enumerations. There appeared to be substantial variations in net undercount rates among places. The fact that the national estimates of omissions and counting errors were close was now seen as accidental.

In 1990, the situation was similar to that in 1980, but worse. The redefined question then asked by the Census Bureau was whether or not the observed geographic distributions of omissions and erroneous enumerations were real or were caused by errors in the data. In 1990, the estimated numbers of omissions (20 million) and counting errors (16 million) were much larger than they had been in 1980 (Bryant, 1993; Ericksen and DeFonso, 1993), and the net undercount (4 million) was slightly, but coincidentally, larger than it had been in 1980.

Given this background, the focus of Leo Breiman's paper seems misdirected. He concludes "The largest part of the original undercount estimate is due to bad data and processing error—80% on the national level." In Breiman's terms, he believes that the correct estimate of net undercount may be as low as 1 million. For this to be true, either the estimated number of omissions would have to be lowered from 20 to 17 million, the estimated number of counting errors would have to be increased from 16 to 19 million or there would have to be some combination of the two. Either way, there would be 30–40 million census errors to be accounted for, and if a decision was made not to adjust the census, one would simply have to hope that the distributions of these errors were so similar that between-area variations in net undercount rates would be minor. In our view, Breiman focused his time and energy on the wrong problem. Rather than trying to show how PES data problems inflated the national estimate of net undercount, he would have better spent his time showing how these errors might have skewed the estimated differentials between places.

Breiman's paper is based largely on the 1990 Post Enumeration Survey evaluation data, which came from three sources: (1) records of quality control procedures; (2) a repetition of matching procedures carried out for a sample of PES cases by more expert matchers at the Census Bureau; and (3) the Evaluation Followup Survey, in which a subsample of PES respondents were reinterviewed. This interviewing occurred in January 1991, fully nine months after census day and five to six months after the PES interviewing period. Using these evaluation data, Census Bureau statisticians had already assessed the quality of the PES data used for the undercount estimates. This evaluation was summarized by Mulry and Spencer (1993). In their best judgment, the national net undercount was *slightly* too high, but the differential undercount among places was substantially as the original PES had indicated.

Eugene P. Ericksen is Professor of Sociology and Statistics, Department of Sociology, Temple University, Philadelphia, Pennsylvania 19122. Stephen E. Fienberg and Joseph B. Kadane are Professors of Statistics and Social Sciences, Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213.

TABLE 1
Weighted proportionate distribution showing how production PES E-sample classifications were reclassified in the evaluation study

Evaluation study classification	Original E-sample classification			
	Correct enumeration	Erroneous enumeration	Unresolved	Total
Correct enumeration	92.8	27.6	90.0	83.1
Erroneous enumeration	3.6	67.2	8.3	13.1
Unmatchable	0.01	—	—	0.01
Unresolved	3.6	5.2	1.6	3.8
Total	100.0	100.0	99.9	100.0
Sum of weights	28,005,586	4,990,529	774,766	33,770,880

Source: Table 35, U.S. Bureau of the Census, 1990 Coverage Studies and Evaluation Memorandum Series, #K-2, July 11, 1991.

Breiman feels that the Census Bureau underestimated the defects in the PES data.

Statisticians evaluating the quality of the adjustment to the 1990 census were generally aware of two shortcomings of the interview data used to evaluate the PES. Secretary of Commerce Mosbacher's Special Advisory Panel, a group whose members had sharply disparate views on adjustment, collectively wrote a letter to the secretary advocating great caution in the use of these data. Since the data were collected so long after census day, there would be uncertainty not just because of sampling but also due to nonsampling error. Just as there was substantial movement between census day and PES interviewing, many PES sample members moved between the day of their PES interview and the evaluation survey interviewing period. This was especially true of the hard-to-count members of the population.

Moreover, there was no direct way to measure correlation bias resulting from the fact that especially hard to count people are missed both by the census and the PES in excess of the numbers expected were the two independent. Breiman, in his evaluation, assumes that the correlation bias is zero. Mulry and Spencer, in accord with other Census Bureau statisticians, regard this assumption as unreasonable and used their best estimate of correlation bias. This dispute matters, because most of the errors cited by Breiman reduce the net undercount. Incorporating an estimate of correlation bias takes us the other way. We believe, along with Mulry and Spencer, that Breiman's assumption of zero correlation bias leads him to overstate his case. This overstatement is extended because he makes no allowance for the uncertainty caused by problems in collecting the evaluation survey data. No doubt those people missed either by the census or the PES were even harder to find in January 1991 during the evaluation survey interviewing period. We can only speculate as to why Breiman chose to believe an assumption that most knowledgeable statisticians have found untenable. In the original version of his paper, prepared for the 1992 New York City census trial, Breiman did present an argument for this position based on a fallacious manipulation of confidence limits across several different methods for estimating the impact of correlation bias (Bell, 1993). He has dropped the discussion but kept the erroneous conclusion.

Because Breiman studied the wrong problem, the national net undercount as opposed to the distribution of this undercount, we do not feel that his conclusions matter greatly. We prefer to rely upon the evaluation made by the Census Bureau, partly because they have studied the questions more thoroughly, but also because we believe them to be more objective. Many of Breiman's judgments appear to be strained, and his evaluations extreme. Here is an example.

In his Table 8, Breiman shows that when additional information was obtained from the Evaluation Followup interviews, 7.2 percent of cases originally classified as correct enumeration and 32.8 percent of cases originally classified as erroneous enumeration had their status changed as a result of the new information. That cases would change status as the result of new information does not surprise us, but we wish that he had presented a more complete analysis. As we show in Table 1, 13 percent of cases changed status; 6 percent went from unresolved to a resolved status or vice versa, while 7 percent went from erroneous to correct enumeration or vice versa. There was a fair amount of cancellation in these changes, and the correct enumeration percentage went from 82.9 to 83.1 percent. The erroneous enumeration percentage dropped from 14.8 to 13.1 percent while

TABLE 2
Estimated original dual system estimates and alternative correction estimates by evaluation strata

Evaluation poststratum	Estimate			
	Original dual system	[P16] corrected		Breiman
		correlation bias	no correlation bias	
1. NE central-city minority	6.83	7.08	5.32	3.7
2. NE central-city nonminority	−0.75	−1.34	−1.34	−2.5
3. U.S. non-central-city minority	5.43	4.51	3.89	2.6
4. NE non-central-city nonminority	0.01	−1.03	−1.03	−1.6
5. South central-city minority	5.68	3.77	3.12	1.5
6. South central-city nonminority	1.94	1.14	1.13	0.3
7. South non-central-city nonminority	1.82	1.78	1.55	0.6
8. MW central-city minority	3.97	5.04	4.17	3.0
9. MW non-central-city nonminority	1.28	0.65	0.46	−0.5
10. MW non-central-city nonminority	0.39	−0.13	−0.13	−0.7
11. West central-city minority	6.14	6.38	5.72	3.9
12. West central-city nonminority	2.13	3.99	3.81	3.1
13. West non-central-city nonminority	1.84	0.68	0.68	−0.2
	Differences between evaluation poststrata			
Northeast (1–4)	6.82	8.11	6.35	5.3
South (5–7)	3.86	1.99	1.57	0.9
Midwest (8–10)	3.58	5.17	4.30	3.7
West (11–13)	4.30	5.70	5.04	4.1

Note: Evaluation poststratum 13 includes a poststratum of American Indian reservations.
Source: Original and [P16] corrected estimates are from U.S. Bureau of the Census, 1990 Coverage Studies and Evaluation Memorandum Series #R-6, July 11, 1991, and have been reproduced in Mulry and Spencer (1993). Breiman's estimates are from his Table 16.

the unresolved percentage increased from 2.3 to 3.8 percent. How much the final rate of erroneous enumeration would change depends on the proportion of unresolved cases imputed to be erroneous, but any way this turns out, the impact of these changes is minimal. Breiman's arguments here seem strained at best.

What Breiman has basically done is to repeat the Mulry and Spencer analysis, to argue that the estimate of zero correlation bias should be used and to add a few additional components of error. Aside from his assumption about correlation bias, his numerical results are not greatly different from those of the bureau, especially when we consider between-area differentials.

Mulry and Spencer (1993) used their total error model to adjust the net undercount estimates for 13 "evaluation poststrata," which were aggregates of the 1,392 poststrata defined by the PES. There were three evaluation poststrata in each of the four regions, one for minorities living in central cities and one for nonminorities living in central cities and one for nonminorities living outside central cities. The 13th stratum included all minorities living outside central cities in the nation. Of greatest practical interest is the comparison between central-city minority poststrata, where the net undercount was thought to be highest, and non-central-city nonminority poststrata, where the net undercount was thought to be lowest. In Table 2, we present the

original estimates of net undercount computed by the Census Bureau; the same estimates corrected by Mulry and Spencer's total error model, with and without the Bureau's best estimate of correlation bias; and Breiman's final adjustments. At the bottom of the table, we present the differences between the central-city minority and non-central-city minority poststrata in each region. To evaluate the effect of assuming that the correlation bias is zero, compare the second and third columns. To evaluate the effect of Breiman's additional analysis, compare the third and fourth columns.

Looking at Table 2, we first see that for the original undercount estimates there were substantial differences between central-city minorities and noncentral-city nonminorities in each region. When Mulry and Spencer adjusted these differences using the results of the total error model, the differences increased in three regions and decreased in just one, the South. Moving to column 3 and assuming zero correlation bias reduces all four differences. In three of the four regions, the differentials were within one percentage point of what they had been originally, but the change in the South was large. Finally, we observe that Breiman's additional adjustments reduced the differentials slightly. In the Northeast, Breiman's adjustment reduced the differential by 1.05%, and in the South, Midwest and West, the reduction was less than 1%. Comparing the original and Breiman's adjusted estimates (columns 1 and 4), Breiman reduced the net undercount differential by 1.5% in the Northeast and 3% in the South, but made scarcely any impact in the Midwest and West. Taking everything together, even if we were to believe Breiman's arguments and were to rely on his results rather than those of the Census Bureau, for the most part we find substantial differences in undercount rates between just those areas where we would expect to find them.

2. FREEDMAN AND WACHTER ON HETEROGENEITY

What do the Breiman and the Freedman–Wachter papers have in common? While both focus on the accuracy of adjustments to census data, they also adopt a shared implicit starting position, namely, that the census counts should be treated as if they are error free until we can show somehow that the use of adjusted counts is warranted. That the census is replete with errors and that the errors have differential impact on minorities never seems to be addressed or acknowledged by these authors.

The Freedman–Wachter paper (FW) addresses the issue of heterogeneity in the census adjusted counts when used for intercensal purposes. What they imply is that heterogeneity is bad and they tell us that "any comparison of error rates between the census and adjusted counts should take heterogeneity into account."

The argument of FW is that poststratum homogeneity is an assumption of adjustment. They find evidence of lack of poststratum homogeneity and seem to want to infer that adjustment is a bad idea. This latter does not follow, of course. Whether adjusted counts are closer to the truth than unadjusted counts is the essential issue and is only tangentially related to heterogeneity. Let us elaborate.

Heterogeneity among substrata is not inimical to adjustment. Kadane, Meyer and Tukey (1992) show that if substrata can be ordered by "catchability" in both the census and the PES simultaneously, the cross product ratio of the stratum is bounded *below* by an average (with respect to a certain probability weighting on the substrata) of substratum cross product ratios. In fact, the ordering is unnecessary: positive correlation (with respect to the same probability weighting) suffices. Thus if one accepts a cross product ratio of 1 in each substratum, as is often done in capture–recapture methods, and accepts positive correlation in catchability among substrata, the implication of substratum heterogeneity in catchability is that the use of a cross product ratio of 1 for adjustment moves in the right direction, but not far enough. A related paper by Darroch, Fienberg, Glonek and Junker (1993) proposes a model for individual-level heterogeneity which incorporates exactly this kind of dependence explored by Kadane, Meyer and Tukey (1992). Implemented in the context of triple system methods, their model when applied to census test data actually illustrates the extent to which heterogeneity in catchability might lead to underadjustment.

The conclusion of FW's analysis of surrogate variables is, as Belin and Rolph observe, at best overstated and they leave us confused about what this has to do with whether or not we might wish to adjust census counts for differential undercount. Fay and Thompson (1993), in the context of loss function analysis, use artificial substitutes for undercount rates as do FW. They conclude that "For most [proxy] variables, the loss function analyses is not seriously distorted in favor of adjustment, but in one case the loss function would overstate the advantages of adjustment." This is a far cry from FW's "almost anything can happen."

How are we to think about FW's proxy variables? Freedman and Wachter note that "The drawback is that the proxies may not behave like undercount rates, in terms of heterogeneity," but then they analyze them nonetheless. Perhaps they have in

mind some model, of the form

$$\text{undercount rate} = f(\text{SUB, ALL, MULT, NM, MOB, POV, Error}).$$

The analysis of DIFF in their Table 9 suggests something of this form with f as a linear function with invented coefficients! What evidence do we have for believing such models and how much credence should we give to the analysis of the component parts?

Freedman and Wachter conclude that the 357 poststrata have too much residual variability on the proxy variables within state. This comes as no surprise to us. In its original analyses following the 1990 census, the bureau used 1,392 poststrata and then smoothed the resulting adjustment factors to remove variability. The later approach, which dropped the number of poststrata from 1,392 to 357, was sure to introduce greater heterogeneity and the "random" features in it are no longer kept under control through smoothing. As Fay and Thompson (1993) note: "Although [census analysts] eliminated smoothing from consideration in 1992, there may have been hidden costs to this decision...."

The statistical literature clearly suggests that, even if FW were successful in showing substratum heterogeneity, by doing so they may have strengthened rather than weakened the case for the use of adjusted data.

Finally, we wonder what FW's analysis really has to say about the wisdom of using adjusted counts for various intercensal purposes. As we move further into the decade and away from the April 1, 1990, census date, there must be more and more error in census counts, both adjusted and unadjusted. Does there come a point when the cumulative errors due to the passage of time swamp the undercount problem? Or does the differential undercount between the white majority and various minority groups that we have observed for over 50 years in decennial census data only become worse?

Comment

Lars Lyberg and Sixten Lundström

The American census adjustment debate must represent the pinnacle of statistical methodological controversy. Usually, statistical discourse is conducted by laconic academics who address technical issues of obscure merit to nonstatisticians. Indeed, it is an anomaly for the profession that an essentially technical issue, such as census adjustment, would attract such widespread and vocal attention.

Our comments should be prefaced with the fact the Swedish censuses are not affected by the type of undercoverage that characterizes the U.S. census undercount. Our approach to census taking is vastly different from that used in America. Sweden is known for its high-quality population registers and uses a register-based approach for the actual count. Since it is extremely difficult to function in Swedish society without a personal identity number (PIN) (and many of the benefits and amenities offered by Swedish society require a PIN), every legal resident is included in the population register. Any undercount is very small (a few hundred) and is linked to lags or delays in the reporting of vital statistics. These delays usually do not last more than 10–13 weeks; so, both in principle and practice, Sweden can conduct an accurate population census any week of the year. A word of caution, though, over the last few years, we have had an increasing problem with overcoverage due to immigrants who repatriate without notifying the authorities.

The United States, on the other hand, lacks Swedish-style population registers and bases its census on a master sample, that is, tracking everyone down by figuring out the number and location of dwellings and then ascertaining who and how many live in a given dwelling. This is obviously a daunting task when multiplied by an entire nation of geographically and ethnically diverse individuals.

When studying the articles by Freedman and Wachter, by Breiman, by Belin and Rolph and by numerous others dealing with the U.S. census adjustment, we have made a number of general observations. First, it should be kept in mind that statistics is the theory and practice of dealing with uncertainty. Second, surveys never produce "true" numbers. What surveys do produce are estimates and every source available should be used to make

Lars Lyberg is Head of Research and Development at Statistics Sweden and is Chief Editor of the Journal of Official Statistics. Sixten Lundström conducts work at the Statistical Research Unit at Statistics Sweden.

these estimates more accurate. Modelling is used at every stage of survey design. When new things are learned, these models are revised to reflect the insights provided by the new information. It can be seen as an iterative process moving toward certainty, but never actually getting there. Third, applied work is very different from theoretical work. Theoretical work is very neat and tidy in that the most important elements are defined by the scientist him/herself. Theoreticians state the premises under which their findings are valid, thereby conducting their research under highly controlled conditions. The practitioner, on the other hand, must do battle with the vicissitudes of daily life. Applied work has the added element that almost anything, regardless of how important, can spin out of control with no regard for the resulting inconvenience or damage.

Fourth, all sciences are subject to controversies; such debate is necessary, and examples from statistics include Fisher versus Neyman, frequentists versus neo-Bayesians, model-based samplers versus design-based samplers and now census adjusters versus census nonadjusters. Furthermore, scientific debates can contain nonscientific ingredients. Perhaps in cases where science cannot reach a consensus, it is best to turn the discussion over to politicians.

The statistical information collected in the U.S. census is used in many different areas, and many different political decisions are based on it. The undercount problem, however, is described only in connection with its use for congress apportionment and the allocation of tax funds. We have not found any description of the "models" used for these purposes, but they are probably not particularly complicated. On the other hand, equitable and fair allocation of funds is another issue with its own difficulties. Perhaps the main problem is not the census undercount, but rather the oversimplified allocation model? If this is the case, the problem really belongs in the political sphere.

In our opinion, the U.S. census must be the most complex survey in the world. There are other surveys that are larger (e.g., the Chinese census), but none comes close to the level of ambition of the U.S. census. The combination of problems associated with population mobility, lack of reliable and complete registers, illegal immigration, homeless people, enumerator security and demands for high-quality data makes the U.S. census an enormous undertaking. The Census Bureau must be commended for its skill and competence in dealing with this task. From the start of the modern survey era, the bureau has been on the forefront, and preeminent statisticians like Morris Hansen, Bill Hurwitz, Bill Madow and Barbara Bailar are among those who have managed the bureau's continuous development. The Census Bureau is the agency where the most extensive and complicated evaluation studies have been conducted. Due to the agency's efforts, the entire survey community has learned a great deal about survey errors. The bureau has led research on, for instance, survey models, processing errors, coverage errors, response variance, estimation and the use of new technology.

It does not make sense to criticize the Census Bureau for collecting bad data. The Post Enumeration Survey (PES), like all surveys, suffers from errors. However, these errors are smaller than those generated in the census. The absolute best a survey practitioner can do is use the preferred procedures. In the case of the U.S. census, the PES is a preferred procedure which provides information that can be used to increase the accuracy of the census count. We believe it is important to appreciate that the PES is not meant to generate true numbers, but rather numbers closer to the true numbers. When discussing the U.S. undercount problem, or the quality of any survey for that matter, it is vital to recognize the limitations inherent in survey practice and view the survey results in this light. It is also germane to this discussion to point out that data of lesser quality than the PES are used every day with excellent results and no complaints.

The article by Freedman and Wachter provides an example of shortcomings associated with evaluation surveys. The PES provides acceptable or perhaps even unbiased estimates for the nation and for some other very large domains. For small domains, model-dependent estimators have to be used. Such estimators include some bias because of model error. For census adjustment, the bureau uses a type of synthetic count estimator. The model error associated with this estimator is due to heterogeneity in poststrata. The heterogeneity, measured by proxy variables, has been shown to be quite large within some poststrata.

At Statistics Sweden, we use a synthetic count estimator regularly when estimating the number of employed persons by different categories of hours worked in each municipality. Variables defining the poststrata are sex, age, industry and income. From a methodology study where the parameter values were already known, we learned that the estimator worked well for the majority of municipalities, but performed poorly for a few municipalities. Despite these results, this method has not received much criticism. We believe that users are aware that this estimate is the best we can do and it is better than no estimate at all.

This leads us to another comment. A census should result in a single published count. Publishing two counts gives the impression that users have a "choice" even though the producer of the cen-

sus never intended such an interpretation. In the 1970 Swedish census, Statistics Sweden presented two numbers: one regular set of estimates with missing data and one with imputed values added. Surprisingly, many users (but perhaps not so surprising after all) knew exactly which estimate to use. In the 1975 census, imputation was not performed, which made comparison to the 1970 census awkward.

We find it both reasonable and natural to use auxiliary information to improve an estimate. After all, this is what survey design is all about. Various model assumptions are made in every design step, but the final result should be expressed as a single count or estimate. We sympathize with Belin and Rolph regarding their general conclusion about the protracted controversy on the undercount problem. An impressive amount of work has been done, but it appears as if we have reached the point where further methodological resources, time and money would be a waste.

Most U.S. statistical agencies have committed themselves to modern quality thinking, that is, various forms of total quality management. It seems as if it would be better to use the "debate resources" to improve the regular census count procedures, thus decreasing the need for extensive and expensive evaluation procedures. This is especially true for the U.S. undercount, where the discussion fails to result in a consensus.

Comment

David Steel

Evaluating and possibly adjusting the census for undercount raises a lot of difficult statistical and general issues. The papers here consider several of these and add to the already large literature on the subject. While the basic questions are now clear, the answers are not. To enable readers to make a judgement about any prejudices I might have on these issues, I should point out that as a former officer of the Australian Bureau of Statistics (ABS) I was involved in the evaluation of the 1981 and 1986 censuses. While the views I have are entirely my own, they are influenced by this past involvement. In terms of my prejudices this could work either way: having been involved in adjustment, I may have a bias to that view to justify my past work; alternatively, detailed knowledge of the many problems involved could lead me to be against adjustment.

In Australia, population estimates based on census counts adjusted for undercount have been released as the official population estimates since 1976. The estimates are produced for states and territories and local government areas. Population estimates are used to determine the number of seats each state has in the federal House of Representatives and the allocation of funds to states and local government areas. The decision to adjust was prompted by the high undercount rate showed by the 1976 Post Enumeration Survey (PES) and the fact that the 1976 census count fell considerably below the population estimates for 1976, which were based on updated 1971 census results. There has been general acceptance and remarkably little controversy surrounding the adjustment. A clear distinction is made between census counts and population estimates. Census counts are produced without any adjustment. There are similarities to the situation in the United States. The level of undercount is basically estimated from a PES which involves an independent household survey and matching between the census and the survey to determine missed people and some categories of erroneous enumerations. Dual system estimates (DSE's) are calculated. The results of the PES are compared with demographic analysis and other population indicators such as school enrolments and Medicare enrolments (Medicare applies to all age groups), primarily at the national level, but with some analysis below this. Synthetic estimation is used to obtain population estimates for local government areas. The procedures for the PES and census evaluation are decided in advance. The view is that quality must be designed into the census and the PES. The estimated level of net undercount is remarkably similar to the United States: 1.9% in 1986 and 1.8% in 1991. In 1991 the state undercount rates ranged from 1.2 to 4.1%. The ranking of the states in terms of undercount has been consistent over time. Further details are given in Choi, Steel and Skinner (1988), Trickett (1992) and Australian Bureau of Statistics (1990).

David Steel is Senior Lecturer, Department of Applied Statistics, University of Wollongong, Northfields Avenue, Wollongong, New South Wales 2522, Australia.

There are also some important differences between the situation in Australia and the United States. The Australian census is conducted by field methods, and collection of census forms is essentially completed in a two-week period, enabling a PES to be conducted approximately three weeks after the census (census forms received after the PES has started are excluded from the DSE's). Conducting the PES as close as possible to the census date is an important factor in improving the quality of the census evaluation. The PES includes about 40,000 households comprising 93,000 people but is much less clustered than in the United States. Over 4,000 clusters are selected, giving a good geographic spread. This can be done since there is no direct equivalent to the E-sample. Attempts to match people selected in the PES sample are made with census forms for the selected dwelling and those in the surrounding areas and at other addresses provided at the PES interview where people may have been enumerated (e.g., usual residence of visitors and the surrounding areas). This provides information on the major categories of erroneous enumerations. Some classes of erroneous enumerations will be missed, such as households which are completely fabricated in the census. In the matching process unresolved cases are finally imputed using a regression-based method.

This size of the PES sample and its design means that direct estimates can be calculated for the eight states and two territories. For each state, separate estimates are calculated for the capital city and the rest of the state. This leads to estimates for 14 major geographic areas, referred to as parts of state. Estimates for local government areas are then obtained using synthetic estimates using age–sex poststrata within each part of state. This means that a lot of the concern about using undercount rates for other states is avoided. However, there will be variation in the undercount rate that is not accounted for by using the age–sex poststrata. The approach has been deliberately conservative, adjusting for factors which have clearly been demonstrated to affect the undercount rate. It is felt that such an approach will bring the population estimates closer to the true distribution, but will probably understate the variation in undercount rates. Hence, while it should lead to improvements, the danger of overadjusting is reduced. Other methods of producing undercount estimates for local areas have been investigated (Steel and Poulton, 1988; Bell, Cornish, Evans and Vincente, 1993) but the current synthetic procedure is considered the most appropriate.

There is little evidence in Australia of difference in undercount rate between different ethnic groups in the community. The exception is the Aboriginal community, whose level of undercount is, to some degree, associated with problems of enumeration in remote areas. Analysis of undercount rates for different groups by birthplace has shown little differences from the Australian born. The group with the highest undercount consists of those born in New Zealand, which is probably due to the relative youthfulness and mobility of this group in Australia. Analysis has shown that the undercount is high for those who were away from their usual residence on census night, with an estimated undercount rate of 16.5% in the 1991 census (Trikett, 1992). Matching procedures for this group are especially important. In 1991 the scope of the areas checked for such people in the matching was expanded, with beneficial results.

Demographic analysis, using the 1921 census as a base, is used to validate the PES results and on occasion make some adjustments to these figures at the national level, which then flow through to lower levels. Reliable birth and death registration systems and a system to measure overseas arrivals and departures have been in operation for a long time. Movements in and out of the country are relatively easy to monitor, with there being limited points of entry. Five yearly gaps in the census also help in the demographic analysis. The view has been taken that the PES is generally more reliable than the demographic estimates, which are mainly used to evaluate the PES. The PES has some problems. A small number of blocks can have undue influence, conducting independent check in remote areas is difficult, matching is not perfect and erroneous nonmatches can inflate the estimated undercount rate. In 1991 there was a downward revision in the estimated resident population for Western Australia. It is thought that the 1986 census count had been overadjusted because of the effect of school holidays on the number of persons away from home at the time of the census (Trickett, 1992). However, based on the evidence and knowledge of the processes used, the judgement has been that population estimates incorporating an adjustment will give a better estimate of total population and population shares than unadjusted census counts. This is a judgement which is finally made by the Australian statistician, whose independence is guaranteed by the ABS Act of 1975. I would expect any concerns about the approaches used would be raised and handled through various mechanisms that exist for state government input into ABS activities, not through the courts.

The Australian experience gives an example where reliable estimates of census undercount can be made and where adjustment is carried out. However, there are sufficient differences not to see this as necessarily endorsing the entire procedure originally proposed for adjusting the U.S. census. The paper by Breiman raises the fundamental issue of what is the

quality of the PES. This a good question to raise and is logically the first to ask. To answer this question involves making an assessment of nonsampling errors, and it is notoriously difficult to get a complete picture of these. The concern is whether the quality of the evaluation is sufficient for its purposes. This is exactly what Breiman is raising concerning the undercount estimates, but the same question can be raised about the quality evaluation of the PES that he reviews and summarises. There are many ways in which a PES could potentially give bad data. While the paper raises some doubts, the case is made in a way that, despite the author's attempt to bring a lot of detailed evidence together, I have no feel for what the errors that might be there would do. The quoting of percentages on what appears to be different bases makes it hard to work through what the likely effect might be. Breiman notes that two 0.5% differences, working in different directions, can affect the estimated undercount rate by 50%. Essentially this point is that, to first order, if 1% of the PES is not matched when it could have been, then the estimated undercount rate will increase by approximately 1 percentage point. Some of the rates that are quoted are calculated on different bases, which may be relevant, but to asses their impact on the undercount estimates the reader needs them expressed as percentages of the entire PES sample. Quoting gross differences also makes independent assessment difficult. A tree diagram of what happen to the PES sample in terms of matching cases and nonresponse would make interpretation easier.

The paper by Belin and Rolph offers some different interpretations. Again we have reasonable statisticians offering a significantly different interpretation of the same data. To help in understanding the situation I look for other evidence. With the adjustment for the processing error mentioned by Breiman the estimated undercount is 1.7%, which is close to the demographic estimate of 1.85% provided by Robinson, Ahmed, Das Gupta and Woodrow (1993). This does not suggest gross overestimation of the undercount rate through the PES. Both estimates are imperfect but their consistency offers some reassurance. Of course it can also be argued that both systems have errors that just happen to give similar results.

The paper by Freedman and Wachter raises valid questions about synthetic estimation. Synthetic estimates are biased; so are census counts. No adjustment is a model. Hence it is desirable to incorporate an allowance for these biases for both estimates. The paper sheds more light on this issue. One result of the analysis is that, based on the proxy variables analysed, the synthetic estimates do explain much of the variation in rates between states. For example, using the root mean squared error in Table 6 and comparing it with the standard deviation across states in Table 5 for the substitution variable, we see a 42% reduction in the variance of the error of prediction when using the synthetic estimates as against the national rate. Applying the national rate would be the same as no adjustment in terms of population shares.

I have sympathy with the comment of Wachter, quoted by Belin and Rolph, on the use of aggregated measure. I do not think this precludes using empirical evidence to resolve issues, but implies looking beyond summary measures. I am therefore surprised that Freedman and Wachter do not provide any plots of the errors or bias of the synthetic estimates. It would have been interesting to see the distribution of the errors and any relationship with the size of states or other factors.

To adjust the census to provide population estimates, the reliability of the adjustment must be better than the error it is trying to correct. There are several ways of assessing the quality of the PES and the associated estimates, none of which is entirely complete or satisfactory, but which in combination should enable a judgement to be made. The process of assessing the quality of the PES has involved several sources. One source is the type of evaluation studies reviewed and summarised, with different conclusions, by Breiman and by Belin and Rolph. These studies themselves will have quality problems and are not complete (but I do not think we need to evaluate the evaluation of the census evaluation—we must stop somewhere). The quality that has been designed into both the census and the PES and the subsequent processing and the quality assurance procedures are factors which decision makers must take into account. Consistency with past results, or at least reasonable explanations of the differences, are also indicators. Comparison with other sources, such as demographic estimates add to the picture. Finally, what is sometimes called face validity, or consistency with what is thought to be known about undercount, is a factor. Some of these factors are "hard" and some "soft," but the overall judgement should be made by combining these assessments together. In Australia that has led to adjustment for the purposes of population estimates. To answer the question raised by Belin and Rolph, I do not think a consensus is possible. The standards of proof required by some are just not achievable. On what major issue would the profession have a complete consensus? Decisions have to be made on imperfect, sometimes contradictory evidence, in the face of strong disagreements by reasonable, well-qualified and well-intentioned professionals.

Rejoinder

Thomas R. Belin and John E. Rolph

NEW DEVELOPMENTS

Since the three articles and the discussions were accepted by *Statistical Science*, a federal appeals court set aside the judgment of a district court judge that had allowed the Commerce Department's decision not to adjust the 1990 decennial census to stand. Although the lower court's ruling was vacated, the ultimate consequences of the appellate decision are far from clear. Actions by the plaintiffs, by the government, by other appellants and ultimately by the courts will determine how the 1990 census adjustment saga is played out. See Fienberg (1994b) for a more detailed discussion of the appeals court ruling.

We hesitate to read the appeals court ruling as an endorsement of our scientific point of view. Nevertheless, the rulings of both the district court and the appeals court reflect a willingness on the part of nonstatisticians to view an adjusted census as a feasible and reasonable approach for improving on the accuracy of an attempted headcount.

Statisticians should understand and appreciate this willingness. We thus hope that this appellate court decision will give a new impetus to the statistics community to help facilitate consensus on how to use estimation methods in census-taking. The Census Bureau's investigation of a "one-number" census is a constructive step in this direction.

RESPONSES TO DISCUSSANTS

We focus here on the discussions by Diamond and Skinner, by Steel, by Lyberg and Lundstrom, and by Ericksen, Fienberg and Kadane; we have not seen either the Freedman and Wachter (FW) or the Breiman rejoinders, although we comment briefly on a point raised in some exchanges with our Berkeley colleagues.

The discussions by Diamond and Skinner, by Steel, and by Lyberg and Lundstrom all provide useful and enlightening perspectives on census-taking practices around the world. The balance in their remarks sets a good example for us to follow here in the United States.

The final paragraph of the discussion by Diamond and Skinner amounts to an excellent summary of our essential points: the debate over census adjustment should emphasize scientific matters, but consensus will require more than just scientific progress. We appreciate their supportive remarks. On a more subtle matter, their point is well taken that the term "heterogeneity bias" could be construed to mean either error in the synthetic assumption explored by FW or error in the assumption of equal capture probabilities discussed, for example, in Alho, Mulry, Wurdeman and Kim (1993). We hope our use of the term was clear from the context. We are also glad that Diamond and Skinner agree with us that the heterogeneity reported by FW is not surprising and that heterogeneity should be reflected in local-area estimates of variability.

Steel's discussion of how statistical estimation is used in the Australian census provides a valuable frame of reference for the debate about the future of the U.S. census. Steel does not attempt to adjudicate our disputes with Breiman, yet we interpret his remarks in his penultimate paragraph as supportive of our point of view: the census-taking process has to stop somewhere, and decisions have to be made. We read Steel's final paragraph as reflecting a semantic difference with our use of the term "consensus." We do not imagine that all statisticians would realistically line up behind one particular census methodology, but we can imagine there being a critical mass of support for a particular approach so that the controversy subsides. Indeed, during the 1980's as a member of the National Academy Census Panel, one of us (Rolph) saw such a critical mass forming on that panel and among the senior staff of the Census Bureau behind a planned adjustment methodology for the 1990 census. Intervening events led to the controversy and adversarial process referred to in these articles, but in our view the process need not be so contentious. For example, the use of postal delivery in the 1960 census was a major methodological change, but one that did not engender much controversy. There may have been some people who opposed this innovation at the time, but the level of support for a mail-out, mail-back census would qualify as consensus from our standpoint.

Lyberg and Lundstrom add a variety of insights that reflect the realities of government statistics practice. In a few places, their statements are stronger than we would make. For example, we do not have a problem with criticizing "bad data"; what we object to is the notion that we should assign a loss of infinity to model-based estimators and then call such an approach good science. We are also

more optimistic than Lyberg and Lundstrom that further investment in methodology will pay dividends. One example of methodology that we cited repeatedly in our paper is Zaslavsky (1993a); this work not only advances the state of the art for census undercount estimation, but it serves as a useful case study that could be adapted to other statistical arenas as well. Overall, however, we appreciate their endorsement of our general perspective on the adjustment controversy.

Ericksen, Fienberg and Kadane make few comments directly about our paper. We would simply point out that some of their recent references (Kadane, Meyer and Tukey, 1992; Darroch, Fienberg, Glonek and Junker, 1993) also serve to illustrate that progress is still being made on undercount-related issues, yielding both new theory and new methods.

We understand that in their rejoinder, FW cite a personal communication from us. We offer the following comment in the spirit of "setting the record straight."

In the initial version of his paper, Breiman made a stronger claim about the increasing proportion of unresolved cases in the Evaluation Followup Survey (EFU) when one reads across his Table 12, which we saw as the kind of nitpicking criticism that deserves to be pushed to the margin. Originally, after we pointed to Breiman's curious claim that one might do just as well in imputing for unresolved cases by flipping a coin with probability 15% of heads, we had written, "The higher proportion of remaining unresolved cases in the higher imputed probability categories is explained in large part by the fact that names were not recorded for many PES individuals." However, it turned out that our explanation was inaccurate; although cases without names constituted approximately 70% of the P-sample cases receiving probabilities of having been enumerated of 75–100%, these cases without names were largely excluded from the EFU and so were not reflected in Breiman's Table 12.

We acknowledge that there were a substantial number of unresolved cases in the EFU and that there is remaining uncertainty about the accuracy of the imputation methods. Our essential point is that there is not much to criticize based on available data, which agree with predicted values extremely well (Belin et al., 1993). To attribute our earlier statement to us as if it is our current view is a misrepresentation.

Overall, although we anticipate that our Berkeley colleagues will continue to support one another, we are pleased at the signs of consensus in this exchange.

Rejoinder

Leo Breiman

I thank the discussants. The descriptions of the methods used in Australia, Great Britain and Sweden were interesting and form a compact introduction to the diversity of methods for estimating population counts. They also underline the difficulty of the census undertaking in the United States. The discussion by Ericksen, Fienberg and Kadane and the Belin–Rolph article contain most of the direct comments about my paper.

BACKGROUND

The effort to adjust the census counts was a complex process. After the initial error evaluation, additional errors were discovered, some of which are discussed in my article. Because the original error analysis has not been updated to take these additional errors into account, the widespread impression remains that the adjustment process was proven to produce more accurate counts than the census.

The validity of any such proof is currently in serious doubt. For one thing, errors of various types are now acknowledged to account for the major part of the original national undercount estimate of 2.1%. The initial loss function analysis used earlier estimates of the bias that, on the national level, were too small by at least a factor of 2. The analysis was also flawed by a significant underestimation of sampling variances (Fay and Thompson, 1993; Freedman, Wachter, Cutler and Klein, 1994). There are also questions about the additional local bias due to heterogeneity (Freedman and Wachter, 1994), the errors resulting from smoothing the adjustment factors (Freedman et al., 1993) and many of the assumptions going into the loss function analysis (Freedman, Wachter, Cutler and Klein, 1994).

This careful scrutiny was possible, in part, due to the availability of three sets of numbers: the census counts, the adjustments and the extensive evaluation data. We view the controversy over

the adjustment as having some healthy outcomes. Methodology and implementation were openly discussed and debated. More statisticians than ever became aware of the problems of carrying out an accurate census. We hope that the planned single-number census for the year 2000 will allow similar careful scrutiny and discussion.

SUMMARY OF MY REJOINDER

In my rejoinder, I first comment on some statements of the overseas discussants, then on the Ericksen, Fienberg and Kadane discussion, and last, on the Belin–Rolph article. The Belin–Rolph remarks are surprisingly angry and include some serious and totally unjustified personal attacks.

The proposed 1991 adjustments will be referred to as the DSE adjustments or the DSE estimates. My article pointed out that major components of the DSE estimate adjustments are data errors stemming from different sources. None of the discussants disagree with my main conclusion concerning the presence of a large amount of error in the DSE estimates, at least at the national level.

Ericksen, Fienberg and Kadane note that there was uncertainty in some of the evaluation data, with the implication that aspects of the evaluation could have been better planned and timed. I agree. Both Ericksen, Fienberg and Kadane and Belin–Rolph maintain that my results do not have any major implications for the distribution of DSE estimates of population at local levels. They also criticize me for "assuming zero correlation bias." These points are responded to in my rejoinder to Ericksen, Fienberg and Kadane.

OVERSEAS DISCUSSANTS

The overseas discussants disclaim familiarity with the DSE adjustment process. However, their comments contain some misunderstandings. For instance, Lyberg and Lundström believe that because the PES was more accurate than the census, the DSE adjustments will be more accurate than the census counts. This does not follow. As shown in Section 4 of my article, small errors in matching can cause the adjusted counts to be less accurate than the census counts.

Steel likes the 1.7% undercount figure obtained by correcting for the coding error because it is consistent with the 1.85% demographic estimate of the undercount and "consistency offers some reassurance." However, the 1.7% number is defensible only if one ignores all of the evaluations carried out by the Census Bureau. If there is anything consistent going on, it is that almost every source of error investigated further lowers the DSE undercount estimate.

Diamond and Skinner like interval estimates and note that the 95% interval (1.00%, 2.25%) given in Mulry and Spencer (1993) agrees with the demographic estimate of 1.85%. However, the Mulry–Spencer article does not correct for the large coding error and other errors found since 1991. Thus, the interval cited is not accurate in terms of current information.

ERICKSEN, FIENBERG AND KADANE

Ericksen, Fienberg and Kadane (EFK) open by making the point that the census had many errors. The first 10 pages of the report of Special Advisory Committee members favoring adjustment also makes this point (Ericksen, Estrada, Tukey and Wolter, 1991). So does a 95-page appendix to this report. About one-third of Fienberg's article on adjustment (Fienberg, 1993) is spent describing errors in the census. Referring to Freedman, Wachter and myself, EFK say: "That the census is replete with errors and that the errors have differential impact on minorities never seems to be addressed or acknowledged by these authors."

The first lines in Wachter (1993a) are: "The 1990 census had flaws. It missed, net, between one and three percent of the population. It missed more men than women. It missed more blacks than whites. These facts are not in dispute." Wachter's article appears in the same journal issue as Fienberg (1993). This issue had four short articles on adjustment. It is hard to imagine circumstances under which EFK could have overlooked Wachter's statement.

Ericksen, Fienberg and Kadane keep raising the issue that the census had flaws, but nobody disagrees with them. No knowledgeable statistician I know denies that the census had many errors and that there was a differential minority undercount. Neither do I. Now that we are all agreed, let us get on to the fundamental question: would the DSE adjustments produce more accurate counts than the census?

The point of my article is that the DSE adjustments are mainly a reflection of bad data. Since an error-filled adjustment will only superimpose more noise on the census counts, EFK have to face the central issue: is it or is it not true that the DSE estimates are mainly a reflection of bad data?

Here are the points that EFK make regarding this central issue:

1. My final estimate is not believable.
2. My analysis is exaggerated and extreme.
3. I used questionable data.
4. I assumed zero correlation bias.

5. The Bureau's evaluation is better.
6. I worked on the wrong problem.
7. My results do not make much difference.

I will take these up, one at a time.

1. My final estimate is not believable.
Ericksen, Fienberg and Kadane begin with a statement from my article, which reads: "The largest part of the original undercount estimate is due to bad data and processing error—80% on the national level." They claim this has the following implication: "In Breiman's terms, he believes that the correct estimate of undercount may be as low as 1 million." Then they go on to show that the true undercount could not be as low as 1 million.

Their logic is wrong and they have overlooked the statement in my conclusion section which reads: "The results of this study should not be taken to mean that I believe that the true 1990 undercount is as low as 0.4% [1 million] or even 0.9%. My focus was on whether the 1990–1991 DSE process produced reasonable estimates of the true undercount. To that, my answer is no, there were simply too many sources of error."

2. My analysis is exaggerated and extreme.
To quote EFK, "Many of Breiman's judgments appear to be exaggerated and his evaluations extreme." They give a single example to support this statement. The example is followed by the sentence "Breiman's arguments here seem strained at best." Now strained is quite a demotion from exaggerated and extreme. Still, since this is EFK's only shot, let us see how strained my argument is.

The example is brief and occurs in my Section 5.5 titled "Reliability of interview data." As part of the Census Bureau's evaluation, a small fraction of the PES households were reinterviewed. This was called the Evaluation Followup Survey (EFU). The example essentially starts with the sentence in my article reading "Match status could be changed from the PES production match status only if new, relevant and reliable information regarding a case was present in the EFU interview." This is followed by a table showing what percentages changed match status as a result of the EFU data.

Ericksen, Fienberg and Kadane go through some computations to show that, although individual cases may change match status, the marginal totals in each match category remain about the same. Having established this, they note, with some triumph, that "the impact of these changes is minimal." This totally misses the point. A person changes match status only if the PES interview information for that person is judged unreliable compared to the EFU information. Thus, the percentages of people changing match status gives a measure of the reliability of the production PES information, and that is what Section 5.5 is about; EFK's example is a complete misreading of a minor issue.

3. I used questionable data.
EFK write that my analysis is based largely on the three sources of evaluation data: the rematching study, quality control and the field reinterviews (EFU). This is only true if "largely" is given a generous interpretation. The one-million-person coding error is not connected with the evaluation data. The later rematching of 104 selected blocks gave another error correction of 250,000 persons.

They note that the EFU data was gathered 5–6 months after the initial PES and that this delay would cause uncertainty in the data. I agree, and I also wish that the EFU had occurred earlier, but there was good reason for the timing. The PES follow-up took place about 3–4 months after the initial PES. The EFU could not be carried out until the PES follow-up was over.

Ericksen, Fienberg and Kadane imply that the effect of the uncertainty would be to inflate the error estimates. There is no evidence supporting this, and some to the contrary. Since each EFU interviewer carried around the previously completed PES interview forms for the households being reinterviewed, disagreements between the EFU and the PES would, if anything, be biased low (see Biemer and Forsman, 1992). At any rate, the EFU provides the only reinterview field data available and was used both by the bureau in their evaluation (which EFK approve) and myself.

4. I assume zero correlation bias.
If some persons avoid official surveys, they will be difficult to count both by the census and by the PES. The capture–recapture assumptions will not apply, and the DSE estimates will be biased low. This is called correlation bias. Now, the argument is that, although data errors tend to cause an overestimate of the undercount, such errors are largely canceled out by the effect of correlation bias (see Ericksen, Estrada, Tukey and Wolter, 1991).

Both EFK and BR are concerned that "zero correlation bias is assumed." This is not correct. Zero correlation bias is not assumed. I willingly admit that in places correlation bias is likely. However, correlation bias is irrelevant to my study. What I look at is how much of the DSE undercount estimates is attributable to bad data. This has no relation to what is assumed about correlation bias.

5. The Bureau's evaluation is better.
Early on in their comments, EFK say, referring to the Bureau's evaluation, "This evaluation was sum-

marized by Mulry and Spencer (1993)." Later, they add: "We prefer to rely upon the evaluation made by the Census Bureau, partly because they have studied the questions more thoroughly, but also because we believe them to be more objective."

The Mulry–Spencer article is not a summary, but a selective extract from the bureau's 1991 evaluation report [P16]. For instance, the estimated bias due to bad data is not reported separately as it is in [P16]. Instead, the only bias figures given are the lower ones obtained by adding a problematic estimate of correlation bias to the data error bias. Thus, the reader is not given the [P16] information that there is a 0.7% bias at the national level due to bad data nor are they given the corresponding [P16] information for the evaluation strata.

The article, although published in late 1993, does not include corrections for the one-million-person coding error discovered by the Bureau two years earlier. It does not correct for other errors detailed in my article that the Bureau later incorporated into its intercensal error analysis (Mulry, 1992b). These include the late late census data and the rematching of 104 blocks.

The bureau has also admitted that they overlooked the new-out-of-scopes problem and to making errors in computing the census day address error (see the Appendix to my article). These latter two, along with many other error sources, were discussed in my article but not in the Bureau evaluation nor in Mulry and Spencer (1993). EFK's view regarding thoroughness and objectivity is not well founded. They are putting their reliance on an analysis known to both the Census Bureau and other statisticians as outdated and erroneous.

6. I worked on the wrong problem.

"... Breiman's paper seems misdirected." "... Breiman focused his time and energy on the wrong problem." "Because Breiman studied the wrong problem... we do not feel that his conclusions matter greatly."

It is the wrong problem because "Breiman studied ... the national net undercount as opposed to the distribution of this undercount...." Not so—my article explicitly gives the undercount estimates and the effects of the errors on them for the 13 evaluation strata (see Table 16). As EFK know, the only evaluation data available was at the level of these evaluation strata. These data were used by the bureau and myself to see how much of the DSE estimates at the national and stratum level could be attributed to bad data.

If I studied the wrong problem, so did the Bureau, but EFK never claim, either in this discussion or in any of their past articles and reports, that the Bureau studied the wrong problem. They do not explain why they believe that the Bureau studied the right problem, but I studied the wrong one. Perhaps EFK think that if they say I am working on the wrong problem often enough, then somehow my conclusions will go away.

7. My results do not make much difference.

Now EFK go to their final line of defense: even if I studied the right problem, my results do not make much difference. Referring to the undercount estimates in the evaluation strata, they say "his numerical results are not greatly different from those of the bureau [DSE adjustments], especially when we consider between-area differentials."

They then try to show that the differences between the DSE estimates and the error-corrected estimates do not matter much; in particular, that the minority–nonminority undercount differentials remain about the same. To do this, they use a somewhat convoluted procedure leading to the conclusion that the differences matter in only one half of the country. One half of the country is a pretty big slice.

Since EFK consider the minority–nonminority differential to be an important indicator, let us see what the effect is using a simple computation. We take as our measure the estimated population proportion of the five minority strata. Using the DSE estimates, the increase in this proportion over the census is 51% larger than the change computed using the error-corrected estimates. A difference of 51% can hardly be called minor.

To see what effect the corrections have on the stratum-level distribution, define the population share of a stratum to be its fraction of the total population estimate. A standard measure of the difference in two population distributions is the sum over the strata of the squares of the differences of the shares. Using this measure, the difference between the DSE adjustments and the census is 74% larger than the difference between the corrected adjustments and the census. By any definition, 74% is substantial.

The right problem was studied and the differences matter. The errors listed in my article significantly affect minority–nonminority differentials and population shares in the evaluation strata. They are likely to have even more effect at lower levels of aggregation. Furthermore, the strata most affected by the errors were the minority strata (see my Table 16). The implication is that the DSE adjustments are likely to be the worst just where we would want them to be the best.

Conclusions about EFK

The above summarizes EFK's comments on my article and gives my rejoinder. The Census Bureau, in its original [P16] analysis, conceded that about 30% of the national undercount estimate was due to bad data. Ericksen, Fienberg and Kadane do not seem to be aware that in 1992 the Bureau's estimate of undercount, corrected for bias due to data errors, dropped to 0.9%, a tacit admission that almost 60% of its original undercount estimate was due to bad data (Mulry, 1992b). My estimate of 80% is more realistic. Regardless of how error-filled the census is, you cannot solve the problem by combining it with another error-filled set of numbers.

BELIN AND ROLPH (BR)

The Belin–Rolph comments on my article are sharply critical. I will go through their remarks section by section before responding to their accusations of professional misconduct. Here is a list of subjects covered in their sections:
Section 5.1, the source of my article;
Section 5.2, a contradiction, national versus differential;
Section 5.3, a contradiction, two ways of looking at matching error;
Section 5.4, obfuscation in analyzing fabrications;
Section 5.5, the assumption of zero correlation bias;
Sections 5.6, 5.7 and the Appendix, two tables of imputation data;
Section 5.8, my dealings with the Census Bureau.

Section 5.1 notes that my present article has its roots in a manuscript prepared for the adjustment lawsuit. Quite true, but so?

Section 5.2 refers to a supposed contradiction in my analysis. The contradiction is that, even conceding that the estimates aggregated to the national level are largely due to errors, it has not been shown that the estimated population distributions at the local level are affected by the errors. This point was raised by EFK and responded to above.

This section has a number of questionable remarks. For instance, in the first paragraph BR think they have discovered one of my big ideas: "One of Breiman's main points is that if every area were undercounted by the same amount, then adjustment would be superfluous and would only add error." Neither this "main point" nor anything resembling it occurs anywhere in my article.

Further on BR ask "Why, then, does he not label evaluation poststrata as minority or nonminority?" Table 4 in my article, which defines the evaluation strata, clearly labels each stratum as minority or nonminority. Belin and Rolph continue: "To allow readers to see the differential minority undercount, we reproduce Breiman's Table 16 with information on minority status added...." Not so! My Table 16 shows the initial 1991 adjustments and the substantial reductions in differential minority undercount when the 1991 adjustments are corrected for errors. This information is eliminated (without comment) in the BR "reproduction."

Next BR say that "after accounting for the computer error, the Census Bureau's loss function calculations favor the adjusted over the unadjusted counts at the state level" and give some references in support. None of their references do what BR claim. Mulry (1992b) deals with the intercensal adjustments. These differ considerably from the original DSE adjustments. Mulry and Spencer (1993) do not account for the computer error; Fay and Thompson (1993) give a review of the Mulry (1992b) work; and Zaslavsky (1993a) is based on the Mulry (1992b) adjustments.

Section 5.3 refers in its title to another contradiction in my work. They state: "...particularly in his discussion of matching error, Breiman does not take into account that errors can cancel." Of course I take this into account and explicitly discuss it in Section 6.1. Then they say: "Curiously ... Breiman then uses the Census Bureau's calculations of the effects of matching errors in his Table 15." As far as I can make out, the contradiction referred to is that I cite both disagreement rates over cases and marginal disagreement rates. The reason for this is carefully explained in the first three paragraphs of my Section 5.1 on matching; BR have either overlooked the technical reasoning or do not understand it.

In brief, the differences in the marginal totals of the match–rematch data for the evaluation strata are direct estimates of the matching error and were used by the Bureau and myself. However, I show that marginal disagreement rates at highly aggregated levels tend to average out and are not good indicators of marginal disagreement rates at less aggregated levels. The disagreement rates over cases are better indicators and are cited for this reason.

Section 5.4 on fabricated interviews refers in its title to obfuscation in my work. The section begins with a quote, supposedly from my article. It is not in my article nor in the original court document. The sentence in my article following the extract from [P6] does not have the faintest resemblance to the quote that BR attribute to me. Then BR imply that I claim the undercount is overestimated by millions due to undetected fabrications in the PES. This is not what

I say, and if readers are in doubt, I would request that they review my Section 5.2.

Further BR comments indicate lack of careful reading. For instance, they state that "quality-control checks suggesting 2–5% interviewer fabrication does not translate to 2–5% of cases entered in the P-sample database being fabrications." Of course not, but they have the 2–5% numbers wrong. The 2–5% numbers come from report [P6] and have nothing to do with quality control. This confusion leads them on a merry chase.

Their concluding statement is: "Breiman discounts the possibility that quality control... would have been able to catch fictitious enumerations." No such discounting occurs. Nobody, least of all the Census Bureau, claims to have a foolproof way of detecting all fabrications. The question is how many got through the PES quality control procedures. What I present is the evidence given in the P-studies concerning this question.

Section 5.5 raises the issue of assuming zero correlation bias. See my response to EFK on this subject.

Sections 5.6 and 5.7 and the Appendix, a third of BR's comments, are about two tables. These two, Tables 12 and 13, appear in my Section 5.7 on imputation and form a minor part of the study. Table 12 contains a summary of the P-sample (PES) imputation evaluation data, and Table 13 a summary of the E-sample (census) imputation evaluation data. They are taken directly from Tables 3.1 and 3.2 of Bureau report [P3].

Section 5.6 is titled "Breiman's Table 13 as 'Bad Data'." Belin and Rolph and I agree that Table 13 does not look good for imputation. They believe that this table is erroneous and have a long story about why I am to be blamed for reproducing the [P3] version. First are two memos Belin wrote to his superiors in the Bureau, but which he admits might have been difficult for me to know about. Next, I should have corrected the error because Belin told a referee about the corrections and the referee was supposed to tell me. There were no such referee comments.

It is remarkable that BR spend so much energy criticizing me for not knowing about and not correcting a supposed error in a table taken from an official Census Bureau report. They are hassling the wrong person. Their argument is with the Bureau and not with me. If the table is wrong, then report [P3] needs correction, and that should be done by the Bureau.

Section 5.7 concerns Table 12 and its collapsed version, Table 11. The section title refers to the "prism" through which I view these tables. They begin with a misreading of my paper. In their first paragraph, BR note that in my Table 11 the percentage of matches among the resolved is 12/39 = 30.8%, while the percentage given in Belin et al. (1993) is 31.6%. Belin and Rolph comment, in a quite annoyed way, that this difference could not be due to rounding. They are right. As stated in my article and in report [P3], the numbers in Table 11 are weighted to the nation. In the Belin article the results are unweighted.

Next BR exhibit a truncated version of Table 12. The evaluation data on P-sample imputations contains almost 60% unresolved data, analogous to nonresponses in a survey. The BR game is to deal with the unresolved data by pretending that it is not there, ergo, the truncated table. Then things look better for imputation. The Bureau does not see it the BR way. Report [P3], in summarizing Table 12, states: "Thus, for P-Sample persons, the imputation process is consistent with EFU results. However, the high percentage of unresolved persons in the EFU (58.55 percent) may limit the utility of this result."

After all the time BR spend on these two tables, this is what we are left with: Table 12 is okay—no questions about its accuracy. Belin and Rolph need to work with the Bureau to resolve their claims about Table 13. They end their discussion of these two tables with the revealing statement that for them the real issue is not the data and substance underlying imputation, but personalities and politics.

Section 5.8 contains a lengthy lecture based on the sentence in my Appendix reading "I have been unable to obtain from the Bureau any more specific information regarding their method for computing census day address error." The section consists mainly of moral disapproval of my efforts to understand how the census day address error was computed and little that is relevant to the substantive issues. I stand by the accuracy of my statements.

In this section, John Thompson, a senior Bureau statistician, is quoted as saying that "[Breiman's] numbers appear to be correct..." but that the Bureau should have the chance to comment and give their interpretation. I agree. It would be instructive to have a Bureau viewpoint represented.

BELIN AND ROLPH'S ACCUSATIONS

Belin and Rolph begin their review of my work by accusing me of "unprofessional practices" and in the conclusion they state "...Breiman is not careful with his facts, and one of the most harmful aspects of Breiman's paper, we believe, is that Breiman has cultivated a sense of distrust of government statisticians that we regard as unjustified and irresponsible."

This is an quite an indictment and should be accompanied by convincing proof of unethical behavior and distortion of facts. Belin and Rolph have not come up with a single fact that I have been careless with, nor any instance of unethical behavior. On the other hand, they have gotten their facts wrong, have been careless in reading, have found contradictions and obfuscations where there are none and have spent most of their time on irrelevant side issues and morality mongering.

Rejoinder

D. Freedman and K. Wachter

1. INTRODUCTION

Census adjustment is not an easy topic. We are grateful to the discussants for their efforts at clarifying the issues. One other idea will not be controversial: Rob Kass and Ram Gnanadesikan deserve thanks for putting this exchange together and bringing it to a successful conclusion.

The commentaries fall naturally into two groups, those from outside the United States and those from inside. It is valuable to have perspectives gained from experience in other countries. We marvel, naturally, at errors measured in hundreds, which Lyberg and Lundstrom attribute to Sweden's PIN-keyed registers. Australia as described by Steel makes an interesting contrast to Britain (Diamond and Skinner), in terms of the trust accorded to results from demographic analysis in Britain and the distrust in Australia—even though Australia has effective monitoring of international migration, which removes one of the chief components of uncertainty in demographic analysis for the United States.

Belin and Rolph (BR) have a free-spirited and wide-ranging commentary which reviews many previous exchanges on the census. Much as we like the authors, we differ with them on readings of the technical and historical record and on matters of scientific principle. With respect to the census, Ericksen, Fienberg and Kadane (EFK) are among the oldest and most familiar of our opponents; but on this occasion, as we shall explain, their critique is off the mark entirely.

According to the rules of engagement, we do not comment on BR's rejoinder and they do not comment on ours, so we get the last word in this exchange—on these pages of this journal. Silence cannot be interpreted as consent: we are sure that BR and EFK will continue the argument in some other forum.

The Census Bureau's latest thinking on the 1991–1992 adjustments is described in Fay and Thompson (1993). Our discussants frequently refer to this paper for arbitration, and we shall too. We hope Bob Fay and John Thompson will not mind such close textual analysis.

Despite the scope of BR's remarks, our paper was not really about the bottom-line question: whether adjustment would have made errors in state population shares better or worse. It was, rather, about a "wild card" in the Census Bureau's assessments of state and local coverage error: heterogeneity. Heterogeneity was omitted from the bureau's loss function analysis. Statisticians on all sides have been arguing ever since what kind of difference that could have made.

In our paper, we measured the difference that heterogeneity does make, in a context that allows an exact answer. We used the same loss function that the Bureau did, with proxy variables instead of undercounts. In our context, the adjustment factors are known with perfect accuracy, so that errors of adjustment are due purely to heterogeneity, and loss itself can be calculated. We found the following:

1. The omission of heterogeneity does bias the estimated risks.
2. Depending on the proxy, the bias can be small or it can be large.
3. The bias can go either way, for or against adjustment.

In particular, we established that loss function analysis can be strongly biased in favour of adjustment; EFK and BR react quite critically to this finding. Before we answer them, let us recall the larger picture in which such arguments take their place.

2. BACKGROUND

Would the proposed adjustment of the 1990 census, or of the intercensal estimates, have improved the accuracy of population shares held by the various states? "Loss function analysis" attempts to balance errors in the census against errors in adjustment, and it seems to be the principal statistical argument that adjustment would improve on the census (BR,

Section 2; Woltman et al., 1991; Mulry and Spencer, 1993; Freedman, Wachter, Cutler and Klein, 1994).

Loss function analysis has several ingredients. The first is estimated sampling error. In the context of census adjustment, the Bureau's estimates for variance (based on their smoothing model) were substantially too optimistic, by a factor of 2 or 3 (Fay and Thompson, 1993, page 83; Freedman et al., 1993); this problem disappears for the intercensals, since the smoothing model was not used. Another input to loss function analysis is the estimated levels of bias in the PES (due to matching error, census day address error and so forth).

The Bureau elected not to estimate bias in the PES for geographical areas, or even for poststrata. Instead, errors were determined for very large aggregates of poststrata, called evaluation strata. With the census, there were 1,392 poststrata and 13 evaluation strata; with the intercensals, there were 357 poststrata and 10 evaluation strata. Breiman shows the Census Bureau's 1991 estimates for bias to be far too optimistic. There seems to be some agreement on these points, the discussants notwithstanding (Section 6 below).

To adjust population counts in small areas, including states, the bureau made the "synthetic assumption": undercount rates are constant within poststrata across geography. Thus, rates are determined by demographics not geography. Failures in the synthetic assumption are termed "heterogeneity." The loss function analysis was done on the basis of the synthetic assumption and therefore could not measure the impact of heterogeneity. To allocate biases from evaluation strata to poststrata, error rates were assumed constant within evaluation strata across poststrata, and then within poststrata across geography—an even stronger form of the synthetic assumption. (For details, see Freedman, Wachter, Cutler and Klein, 1994, pages 260–261.)

Section 5 below indicates how the results of loss function analysis change if we correct the variance and bias estimates. We also show that the scheme for allocating errors from evaluation strata to individual poststrata has considerable influence (Freedman, Wachter, Cutler and Klein, 1994, pages 262, 264). Our present paper, however, dealt with another topic in the adjustment debate: the impact of heterogeneity on loss function analysis. Does heterogeneity create bias in estimated risks? ("Risk" is expected loss, and the goal of the loss function analysis was to obtain unbiased estimates of risk.)

The modelers have taken several positions:

1. Heterogeneity is trivial.
2. If not, the impact on loss function analysis is trivial ("robustness").
3. If neither 1 nor 2 is admitted, then loss function analysis is "conservative": the proadjustment position is even stronger than the data say it is.

Belin and Rolph now concede that heterogeneity is substantial, but stand on point 2; EFK concede nothing, but insist on point 3. The reasoning will be discussed below. We stress that the issue is bias in estimated risks, not bias or variance in estimated counts—a distinction that some discussants seem to have found quite subtle.

Our paper uses proxies for the undercount: variables (such as substitutions or imputations) thought by the Census Bureau to resemble the undercount with respect to heterogeneity. The basic setup involves a matrix; the row index i corresponds to poststrata, running from 1 to 357 in our examples; the column index j corresponds to geographical areas, for instance, the 50 states and Washington, D.C. In each cell, we have the census count c_{ij} and the proxy undercount t_{ij}. The "true" population of area j is the column sum of $c_{ij} + t_{ij}$, with i running from 1 to 357.

Take this from the perspective of statisticians who know the census counts c_{ij} in each cell, but not the t_{ij}. Although the undercounts are unknown, row totals are given:

$$\sum_{j=1}^{51} t_{ij} \quad \text{for } i = 1, \ldots, 357.$$

Now the t_{ij} can be estimated by the "synthetic method,"

$$\hat{t}_{ij} = f_i \times c_{ij}.$$

The factors f_i are computed as follows:

$$f_i = \frac{\sum_{j=1}^{51} t_{ij}}{\sum_{j=1}^{51} c_{ij}}.$$

The row totals of the proxy undercounts are given, so the factors f_i are free of error. (We do not believe that all our discussants paid due attention to this point.) In the adjustment literature, $1 + f_i$ is an "adjustment factor" which adjusts the census count in a cell to its estimated true count.

The "adjusted" population of area j can be obtained as a column total of $c_{ij} + \hat{t}_{ij}$, and shares can be computed in the obvious way. Moreover, the squared error for the raw census and the adjustment can be estimated using the synthetic assumption; comparing these estimates to the true errors enables us to measure the bias in loss function analysis due to heterogeneity.

Risks are estimated as follows. On the basis of the synthetic assumption, adjustment "must" get the

right answer, so its squared error is zero; and the squared error of the census can be calculated with the adjusted population taken as "truth." Of course, in the presence of heterogeneity, such estimates may be quite misleading.

The setup is rather like real census adjustment. The whole object of the PES, poststratification, smoothing and so forth is to obtain estimates of adjustment factors $1 + f_i$. However, our setup differs from adjustment in two major respects: (i) The row totals of the proxies (and hence the adjustment factors themselves) are free of error. (ii) We are working with proxies rather than undercounts. The advantage of (i) is that we can focus on heterogeneity, pure and simple; (ii) is the price.

3. ERICKSEN, FIENBERG AND KADANE

The chief technical argument offered by EFK is that "Heterogeneity... is not inimical to adjustment," and they cite theorems to support their position. However, EFK have utterly missed the point. Their theorems say that, in some circumstances, estimated adjustment factors are too small. In our context, the adjustment factors are not estimated. They are known exactly. There is zero bias and zero variance. The theorems cited by EFK are irrelevant to the calculations that we present. EFK also complain that our adjustment factors are not smoothed. Of course not—smoothing would (if all went well) reduce sampling error in the factors. Our factors are known and are not subject to sampling error. Smoothing them would just add bias.

Even in a context where adjustment factors are only estimates, a distinction must be drawn (EFK always resist this) between (i) adding more people to rows of the matrix and (ii) changing the population shares of columns. The mathematics cited by EFK suggest that, in some circumstances, people should be added to a poststratum beyond what is done by the dual system estimator; in this sense, adjustment is "conservative." That, however, has no direct implications for population shares of states, because the theorems do not say where the additional people live. State shares are the focus of the analysis, not population counts for poststrata.

When discussing our proxy "DIFF," EFK impute to us a regression model for undercount rates, and demand justification. "What evidence do we have for believing such models," they ask. They may have woken up to the idea that assumptions need to be justified—a welcome development. However, they seem to misunderstand DIFF. The Census Bureau's proxies all have a somewhat unrealistic feature: they are positive everywhere. We constructed DIFF, not by regression, but to have a positive part like gross omissions (the undercounts) and a negative part like erroneous enumerations (overcounts). Indeed, the correlations reported in our Table 10 should warn against regression modeling.

Proxies are used (by the Bureau and by us) not to model undercounts in the sense of regression, but as *proxies*: variables that are analogous to undercounts and that stand in for them (Fay and Thompson, 1993, Section 4.3). If the analogy is a bad one (heterogeneity in undercounts does not behave like heterogeneity in the proxies), then the Bureau's research effort on the proxies casts no light on adjustment, and neither does our paper. We made the point earlier; it bears repetition. Conversely, if heterogeneity in undercounts behaves like heterogeneity in the proxies, our work says something about loss function analysis: appreciable bias in risk estimates is a distinct possibility.

Ericksen, Fienberg and Kadane do have one serious point. The 1,392 poststrata used for adjusting the census are different from the 357 for the intercensals. Our findings on the 357 poststrata are most relevant to the bureau's loss function analyses on the intercensals (Bureau of the Census, 1992c, 1993). According to EFK, the fact that "the 357 poststrata have too much residual variability...comes as no surprise"; indeed, dropping the number of poststrata from 1,392 to 357 "was sure to introduce greater heterogeneity." Since we have looked at heterogeneity for the 1,392 poststrata, EFK's theory can be tested.

Results for the 1,392 poststrata are based on a sample covering 200,000 blocks, drawn from the 1990 census by the bureau for its P12 Evaluation Project (Kim, 1991). Unlike calculations for the 357 poststrata, which were based on the whole census and therefore not subject to sample variability, our estimates for the 1,392 poststrata include a correction term for sampling error (Wachter and Freedman, 1994). Results are shown in Table R1; the proxies are substitutions, allocations, multiunit housing and non mailbacks.

The first two columns correspond to columns 3 and 4 in Table 5 of our paper. The last column measures heterogeneity across local areas rather than states. The standard deviations within poststrata across states are all of roughly the same magnitude as before. The value for multiunit housing is bigger, the others are smaller.

Ericksen, Fienberg and Kadane may be right, that 1,392 poststrata do have less residual heterogeneity than 357, although the difference seems minor. Their main idea, that the levels of heterogeneity uncovered by our paper arise from dropping the number of poststrata from 1,392 to 357, turns out to be mistaken. Even with 1,392 poststrata, there was plenty of het-

TABLE R1
Heterogeneity with 1,392 poststrata

	Standard Deviations (%)		
	Across post-strata	Within poststrata across states	Within post-strata across local areas
SUB	0.7	0.6	2.3
ALL	8.0	2.9	7.1
MUH	23.7	10.4	22.3
NMB	12.0	4.3	10.7

Notes: "Local" areas have populations of about 10,000; r.m.s. standard deviations across poststrata are reported, with 12 Indian poststrata excluded. Standard errors are relatively small. For additional details, see Wachter and Freedman (1994).

erogeneity at the state level, while the level of heterogeneity for substate areas was striking.

4. BELIN AND ROLPH

Model Validation

As we read them, BR simply do not accept the idea that modelers have a responsibility to validate the models (Section 2.3). The closest they come is the idea that models are better when "violations of the assumptions . . . appear to be harder to find" (Section 2.4). Once again, the burden of disproof is placed on the critic.

Loss Function Analysis

Belin and Rolph ask (Section 2.6), "What is the right loss function?" This a question without an answer, because loss functions do not measure real losses; they are only summaries of error distributions, which may or may not be useful. The papers cited by BR at the end of Section 2.6 do not face up to that central difficulty. For example, here is Zaslavsky (1993a, page 1092), arguing for squared errors: "If there is a social good (such as a government expenditure) to be allocated, and aggregate utility is a smooth (twice differentiable) function. . . ." But why is there an aggregate utility function in the first place, rough or smooth? Over the years, the attempt to construct aggregate utility functions has met insurmountable technical and conceptual problems; Sen (1988) reviews work in this area.

For such reasons among others, we are skeptical of loss function analysis (Wachter, 1991; Freedman, Wachter, Cutler and Klein, 1994). Now, BR claim (Section 2.2), our position is self-contradictory. On the one hand, we believe "that no index for summarizing the evidence from data is an unambiguous measure of whether one estimate is better than another." On the other hand, we think census adjustment is an empirical issue. But where is the contradiction? Their argument depends on an implicit assumption: the only way to make an empirical assessment is to pick a loss function and run the numbers. They cannot be right. For one thing, optimal estimators—optimal by all sensible criteria—seldom exist. That would seem to be the central lesson from 50 years of work on estimation theory. Indeed, a few pages later, BR concede that "it is impossible to determine a single loss function that is appropriate for evaluating every effect of an error in census numbers" (Section 2.6, item 1).

No single loss function can do the job, for reasons given above: loss functions do not measure the real social costs of errors, and different summaries of error distributions may give the advantage to different estimators. Even more to the point, when risks (expected losses) have to be estimated, different statistical assumptions about the data may give radically different conclusions. At best, loss function analysis is only part of an empirical assessment. Indeed, BR's bottom line (at the end of Section 2.2) is quite modest: they "see no reason to ignore the evidence from research on loss function analysis." Neither do we. Section 5 below reviews this evidence, as it bears on the key issue—the accuracy of state population shares.

We turn to the narrower question in our paper: does heterogeneity create bias in estimated risks? Our Table 9 covered seven proxies and showed that the biases could be large or small, proadjustment or antiadjustment. We conclude that the data cannot decide the issue. There is a critical parameter that does not seem estimable: the correlation between errors in the adjusted shares and the adjustments themselves.

Belin and Rolph (Section 4) make two arguments on this topic: (i) on average, over the seven proxies in the table, the results "favor neither adjusted nor unadjusted figures" and (ii) this is to be expected on theoretical grounds because "it does not seem obvious why such a correlation [between errors and adjustments] would occur." The structure of these arguments is illuminating. In (i), lack of knowledge about which proxy best represents the undercount is replaced by a model for ignorance (the uniform distribution); then strong conclusions are drawn from the model. In (ii), an unknown correlation is replaced by 0. The unknown is made known by models, whose conclusions are to be accepted unless they can be disproved.

Belin and Rolph describe our position as "adversarial," preferring "the more balanced interpretation in Fay and Thompson (1993)"; EFK appeal to the same paper for the same reason. But what do Fay and Thompson say? "Failure of the homogeneity as-

sumption is potentially a larger source of error than all errors explicitly included in the total error model and loss function analysis"; Fay and Thompson go on to ask, "what sense can then be made of the loss function analysis?" After reviewing the bureau's research program, they conclude:

> A much larger base of experience along the same lines may suggest principles or test procedures to distinguish the circumstance under which the loss function analysis is robust [against heterogeneity] compared to instances leading to its breakdown .... In 1990, the issue of heterogeneity affected the most constitutionally important statistics: the population of the states ... future designs should set realistic and clearly defined reliability goals for direct estimates [not using the synthetic assumption] for states. (Fay and Thompson, 1993, pages 81–83)

Will EFK and BR accept that formulation?

Imputation Models

The 1991 estimated undercount from the PES was about 5.3 million persons (net, nationwide). However, 4.1 million persons (weighted to national totals) were "unresolved" in the P-sample: it could not be determined from the PES fieldwork whether or not these cases matched to the census. Their match status was imputed by—you guessed it—a model. This imputation model has a powerful effect on estimated adjustments at the state level (Wachter, 1991), and the model has two very peculiar features (Wachter, 1993b). To explain these, we refer to the Evaluation Followup Survey (the EFU), a second survey that tried much later to reinterview a sample of the unresolved P-sample cases.

(i) Nearly 32% of the unresolved P-sample cases fell into a special class, which we shall call the Q-class (Q for "question marks"). This Q-class consisted of cases about whom only minimal information was obtained in the first wave of PES interviews. It was not judged cost-effective to send such cases to PES follow-up, still less to the Evaluation Followup many months later. Instead, match status was imputed by assuming that cases in the Q-class were like PES respondents with the strongest information, namely, those who could be matched by the computer after the first wave of PES interviews. This is a peculiar assumption.

(ii) The remaining 100% − 32% = 68% of the unresolved PES cases were in the EFU sampling frame: fieldwork was done by the EFU on a sample of these cases, to validate the imputation model. However, the EFU could resolve only 41% of the cases that were sent to EFU. Thus, 59% remained unresolved (match status to the census remained indeterminate). The unresolved group must consist of cases with weak data. However, the imputation model says that these weak cases match to the census at a much higher rate than the cases resolved in EFU—the cases with strong data; details are in Appendix 1. This is equally peculiar.

In Sections 2.3 and 5.7, BR now try to defend the imputation model, using data from the EFU. However, the EFU resolved only $(1 - 0.32) \times 0.41 = 28\%$ of the unresolved PES cases. What about the remaining 72% of the PES imputations? Were these right or wrong? The EFU results cannot tell us, because the EFU could not decide the match status of those persons. This difficulty has been explained in the journals (Wachter, 1993b) and in private correspondence, a snippet of which BR reproduce. BR respond by calling our position "very extreme," but where do they draw the line? If 90% of the data are missing? 95%? 99%? Or do they think that any amount of missing data can be filled in, just by making up models?

The model's assumptions are silly. It is even sillier to claim these assumptions have been validated by the EFU, when the groups in question were either not sent to EFU or remained unresolved in EFU.

When Leo Breiman looks at the EFU data—all the EFU data—he finds another paradox: the rate of unresolved in EFU goes up with predicted match rate. The Census Bureau did the same analysis (Gbur, 1991c). Belin and Rolph say that our friend Leo is "play[ing] games with ... percentages" by looking at all the data: BR insist that percentages should be based only on cases resolved in EFU. That is because BR look at the data only through the prism of their models. BR are playing games with models; Leo is blowing the whistle.

Some Points of Detail

We think BR are wrong on many points of factual detail, and their interpretations of the scientific literature are often quite strained. We give three examples.

Documentation. According to BR (Section 2.3), the adjustment process should have been, and was, "fully documented with its assumptions spelled out." We have replicated many parts of the Bureau's smoothing model and loss function calculations (Freedman et al., 1993, page 416; Freedman, Wachter, Cutler and Klein, 1994, page 277). However, we had a lot of friendly help from Bureau personnel. An investigator who seeks to do such work just on the basis of the printed record will have quite a frustrating time: the

documentation is maddeningly cryptic about many critical issues.

Presmoothing. Before using estimated variances to smooth the adjustment factors, the Bureau "presmooths" the variances: indeed, smoothing without presmoothing would have cut the estimated undercount from 2.1% to 1.2% (Ericksen and Tukey, 1991, page 2). Our opponents have defended presmoothing on technical grounds, but we believe their arguments are essentially circular (Freedman et al., 1993, pages 383–385). In response, BR assert (at the end of Section 3.8) that the Census Bureau decided "to carry out presmoothing for bias-reduction reasons" around 1988 or 1989. They are mistaken. Some form of presmoothing was indeed under active consideration by the Bureau, at least since the 1988 test census in St. Louis (the "dress rehearsal"). However, the key decisions—whether to presmooth and how to do it—were still being debated in 1991. For example, one of us participated in such discussions with senior personnel from the Bureau and the Special Advisory Panel on May 16 of that year.

Hindsight is 20/30. There was an adjustment to the census proposed in 1991 and evaluated in 1991, based on data available in 1991, including data on estimated errors in the PES. Subsequently, additional errors were discovered (Section 6 below). Some of these were corrected, leading to another adjustment proposed in 1992 for the intercensals, and evaluated based on data available in 1992 about the remaining errors.

In Section 5.2, BR suggest that evaluating the 1991 adjustment based on the 1992 error estimates (rather than the 1991 error estimates) makes the case for the 1991 adjustment even stronger. That just misreads the literature. Shown below are three possible evaluations:

(a) the 1991 adjustment evaluated using 1991 error estimates;
(b) the 1992 adjustment evaluated using 1992 error estimates;
(c) the 1991 adjustment evaluated using 1992 error estimates.

The papers cited by BR focus on (a) and (b). Some of the work for (c) has been done; but BR ignore the results, which are summarized below.

5. LOSS FUNCTION ANALYSIS

We now consider loss function analysis for the 1991 proposed adjustment to the 1990 census; 1,392 poststrata are in full sway. The focus is on state population shares; "loss" is squared error in state shares (Woltman et al., 1991; Mulry and Spencer, 1993). Let G be the Census Bureau's estimated covariance matrix for the adjustments, as derived from their smoothing model; let H be the estimated covariance matrix for estimated biases in those adjustments. (Although H does not affect point estimates of risk, it does come into the estimated standard errors.)

As noted earlier, G is biased downward, by a factor of 2 or 3. It is shown in Freedman, Wachter,Cutler and Klein (1994, pages 268ff) that H too is biased downward, by a factor on the order of 50 or 100. That is a remarkable claim, and we stand behind it. In brief, the Bureau estimated biases on the basis of an Evaluation Follow-Up sample that was perhaps 7% of the size of the PES (in terms of households, at any rate). The Bureau is claiming variances about 6 times smaller than raw variances in the PES, rather than $1/0.07 = 14$ times larger: $6 \times 14 = 84$ is a big factor. The bureau achieved its reduction in apparent variance by computing H on the basis of an allocation scheme, omitting any uncertainty due to variation in error rates across poststrata or geography. (For a tantalizing hint on the existence of this difficulty, see Fay and Thompson, 1993, page 79.)

What are the implications for loss function analysis? The first four lines of Table R2 allocate bias from evaluation strata to individual poststrata using the bureau's "PRODSE" method. The last four lines of the table increase the level of bias to 50% of the undercount (Section 6) and allocate in proportion to the undercount. The effects of correcting G and H are shown too. For details, see Freedman, Wachter, Cutler and Klein (1994).

With the Bureau's way of doing things, reported in the first line of the table, the estimated risk difference (census risk -- adjustment risk) is 667 parts per 100 million, with an SE of 281: adjustment is a winner. In the last line of the table, which strikes us as the most realistic, the census comes out ahead: the difference is not significant. With intermediate lines, the case for adjustment is hardly convincing. In sum, loss function analysis is driven by the models that underlie it not by the data.

Line 1 in Table R2 may also be contrasted with line 5, where bias is allocated as 25% of the undercount. The contrast demonstrates that the allocation scheme determines the outcome even if we grant the Bureau their estimates for G, H and overall level of bias. (The Bureau's 1991 loss function analysis included an allowance for bias amounting to nearly 25% of the net undercount; Mulry, 1991, Table 14.)

Table R2 is computed on the basis of the synthetic assumption. If undercount rates are heterogeneous within poststrata across states, that would be another source of bias in the estimated risk differences. How large is the effect? We do not know, and no-

TABLE R2
Impact of allocation schemes for state-level biases, correction of final variances and correction of variances in estimated biases: Estimated risk difference, census risk − adjustment risk, and standard error; units are parts per 100 million

Allocation of bias	Correction factor for G	Correction factor for H	Estimated risk difference	SE
PRODSE	1	1	667	281
PRODSE	1	50	667	890
PRODSE	2	1	542	371
PRODSE	2	50	542	885
0.25 × undercount	1	1	193	199
0.50 × undercount	1	1	−125	156
0.50 × undercount	1	50	−125	859
0.50 × undercount	2	1	−250	169
0.50 × undercount	2	50	−250	321

body else does either. For example, if ALL is a good proxy for undercounts and we can extrapolate from 357 poststrata to 1,392, the bias in estimated risks will be small. If DIFF is the better analog, the bias is of the same order of magnitude as the estimated risk difference itself, and favors adjustment.

6. BIAS IN THE PES

We distinguish between "measured" and "unmeasured" bias in the PES. Measured bias is caused by matching error, census day address error and so forth. In principle, such biases can be estimated by reinterviewing and rematching studies, although the difficulties are numerous. Generally, the measured biases cause the PES population estimates to be too high. Correlation bias, on the other hand, is unmeasured. Typically, this bias occurs when (even within a poststratum) people who are missed by the census are also more likely to be missed by the PES. This sort of bias makes the PES estimates too low. There are "unreached people," missed both by the census and by the PES adjustment.

We discuss the measured biases first, then return to correlation bias. In July 1991, the estimate for the net national undercount was 2.1%, with measured biases thought to total 0.7 percentage points. Additional errors were discovered in the PES (BR, Section 5.2; Fay and Thompson, 1993, page 74; Bureau of the Census, 1993, page 75.) These errors reduced the undercount estimate by 0.5 percentage points. The measured biases were still thought to total 0.7 percentage points, leaving $2.1 - 0.5 - 0.7 = 0.9\%$ for the undercount. In other words, on the bureau's latest figures, measured biases amounted to 57% of the 1991 estimated net undercount: $(0.5 + 0.7)/2.1 = 0.57$. On Breiman's figures (Section 6.1), measured biases amount to 80% of that estimated undercount. Either way, most of the 1991 estimate represents bad data rather than undercount.

Correlation Bias

Correlation bias cannot be measured directly when the census and the system designed to correct the census both miss the same people. There can be no direct evidence about people if the surveys cannot find them. Thus, correlation bias is measured indirectly, using demographic analysis to make a second estimate of the national population (but see, e.g., Darroch, Fienberg, Glonek and Junker, 1993).

Commenting on this rather obvious point, Breiman says that estimates of correlation bias have only a "tenuous connection with any data." Belin and Rolph respond (Section 5.5):

> Correlation-bias estimates have more than a "tenuous connection with any data"; on the contrary, estimates are based on combining PES data with evidence from vital records about the ratio of the size of the male population to that of the female population. These data are used to estimate a parameter that characterizes dependence in omission rates between the census and PES (Bell, 1993).

We defer to Fay and Thompson (1993, page 76), who seem to agree with Breiman:

> "An important component of the total error model, correlation bias, could not be directly measured at any level, and was only indirectly inferred nationally by use of sex ratios derived from demographic analysis. Distribution of the national results relied entirely upon models that could not be checked against any direct evidence. . . ."

Numerical results from Bell's model give some insight into plausibility of assumptions. At the national level, the model says:

1. the PES estimates missed 890,000 white males total, of whom 13 (this is not a typo) are in the prime age group between 20 and 30;
2. the PES estimates missed 760,000 black males, of whom −28,000 are under age 10.

These results are incredible. Indeed, a direct comparison of demographic analysis with the PES estimates shows the latter found 190,000 too many white males, not 890,000 too few.

Likewise, when EFK say that Mulry and Spencer "used their best estimate of correlation bias," EFK mean that Mulry (1991) took Bell's national totals and disaggregated them to the 13 evaluation strata

and the 1,392 poststrata. (It is the latter numbers that feed into the loss function analysis, through the "*total error model.*")

The rationale for Mulry's algorithm was not provided. EFK and BR may not have examined the calculations; if they do, they will find that:

1. Mulry uses Bell's numbers as if they stood for correlation bias alone when in fact they are estimates of net error (correlation bias offset by the measured biases).
2. Mulry's detailed numbers do not add up to Bell's totals. Bell starts out with 1,652,000 men and no women; Mulry ends up with 285,000 men and 307,000 women: massive numbers of men disappear or are converted into women.

(These figures are not reported by Mulry, but can be estimated with reasonable precision from the available data.) Fay and Thompson (1993, page 76) sum up as follows:

> Unfortunately, the integration of [Bell's] results into the total error model was highly problematic . . . . Thus, it is impossible satisfactorily to characterize by any explicit model the nature of the realized estimates incorporated into the total error model.

To defend the Census Bureau's model for correlation bias, EFK and BR brush aside these facts, blurring the distinction between measurements and allocation schemes. Correlation bias was not measured, and the estimates have only a tenuous connection with reality, just as Breiman said.

To say a bias is unmeasured is different from saying that it is zero, although EFK pretend not to understand that distinction either. Despite wide uncertainties in all the figures, direct comparisons between demographic analysis and the PES, with measured biases taken into account, suggest that correlation bias is substantial (Wachter, 1991). At the national level, large numbers of black males seem to be missing from the adjusted census counts. What geographical areas are they missing from? No one knows. A reasonable opinion is that they are missing from inner cities with large minority populations, in the northeast and midwest.

Had the census been adjusted, the state population share of New York, for example, would have gone *down*: down, not up. Why did the 1990 adjustment bring down the population shares of New York and other such states? Correlation bias is a prime suspect. The bureau's total error model and loss function analysis cannot detect salient errors in the proposed adjustment to state shares, because the bureau's stylized rules for allocating unreached people to geography have little connection with facts on the ground.

7. DISCUSSION

Would the 1991 adjustment have improved on the census? Proponents of adjustment give strongly positive answers to this question in the academic literature, in the administrative record, and in the courtroom. For example, Ericksen (1991, page 3) writes:

> The only reasonable conclusion is that the adjusted count is more accurate than the unadjusted count. . .under any reasonable basis of comparison, the PES-adjusted enumeration is more accurate than the unadjusted census enumeration. Those adjusted results have also been shown to be robust to variations in reasonable alternatives to the PES "production procedures," and to variations in the statistical models used to generate the adjusted figures.

According to Fienberg (1992a, page 28), "the results of the Census Bureau's evaluation studies clearly supported the use of adjustment for the 1990 census results." Here is Rolph (1993, page 97), summarizing the evidence on loss function analysis, as presented by him and other plaintiffs' witnesses in *New York v. Department of Commerce*: "the Bureau's analysis clearly demonstrated that the adjusted counts were an unambiguous improvement on the original enumeration." These are sweeping claims, especially when compared to the results in Breiman or in Table R2 above.

The first line of Table R2, replicating the Census Bureau's analysis for the 1991 adjustment, is worth another comment. The 667 is the difference between an estimated risk (squared error) of 734 for the census, and 67 for adjustment; units are parts per 100 million. Over the 50 states and Washington, D.C., the r.m.s. error in population shares from the census is estimated as $\sqrt{734 \times 10^{-8}/51} = 0.04\%$. (The error distribution is quite skewed, which creates additional complications; Freedman, Wachter, Cutler and Klein, 1994, pages 255–259.) Our opponents contend that these errors can be reduced by an order of magnitude in size, if only we would agree to use their adjustment technology. Given the scale of the errors, that is—or should be—an astonishing claim.

EFK invent for us the position that the census is perfect and that only perfect models are usable (Fienberg, 1992a, page 27); BR have us assuming that undercount rates are constant across poststrata (Section 2.5). Thus, instead of showing that their models are accurate enough to correct *miniscule* errors in census shares, they remind us that the world is imperfect. That sets up their favorite rhetorical trick, which they play over and over again: justifying their assumptions as being less imperfect than the ones they have created for us.

The imperfection of the world is an argument much loved by modelers. The work must be done; the lesser evil must be chosen; and the best is the enemy of the good (Fienberg, 1992a, page 27; BR, Section 2.3). This argument has little force in the present context, when "the work that must be done" is only the creation of a technical record to defend a prior set of models.

Listen to them. They can adjust Stockton, with almost no data from that city, just by making the right assumptions (that demography overrides geography). They can validate their imputation model, when 75% of the validation data are missing, just by making the right assumptions: cases with weak data are easier to match than cases with strong data. They can measure correlation bias in New York, just by making the right assumptions, although they cannot quite explain what those assumptions are.

The best we can say for adjustment is that (i) it can fix the estimated differential undercounts at the national level and (ii) its impact on the accuracy of state population shares cannot be determined with any great confidence. In our opinion, however, adjustment is likely to degrade the accuracy of the state shares, and for substate areas adjustment seems even worse.

What will happen in the next census? Unless the analytical mistakes of 1990 can be recognized, they are likely to be repeated on an even larger scale in the year 2000. Efforts by BR and EFK to defend the mistakes of the past may cast a long shadow.

APPENDIX 1: VALIDATING THE IMPUTATION MODEL WITH EFU DATA

We document our calculations for missing data in the EFU as follows. Gbur (1991c, Table 3.1) shows 41% resolved in EFU out of the 2.8 million unresolved-in-P-sample cases sent to EFU, weighted to national totals. There were a total of 4.1 million unresolved P-sample cases in the bureau's "Advisory Use File": $(4.1 - 2.8)/4.1 = 0.32$, that is, 32% of the unresolved P-sample cases fell into the "Q-class, page 5 above." Our discussion focuses on the unresolved cases in the P-sample; other issues would arise for the E-sample (Breiman, 1994, Section 7).

Our next object is to state the issue in dispute between Breiman and BR (Section 5.7). Some notation will be helpful, although the argument will be informal. Let M be match status in the census (1 is a match, 0 is a nonmatch). Let R indicate inclusion/exclusion and resolved/unresolved in EFU: -1 is excluded, 0 is included but unresolved, 1 is resolved; the Q-class corresponds to $R = -1$. Let Z be the covariates in the imputation model that BR are defending; let $\hat{p}$ be the match probability imputed by the model; $\hat{p}$ is a function of Z, and Z is computed from PES data. For cases unresolved in the PES, match status is unknown; for a sample of these cases, M is determined by the EFU, but not for cases excluded from the EFU sampling frame or cases left unresolved in the EFU.

To validate their model, BR need to show that

$$P\{M = 1 \mid Z\} \approx \hat{p}. \tag{1}$$

Their claim is weaker:

$$\textit{BR's claim } P\{M = 1\} \approx E\{\hat{p}\}, \tag{2}$$

where "P" stands for weighted relative frequencies and "E" for weighted averages. What the EFU data show is weaker yet:

$$P\{M = 1 \mid R = 1\} \approx E\{\hat{p} \mid R = 1\} \approx 0.32. \tag{3}$$

It is this agreement that BR emphasize so strongly in Section 2.3. (There is an irritating numerical coincidence: 0.32 is also the fraction of Q-class cases.) To get (2) from (3), BR need

$$P\{M = 1 \mid R < 1\} \approx E\{\hat{p} \mid R < 1\}. \tag{4}$$

Of course, $P\{M = 1 \mid R < 1\}$ is not known: that is one implication of $R < 1$. You might think the trail ends here, but BR do not give up.

The next calculation is slightly indirect, because there are gaps in the documentation. As usual,

$$\begin{aligned} E\{\hat{p}\} &= E\{\hat{p} \mid R = 1\} \times P\{R = 1\} \\ &\quad + E\{\hat{p} \mid R < 1\} \times P\{R < 1\}. \end{aligned} \tag{5}$$

We know from the Advisory Use File that $E\{\hat{p}\} = 0.53$, and $P\{R = 1\} = 0.41 \times 2.8/4.1 \approx 0.28$, as before. By (3), $E\{\hat{p} \mid R = 1\} \approx 0.32$. Thus,

$$\begin{aligned} E\{\hat{p} \mid R < 1\} &= \frac{E\{\hat{p}\} - E\{\hat{p} \mid R = 1\} \times P\{R = 1\}}{P\{R < 1\}} \\ &\approx \frac{0.53 - 0.32 \times 0.28}{1 - 0.28} \\ &\approx 0.61. \end{aligned} \tag{6}$$

That $E\{\hat{p} \mid R < 1\} \approx 0.61$ is an empirical fact. Via (4), BR's claim (2) entails $P\{M = 1 \mid R < 1\} \approx 0.61$. By (3), $P\{M = 1 \mid R = 1\} \approx 0.32$; this is another empirical fact. Thus, BR's claim (2), coupled with the data, entails

$$P\{M = 1 \mid R < 1\} \approx 2 \times P\{M = 1 \mid R = 1\}. \tag{7}$$

You might think, as we do, that $P\{M = 1 \mid R < 1\}$ should be substantially lower than $P\{M = 1 \mid R = 1\}$.

But it must be nearly twice as high if you accept BR's claim (2). By their account, the cases that were excluded from EFU or were unresolved in EFU are much more likely than the resolved cases to match to the census. Belin et al. (1993, page 1164) actually say this, after presenting a calculation similar to ours. There is simply no evidence to support their bizarre story: M is unknown for $R < 1$. (Belin et al.'s arithmetic is wrong because they include $R = -1$ in one place and exclude it in another; even if they had gotten the arithmetic right, however, they would still be in the same logical mess.)

So far, the discussion has been about all the P-sample imputations, those excluded from EFU and those included in EFU. Breiman's paradox is only about cases included in EFU:

(8) $$P\{R = 1 \mid \hat{p} \text{ and } R > -1\} \quad \text{decreases as } \hat{p} \text{ increases.}$$

To explain (8), BR considered a group of cases with weak data, whose "names are not recorded." Let $W = 1$ for cases in this group, and $W = 0$ for other cases. Belin and Rolph have argued (personal communication) that

(9) $$E\{\hat{p} \mid W = 1\} \text{ is high}$$

and

(10) $$P\{R = 0 \mid W = 1\} \text{ is high.}$$

If (9) and (10) were right, they could explain Breiman's paradox:

(11) the weak cases do not get resolved in EFU, and that *lowers* the match rate among cases with big $\hat{p}$'s.

We consider these steps in turn. Assertion (9) implies (for the EFU universe) that unresolved P-sample cases with weak data match to the census at higher rates than other cases. Our view, of course, is that PES forms have weak data because there is only weak evidence that the corresponding people exist in the first place. Then cases with weak data will match to the census at lower rates, and (9) seems questionable at best. Assertion (10) seems right in general, but problematic for BR's chosen group of weak cases (see below). If you grant (9) and (10), then (11) is a good argument.

None of this can have much numerical impact: the group of weak cases considered by BR (whose "names are not recorded") was, with minor exceptions, not sent to EFU at all. These are cases with $R = -1$, not $R = 0$. They have nothing to do with Breiman's finding. The only additional argument made by BR is that they "wholeheartedly disagree" with us (personal communication). Their hearts are in the right place. Now, can we appeal to their heads?

APPENDIX 2: LEGAL PROCEEDINGS

In 1988, New York City (among others) filed suit in federal district court to compel the Department of Commerce to adjust the census. In 1992, the court ruled against adjustment, on fairly narrow grounds (Freedman, 1993b, page 106). New York went to an appeals court, the "Second Circuit," which vacated the judgment of the district court and remanded for further proceedings. The Supreme Court may yet determine the issue, because the Department of Commerce prevailed in the Sixth and Seventh Circuits: *Detroit v. Franklin* and *Chicago v. Department of Commerce*. Legal commentaries by statisticians should be taken with several grains of salt—and a dash of vinegar—but here we go.

Generally, an appeals court will rely on the district court's findings of fact. The Second Circuit held that adjustment was more accurate than the census, paraphrasing the district court to say that "for most purposes and for most of the population...adjustment would result in a more accurate count than the original census" (page 37 of the Second Circuit's typed opinion, dated August 8, 1994). However, the Second Circuit seems to have picked and chosen among the findings of the district court. For example, according to the district court, plaintiffs had failed to "illustrate affirmatively the superior accuracy of the adjusted counts [at the state and local level] for any reasonable definition of accuracy" (*Federal Supplement* **822** 924). The Second Circuit simply ignored the findings that did not suit.

The Second Circuit emphasized numeric accuracy and criticized the Secretary of Commerce for giving priority to "distributive accuracy," that is, accuracy of population shares for geographic areas (Second Circuit, page 42). Curiously enough, the Second Circuit based its own legal argument on cases (including *Baker v. Carr*, *Wesberry v. Sanders*, *Reynolds v. Sims* and *Karcher v. Daggett*) that deal with distributive accuracy (Second Circuit, pages 29ff). There is some difficulty, discussed briefly by the Second Circuit on pages 38ff, in applying these cases to the federal government; however, such legal issues are grist for another article by different authors.

Coming back to statistics, the Second Circuit's premise seemed to be that minority groups with large undercounts at the national level are concentrated in states whose shares would be adjusted upward. This is a fallacy. The same area often contains members of relatively overcounted groups along with mem-

bers of relatively undercounted groups. Therefore, an area's share often goes down as a result of adjustment, not up, despite a concentration of minorities. Urban blacks have an undercount three times that of the rest of the population, according to the PES; but 55% of them live in states that would lose population share if the adjustment were implemented.

The state with the largest number of blacks in the country, New York, would have its share adjusted downward. Pennsylvania, with nearly a million blacks, would lose a seat in Congress. The impact of adjustment on the constitutional right to equal representation has to be assessed area by area, in terms of demographic makeup and proposed adjustment factors. The Second Circuit did not come to grips with the statistical facts.

ADDITIONAL REFERENCES

AUSTRALIAN BUREAU OF STATISTICS (1990). *Census 86: Data Quality Undercount*, Catalogue No. 2607.0.

BELL, P., CORNISH, J., EVANS, J. and VINCENTE, W. (1993). Small area estimation of census undercount. Paper presented at the 49th Session of the International Statistical Institute.

BIEMER, P. and FORSMAN G. (1992). On the quality of reinterview data with application to the current population survey. *J. Amer. Statist. Assoc.* **87** 915–923.

BUREAU OF THE CENSUS (1975). Coverage of population in the 1970 census and some implications for public programs. Current Population Reports, Series P-23, No. 56, Government Printing Office, Washington, D.C.

CHOI, C. Y., STEEL, D. G. and SKINNER, T. (1988). Adjusting the 1986 Australian census count for under-enumeration. *Survey Methodology* **14** 173–189.

DIAMOND, I. (1993). Where and who are the missing million? In *Regional and Local Statistics: Statistics Users Council Conference*. 132–145. IMAC Research, Esher.

ERICKSEN, E. (1991). Letter report to Secretary Mosbacher, dated June 21/91. Reproduced in Department of Commerce (1991b). Office of the Secretary. Decision on whether or not a statistical adjustment of the 1990 decennial census of population should be made for coverage deficiencies resulting in an overcount or undercount of the population; explanation. Report dated July 15, Washington, D.C. Vol. II. Also see *Federal Register* **56** 33,582–33,642 (July 22).

ERICKSEN, E. and TUKEY, J. (1991). Letter report to Secretary Mosbacher, dated July 11/91. [Reproduced in Department of Commerce (1991b). Office of the Secretary. Decision on whether or not a statistical adjustment of the 1990 decennial census of population should be made for coverage deficiencies resulting in an overcount or undercount of the population; explanation. Report dated July 15, Washington, D.C. Vol. III. Also see *Federal Register* **56** 33,582–33,642 (July 22).

FIENBERG, S. E. (1994). An adjusted census in 1990? Trial judgment set aside *Chance* **7** (4) 31-32.

HIMES, P. and CLOGG, C. (1992). An overview of demographic analysis as a method for evaluating census coverage in the United States. *Population Index* **58** 587–612.

KADANE, J. B., MEYER, M. M. and TUKEY, J. W. (1992). Correlation bias in the presence of stratum heterogeneity. Technical Report 549, Dept. Statistics, Carnegie Mellon Univ.

KIM, J. (1991). 1990 PES evaluation project P12: evaluation of synthetic assumption. Technical report, Bureau of the Census, Washington, D.C.

MULRY, M. (1991). 1990 Post Enumeration Survey evaluation project P16. Total error in PES estimates for evaluation poststrata. Technical report, Bureau of the Census, Washington, D.C.

OFFICE OF POPULATION CENSUSES AND SURVEYS (1993). How complete was the 1991 Census? *Population Trends* **71** 22–25.

ROBINSON, J. G. (1994). Use of analytic methods for coverage evaluation in the 2000 census. Paper presented at Population Association of America Meetings.

SEN, A. (1988). Social choice. In *The New Palgrave: A Dictionary of Economics* **4** 382–393. MacMillan, London.

SKINNER, C. J. (1991). The use of synthetic estimation techniques to produce small area estimates. Social Survey Division New Methodology Series NM18, Office of Population Censuses and Surveys, London, U.K.

STEEL, D. and POULTON, J. (1988). Geographic estimates of under-enumeration. In *Proceedings of the Survey Research Methods Section* 119–128. Amer. Statist. Assoc., Alexandria, VA.

TRICKETT, P. (1992). Preliminary assessment of the 1991 census undercount. Paper presented to the Sixth National Conference of the Australian Population Association.

WACHTER, K. (1991). Recommendations on 1990 census adjustment, dated June 17/91. Reproduced in Department of Commerce (1991e). Office of the Secretary. Decision on whether or not a statistical adjustment of the 1990 decennial census of population should be made for coverage deficiencies resulting in an overcount or undercount of the population; explanation. Report dated July 15, Washington, D.C. Vol. II. Also see *Federal Register* **56** 33,582–33,642 (July 22).

WACHTER, K. (1994). Discussion of "Use of analytic methods for coverage evaluation in the 2000 census," by J. G. Robinson, Population Association of America Meetings. Unpublished.

Mrs. MALONEY. I would like Dr. Breiman, you believe or so you so say in your testimony that the process for measuring census accuracy through a post census survey that is matched to the original census count results in an estimate that represents "bad data" more than the true undercount.

Mr. BREIMAN. Yes.

Mrs. MALONEY. Under your theory, the net national undercount in 1990 was only 0.4 percent to 0.9 percent, not 1.0 percent as the Census Bureau believes. Is that a fair analysis, what I said?

Mr. BREIMAN. Well I want to distinguish two things. One is my paper certainly does not prove that the undercount was say 0.9 percent or 0.4 percent. What it shows is that if you subtract the effect of data errors from the estimates, you get down to 0.9 or 0.4 percent.

Do I believe that the true undercount was 0.9 or 0.4 percent? No. I believe that the demographic estimates were more on the point.

Mrs. MALONEY. To translate it into terms for the people instead of percentage points, how many people do you think were missed in 1990?

Mr. BREIMAN. I think the most reliable estimate we have would be the demographic estimate, which is up I think at around 5 million people.

Mrs. MALONEY. So you think 5 million people?

Mr. BREIMAN. Yes, 4 to 5 million people. But you have to understand I am not a demographer, so I don't know what went into that estimate.

Mrs. MALONEY. Just to clarify. Do you believe that the differential undercount between whites and blacks was smaller than the 3.8 percent measured by the post enumeration survey or the 4.4 measure by the demographic analysis? You believe it was the 4.4, am I correct? I don't want to put words in your mouth.

Mr. BREIMAN. I am not sure, but I will say this. That I don't doubt that the undercount was in the neighborhood of 5 million and that probably blacks and minorities were undercounted more than say whites who lived in stable areas. The demographic analysis shows that and I have no reason to doubt it. The real question is what is the PES measuring. Is it measuring that undercount or is it measuring error?

Mrs. MALONEY. The Census Bureau and the U.S. General Accounting Office reported that the total number of errors in the 1990 census was between 14 million and 25 million. My time is up. Very quickly, how many errors do you think there were?

Mr. BREIMAN. I would take the GAO's estimate. You see I don't deny that the census had many errors. All I am saying to you is are you going to cure the errors by adding more errors to it?

Mrs. MALONEY. My time has expired.

Mr. MILLER. Before I call Mr. Snowbarger, I express great disappointment in the ranking member, to try to follow the policy that's so often used by the minority, to try to impeach the credibility of witnesses. Dr. Breiman is a very distinguished full professor at the University of California at Berkeley, author of many textbooks, dozens of articles. I believe anybody in the statistical community would consider you really one of the outstanding thinkers in that area. So I feel very disappointed in the ranking member.

Mr. Snowbarger.

Mrs. MALONEY. Point of clarification and point of personal privilege since my name was mentioned.

Mr. MILLER. Yes.

Mrs. MALONEY. I in no way want to criticize any of our witnesses with their professional credentials and their accomplishments and their dedication and their sincerity. But I do think that it is a point fact when a court rules that one is not an expert witness. I don't think that that is personally attacking someone. That is putting a fact into the record to clarify a situation. It was never meant in a personal way. But when one is put forward as an expert witness and a court rules that one is not, with documentation several pages long, I don't think that that is a personal attack or an unfair attack. I think that is just putting forward actual data, fact.

Mr. MILLER. Let's move forward. You said you are not interested in this particular hearing, that you think it's not worth your trouble. I am sorry that you have to go attacking.

Mr. Snowbarger.

Mrs. MALONEY. Mr. Chairman, everything is worth my time. Any meeting you call, I want to be there. I have just said that we should have some hearings on these alternatives that you are proposing. I think that's a fair thing to bring up.

Mr. MILLER. It's Mr. Snowbarger's time.

Mr. SNOWBARGER. Thank you, Mr. Chairman. I want to echo your sentiments. It seems like any time this administration is attacked, the only way that they can make their case and make their argument is to attack back and attack those who disagree with them.

With that, Dr. Breiman, frankly if you would like an opportunity to respond to the earlier question from our colleague about the court case, I would be happy to give you some of my time to do that. If you don't want to, I'll carry on with questioning. Would you like to have some time to respond?

Mr. BREIMAN. Well, I would say two things. One is that I think by the time we got to me as an expert witness, the judge was so overwhelmed by expert witnesses that clearly he was out of the mood to listen to another one. My lawyers or the lawyers in the Department of Justice were going to appeal this decision up, they said if necessary, to the Supreme Court, because they were very certain that it was an incorrect decision on the part of the judge.

But the other thing, Mrs. Maloney, Congresswoman Maloney, is that if you will read my paper and the commentaries on it, I would rather stand by what I have produced rather than worry about whether the judge was right or wrong.

Mr. SNOWBARGER. Let me indicate that I had some experience with the post-enumeration sampling because as a State legislator, I was involved in the legislative redistricting for the State of Kansas. As we were asked by the Census Bureau for our input on PES and on the data that came out of that, we were very strongly opposed to it. We felt like it had an adverse impact on rural States like Kansas. So I really have to admit a bias here, coming at it in that direction. What you two gentlemen have indicated today also kind of bolsters that feeling of mine.

You are talking a lot in statistical language. I am a layman to all of this. Listening to all of our statisticians argue today, both pro and con, is probably going to be about as helpful to me as when Secretary Rubin and Alan Greenspan come before the Banking Committee and give their different opinions of various economic issues. Again, we can find probably basis for a decision both ways.

Let me talk about the idea of sampling in general. I presume that both of you would support the general concept of sampling? That it has value in certain circumstances?

Mr. BREIMAN. Yes.

Mr. YLVISAKER. Absolutely. Statisticians are out of work without sampling, by and large.

Mr. SNOWBARGER. That's an interesting comment that we may want to make a point of in the record, that statisticians would be out of work if we didn't have it, particularly when we are talking about one of the largest sampling projects that we can think of.

Well, it has been a useful tool. So what you gentlemen are trying to say is that even though it is a useful tool under certain circumstances, you don't see the current proposed sampling method as that useful a tool as it might be in other circumstances?

Mr. YLVISAKER. Well, there's the question between the sampling which can be carried out assiduously and very correctly and so on. But we're not talking about sampling now. We're talking about things like matching. This is another kind of operation.

Mr. SNOWBARGER. Someone needs to go further and explain that. Dr. Breiman, you mentioned matching a number of times. Again, for this poor layman, can you explain what that process is and why it is necessary? That's what I thought. This is difficult.

Mr. BREIMAN. It's quite simple, Congressman. I guess the best analogy is to suppose you got a pond and you want to find out how many fish are in this pond. So you go, and you fish and fish and fish and you catch a fish. You mark a red cross on its back and you throw the fish back in. So maybe you catch 1,000 fish and so on, and you think well, have I caught them all? Now what you do is you come back a week later, and you catch 100 fish. Two of them do not have red crosses on their backs, so your undercount estimate is 2 percent.

OK? But suppose one of the red crosses was just rubbed off. Then your undercount estimate goes down to 1 percent. You see? So the problem is in the matching. In the second catch, you have got to match the second catch of 100 to the first catch.

Now people, you can't paint red crosses on people. They are even harder to match than fish, much harder, because you will get different names. You will get different ages. People are moving around and so on. The matching process is terribly risky, particularly in those areas where you suspect there is most undercount.

If you go to an area that's filled with say recent immigrants that speak different languages and they are highly mobile, the census is going to count them. Two months later you are going to come in and recount them in the mini-census and try to match them. Well there are going to be a lot of people you can't match. Does that mean that the census is undercounted? Not at all.

Mr. SNOWBARGER. Thank you, Mr. Chairman. My time is expired.

Mr. MILLER. Mr. Blagojevich.

Mr. BLAGOJEVICH. Thank you, Mr. Chairman.

Dr. Breiman, let me just first of all say that you know, those of us who have taken the other position, and believe sampling is an available method that ought to be explored, are concerned about some of the attacks about the scientific methods proposed by the Census Bureau to supplement the traditional direct counting methods for the year 2000. Now the descriptions the proposed methods will not, in my opinion, help the public understand what are essentially complex scientific procedures. In particular, how they apply in a public policy setting.

Since you are all credible professionals, albeit one court has evidently said you are not an expert, and I am saying this somewhat tongue in cheek, but since you are credible professionals, and I'm sure the court would acknowledge that, Doctor. If you could, could you comment on some of the terms that have been used to describe the Bureau's plan and tell me whether you think these terms are appropriate or accurate?

No. 1, the use of sampling, to estimate a portion of the population not counted directly, has been described as cloning. Are you comfortable with the term cloning?

Mr. BREIMAN. Could you repeat that term?

Mr. BLAGOJEVICH. Cloning. Cloning.

Mr. BREIMAN. In the sense that this procedure will produce imaginary people, is that the sense in which you mean cloning?

Mr. BLAGOJEVICH. It's not a word I have used, but it has been used by others. Others have described the use of sampling as a method to estimate a portion of the population that we would miss in a direct count. Some have called it cloning. You are cloning people essentially. Would you say that that's an accurate description of what the sampling method is?

Mr. BREIMAN. Yes. They are cloning people in the sense that they are—what they come up with is, for instance, an estimate of the undercount. So what do they do? They raise the population of certain areas by their estimate. Now where are these people that are being tossed into the pot? No one knows exactly what houses they live in or anything of that sort. The population is just raised.

Mr. BLAGOJEVICH. So you are comfortable with the term cloning?

Mr. BREIMAN. In that sense.

Mr. BLAGOJEVICH. OK. Now is this a viewpoint that is shared by other individuals in the scientific disciplines? In other words, in the scientific community dealing with these issues of statistics and sampling and the like, is cloning actually used as a term to describe estimation procedures based on sampling?

Mr. BREIMAN. No. It's—I have a little difficulty with the term because I haven't heard the word cloning used before in reference to this procedure. So I am only guessing at what is meant by it.

Mr. BLAGOJEVICH. OK. I mean I am hardly an expert in anything. I am sure every court is probably willing to say that. Science is certainly not my area of any knowledge. But as I understand it, cloning is not something, the term is not used in the scientific community to describe this method of sampling. Is that a fair statement, Doctor?

Mr. BREIMAN. I would say that in statistics, we would use the terminology we're imputing people.

Mr. BLAGOJEVICH. OK. The use of sampling both to complete non-responsive followup in the census and then to measure undercounts and overcounts in the initial count has also been called by some guessing. This is the phrase that's been used or the term that has been used. Would you say that is an appropriate way to describe the basic procedures in the Census Bureau's plan? In other words, the Census Bureau has its plan. Would you say it's fair to characterize their basic procedures as guessing? Let us know what you think about that.

Mr. BREIMAN. Well, I would call it, for instance, the non-response follow-up, the sampling for non-response follow-up, I would call that better than guessing, I would call it an educated guess.

Mr. BLAGOJEVICH. OK.

Mr. BREIMAN. What you are doing is that you are using what the census would call the hot deck procedure, which means the non-response households, you are going to judge what is in them by the closest household to it where you know who is in that closest household.

So if you have a non-response house and you don't know who is in it, but the house next door has got five people in it of certain ages, you are going to assume that the house you haven't gotten a response from is similar to the house you do know about.

Mr. BLAGOJEVICH. OK.

Mr. BREIMAN. Basically that's how they are going to proceed. Now whether you want to call that cloning, it's actually called imputation in statistics.

Mr. BLAGOJEVICH. OK. But you wouldn't call it guessing?

Mr. BREIMAN. Pardon?

Mr. BLAGOJEVICH. I mean to describe it as guessing would be over simplifying it?

Mr. BREIMAN. Right. You usually call it educated guesses or we use the word imputation.

Mr. BLAGOJEVICH. What about this non-responsive followup. Would you agree with this method as being one that can produce accurate results?

Mr. BREIMAN. Let me put it this way, that I think the best thing we can do is to come up with a census that enumerates people one by one as well as we can, without doing any sampling. I think one of the problems is that the proposed ICM has diverted a lot of money and energy from the process of just going out and enumerating. I would urge that Congress simply appropriate enough money to the Bureau to let them do the best job they can of counting one by one.

Mr. BLAGOJEVICH. OK. I'm out of time, Doctor, but the non-responsive method is something that you don't dispute works. Right or wrong?

Mr. BREIMAN. The non-response sampling works in the sense that it's an educated guess. But you really don't know who is in that house.

Mr. BLAGOJEVICH. Thanks. I ran out of time.

Mr. MILLER. Mr. Davis.

Mr. DAVIS OF VIRGINIA. Thank you very much. Thank you both for being here. Sometimes the questioning up here gets unduly personal and guttural. This is a very political process, counting people.

Sometimes it arouses the most partisan passions in Members. I know there's one member of the committee who would apologize for anything that might have been said to somehow impugn either of your, what I would call, expertise in these areas. For example, Dr. Breiman, I don't know what court ruled, but I know you know more about statistics and have more expertise than Mr. Quelo, for example, who is the administration's appointee on this.

So I have a few questions. I am correct, aren't I, in assuming that you all are members of the Statistical Association, the American Statistical Association?

Mr. BREIMAN. I'm a fellow of the American Statistical Association.

Mr. YLVISAKER. Yes. Same.

Mr. DAVIS OF VIRGINIA. OK. And given your experience with both statistics and the census, there can be no doubt that you would qualify as experts on the issue of census sampling?

Mr. BREIMAN. I'm sorry. Could you repeat that?

Mr. DAVIS OF VIRGINIA. Given your experience in both statistics and the census, and Dr. Breiman, you have a bachelors from Cal Tech, a masters from Columbia, which is not near Mrs. Maloney's district, a Ph.D. from Berkeley. Would you consider yourself an expert in the issue of census sampling?

Mr. BREIMAN. Yes, after my studies of it.

Mr. YLVISAKER. I have looked at certainly the plan of samples and so on. Yes, these are intelligently planned samples. They do what they are supposed to do and so on. I would like to separate, as could be done, things like terminology of cloning and things of that nature from the process of sampling. Sampling is quite apart from what we are talking about in many cases here.

Mr. DAVIS OF VIRGINIA. Were either of you ever consulted in your capacity as a member of the ASA in terms of whether or not you endorsed the specifics of the plan for census 2000?

Mr. YLVISAKER. No. I was not.

Mr. BREIMAN. No.

Mr. DAVIS OF VIRGINIA. OK. Were any of you asked to be a part of the blue ribbon panel to study the sampling of the census?

Mr. BREIMAN. No.

Mr. YLVISAKER. No.

Mr. DAVIS OF VIRGINIA. Do you know how the panel was chosen?

Mr. BREIMAN. I'm sorry?

Mr. DAVIS OF VIRGINIA. Do you know how the blue ribbon panel was chosen by the ASA?

Mr. BREIMAN. No. I don't.

Mr. DAVIS OF VIRGINIA. Can you tell us if the entire ASA ever voted to endorse the Bureau's sampling plan?

Mr. BREIMAN. No.

Mr. YLVISAKER. The ASA stops very much short of endorsing such plans.

Mr. DAVIS OF VIRGINIA. But in a generic sense, wouldn't any statistician say that using sampling techniques is a bad idea? You wouldn't say that, would you? In a generic sense.

Mr. BREIMAN. No, no. I mean there are many places in which sampling is——

Mr. DAVIS OF VIRGINIA. It's a fundamental tool of the trade, isn't it?

Mr. BREIMAN. Right.

Mr. YLVISAKER. Absolutely.

Mr. DAVIS OF VIRGINIA. So even though sampling in many situations has been a useful tool, that doesn't mean it's appropriate all the time, does it?

Mr. YLVISAKER. We have to talk about what we're going to do with the sample, for one thing.

Mr. DAVIS OF VIRGINIA. You both agree with that. The minority constantly claims that the American Statistical Association has endorsed sampling in the 2000 census. But I think their claim or that claim would be misleading. Let me read you a line from the ASA's recent court brief on sampling. "ASA also takes no position in this brief on the details of any proposed use of statistical sampling in the 2000 Census."

Do you think that quote from the ASA is saying equivocally that sampling should be used in the census, or that in a very general sense, sampling is a useful tool of the social sciences that may or may not be appropriate for the census?

Mr. BREIMAN. Yes.

Mr. YLVISAKER. Statisticians are protective about sampling. They have to be. But they don't have to endorse the plan of the census. They certainly couldn't and didn't.

Mr. DAVIS OF VIRGINIA. So basically it looks like ASA would endorse sampling as a general principle of science, but applying it in this particular case at this point, to your knowledge, they have made no endorsement.

Mr. YLVISAKER. That's true.

Mr. DAVIS OF VIRGINIA. Correct from both of you. So while every statistician probably would endorse sampling in the abstract sense, would this endorsement have anything to do with the 3,600 separate components of the Bureau's plan for the 2000 census?

Mr. YLVISAKER. Not directly.

Mr. DAVIS OF VIRGINIA. So you would say that when the minority constantly goes around stating that the ASA has endorsed the Census Bureau's 2000 census plan, that that would be an exaggeration of the nature of ASA support and that Members of Congress and the media should be skeptical of "an endorsement" of this plan?

Mr. BREIMAN. Very much so.

Mr. DAVIS OF VIRGINIA. You both agree with that?

Mr. YLVISAKER. Yes.

Mr. DAVIS OF VIRGINIA. Charles Jones, who is the former associate director of the decennial census and a 32 year Bureau veteran said of the administration's plan, "You are replacing one set of errors you don't quite like with a set of errors you like better."

Do you agree with this quote? What does it imply in terms of the accuracy of the census?

Mr. BREIMAN. Let me say that I know Charlie and I respect his judgment highly. I would certainly go along with his quote.

Mr. YLVISAKER. I know Charlie Jones as well. I am not going to argue with him on these issues.

Mr. DAVIS OF VIRGINIA. It looks in some ways like you can improve the process, but you could make it worse in other ways; is

what I think he was saying. Do you think that would be a fair reading when you apply sampling?

Mr. BREIMAN. Yes.

Mr. YLVISAKER. We're in the position, we don't know.

Mr. DAVIS OF VIRGINIA. All right. Mr. Chairman, I yield to you if you have any additional questions.

Mr. MILLER. I have a question, but I think the time is about up.

Mr. DAVIS OF VIRGINIA. OK. Thank you both.

Mr. MILLER. Mr. Davis.

Mr. DAVIS OF ILLINOIS. Thank you very much, Mr. Chairman. You know I just saw an advisor of mine, who just left when I did my dissertation. I am glad he left before we got to the questions. Statistics was never my best subject, as I'm glad Dr. Leon Dingle left out.

Well let me ask you, Dr. Breiman, who determines what an expert is or who an expert is?

Mr. BREIMAN. I'm sorry? Could you repeat that?

Mr. DAVIS OF ILLINOIS. Who determines who is an expert?

Mr. BREIMAN. Well, I would say that varies depending on the situation. I can't give you a universal rule.

Mr. DAVIS OF ILLINOIS. Oh, OK. So the situation kind of dictates the action. So one can be an expert today, and then not an expert tomorrow?

Mr. BREIMAN. Well, I would say that for instance, a good biologist is an expert in biology today and tomorrow.

Mr. DAVIS OF ILLINOIS. OK. Well, I just was wondering about the definition in terms of the whole business of the expert.

Well let me ask you, you indicate that in your conclusion, 70 percent of the initial 1991 undercount estimates came from data errors.

Mr. BREIMAN. At least 70 percent.

Mr. DAVIS OF ILLINOIS. How would you propose that those be corrected?

Mr. BREIMAN. I don't believe they can be. As I said, I think that they are inherent in the nature of the U.S. population, which is a highly mobile diverse population.

I mean for instance, to give you an idea, you have census day. On census day you count everybody. The count is taken as of that day. Two months later, you come along and you try to see if the people in this house were actually counted by the census. But the people have moved out. Can you get any information about who they are? So that means that you have to decide, did the census match them or did they not match them?

Mr. DAVIS OF ILLINOIS. Are you suggesting that it is virtually impossible to get an accurate census count?

Mr. BREIMAN. No. What I am suggesting is that it's virtually impossible to correct a census for undercount by the methods used in the 1990 or proposed for the year 2000.

Mr. DAVIS OF ILLINOIS. Then how can we correct the undercount?

Mr. BREIMAN. I'm not sure. I'll just give you my best guess at this time. That is, to put as much of the resources as we can into getting an accurate enumerated census.

Mr. DAVIS OF ILLINOIS. Well let me ask, we have already demonstrated with enumeration that we're going to miss large numbers

of people, as evidenced by having missed them. So how then does this correct what we have already attempted to use and didn't get any real recourse?

Mr. YLVISAKER. I would submit that there are political solutions. There are political solutions I should think.

Mr. DAVIS OF ILLINOIS. I would be pleased to hear those.

Mr. YLVISAKER. You should decide that undercount is serious and that you wish to do something about it. Don't ask statisticians to find out precisely what it is.

Mr. DAVIS OF ILLINOIS. Let me ask this then. If there had to be errors and we concluded that there had to be an undercount, where should it be?

Mr. YLVISAKER. A political question. There are simple answers that can be given. We take the proportion of blacks in a State and we multiply it up by the demographic undercount estimate and let that run across States.

Mr. DAVIS OF ILLINOIS. So you'd say we want to make sure that the undercount is not among blacks and other minority groups.

Mr. YLVISAKER. One is taking some steps to do something about a known undercount, and simple steps.

Mr. DAVIS OF ILLINOIS. I guess the problem that I have is all of the data that we have looked at in terms of undercount and where the undercount has occurred, it's among those population groups that can least afford to be undercounted as a group, African-Americans, Hispanics, poor, rural, urban, inner city.

Mr. YLVISAKER. I agree with you.

Mr. BREIMAN. The difficulty is, one difficulty is, Mr. Congressman, that for instance, it seems to be easier to count the blacks in New York City because they seem to be more settled and less mobile than say the blacks in Los Angeles. So if you do across the board thing, how do you account for that?

Mr. DAVIS OF ILLINOIS. I think in terms of—Mr. Chairman, I know that my time is up.

Mr. MILLER. We're going to do a second round if you want. I have some questions and I would like to do a second round, but if everybody doesn't want to use up all their time, we'll of course go right onto the second panel. We are going to expect a vote in a few minutes.

Mr. Davis made a comment that we know that we have already demonstrated we are going to miss some, which we understand is an undercount. But your statements and the statements, I think, of the next panel, will say that we also have demonstrated that sampling was a failure in 1990. The question we have is, do we have to accept an undercount or do we look at a failed sampling scheme and without any empirical evidence that it is going to work. Is that a fair statement?

Mr. YLVISAKER. It is.

Mr. BREIMAN. Yes.

Mr. MILLER. The reason I would like to have a second round is, Dr. Ylvisaker, you were a part of the CAPE Committee, right?

Mr. YLVISAKER. I was.

Mr. MILLER. Would you very briefly explain what that is? Dr. Bryant, I think, was there at the time.

Mr. YLVISAKER. I was on the panel of experts, so-called panel of experts of the Committee on the Adjustment of Post-Censal Estimation. That is, the census must be somehow carried out between censuses. That is, populations have to be adjusted. So the committee began its deliberations I believe in 1991. There was a workshop in 1991 which I attended. Generally the committee took care of its own business. They did talk to a panel of experts, of which I was one. They met in the summer of 1992. They set out about five different issues, roughly to review the situation. That is, we had a census count. We had an adjusted census count. What should we be doing in between the census of 1990 and the 2000 census. The committee therefore reviewed the PES, the errors in the PES, a variety of things that had taken place between 1990 and 1991 to come up to a conclusion.

I would invite everybody to look at the CAPE report. I think it is actually a very fine report. It sets out the issues very well. A decision was to change basically the methodology of the PES as it was practiced at the time. It was too complex. They made it simpler. They found errors. They corrected some of the errors. They knew more errors existed. They decided they couldn't correct those. They decided to adjust the State populations and the national population according to the errors that they found. They did not carry this down to lower levels.

That was basically the work of that. They looked into a variety of things, whether they could settle very much about correlation bias, heterogeneity of poststrata, things of that nature. I thought it was a worthwhile committee.

Mr. MILLER. How many times did the Census Bureau change the adjustment, do you know? Their plan was to use the full enumeration and then adjust the data. There were several adjustments. My understanding is there were three separate adjustments that were put forth. Each one had its own type of problems.

Mr. YLVISAKER. I'd hesitate to count them myself. The ones that I am aware of are of course the initial census, the initial adjustment, the adjustment that was taken when the number of poststrata was cut by a large amount, which eliminated the smoothing steps. So that would make a total of three counts that I think are present in many documents.

Mr. MILLER. But the conclusion was that the adjustment was not.

Mr. YLVISAKER. The initial adjustment certainly went by the boards, yes. There was an improved adjustment at State levels and so on, which was found in 1992.

Mr. MILLER. The plan for this coming year is that they are only going to allow, I mean for the 2000 census, only 5 weeks to do this. Is that realistic? The GAO has been very critical of the timelines involved. All of you took 2 years to analyze it. The GAO says it won't work. They studied 1990 in the attempt.

Mr. YLVISAKER. I guess I would be very nervous were I in charge of getting this done, yes.

Mr. MILLER. Because I think you had 2 years to go through that.

You asked about complexity. You said that the PES was too complex. Right?

Mr. YLVISAKER. The way the data were used was very complex. There was a smoothing model used.

Mr. MILLER. I assume it's much simpler in the year 2000?

Mr. YLVISAKER. My understanding of what is planned, it's much closer, it's much simpler to understand, much more easy to describe, the smoothing that may be used. It's not clear it is going to be used. This makes sure that say a given poststrata number might be estimated. This undercount might be estimated by using other borrowing strengths from other poststrata, like the size.

Mr. MILLER. Which is more accurate, especially at the census block level or census tract level. Which is more accurate, the adjusted number back in 1990 or the actual enumeration? What is the consensus in that?

Mr. YLVISAKER. I have no idea to tell you the truth. That's my problem.

Mr. MILLER. No idea. Do you? I know Dr. Breiman will definitely say that it's going to be less accurate. Is that right?

Mr. BREIMAN. Pardon?

Mr. MILLER. Comparing the accuracy between the adjusted numbers in 1990, especially the census tract level, the adjusted numbers versus the actual enumeration numbers; which are more accurate?

Mr. BREIMAN. We don't know. We have no way of telling because so much of the adjustment was noise that you know, just due to errors, that we have no idea. In some places it could have made it more accurate, in some places less accurate. It's undetermined.

Mr. MILLER. Thank you. The next panel is going to be able to discuss the issue of trust. When you trust these big computers in Washington and all the smart people doing it. Which one can the American people trust?

Mrs. Maloney.

Mrs. MALONEY. Yes. I would like to ask that Members be allowed to submit questions for the record because we have many more panelists to speak, if that would be appropriate. Because I do have more questions. Is that all right?

Mr. MILLER. Yes. Without objection.

Mrs. MALONEY. I'd just like to close with one general question, although I have many others that are a little more technical. Everyone admits that there was an undercount. Now we may disagree on what that undercount is, 8.4 million is what GAO says and other sort of independent groups. You today, Dr. Breiman, said you believed it was more, like 5 million. But there is an undercount.

Getting away from the numbers, the question is, what do you do about the undercount? The Census Bureau has a plan to count the undercount using modern scientific methods. In 1990, they used enumerators and they didn't get the undercount. There was no lack of money. There was no lack of people. They went out and there was an undercount for reasons that really both of you outlined in your testimony, because of the society that we live in.

So my question is, what do we do about the undercount? Do you just let 8 million people not be counted, not be represented, not funded, et cetera? Or do you move forward with the best scientific methods available to correct it? My question is basically what do you do about the undercount?

We had a plan in 1990 that produced, I'm using the GAO number, 8.4 million people were missed, 4.5 million people were counted twice, and the question is how do we get the most accurate count? Many scientists and professionals have said that you do that by using scientific methods. You are saying, I'm maybe hearing you incorrectly, that the way to correct it is to do more enumeration.

It is my understanding from the reports that I read, they did enumeration. But for whatever reason, those people could not be counted. You have testified they moved, they don't open their doors. I have read reports by some enumerators who on their own reports said they wouldn't go into the neighborhoods, so they guessed or they knocked on the doors five times. So then they started guessing. This is enumerators writing up how they "guessed" in their final report.

I read one scientific journal where the woman said that one of the people in her team came in and—I won't get into hearsay. But anyway, so what do we do to correct the undercount?

Mr. BREIMAN. You know, this is something that I really haven't chewed over. What I know is what doesn't correct the undercount. That I know very thoroughly because I have looked at that for a long time. I can tell you the PES or the ICM is not a thorough scientific method for correcting the undercount. It is filled with errors. That is what I have tried to tell you. So we know one way that doesn't work. We have not studied alternative methods that may work. If you give me that as homework, I would be glad to go back and chew on it.

Mrs. MALONEY. Well we have, we know without using scientific methods, that it will produce at best an undercount of 1.9 percent.

Mr. BREIMAN. I'm sorry, what?

Mrs. MALONEY. Well, I will just put it in—I'll just put a series of questions into the record. But I would like your comment, Doctor.

Mr. YLVISAKER. Ylvisaker.

Mrs. MALONEY. Your name is difficult for me. Ylvisaker. Dr. Ylvisaker, there is an undercount everyone acknowledges it.

Mr. YLVISAKER. Yes.

Mrs. MALONEY. They may disagree, but some have said the undercount was 10.8. Others, we have been using 8.4, but whatever. The undercount is there. What do we do about it? We know that using enumerators that are funded to knock on doors 20 times, we are still going to have the undercount. So do we just accept the undercount or do we move forward with scientific methods that give us a better—no plan is perfect, but a better count, a fairer count, a more accurate count.

Mr. YLVISAKER. I would put in one item. That is, the best scientific methods were used in the PES in 1990. That did not work. Now we're talking about the best methods in the year 2000. The question is whether that might work. I have no real reason to believe that it will work based on what I have seen from 1990 and the complexity of what must be done.

What I am suggesting is that statistics may not be the best possible answer for you to correct the undercount. It may. I don't see the methods on the table which will do it. I suggest that there are political and simple-minded methods which will correct for

undercount, acknowledge undercount and make some sort of effort in this direction.

Statistics and sophisticated statistics, calling everything great scientific method does not necessarily correspond to the situation that currently exists. So we can accept this and say it's a fine thing to do, but——

Mrs. MALONEY. Yet is it not true that many scientists say that scientific methods would correct the undercount?

Mr. YLVISAKER. Oh, we disagree. Absolutely.

Mrs. MALONEY. I would, for one, prefer to rely on scientific methods as opposed to a political solution as you suggest.

Mr. YLVISAKER. Oh, well I would too.

Mrs. MALONEY. I find it surprising that a scientist would recommend a "political" solution to solve a scientific problem.

Mr. YLVISAKER. If it's required, then I say—there are things that science cannot help us with I'm afraid.

Mrs. MALONEY. My time is up.

Mr. MILLER. Mr. Blagojevich.

Mrs. MALONEY. We're going to put our questions in the record, and hopefully both of the gentlemen will respond, as will the other panelists. But right now it's almost 12. We have two more panels.

Mr. MILLER. We're getting ready to have a vote in a few minutes, is my understanding, on the CR.

Let me thank Dr. Ylvisaker and Dr. Breiman. Thank you very much for being with us today. We appreciate your contribution to this debate.

We'll have the next panelists come forward please. We're told that there is going to be a vote any minute now. So rather than getting started, maybe we better go vote. It's unfortunate. We have a vote on a CR. I might want to share with everybody that the CR will be for the first 10 days in October. Our fiscal year ends on September 30, but the decennial census is being taken care of for the first 10 days with increased funding because it has greater needs at the beginning of October. So that was a concern that was raised at a hearing last week. It is my understanding that it has been resolved to the satisfaction of both OMB and the Census Bureau, and was included in the CR that's going to be passed momentarily.

So rather than beginning, let's just go ahead and take a recess and come back immediately after the vote.

[Recess.]

Mr. MILLER. Let's begin with the next panel. We have Mrs. Maloney here. If the next three panelists would stand.

[Witnesses sworn.]

Mr. MILLER. Let the record show that they have all stated and answered in the affirmative. We'll proceed as we have before with the opening statements by the three Members.

I have a unanimous request, but I'll wait until Mrs. Maloney comes back. So let's proceed. Who would—I have a unanimous consent to enter into the record, I referred to in my opening statement an AP article I would like to have included and this Amstat News from the Statistical Association. Since Mrs. Maloney brought up the issue of all the editorials, I would like to enter editorials that state that they are against sampling, which includes the Wall Street Journal, the Washington Times, the Detroit News, the Indi-

anapolis Star, the Arizona Republic, the Atlanta Constitution, the New York Post, and the Cincinnati Enquirer, just to name a few. As we have said, the minority does not have a monopoly on statisticians and they do not have a monopoly on the editorial pages of this country either.

Is there objection? If no objection, we'll enter these in the record.

[The information referred to follows:]

1/2

KENNETH BLACKWELL W. TIMES 6/30/98 B4

Security. . . or chicanery

Last month, Vice President Al Gore told thousands of members of the National Association for the Advancement of Colored People that the debate over the 2000 Census was about race, not methodology. It was a simple and irresponsible statement, meant to raise old enmities and stereotypes rather than thoughtful discussion. Had he chosen the high road, though, Mr. Gore would have been forced to address something totally new: the administration's plan for Census 2000.

Less than two years from now, the Clinton administration proposes — for the first time ever — to blend estimates and actual counts into a single population number for the decennial census. That mixed bag of counted and estimated people will determine the allocation of more than $182 billion in federal funds to the states. If the Census Bureau chooses to disregard current law and implements its plan to produce this one-number census, the ability of states and localities to legally challenge the results will be severely compromised.

As an example, in 1990 an entire ward was missed in one town in the district of U.S. Rep. Tom Petri, Wisconsin Republican. Community leaders found and corrected this mistake during the post-census review. For 2000, the Bureau does not plan a post-census review — removing from states and localities both an avenue to legal remedy and a chance to correct mistakes.

That concerns me, both as a former mayor of Cincinnati, and as co-chairman of the Census Monitoring Board, a federal, bipartisan oversight board reviewing Census 2000. In this extremely fast-moving process, I am concerned that decisions are being made, and actions taken, without substantial input from some of the most important players in the census process — specifically, state and local leaders, and Mr. Gore should be especially concerned with two things:

The Master Address File, the address list used to mail census forms across the country, is likely to be inaccurate in many poor and rural areas, as well as high-growth areas, regardless of their racial makeup.

Currently, there is no plan to conduct a post-census review after Census 2000. In 1990, the post-census review "found" more than 47,000 people in and around Detroit — almost all of whom were minorities. After 2000, cities will be stuck with the number of bureau calculates.

Last week, the board heard from local officials in Columbia, S.C., who reported several examples in the recent census dress rehearsal where the bureau was unresponsive.

The U.S. House of Representatives recently passed an appropriations bill that provides for a review of this and other major changes planned for the 2000 census.

Columbia's community development director said: "One of the things that we ran into with the Census is that it's easy to say to one thing but do another. For example, this idea of, 'We're going to use multiple resources.' But, in fact, they used singular resources in the dress rehearsal."

South Carolina had a number of resources available — more than most states. For one, the state has one of the leading address database programs in the country. Jack Maguire, a state database expert, testified that South Carolina officials found more than 26,600 addresses missing from the Census Bureau's Master Address File.

If the bureau and Mr. Gore follow through with their plan to throw out the post-census review, states and localities would lose that avenue to appeal. With limited local input on the front end, and no review on the back end, state and local governments would be effectively shut out of the process.

That must not happen. In the words of Mr. Maguire, "If it comes to accurate data and precise data and data that makes a difference, the Feds are the furthest away. And states are a little closer, but the local government is where it happens."

The U.S. House of Representatives recently passed an appropriations bill that provides for a review of this and other major changes planned for the 2000 census. The bill would buy time for a review without shutting down all other federal agencies funded by the same bill. President Clinton has threatened to veto.

He shouldn't. If we can slow down the census process and take a look at it without shutting down government, we should. We need to examine some important issues, like the ability of state and local governments to participate in a process that determines how much federal money comes their way.

As a former mayor and undersecretary of HUD, I know and believe in the power of local government. I also believe the U.S. House of Representatives did state and local government a service, by providing for our review of major changes planned for the 2000 Census — changes that could have major repercussions in neighborhoods of every color. On the other hand, I believe Mr. Gore did us all a disservice by playing the race card.

J. Kenneth Blackwell is treasurer of the State of Ohio and co-chairman of the U.S. Census Monitoring Board, Congressional Members. He is a candidate for the office of secretary of state of Ohio. This article was distributed by Scripps Howard News Service.

OUR VIEWS

Every American, not just a sample

The Constitution requires the government of the United States to conduct an "actual enumeration" of the population every ten years, "in such manner as they (Congress) shall by law direct." The Census has been tallied every decade since 1790, when it was awfully hard to get an accurate count.

In those days, it sometimes took weeks to travel from one end of the 13 states to the other. There were fewer than 4 million Americans, we think, because the count missed a good many. The counters, or "enumerators" in Census Bureau lingo, rode horseback from farm to farm; many people refused to cooperate because they didn't know why the government needed the information. Even the election of George Washington as president did not fill that early credibility gap.

After 200 years, the accuracy of the Census is not really in doubt: Some people are missed, despite the efforts of mail-in ballots and bilingual enumerators and a massive advertising campaign. Every Census since 1970 has been challenged as particularly inaccurate in counting urban residents and minorities, including the urban poor or non-English-speaking Americans.

The Clinton administration proposed last year to reduce the undercount in 2000 by using a sophisticated sampling method to project how many people actually lived in hard-to-count areas. The Census Bureau said that at least 90 percent of the people living in those Census tracts would have been "actually enumerated" before any fiddling with the numbers, but that wasn't good enough for Speaker Newt Gingrich, who challenged the idea in court.

Republicans generally think that sampling adds predominantly Democratic voters to the rolls. The Democrats accused the Republicans of being anti-minority, but a federal court agreed this week that federal law governing Census Bureau operations requires an actual count for reapportionment purposes.

The federal court ruled based on federal statutes, not the Constitution's mandate; the Census Bureau's appeal next goes to the U.S. Supreme Court to settle the issue.

We believe that the Constitution says what it says.

The federal laws governing the Census, written in 1957 and 1976, contemplated using sampling methods for the vast amount of other data collected by the bureau — how many people live in what kind of houses, have whatever kind of heating, and so on and so forth.

All that information is incredibly useful to businesses, to the government and other users of Census data; it can be extrapolated and interpolated and whatever other kind of massaging is necessary and legitimate. Those purposes, though, aren't the weighty subject that the Constitutional Convention had in mind with its requirement, in 1787, of an "actual enumeration" to apportion the Congress of the United States.

For that purpose, and for other official uses, such as distribution of federal aid, the words of the Constitution ought to be the words the Census should live by.

BB Baton Rouge, Louisiana
The Advocate 8/29/98

Give up on sampling the census

The Constitution won a major victory this week when a three-judge panel ruled unanimously that the Census Bureau can't use statistical sampling to count Americans in 2000. The Founders included a requirement for an "actual enumeration" in the supreme law of the land, and the judges said the Constitution means what it says.

Unfortunately, the Clinton administration doesn't appear willing to accept defeat. It appealed the ruling to the Supreme Court. And the Census Bureau insists on forging through the roadblocks, preparing to use sampling even though it is now illegal and will remain so at least until the high court rules. Incredibly, this issue could end up bringing the government to a partial halt. Republicans are prepared to pass an appropriations bill that would ban sampling, and Clinton says veto it.

We have no quarrel with the science of statistical sampling. It is a proven way to accurately take stock of large numbers. What we quarrel with is its use by politicians on a matter as vital as the apportionment of congressional districts. Official pressures could come to bear on decisions as to which districts should be counted with sampling and which should not. Americans have suffered through the military base closure fiasco, the administration's pressure to make citizens out of potential Democrats who never had never undergone required FBI checks and a host of other recent incidents that ought to make them wary of politicizing the census.

Democrats say the GOP ought to show it is prepared to pay the enormous cost of conducting an accurate door-to-door count. That may be true. It also may be true that the underlying reasons for this fight are political on both sides. Democrats want more congressional representation and Republicans are trying to prevent it.

But that is precisely why it's best to rely on the wording of the Constitution. Yes, census counters may have missed millions of people in 1990, including some ethnic minorities. But the census cannot subject itself to political pressures, and the president should think twice before vetoing an appropriations bill in the name of a cause that has been rendered illegal.

The Desert News, Salt Lake City, UT
8·29·98

The 2000 census

A special three-judge federal panel, which Congress specifically empowered to rule on suits involving census issues, has dealt a major blow to the Clinton administration's unprecedented plans to employ statistical sampling techniques to change the head count obtained in the 2000 census. In a ruling handed down this week, the three judges —Circuit Judge Douglas Ginsburg and District Judges Royce Lamberth and Ricardo Urbina —unanimously ruled that the Clinton administration had misinterpreted the Census Acts of 1957 and 1976. The judges issued an injunction forbidding the Census Bureau from using statistical sampling to change the overall head count achieved in the 2000 census.

Statistical sampling can be used for the collection of subsets of general demographic data involving income, education, occupation and the like, as it has been used in the past, the court ruled. But where the census is used to "determine population for the purposes of apportionment" in the House of Representatives, the judges ruled, the Census Bureau has been expressly prohibited by Congress from using statistical sampling.

The sampling issue has indisputably political origins traceable to the census' inability to count everyone. The Census Bureau has reported that post-census evaluations revealed that the undercount for the 1990 census amounted to 1.8 percent, a 50 percent increase over 1980's undercount. But the undercount was not evenly distributed racially or geographically — and, thus, not politically, either. The Census Bureau estimated that 4.4 percent of blacks, 5 percent of Hispanics and 12.2 percent of American Indians were not counted in 1990, compared to only 0.7 percent of non-Hispanic whites. Also, much undercounting occurred in the nation's big cities, whose congressional representation is decidedly Democratic. Because congressional district boundaries were determined by unadjusted census data, the undercount resulted in a malapportionment of Congress, the administration argued. Moreover, because census population data are used to allocate about $180 billion in federal spending, any inaccurate head count would result in financial misallocations. Whatever the problems with undercounting, Republicans have argued that the constitution nonetheless required Congress to conduct an "actual enumeration" of the population, which they have insisted precludes sampling. The attorney for the Republican-controlled House of Representatives, which was the plaintiff in the case, argued before the court that "the word 'enumerate' in every dictionary at the time said to count one-by-one or to reckon singularly." In its decision this week, however, the panel of federal judges did not get to the constitutional issue, finding instead that sampling was, as a matter of law, specifically forbidden by Congress for apportionment purposes.

Justice Department lawyers, who argued the administration's position, implausibly asserted that Congress' permission for the Census Bureau to use sampling in some areas automatically extended to the enumeration for apportionment purposes. The judges found the exact opposite to be true. As the court noted, "[W]hen Congress first authorized the use of sampling in 1957, [the congressional apportionment function] was expressly excepted." In 1957, Congress approved the use of sampling "except for the determination of population for the purposes of apportionment of representatives in Congress." When the use of sampling was expanded by the 1976 amendments, Congress again retained what the court called its "ubiquitous qualifier — 'except in the apportionment of the U.S. House of Representatives.' " The court reasonably concluded, "The inclusion of this 'except' language must be read to mean that one area was affirmatively carved out from the general desire to augment the use of sampling — the area of the congressional apportionment enumeration."

For now, pending a certain appeal to the U.S. Supreme Court, the country is spared the unholy possibility of a politicized Clinton administration manipulating the 2000 head count to maximize political benefits for the Democratic Party.

The Washington Times, 8-27-98

Newsday (New York, NY)

August 27, 1998, Thursday, ALL EDITIONS

SECTION: VIEWPOINTS; Page A45

LENGTH: 749 words

HEADLINE: LET'S NOT 'INVENT PEOPLE FOR THE CENSUS

BYLINE: James P. Pinkerton. James P. Pinkerton's e-mail address is pinkerto@ix.netcom.com.

BODY:

FEDERAL OFFICE Building No. 3 in Suitland, Md., is the home of the U.S. Bureau of the Census, ground zero for what will prove to be the largest transfer of power in American history. The decennial census for the year 2000 will redirect trillions in federal revenues, as well as shift a score or more of congressional seats from one state to another.

The bureau, an arm of the Commerce Department, has a constitutional mandate to conduct an "actual enumeration" of the population without regard to politics. But given the controversy over the upcoming census - which has caused the Republican-controlled legislative branch to sue the Democratic-controlled executive branch - it is fitting that across Silver Hill Road from the census headquarters is the local Democratic headquarters.

For its part, the bureau estimates that the 1990 census missed 1.8 percent of the population - 4.7 million people. As a result, many Americans, especially members of minority groups, "were denied an equal voice in their government." Moreover, "federal spending employing population-based formulas - for schools, crime prevention, health care and transportation - was misdirected."

And so the bureau proposes, for the first time in U.S. history, to speak for the speechless. In its "Census 2000 Operation Plan," it pledges to "contact at least 90 percent of the households in each census tract" and then to use "widely accepted statistical methods" to calculate the remainder of the population. Such "sampling," the bureau promises, will "produce better results at less cost."

Who can be against progress? Charles D. Jones, who retired after 32 years at the bureau as its associate director for the decennial census, says of sampling: "You're replacing one set of errors you don't like with a set of errors you like better."

Jones points to an analysis of the 1990 census, showing men were three times less likely to be counted than women. Deadbeat dads and criminals loomed large in this category. "These people," Jones adds dryly, "have an inherent interest in not being counted. But isn't it the bureau's duty to find everyone? Sure. And when Jones oversaw the last census a "special-coverage" program enumerated 457,000 prisoners and parolees.

But making an extra effort at finding folks is not the same as fictionalizing them. Jones cites the 1990 analysis, which surmised that 12.7 percent of black men between the age of 25 and 29 were not counted. The bureau could, he allows, simply create an eighth black male in that age bracket for every seven that it counts, but where would it situate those hypothetical individuals? After all, figuring out precisely where real people live is the whole point of the census; it determines congressional reapportionment and federal funding flows.

With just a slim 11-vote margin, it's little wonder the House Republicans sued to block the bureau's plan. They worried that a sampled census would conjure up 5 million or 10 million people, thereby transferring a dozen congressional districts out of Republican suburbs and into Democratic inner cities - and thereby transferring the speakership. On Monday, a three-judge panel sided with the Congress; Bill Clinton's administration will appeal.

In the meantime, a nervous Congress established a Census Monitoring Board, split between Capitol Hill-appointed Republicans and White House-appointed Democrats, with offices in Suitland. Jones, the census veteran, is a consultant for the Congress, but he retains a civil servant's aversion to sharp language. By contrast, Fred Asbell, who directs the GOP's observation team, is more cutting. Sampling, he argues, permits "all manner of malfeasance." The right approach, Asbell insists, is the old-fashioned method of counting people - not concocting them in a computer.

All this year, first in regard to Iraq and now the embasssy bombings, political junkies have speculated about a "Wag the Dog" scenario: The president manipulates a foreign crisis for his domestic benefit. Yet, the truth is closer to home: Clinton isn't wagging the dog; he's wagging the data.

If the Democrats can rejigger the census, they can recapture the House. And, while their task would be easier if the courts were to allow sampling, even if they don't the bureau is a big place. The few Republicans wandering around Suitland are unlikely to find all the statistical sleights that the Clintonians have up their sleeves.

LANGUAGE: English

August 27, 1998, Thursday, Late Edition - Final

SECTION: Section D; Page 2; Column 1; Business/Financial Desk

LENGTH: 911 words

HEADLINE: Economic Scene;
Politics aside, census sampling still raises questions of accuracy.

BYLINE: By Michael M. Weinstein

BODY:

A PANEL of three Federal judges delivered a blow to Democrats this week when it ruled that sampling techniques planned for the 2000 census violated Federal law. The issue is far from dead, though, as the Administration now plans to appeal to the Supreme Court.

Democrats embrace sampling -- data derived from a small set of households -- to supplement census information gathered door to door and through the mail. They believe it would add millions of blacks, Hispanic people and poor to the census rolls. That would lead to billions of dollars of additional Federal spending in Democrat-controlled districts and perhaps a dozen or more additional Democratic seats after Congressional districts are redrawn. The Republicans loathe sampling for the same reasons.

But the debate is about something more difficult, and less partisan, than politics. The handful of statisticians who have mastered the ferocious mathematical and practical complexities are split over the usefulness of sampling. William Kruskal, former chairman of the statistics department at the University of Chicago, speaks for many of his colleagues when he admits that "no one really knows."

A middle ground, using sampling techniques for parts of the census and not for others, makes perhaps the most sense.

The bureau proposes three separate uses of sampling. One sample would focus on a small group of households that fail to mail in their census forms or to respond to repeated visits by census takers, and then use statistical techniques to reach a description of all nonresponding households. Most experts appear to agree that this use of sampling is a sensible way to cut the soaring costs of conducting the census without sacrificing accuracy.

The second proposed use, to correct for the undercount of minorities and poor people, draws more criticism.

The 1990 census underestimated the population by less than 2 percent. But it underestimated the number of blacks by more than 4 percent and the number of Hispanic people by about 5 percent. Likewise, it missed more renters than homeowners and more poor families than rich.

The bureau proposes to fix this problem by conducting a follow-up survey of 750,000 households. The bureau would compare the two surveys to determine, for example, how many urban blacks appeared in the follow-up survey but not the first -- an indication of how many were missed in the initial census. The bureau would then use a complicated series of calculations to estimate what portions of several dozen demographic groups were initially missed.

Critics, like Prof. David Freedman of the University of California at Berkeley, say the sampling procedure is fraught with problems. What if someone moves, shows up on the rolls of the second survey but appears nowhere in the country in the initial census? Was he one of the millions of uncounted Americans? Perhaps not. He might have been accounted for by the proposed statistical correction for nonresponding households (the first use of sampling) even without his ever having filled out a form.

Here's an even trickier problem. Sampling could improve the census count for every state, yet damage the overall value of the census. That's because the important census figures, in terms of funding and Congressional representation, are not the absolute numbers for each state and county, but rather their respective shares of the national population. After all, there will be 435 House seats no matter how many people are counted. If sampling does a better job locating missing households in one state than in another, the distribution of House seats and Federal spending could be made less fair.

Imagine that the initial census misses 1,000 black households in Harlem and 100 black households in Phoenix because of evasive households and less-than-diligent census takers. The follow-up sample will be plagued by similar problems. What if it turns up, say, only 50 of the missing black households in Harlem but 75 of the missing households in Phoenix. Then each state's population count will improve, but Arizona's share of the total population will be overstated.

Polling only works when you have a reliable baseline 8/27/98

By CINDI ROSS SCOPPE
The State (COLUMBIA, SC)

The insidious nature of the twice-removed guesstimate called weighting jumped out at me while I was analyzing the differences in white and black responses to one of The State's polls. Several of the numbers reported for African-Americans were significantly different from previous polling. When I asked our pollster if there was something strange about our sample, he said our unweighted sample of 492 people included only 41 black people.

That meant that while we reported an overall margin of sample error of 4.4 percent, the error margin for black voters was 15.3 percent. And then we took that sample with a 15 percent margin of error and multiplied it three times so that black people made up about a third of our sample. This multiplying effect meant that if we had a bad subsample, our overall sample could be off by significantly more than 4 points.

That problem haunts me every time I think about the Census Bureau's plans to use sampling for the 2000 census. The proposal was struck down by a special three-judge panel Monday, but the Clinton administration plans to appeal to the Supreme Court. That means the question is not closed.

The courts are no place to decide how to accurately count the U.S. population. That question is not, as statistical experts would have us believe, a simple matter of mathematics. Sampling is based on statistical equations, but at its heart are assumptions and judgment calls. And in this case, the sampling would be based on layer atop layer of assumptions.

There's no question that we must find a way to count under-counted minorities: Officials estimate that the 1990 census missed 1.6 percent of the U.S. population. They say that while just 0.7 percent of the non-Hispanic white population was uncounted, the figure was 4.4 percent of African-Americans and 5 percent of Hispanics.

And I think Democrats are probably right in charging that congressional Republicans oppose sampling because the people we're missing are more likely to be Democrats than Republicans.

But as one pollster I talked to noted, census data often provide the baseline that pollsters use to determine what the racial or gender or educational or income make-up of a certain population is. If their polling results don't match up, they simply weight them to match. But what if the baseline is weighted?

The Census Bureau wants to actually count everyone in 90 percent of the households in each census tract (which has about 1,700 dwellings). Data from the 90 percent would be used to define the other 10 percent. The results of that count and extrapolation would then be modified, based on the results of a survey of 750,000 households.

But each step of that process -- and there are at least four steps -- requires census officials to make assumptions. Assumptions that cannot be checked against anything except other assumptions.

We assume the 1990 census was off by 1.6 percent because later sampling told us so. But it is inherently illogical to conclude that people who can't be found and counted in a census will be found and counted in a survey. So assumptions had to be made, and weighting had to occur, to produce numbers that matched somebody's preconceived notions.

We assume that we can count 90 percent of the population to create a baseline. But we don't know what 90 percent of the population is unless we know what 100 percent of the population is. Or unless we guess.

We assume that census officials can randomly pick 750,000 people who will look like America. But the people who go uncounted in a full-blown headcount are even more likely to go uncounted in a scaled-down survey. So the sample looks like America only if it's somehow weighted to match our notion of what America looks like.

Consider the standard language The State uses to describe what a poll of 500 registered voters represents:

"In theory, there is a 95 percent probability that the results of the poll are within 4.4 percentage points of what would be obtained by talking to all adults in South Carolina who are registered to vote, have a telephone, are home when the poll is conducted and are willing to participate."

Notice that we don't claim that our results reflect the opinions of all registered voters -- only those who have a phone, are home when the survey is conducted and are willing to participate. The results of a survey can only be assumed to represent the people who share the characteristics of those surveyed.

Of course, by manipulating your assumptions, you can guarantee to count more women or more African-Americans or more 56-year-olds or more divorced men. But you cannot guarantee that your count will be a more accurate reflection of the total population than an actual count of the total population.

Only that it will be different

REVIEW & OUTLOOK

Sampling Error

Monday's court ruling against the use of statistical sampling in the Census is scarcely a run-of-the-mill judicial decision. Rather, it's another rejection of the Clinton Administration's bending of the law and, perhaps more importantly, a rebuff to the latest Clinton tactical political brainstorm.

The Clinton Commerce Department has insisted it already has legal power to institute statistical sampling as the centerpiece of the year 2000 Census. The House of Representatives sued, at the behest of Speaker Gingrich, contending Commerce did not have such power without legislation Congress had no intention of passing. To speed the decision, the suit was heard by a special three-judge panel of two district judges and one from the D.C. Court of Appeals, two appointed by Ronald Reagan and one appointed by President Clinton. They came down 3-0 against sampling.

Judge Royce Lamberth's decision told Commerce to read the law. Congress had indeed authorized Census to use sampling, but the same law prohibits its use for the vital purpose of apportioning House seats. The whole thrust of the sampling argument, of course, has been that the traditional head count misses some people who should be statistically reinserted for the purposes of apportionment. No one cares about using sampling for scholarly questions, or for that matter about extra efforts to locate hard-to-reach people. But the statutory exception conforms with the Constitution, which specifies reapportionment based on an "actual enumeration" of all Americans, meaning, as the dictionary defines it "to count off or name one by one."

As we've repeatedly seen in the past 200 years, the Founding Fathers were not fools. Yes, statistical sampling has been scientifically developed since their time, and we haven't the least doubt that in scientific hands it's a valid tool. But what hasn't changed since the Constitution was written is human nature, particularly that of political humans. The reason for an actual enumeration is not that sampling is scientifically flawed, but that politicians cheat.

In designing statistical sampling, the Census Bureau would have to make a whole series of decisions about what kind of districts to sample, and how much to adjust different results. Though there may be some defensible scientific basis for doing this, the decisions would be subject to political pressure every step of the way. Who would trust Census bureaucrats to stand up to this Administration if it tried to twist the results?

In the last Presidential election, the Immigration and Naturalization Agency let itself be used when a record 1.1 million people were made U.S. citizens in a clear attempt to add to the ranks of Democratic constituencies. Post-election it was learned that 180,000 of these people had become citizens without formal FBI approval. Of those 180,000, the FBI had criminal records for 71,000, including 16,400 who had been arrested on felony charges.

Congressional probes have documented how Vice President Gore's office pressured the INS to take "drastic measures" to speed up its "Citizenship USA" program. A memo indicates that White House deputy chief of staff Harold Ickes was briefed on "new-citizen voter registration" at a September 1995 meeting held only one month after "Citizenship USA" was created. "[T]he pace of naturalization will limit the number of new voters," the memo warned. A later memo lamented that INS lethargy would fail to "produce a million new citizens before election day."

After this episode, the Administration wants Congress to drop the Constitutional "enumeration" for something more "modern," and more flexible. At the same time Attorney General Reno has been flouting the law by refusing to appoint an Independent Counsel for the President and Vice President on the subject of campaign contributions. For that matter, the same Commerce Department that insists it can conduct sampling fairly brokered seats on foreign trade missions in the first Clinton term as a crude campaign fund-raising device. At least we still have a judiciary willing to blow the whistle.

Not only that, but the Administration has made the sampling issue a centerpiece of its "close the government" tactic. As recently as August 3, the President vowed to veto any "bare bones" continuing resolution to keep the government funding at current levels pending resolution of policy disputes in the budget. In 1996, he pulled off a public relations coup by vetoing a bill that would have kept the government open and then blaming the Republicans for having closed it. The Republicans have still not recovered, one reason why the President hopes he might pull the same trick again.

In particular, President Clinton threatens to veto appropriations for Justice, State and Commerce if they include limits on the use of sampling. How convenient to veto Justice, including money for Ken Starr's investigation. The court ruling leaves this strategy about as phony as the sincerity-dripping denials that the President had a sexual affair with Monica Lewinsky. We await smoke signals from the White House telling us what the next veto excuse will be.

The Wall Street Journal, 8-26-98

August 26, 1998, Wednesday, JOURNAL EDITION

SECTION: EDITORIAL; Pg. 12A

LENGTH: 435 words

SERIES: Today's News

HEADLINE: Census 'estimate' wrong and unfair

BODY:

NOW WE KNOW how desperately the Clinton administration wants the chance to fiddle with the outcome of the 2000 census. Officials said this week that they don't care what the law says --- as described flatly and unequivocally by a special panel of federal judges --- and they're going to keep on trying to change the process so they can "estimate" some 10 percent of the population.

The three-judge panel said there was no uncertainty about the intent of the law, which allows sampling for census purposes "(e)xcept for the determination of population for purposes of apportionment of Representatives in Congress among the several States. . . ." Commerce Secretary William Daley announced Tuesday that the government will appeal that ruling, hoping to get the Supreme Court to accept the administration's grammatical juggling act that contends the statute means the opposite of what its words say.

We're not surprised at the decision to appeal, because the White House and Democratic leaders see the opportunity to manipulate the census count as a cheap way to dig up --- or create out of thin air --- a few million more Democratic voters and carve out a few more Democratic districts for congressional and local government seats. The assumption is that those most of those missed by traditional counting methods are minorities and poor, and thus more likely lean to the left.

Frankly, we don't think the political proclivities of these invisible residents is the key point in this debate. The biggest problem with the sampling plan is that it would simply be wrong and unfair.

We might have less objection if the backers of the "new and improved" method had proposed using it to help account only for the 1.5 percent or so who would probably be missed by the "actual enumeration" demanded by the Constitution. But the administration decided instead to do a real count for just 90 percent of the population, and to estimate a whopping 10 percent. That introduces way too much opportunity for, shall we say, creativity in decisions about where to count and where to guess, and about which locations to use as controls for making final adjustments in the sampled areas.

It is, of course, desirable to have a census that is as nearly complete and accurate as possible. There are some things under consideration that could make traditional counting a bit more accurate --- maybe 99 percent, instead of 98.5 percent. That's not perfection, but it has one critical thing in its favor: It would be a lot more trustworthy than a system in which a tenth of the people were imagined into existence.

LOAD-DATE: August 27, 1998

Let them count, 1 at a time

Statistical sampling to determine U.S. population a bad idea

Since 1790, census-takers have counted heads and collected volumes of data on how we live. These census-takers don't take a poll and, based on a sampling of 1,800 people, determine that three people live in your house, four next door and 4.6 million in all of Arizona.

EDITORIAL: OUR VIEW

No, they count heads — one by one, as the Constitution requires in its reference to an "actual enumeration" of Americans every 10 years. In the same way, we hold elections instead of relying on a sampling of voters to determine our elected officials.

The traditional method of taking a census is being challenged by President Clinton and most Democrats. They want to use statistical sampling — that is, polling — in the 2000 census. It's a bad idea that invites bias, fraud and all sorts of other problems.

At stake in this controversy — as is usually the case when politics rears its head — are money and power.

Data collected from the 2000 census will be used to divvy up nearly $200 billion a year in federal funds among cities and states. The data also will determine which states pick up seats and which lose them in the House of Representatives.

Local governments and businesses depend on the census data being honest and reliable. Cities use the information in their planning for police and fire stations, among other facilities. Businesses rely on it to target customers and help decide where to locate everything from restaurants to medical centers.

The debate boils down to this: Republicans believe census-takers should count heads for reasons of accuracy, fairness and because the Constitution requires it. Democrats, saying too many Americans were missed in the 1990 census, particularly in larger cities, propose using a statistical sampling method to supplement the standard head count.

Under the Democratic proposal, the census would count 90 percent of the population. Then, a statistical sampling would be used to make estimates of the other 10 percent — or about 27 million people, give or take a few million.

All of which sparks the question:

"How do you know when you're at 90 percent if you don't know what 100 percent is?" asks U.S. Rep. John Shadegg, R-Ariz., who serves on the newly-formed House subcommittee on census.

Yet another problem is that sampling becomes less accurate as an area gets smaller. For example, sampling a neighborhood block is subject to great error — plus or minus 28 percentage points, according to one study. That has enormous implications because redistricting starts from very small census blocks.

For all their stated concerns about undercounting among inner-city Blacks and rural Hispanics, Democrats don't propose sampling to tally the most undercounted minority in the last census: Native Americans.

Their rationale is that counting heads is the most accurate method on reservations — which doesn't say much for the polling argument. What's really going on, of course, is that there aren't enough undercounted Native Americans to interest Democrats. For Democrats, the next census is largely about finding those groups traditionally viewed as Democratic voters.

While the census isn't perfect, it can be made to work by returning to the basics of knocking on doors to make sure everyone is contacted and completes a survey in this $4 billion process. Many of the 4 million who were missed in the last census were poor. Half were estimated to be children. If all else fails, let's explore reaching the undercounted when they come into contact with the government for medical and other services.

Polls, as any politician knows, can be inaccurate. They can be biased, depending on how you decide to conduct them. They can also be easily manipulated for political purposes. In 1990, the Census Bureau suggested five formulas for adjusting numbers, and each would have produced a different apportionment of House seats.

Under the Democratic method of counting, there is even talk that some surveys would be rejected if they didn't fit the statistical scenario. That's wrong. Americans participating in the census have every right to expect that their surveys will be counted — and that they count, too.

Arizona Republic 8·17·98

Maintain opposition to sampling scheme

THE BATTLE to keep the 2000 census count from becoming part of a Democratic Party get-out-the-vote campaign is rapidly approaching a critical showdown between Congress and the White House. Word is that President Clinton may veto a funding bill for the Commerce, State and Justice departments, hoping that he can get the American people angry at Republicans for "shutting down government."

The confrontation was set up this way: The Senate passed an annual appropriations bill for the three government agencies, but the House voted for six months of funding. The House goal was to hold the rest of the funds in abeyance until agreement is reached over the administration's controversial plan to use "statistical sampling" to come up with a number for people missed by traditional census-taking methods.

We think that's a proper tactic because we don't believe the Democrats' plan for guessing — scientifically or otherwise — how many people were missed is good for the country. The issue is not whether Democrats or Republicans would benefit from a more complete census count; the issue is the honesty and accuracy of the count that would result.

At the theoretical level, statistical sampling may be a reasonable approach to a real problem — the unargued fact that several million people are uncounted in each census. In practice, however, the estimating and calculating would be left in the hands of an administration that has already demonstrated its willingness to manipulate the meaning of words, and its considerable skill in doing so; we have no doubt it would be equally clever at juggling numbers.

It's not clear whether the House effort will survive the conference committee process, in which its version of the funding bill must be reconciled with the Senate version. Clinton's veto threat may cow some Republicans, who remember the damage the party suffered in government shutdowns in 1995. We hope that doesn't happen, though, because it's important that Congress continue to fight against the sampling scheme. If that plan is carried out by the Clinton administration, the damage to the party — and to fairness in the American political system — will be much worse.

The Atlanta Journal-Constitution

8·17·98

Las Vegas Review-Journal (Las Vegas, NV)

July 28, 1998 Tuesday, FINAL EDITION

SECTION: B;

Pg. 6B

LENGTH: 518 words

HEADLINE: Squabble over 'sampling'

BODY: Congressional Republicans now threaten to withhold or limit $952 million in money earmarked to prepare for the Year 2000 census, in a dispute over administration plans to use computer 'sampling' to supplement or (in some cases) replace more traditional methods of head-counting.

'Sampling' is a process by which statisticians gather detailed data about a limited group of people, and then extrapolate that data onto a state or nation.

Both sides of the debate seem to believe such computer extrapolations will increase the count of racial minorities and the economically deprived, particularly in urban areas.

The purely partisan concern is that beefing up inner-city population counts via 'sampling' may divert more federal funds _ and perhaps a few more congressional districts _ to urban areas, which tend to vote Democratic.

But that's the wrong way to look at the issue. If more blacks or Hispanics _ or Eastern Europeans, for that matter, or homeless hobos _ can really be found, they should indeed be counted and represented.

The problem is that sampling, by its very nature, takes any error in the initial sample _ or in the process by which the sample populations are selected _ and multiplies it.

Although the result will tend to look far more 'scientific' than a number picked out of a hat, programmers have long referred to this effect as GIGO _ 'Garbage In, Garbage Out.'

Let us suppose that errors creep into a sample count of the population of Corpus Christi. Large numbers of undocumen- ted immigrants appear to be present in the metropolitan area, though many run away or hide from the census-takers. In the end, the census-takers' 'best guess' is that the Texas city of 300,000 has an Hispanic population of 100,000 (somewhat higher than the Texas state average of 26 percent), and that there are 20,000 undocumented residents _ a guess which happens to be off by 12 percent.

The computers are then set to whirring, and dutifully inform us _ down to a specific-sounding decimal point _ that Wisconsin, with a population of 4.9 million, must have 1.63 million Hispanic inhabitants, of whom 326,667 are undocumented seasonal agricultural workers.

The example is exaggerated _ hopefully the Census Bureau knows there aren't many avocado groves in Eau Claire. But the fact is, the Constitution calls for an 'actual enumeration' _ a counting. If we're going to rely on computer projections, why not really save some dough? Just run out existing trends for another decade, and issue the 2010 census at the same time?

Americans are counted and tracked today as never before. Working from birth and death certificates, school and vaccination records, voter registrations and motor vehicle data, Washington should be able to get a line on 96 percent of us, pretty quickly.

For the 4 percent where questions remain, let the census takers go forth and count, in person.

Virtual Census

Clinton, Democrats Would Gerrymander Count

The Union Leader (Manchester NH) 7/21/98

Bill Clinton continues to flaunt his contempt for the Constitution that he took a sacred oath to defend. Now King Clinton seeks to politicize the U.S. Census and manipulate it in the largest gerrymander in the nation's history.

Since 1790, the United States Census, mandated by the Constitution for the purpose of apportioning representation in Congress, has counted the American people. Actually counted them. The Constitution, in fact, requires an "actual enumeration," a literal head count. Not a guess, not an estimate, not a projection. A count. Period.

Last week Vice President Al Gore appeared before the NAACP and declared preposterously that Republicans have so little regard for black Americans that "they don't even want to count you!" Let us be clear: It is the Democrats who do not want an actual person-by-person count.

Ignoring the plain language of the Constitution and 210 years of precedent, Mr. Clinton proposes that the Census count only 88 percent of the population. The balance of the enumeration would rely on a "sampling" method to "adjust" statistically the final numbers — three separate times. The Clinton-Gore administration claims this fancy guesswork would make the Census "more accurate."

More accurate? Than an actual count? Oh, please! This amounts to a virtual Census counting virtual people.

What is really going on here is a massive effort by the Clinton administration to rig the Census. The Dems hope to use this sampling scheme to increase the proportion of House seats in core urban centers of the country, where Democrats tend to dominate, at the expense of the more Republican suburban and rural areas.

The temptation to manipulate this sampling sleight-of-hand will be hard to resist. One need only to look at the gaggle of liberal organizations trumpeting this statistical guesswork to get a pretty good idea of what is really going on here. Moreover, Bill Clinton has put such partisan political hacks as John Podesta, White House deputy chief of staff, and the disgraced former congressman and legendary Democrat arm-twister Tony Coehlo in charge of monitoring the sampling scheme.

How can this usurpation of the Constitution be stopped? It will not be easy. Twice Congress has attempted to condition funding for the Census on the bureau taking an actual head count and twice Bill Clinton has vetoed the measures.

House Speaker Newt Gingrich has filed a lawsuit against the administration to force it to obey the requirements of the Constitution and federal laws that prohibit the use of statistical sampling. Unfortunately, the courts may not rule on the Gingrich lawsuit in time to stop the rigging of the Census. Given this President's great talent for stalling, delaying, stonewalling and manipulating the judicial system, this remains a distinct possibility.

Americans may not care much about the Census and esoteric arguments about sampling, but they should. If Bill Clinton and the Democrats can run roughshod over any part of the Constitution, then none of our liberties is safe. And if the apportionment of Congress can be rigged, then representational government will have become extinct.

—Richard Lessner

KENNETH BLACKWELL W TIMES 7/12/98

Sense of a census to count on

There's a war raging on Capitol Hill. The media report blood has already been spilled in the partisan fight over the best way to count the U.S. population in the 2000 census. Democrats are lobbing explosive statisticians at Republicans bivouacked behind the Constitution.

Charged to observe and report on the 2000 census, we eight members of the U.S. Census Monitoring Board are walking into no-man's land. We are caught squarely between partisan forces. Both sides are clamoring for our support. But I would like to think that we are here to provide relief to the biggest casualties in this war of words: the 2000 census, and the people at the Census Bureau who conduct it.

The Census Bureau's main facility is in Suitland, 15 minutes and a lifetime away from the Hill. There, a few thousand brave souls face the daunting task of collecting detailed information about more than 270 million Americans. It's a needle-in-the-haystack job, with the added burden of counting and describing every piece of hay. Aside from the sheer size of the assignment, they contend with public apathy and, increasingly, political pressure.

The situation is this: The 1990 census — the decennial count of all U.S. citizens used to apportion seats in Congress and guide the flow of federal dollars — came up short by a net undercount of more than 4 million people. Although that was a successful count of more than 98 percent of the population, the undercount was bigger than in 1980, which is cause for concern. Of even greater concern, however, is the fact that the 1990 census missed a much larger proportion of minorities than whites. That is unacceptable. It is also a starting point for improvements.

The partisan war rages over how those improvements ought to be made. One side believes statistical estimates can and should be used to add or subtract people from the rolls, to correct for undercounts. The other side believes nothing but a conventional head count will satisfy the constitutional mandate of an actual enumeration.

The unconscionable part about using statistical methods is that real people who fill out census questionnaires will necessarily be subtracted from the census to round out the calculations. But missing people through an inefficient enumeration also seems unconscionable.

A reasonable alternative exists, but it may be difficult to hear above the shouting. There should be a way to improve the fairness and accuracy of the census, and still ensure that every American who fills out a census form will be counted.

The 1990 effort was a good start, but it does not take a mathematician to see room for improvement.

One common sense solution would be to use administrative records. Filed with government agencies that administer public programs, administrative records offer generally up-to-date information such as name, address, Social Security number, age, race and gender.

Using administrative records is not unprecedented. Recognizing their potential benefits to improve counts of minority populations, the 1990 census added about 1.3 million people to the rolls using parole and probation records. Five years later, the Bureau of Justice Statistic showed 3.8 million people on parole or probation. The 1990 effort was a good start, but it does not take a mathematician to see room for improvement. That leaves nearly 2 million additional people, more than one out of four African-American, with an address on file.

Those 2 million are particularly significant, considering the net undercount in 1990 was 4 million. An additional 2 million records would reduce the undercount by half, with a substantial increase in minority coverage.

Another substantial increase in coverage would come from using Medicaid records. Children under the age of 18 represented more than half (52 percent) of the undercount in 1990. In 1996, Medicaid had records on 18.3 million people aged 20 and under. Many of those young people are in the hard-to-count populations: low-income minorities in dense urban areas. Again, it doesn't take a scientist to figure out that a single mother struggling to make ends meet and raising a family might not have time to sit down and fill out a census form. But that same mother will take the time to enter similar information on a Medicaid form, because that's time spent toward medicine for her children.

These are common sense solutions that can be obscured by bureaucracy, tied up with red tape and shouted down by partisan arguments. But most Americans don't care about partisanship — they care about what works.

Using administrative records works. The method adds people frequently missed in a conventional enumeration, without subtracting people who were not. Meanwhile, statistical inference can be used to fill in the blanks of missing information — about race, age or gender, for instance — without raising questions of constitutionality.

By highlighting common sense solutions like the use of administrative records, the Census Monitoring Board has an opportunity to help negotiate a peace between warring factions. That might spare some hardworking people caught in the crossfire — like those brave souls in Suitland, Maryland.

J. Kenneth Blackwell is co-chairman of the U.S. Census Monitoring Board and treasurer of Ohio.

The Indianapolis Star, 7·10·98

The census dispute

The battle over using statistical sampling in the census of 2000 is now in the courts, specifically U.S. District Court in the nation's capital.

On June 12, a three-judge panel held the first hearing on the legality of sampling. A lawsuit filed by House Speaker Newt Gingrich and House Republicans claims sampling violates the Constitution and the Federal Census Act and should be prohibited.

The suit asks that the decision of the court, whatever it is, be immediately appealed to the Supreme Court. A second suit, filed by the Southeastern Legal Foundation, is pending. It also seeks to block sampling.

The "actual enumeration" called for by the Constitution may not be precise, but neither is census guesswork.

"The future of the republic depends on the just resolution of this issue," said Matthew Glavin, foundation president. "It will decide whether or not we can move forward as a representative democracy."

If the legal action is new, the controversy is not. For decades, Democrats have pushed for changes in census procedures and recruited liberal organizations to the cause. They claim the census seriously undercounts minorities and the urban poor — groups traditionally disposed to vote Democrat.

Democrats argue that the population shortfall results in fewer Democratic seats in the House of Representatives and a smaller share of federal largess for the Democrats' constituency.

Those arguments are wholeheartedly embraced by the White House and the Commerce Department, which oversees the census. Thus the Clinton administration is determined to ram sampling through, insisting it is the only way to produce a fair and accurate count.

Not every skeptic is a Republican. David A. Freedman, a University of California statistician, says sampling is questionable and risky. He notes that post-census surveys used to estimate the 1990 undercount turned out to be wrong. A computer coding error falsely inflated the undercount by about a million people. The more complex sampling procedures devised for the 2000 census could add more flaws than they fix, Freedman maintains.

Some Republicans charge sampling is vulnerable to political abuse, manipulation and exploitation. "Cooking" the numbers, they say, is not unthinkable for an administration that perverted naturalization laws in order to swell Democrat votes in the 1996 election and harvested illegal campaign contributions by the bushel load.

The "actual enumeration" called for by the Constitution may not be precise, but neither is guesswork, however sophisticated. A real person is better than a virtual one any day and a lot less subject to electronic error, deliberate or otherwise.

The Detroit News

July 03, 1998, Friday

SECTION: Editorial; Pg. Pg. A8

LENGTH: 547 words

HEADLINE: The Sampling Ploy

BODY:

President Clinton wants ongress to change the way the government conducts the decennial census. Under the old method, census takers went door-to-door interviewing people. They inevitably missed a few, but not many. The administration guesses that a little more than 1.5 percent of the public doesn't get counted.

In order to rectify that shortcoming, the president suggests using statistical guesses in such potential trouble spots as cities, sparsely populated rural areas and Indian reservations. He proposes using sampling methods, not unlike the ones employed by pollsters.

Under the procedure recommended by the administration, statisticians would correct for "improper" counts by adding presumably forgotten souls in some areas and reducing the estimates of residents elsewhere.

This is a high stakes game. Detroit stands to lose millions of dollars in federal and state aid if the nose counters don't place the city's population at more than 1 million. Mayor Dennis Archer is so eager to make the numbers work that he has received a $60 million grant to rip down 8,000 decrepit houses within city boundaries and has chosen to spend another $70 million sprucing up other homes -- the better to tempt newcomers to Motown.

Democrats like the idea of sampling because it tends to beef up the population estimates in Democratic strongholds. Unfortunately, however, sampling has a rotten record.

The Census Bureau tried it out in 1990 and generated wildly inaccurate results. It didn't get proper ratios of men to women. It discovered that sampling was especially unreliable in localities with fewer than 100,000 residents -- that is, most American jurisdictions.

A congressional panel recently heard testimony that the sampling regime would have improperly stripped Pennsylvania of one congressional district in 1990 and given Arizona one it didn't deserve.

Undaunted, the White House has adopted a two-pronged strategy on the census issue. First, appeal to greed. The president traveled recently to Houston and promised that the city could get millions of dollars under his plan. (He didn't point out that it would come at somebody else's expense.)

Fortunately, House leaders plan to propose legislation that would take the sting out of this divide-and-conquer strategy. The bill would prevent the government from deleting the names of anybody who had filled out a census form -- unless the documents were fraudulent or duplicative. This would give already-counted Americans a direct interest in census accuracy.

Second, the administration is seeking to set up a confrontation. This would involve pushing Congress to the brink of a government shutdown next year and folding the census changes in with such vital business as funding the Federal Bureau of Investigation. But GOP leaders have drafted a budget that provides a full year's funding for every agency but the Census Bureau, which gets only six months.

This maneuver will force Congress to debate sampling soon. Based on the evidence available, the "reform" would hold the census hostage to the ever-shifting whims of interest-group politics. Unless somebody can prove otherwise, Congress ought to reject sampling and stick with the old-fashioned enumeration required by the Constitution.

LOAD-DATE: July 03, 1998

NEW YORK POST
America's oldest continually published daily newspaper

Count This Nominee Out

President Clinton has appointed a New York statistician, Kenneth Prewitt, to be the new director of the U.S. Census Bureau. Confirmation to such a traditionally low-profile job would be routine in normal circumstances. But the job is no longer low-profile, and the circumstances are anything but routine. The Senate should ask Prewitt some tough questions.

Prewitt would succeed Martha Farnsworth Riche, who resigned in January as controversy grew over the Clinton administration's declared intention to use for the first time "sampling" — rather than the Constitution's mandated "actual enumeration" — in the 2000 census.

Prewitt sought to sidestep the issue when questioned by reporters, but made it clear he supports sampling.

"The weight of professional opinion, which I share, is that sampling, if properly used, can contribute to a more accurate census," he said. "Reasonable people can come to a different conclusion. This is a matter of scientific and technical review. It is not a matter of partisan disagreement. It is not a matter of ideological positioning, and it is not a question of whether data can be manipulated or not."

On the contrary, Dr. Prewitt. It is all of those things, and it is the man who appointed you who has injected partisanship into what had been an apolitical process.

The Republicans are defending the integrity of one of the few specific commands given to the federal government by the Constitution, which is to carry out an "actual enumeration" (the precise words) of the country. That means an actual head count, not a slippery estimate.

The Clinton administration has a track record of politicizing things its predecessors never would have dreamed of politicizing: the White House Travel Office and the Immigration and Naturalization Service, to name but two. It is the height of folly to assume that census sampling would be any different.

Let's be clear here: It is not the GOP that is being "partisan" in defending the Constitution; it is the Clinton White House that is seeking to subvert it.

Sense and the Census

By J. KENNETH BLACKWELL

A partisan war is under way on Capitol Hill over the 2000 census. Democrats want to use statistical sampling techniques to compensate for undercounting, especially of minorities, that has characterized past censuses. Republicans believe that nothing but a conventional head count will satisfy the constitutional mandate of an "actual enumeration."

The unconscionable part about using statistical methods is that real people who fill out census questionnaires would necessarily be subtracted from the census to round out the calculations. But missing people through an inefficient enumeration also seems unconscionable.

Fortunately, there is an alternative. The Census Bureau should use administrative forms to fill in gaps in the census. Filed with government agencies that administer public programs, such records offer generally up-to-date information such as name, address, Social Security number, age, race and sex.

There is a precedent. The 1990 census added about 1.3 million people to the rolls using parole and probation records. This is a substantial number: The net undercount in the 1990 census is estimated at four million.

Another substantial increase in coverage would come from using Medicaid records. Children under 18 represented 52% of the undercount in 1990. As of 1996, Medicaid had records on 18.3 million people 20 and under. Many of these young people are in the hard-to-count populations: low-income minorities in dense urban areas. It doesn't take a mathematician to figure out that a single mother struggling to make ends meet might not have time to sit down and fill out a census form. But she will take time to enter similar information on a Medicaid form, because that's time spent on her children's health.

This is a common-sense solution that shouldn't be obscured by bureaucracy, tied up with red tape or shouted down by partisan opposition. Most Americans don't care about partisanship—they care about what works.

Mr. Blackwell is co-chairman of the U.S. Census Monitoring Board and treasurer of Ohio.

The Wall Street Journal
6·26·98

The Florida Times-Union
JACKSONVILLE, FL
Monday, June 22, 1998

CENSUS

Count, don't guess

A federal court hearing degenerated into rhetorical fancies when lawyers argued over the constitutionality of "statistical sampling."

The Clinton administration wants to supplement the census with estimates that would assume the existence of 4 million more people, mostly minorities. That would increase funding for some areas that typically vote Democratic.

Congress is opposed, noting that Article I of the Constitution mandates an "actual Enumeration" — not guesswork.

An administration lawyer told the court enumeration means "numbering up," not a one-by-one head count.

But, as congressional lawyers pointed out, dictionaries at the time defined enumerate as "counting one by one."

Nor are modern dictionaries any more favorable to the administration. Webster's defines enumerate as "1. to count 2. to name one by one."

But, the administration insists, estimates are more accurate.

The Constitution is the supreme law of the land. Therefore, even if that were true, it would be irrelevant. But how could it be more accurate?

As President Clinton once said, "Most people understand that a poll taken before an election is a statistical sample. . . . And sometimes it's wrong."

True. That is why the Constitution requires an "actual Enumeration," not a statistical sample.

June 15, 1998, Monday, JOURNAL EDITION

SECTION: EDITORIAL; Pg. 10A

LENGTH: 418 words

SERIES: Today's News

HEADLINE: CENSUS METHOD IN DEBATE;
Count sampling out

BODY:

APPELLATE JUDGES earn their keep by trying to figure out just what the Constitution says about various legal issues. Jurists of the U.S. District Court in Washington have a new task: trying to decide what the dictionary says, too.

A three-judge panel heard arguments last week in a challenge to the Census Bureau's plan to use a new system of statistical sampling in the 2000 count of America's people, in an effort to account for people missed by the traditional methods --- mailed forms and personal visits by census-takers. An attorney representing the U.S. House of Representatives, which is trying to quash the plan, says that the "actual enumeration" required by the Constitution means counting people one by one; a Justice Department lawyer retorts that it means only "numbering up," by whatever method.

When they get finished with that debate, they can take up angels and the heads of pins. Meanwhile, we'd say they're spending too much time on the wrong argument. Whether the sampling process meets the definition of "enumeration" or not, we believe it's the wrong way to go.

That's not because we think it is mathematically unsound, or that it would necessarily produce inaccurate results. Experts, including the National Academy of Sciences, say it could be a reliable method, one that could at least be nearer a true count than the old methods, which are known to miss millions of people.

Nor is it because, as many Republicans fear, even an accurate count might result in redistricting that would increase the number of Democrats in Congress and in some state legislatures. If that's the truth of where people are and what they believe, so be it.

We object to sampling because, frankly, we just don't think a highly partisan administration can be trusted to be coldly honest when dealing with a process that, by its nature, involves interpretation, estimation and manipulation of figures. The temptation to "help" the results along would simply be too great --- whether the administration was Democratic or Republican.

The Economist magazine, in an article on the use and misuse of statistics, put it as well as we've ever seen it said: "Data tortured long enough will almost always confess."

No matter how sound the theoretical base of the sampling system, it will be human beings --- in the indirect employ of the White House --- who put the numbers in and take them out. Only the most blissful optimist could imagine that politics could be kept out of that process. .

LOAD-DATE: June 16, 1998

The New York Post

June 14, 1998, Sunday

SECTION: PostOpinion; Pg. 073

LENGTH: 842 words

HEADLINE: WOULD YOU BUY A USED CENSUS FROM THIS PREZ?

BYLINE: GEORGE F WILL

BODY:

ON July 11, 1787, the Constitutional Convention discussed what would become the requirement of a census by "actual enumeration" every 10 years, to revise the allocation of congressional seats among the states. The Convention made this the national government's responsibility because, said Edmund Randolph, the states would be too "interested" to be impartial. It made it mandatory because, said George Mason, "those who have power in their hands will not give it up while they can retain it."

In that July 11 discussion, James Madison said there should be a constitutional limit on discretion in the conduct of a census because "all men having power ought to be distrusted to a certain degree," given "the political depravity of man." Speaking of which, President Clinton has found yet another law to disregard, this time a provision in the Constitution.

For the 2000 census he wants to disregard the Framers' stipulation of an "actual enumeration" - actually locating actual people. Instead, he wants the census to be an enumeration supplemented by statistical "sampling" - by more or less sophisticated guesses.

The 1990 enumeration - mailed questionnaires, supplemented by door-to-door interviews - undercounted an estimated 4 million Americans, 1.6 percent of an increasingly large and mobile population. Democrats, who assume that most of the undercounted were poor or otherwise "marginalized" and therefore prospective Democratic votes, propose for 2000 a census that would enumerate 90 percent of the population (down from 98.4 in 1990). The rest would be estimated by complex extrapolations from a survey of 750,000 households (0.75 percent of all households).

The central problem is the political temptations in sampling: Political objectives can shape the assumptions that must be made to frame any formula for making the final extrapolation. Politics can govern the selection of the final 750,000 households who are surveyed, and of the surveyors.

No census is going to be completely accurate. In enumeration, some people will not be found - indeed, some will not want to be found. In sampling, notes David Murray of the Statistical Assessment Service in Washington, accuracy is a function of inherently imprecise processes of "estimation, imputation and probability." So, Murray says, the quest for accuracy must involve one of two risks - an undercount or a "fuzzy count" from sampling.

The assessment of what risks are worth running for improved accuracy occurs in the context of the Clinton administration's contempt for law. Murray sees the birth of a powerful new political temptation: "The ability to "create' or "eliminate' millions of strategically placed citizens with the stroke of a pen introduces a potent and disturbing new political weapon."

No administration, and certainly not this one, should wield such a pen. Michael Barone, author of the "Almanac of American Politics," notes: "This is a White House that had no scruples about getting the Immigration and Naturalization Service to drop criminal checks on applicants for citizenship so that more Democrats could be naturalized in time for the 1996 election; why would it suddenly develop scruples about adjusting census numbers for political purposes?"

Why indeed expect that? This president has written a remarkable record of lawlessness, of three sorts.

One sort arises from his glandular life. A second arises from his lust for money to finance his life's work, campaigning. The third sort, the most serious, involves constitutional damage. It is contemptuous disregard for clear laws governing institutional relations. Three examples:

Clinton could not win Senate confirmation for Bill Lann Lee as assistant attorney general for civil rights, so he has Lee exercising power illegally as "acting" holder of that office. Clinton will not submit for Senate ratification the Kyoto Treaty, which would make America pay for global warming. Rep. John Dingell (D-Mich.) calls this exercise in American masochism "the most asinine treaty I've ever seen." The Senate dismissed it in advance with a 95-0 resolution of disdain. Nevertheless, Clinton proposes to go on to negotiating expansions of the treaty.

In a third derogation of the Senate's right to advise and consent, Clinton, without seeking Senate approval, has extended the scope and changed the substance of the ABM Treaty, a 1972 agreement with a now nonexistent entity, the Soviet Union. He has made several independent republics from the former Soviet Union parties to the treaty. And in the guise of "interpretation," he has made the treaty applicable to theater as well as strategic systems, in a way that makes it an even larger impediment to developing ballistic missile defenses.

Clinton's proposal for census sampling - forever severing this constitutionally mandated exercise from its anchor against politicalization - comes in the context of Clinton's lawlessness. Regarding the undeniable potential for political abuse of sampling, Clinton's position is: Trust me.

No.

LANGUAGE: ENGLISH

LOAD-DATE: June 14, 1998

The Tampa Tribune

June 14, 1998, Sunday, FINAL EDITION

SECTION: COMMENTARY,

Trust the Head Count

President Bill Clinton was stumping in Texas last week for the use of statistical sampling as the Census Bureau prepares for the head count of Americans for the new millennium. Congress should reject such interference. Neither the president nor the Census Bureau have provided assurances that sampling would produce impartial figures for congressional apportionment numbers or federal funding allocations.

The Census Bureau admits that an estimated 4 million people were not counted in 1990. The higher rates of undercounting were among minorities. President Clinton is convinced that these hard-to-count people will continue to be elusive, thus requiring the use of scientific adjustments.

Clinton

Under the Census Bureau's proposed sampling method, the agency would contact, either by mail or in person, a selected sample of households that fail to respond to the census until at least 90 percent of the residents in each census tract are counted. Estimates of the remaining 10 percent would be based on statistical calculations. The final results would be released as a "one-number" census.

The local Census Bureau office assigned additional residents to Detroit by making an arbitrary change in the formula for allocating the projected population growth of Wayne County — and after pressure from the Wayne County government.

This issue understandably has Congress and the Clinton administration at an impasse. Congressional Democrats, who stand to benefit if census results show higher counts in traditionally Democratic urban areas, are supportive of sampling. Republicans are doubtful.

Wrangling over census-taking also has drawn intense scrutiny from such groups as the Southeastern Legal Foundation (SLF), an Atlanta-based public interest law firm that argues that the Constitution requires that census takers literally count, not guess, how many Americans exist. The SLF last year filed a federal court lawsuit questioning the practice.

With good reason. Statistical sampling leaves the census process vulnerable to all sorts of partisan manipulation.

Metro Detroit residents saw evidence of that last year when Wayne County mounted a successful challenge to its estimated population figure, which had the effect of pushing Detroit's estimated population above one million. Data from the Southeastern Michigan Council of Governments had Detroit's population well below the one million mark by the year 2000. However, the local Census Bureau office assigned additional residents to the city by making an arbitrary change in the formula for allocating the projected population growth — and after pressure from the Wayne County government.

Skeptics also have cause to believe statistical adjustments have the potential to complicate and compromise census data. At stake, for example, is how many members of Congress are apportioned to each state. In this instance, adjustments can create people. They can't, however, create income or other demographic data needed to determine a locality's eligibility for hundreds of million of dollars in federal funds for an array of social programs. Using two sets of numbers would create inconsistencies not easily accepted or understood by the public.

No one suggests that any of the available options provide a perfect reflection of America. But Americans have grown to trust and rely on the best and most thorough head count possible.

Detroit News June 7, 1998

Seattle Post-Intelligencer
THE VOICE OF THE NORTHWEST SINCE 1863
Friday morning • June 4, 1999

Founding Fathers figured America needed a census it could count on

Bruce Chapman

When the public ...s evidence of election ...aud – as in the Louisi...a and California con...ressional races last fall – confidence in the very integrity of democracy is ...ventened. But imagine ...hat would happen if ...e whole U.S. statistical ...ystem, including the ...opulation count upon ...hich Congress itself is ...pportioned, and by ...hich billions of dollars ...f federal programs are allocated, became ...uspect. Such a nightmare could happen in ...ess than three years' time. A revolutionary ...lan is before Congress to change the U.S. ...nsus for 2000 from an easily understood ...nd exact count of the population to an arcane ...rocess adjusted by survey samples. A crucial ...nal run of the census is planned for next year.

But the Clinton administration and Congress are still locked in disagreement over the sampling proposal. If a resolution is not reached soon, the results could be questionable population figures, an increase of state and municipal lawsuits over the census process and another drop in public trust in government. The decennial census is a core institution of representative democracy and one of the few federal functions specifically mandated in the Constitution. It is described in Article I as an "actual enumeration." Dictionaries and Congress alike indicate that enumeration means "to count off or name one by one; list" (American Heritage) and "to ascertain the number of; to specify one after another" (Webster's). Yet the Clinton administration believes the Supreme Court will be willing to stretch such a definition to include imputation of fictional persons from survey samples into the census.

It is true that sampling would save money (about $400 million), but the same could be said for elections, where we also demand a hard count. Yet which of us would like to let pollsters elect public officials instead of going through the "expensive" process of an actual election? Sampling also is supposed to provide a more "accurate" population total. The hard count has always failed to include some residents who do not fill out their census forms or who cannot be found. This "undercount" was an estimated 1.8 percent of the national total in 1990, up from 1.2 percent in 1980, but well down from the 5.4 percent, for example, in 1940. The undercount is higher for rural dwellers and inner-city minorities, including illegal aliens, people working odd hours and the homeless. With sampling, it is argued, the adjusted census totals for the nation, states and even counties and most cities could be slightly more accurate. But at the level of small areas such as city blocks, where the virtual people would be fitted, the resulting numbers could be in error by as much as 28 percent (according to one analysis) or even 35 percent (another analysis). It is these small area details that are the stuff of sensitive political redistricting. As with elections, a small number in a specific area can lead to big changes.

Therein lies the prospect of increased lawsuits. With all its human flaws, the census enumeration has stood up to every court challenge. But, survey data representing only statistical abstractions standing in for flesh-and-blood residents could easily be unmasked in a court. Any citizen could investigate and prove his case when the official number of people supposedly living on his block was off by 28 percent or more! Unfortunately, the sampling issue is already politicized. Republicans and states in the East and Midwest that stand to lose if sampling is deployed tend to oppose sampling and adjustment on those grounds, while Democrats and Western and Southern states that stand to gain, tend to support it.

But there is great wariness in Congress, nonetheless, including among Democrats. Most of the states that would gain are in Republican parts of the country (Texas, Florida, Arizona, notably), while predominantly Democratic states that in the past have complained most about the undercount (such as New York) are suddenly discovering from computer simulations that an undercount adjustment actually might hurt them.

Some African Americans are worried – correctly – that once the Census Bureau knows it can rely heavily on samples to improve accuracy in a county or state, it may spend less time, money and effort on outreach in central cities, leaving the base upon which the sample is taken as flawed as today.

In congressional testimony and reports representatives from both parties express concern that when Americans as a whole find out that sampling will take the place of the 1 percent or so of individuals who fail to return the mailed census forms or otherwise cannot be counted, the voluntary census response rate will fall still further – leading to a need for still more sampling.

The U.S. statistical system could come to rest increasingly on the soft base of sample data rather than on the base of hard data that the census has always provided. Subsequent government surveys effectively would be samples of a sample. Computer models show that even national population totals can change dramatically depending upon what assumptions go into formulating samples and adjusting the data. And a small error can lead to major problems. In a simulated adjustment of the 1990 census such an error amounted to a million-person population shift, and had it been implemented, the unwarranted loss of a House seat by Pennsylvania.

Many statisticians remain sanguine about such dangers. But they tend to see the census mainly as a statistical process and not the crucial democratic institution that the American founders intended. If the whole population count becomes another abstract, complicated activity understood only by experts, the census will lose much of its historic role and our public support.

Congress, therefore, should act now, to limit sampling and provide the extra funds necessary for improved outreach and promotion of a standard census in 2000. The alternative means trouble.

■ Bruce Chapman, director of the U.S. Census Bureau from 1981 to 1983, is president of Discovery Institute in Seattle. His column appears each Friday.

Chicago Sun-Times
Dennis Byrne

Constitution goes down for the count

June 3, 1998

BY DENNIS BYRNE SUN-TIMES COLUMNIST

The U.S. Constitution makes no bones about how Americans should be counted: They must be ``enumerated."

It makes clear that congressional representatives must be apportioned among the states, ``counting the whole numbers of persons in each state." Enumerate, says my dictionary, means: count, list, specify one after another.

No where does it mention estimate or guess.

But that is precisely the way President Clinton proposes to conduct the 2000 census. In a ``roundtable" in Houston, Clinton on Tuesday made yet another pitch to toss out the Constitution, and count Americans by ``sampling."

He says it's because too many people--the poor, minorities, urbanites and (most important, but he won't say it) Democrats--are undercounted in the decennial census. How do we know? Did someone actually go out and lay eyes on the supposed millions who have been undercounted? Actually, no. The uncounted haven't been counted; they've been estimated, too.

This, however, is reason enough for Clinton to simply void a part of the Constitution. Might as well also toss out the part about having to be at least 25 and a U.S. citizen for seven years before you can be elected to the House. Why not? I'm sure that within the penumbra, the fringe or the dark side of the Constitution, the Supreme Court eventually will discover that the Founding Fathers also meant that 13-year-olds from Brunei can serve in the House.

Just so we all know what we're fighting about, here is how the census sample would be conducted: Census Bureau reps would call on a ``scientifically selected" sample of households that failed to return their questionnaire, until they contact about 90 percent of the non-responding households in each census tract. From these visits, the bureau would extrapolate the characteristics of the households they couldn't reach, figuring that the missing folks are pretty much like the ones they reached. Then the census counters would go back to 750,000 randomly selected households nationwide to determine whether they were in the initial count. By comparing the two counts, the bureau would determine how many people were missed, counted twice, counted in the wrong place and the total undercount.

Follow that?

What's going on here is an elaborate cooking of the books to come up with inflated numbers that partisans assume--rather, want to--exist. Among them are Democratic Party minions of the poverty, social service, big city political and other self-interest groups that drink at the federal spigot. It's in their interest to hike up the numbers not just to increase their representation in Congress, but to get a larger

share of federal subsidies, from road building to health care, that flow to localities based on their population size and characteristics.

If you think this whole thing is loopy, here's what we might as well expect next: voting by scientific sample. I can imagine the pitch from Clinton or some like-minded compassion warrior: A lot of people would get out to vote, if they weren't poor, ignorant, isolated or alienated. But they, too, deserve a voice. So here's what we'll do. We'll take a sample of those who do show up to vote, extrapolate from that the characteristics of those who don't, then compare the results with a random survey of state voters, and from that determine whom the no-shows would have voted for.

This is no more stupid than asserting that we should likewise guess the size and characteristics of the U.S. population. See, nowhere does the Constitution talk about counting groups; it calls for the counting of individuals, person by person.

So what about those who are assumed to be missing and uncounted? If they exist, let's not rule out the possibility that they're not being counted because they don't want to be--because they're too lazy to return the questionnaire, or because they're hiding. Poor things, they have to hide, because they don't trust the government, or they're here illegally, blah, blah or blah.

So here's the deal: Tell all the ``leaders'' and politicians who are pushing for a censguess, that if they want their ``group'' counted and their ``interests'' represented, they should make sure their constituents send back the questionnaire, answer the door, or--gosh, imagine this--take it upon themselves, if it looks as if they'll be missed, to be counted.

Dennis Byrne is a member of the Sun-Times editorial board. E-mail: dbyrne @suntimes.com

Back to Byrne Page

THE NEW YORK POST - 4/12/98

EDITORIAL

PLAYING POLITICS WITH THE CENSUS

President Clinton and the Democrats have been blissfully reassuring for months about their plans to institute statistical sampling in the upcoming 2000 census instead of the actual head count that the Constitution has mandated since the first census in 1790.
Republicans and conservatives point out that a painstaking person-by-person count is one of the few specific obligations the Founding Fathers assigned to the federal government (the words the Constitution uses are "actual enumeration").

This is more than a difference of opinion; it is a matter of utmost importance. The purpose of the census is to determine Congressional representation, and so an honest and complete count of American citizens goes to the heart of the American democratic experiment. It's vital that it not be politicized.

Statistical sampling is necessary for the sake of accuracy, Clinton and the Democrats say. An attempt to gain partisan advantage? Heaven forbid!

So why, then, did the Democrats last week appoint as one of their four overseers of this tricky new process one of the most partisan Americans of the second half of the 20th century - former Congressman Tony Coelho? He masterminded a brutally cynical (and successful) effort, beginning in 1982, to force corporate America to underwrite the Democratic Party in the 1980s and early 1990s. It would be no exaggeration to say that Coehlo's efforts delayed the GOP takeover of the House for more than a decade.

That this hyper-partisan (who resigned from office under a cloud in 1989) should be put in so sensitive a position on a matter of such gravity should set alarm bells a-ringing.

The GOP now has hard evidence to back up its concern that Democrats secretly intend on using the 2000 census - and the sampling process they are trying to deploy on the Q.T. without a national debate - as a means of winning back control of the House of Representatives.

It works like this: The areas that are the hardest to count by traditional methods -filling out census forms and paying house visits - are large cities like New York. They have substantial numbers of transient residents and a large population of poor and uneducated people. It is these people who are supposed to be "found" via sampling estimates.

Sampling is the Democratic strategy for increasing the proportion of House seats concentrated in core urban areas at the expense of more Republican suburban and rural residents. Republicans fear that sampling could cost them as many as 22 seats in the House.

But wait. If that is the case, won't New York gain if sampling comes into effect? The answer is: not necessarily.

The House Subcommittee on the Census has done some preliminary studies of what the nation can expect if sampling is the method the Census Bureau uses to conduct the 2000 count. While the city itself (or, more accurately, all the boroughs except Staten Island) would be credited with a larger population number under sampling, the statistical theory underlying the method will actually subtract people from places like Nassau and Suffolk Counties on Long Island and rural areas upstate whose residents are meticulous about filling out census forms.

While the city might gain, other areas of the state will lose proportionately. And that will have a ripple effect when it comes to state-level funding as well.

A few Republican members of Congress, notably Connecticut's Christopher Shays, have been seduced into going along with the Democratic strategy because they apparently believe sampling will help the northeastern part of the country. This is ironic given that its likely effect will be to punish areas of high compliance with the census, like Shays' own Fairfield County.

House Speaker Newt Gingrich is suing the Clinton administration to force it to hew to the Constitution's demand of "actual enumeration" and current federal law, which forbids the use of sampling.

If people like Nassau County Executive Thomas Gulotta and Suffolk County Executive Robert Gaffney want to file friend-of-the-court briefs to stop this travesty that will hurt their budgets, they have *until* May 4 to do so.

END

THE NEW YORK TIMES **OP-ED** SUNDAY, DECEMBER 7, 1997

Essay

WILLIAM SAFIRE

Sampling Is Not Enumerating

WASHINGTON

As elections demonstrate, a poll is an educated guess and not a hard count. A sampling is not an enumeration. *Often pollsters are mistaken.*

Here's how polling warps politics. In the 1996 Presidential campaign, most major polls showed Bob Dole trailing throughout by a whopping 12 to 20 percentage points, throwing the G.O.P. into despair. Even right-wing publications hooted at their "sure loser"; *dispirited Republicans* stayed home in droves.

Toward campaign's end, the New York Times/CBS poll — run by the most respected statistical geniuses and random samplers alive — showed Bill Clinton ahead by a likely landslide of 18 points. ABC and NBC/ *Wall Street Journal* said 12. But one poll — the Zogby poll for Reuters — was out of step. It showed only a seven-point gap.

On Election Day, the actual enumeration showed Zogby alone to be within one point of accuracy. The other polls that made the Dole campaign a laughingstock in the media — and helped confer a false inevitability to Clinton's re-election from the start — had been grievously misleading. (In the Times/CBS case, fully 10 points off.)

Thus can reliance on samples distort our politics. Sampling is no science; ask President Dewey and Prime Minister Peres.

Democrats want to gain a political edge in 2000 by changing our method of counting American noses. This flies in the face of the U.S. Constitution, which in Article I calls for an "actual Enumeration," with a capital E —

Time to cut the non-census.

which means "counting one by one."

Democrats led by President Clinton say the founding fogies did not have available the blessings of modern sampling and their stricture about counting everybody can be ignored. Liberals want to replace, or "augment," laborious counting with the educated guesswork of sampling.

Reason: Census takers don't like to climb five stories in crummy neighborhoods, where residents are more likely to vote Democratic. This assumes most of the people who hide from Government bell-ringers, or don't have phones, or can't read their mail, or are recent immigrants, are likely to be Democrats.

That stereotype strikes me as insulting. But the Democratic minority in Congress, backed up by the White House, insists Democrats have been getting a "short count" by letting the Census Bureau do it the constitutional way. By having a statistician put a thumb on the scale, liberals figure they can pick up a dozen House seats and increase spending on the poor.

Was the 1990 Census accurate? Probably not; sloppiness in planning, fearfulness about interviewing and poor mail service failed to count millions — including libertarian moonshiners in the Blue Ridge Mountains and Republicans in Beverly Hills mansions, all threatening armed response.

To do better next time, Democrats want to (1) do another slapdash nose count; (2) redo selected slums with a vengeance, and (3) extrapolate those redone samples to skew — or "weight" — the earlier count. An oversight committee (Charlie Trie, chairman?) would watch out for political manipulation. This is supposed to increase people's trust in government.

The right answer is to improve enumeration before 2000: (1) advertise to reassure the reluctant and gain their cooperation; (2) improve mailing lists to reach everyone with an address, and (3) train census foot soldiers to make them more effective in finding the homeless.

The wrong answer is the cockamamie compromise that Clinton and Republican leaders have just perpetrated: a bill passed with a straight face by the Congress, solemnly signed by the President, requiring the Supreme Court to rule in advance on the constitutionality of sampling.

Here's what will happen. The Justices will respond unanimously to the legislative and executive branches: "Whaddaya, out of your minds? Didn't we just teach you on the legislative veto that we're not in the business of giving advisory opinions? We will decide when an issue is ripe for our decision and who has standing, so take your silly bill and gedoudahere." (Jurists call this "denying cert.")

Justice Hugo Black read the line in the Constitution that begins "Congress shall make no law" and opined "No law means no law." In the same way, "actual Enumeration" means what it says. Start counting. This sample is not for sale. □

Mr. MILLER. Let's begin. Dr. Koyak, if you would begin.

STATEMENTS OF ROBERT KOYAK, ASSISTANT PROFESSOR OF OPERATIONS RESEARCH, NAVAL POSTGRADUATE SCHOOL; LAWRENCE BROWN, STATISTICS DEPARTMENT, THE WHARTON SCHOOL, UNIVERSITY OF PENNSYLVANIA; AND MARTY WELLS, PROFESSOR OF ECONOMIC AND SOCIAL STATISTICS, CORNELL UNIVERSITY

Mr. KOYAK. Thank you. I want to thank Chairman Miller and the members of the Census Subcommittee for inviting me to testify today on issues facing the 2000 census. The views that I express are my own and do not reflect official policy or position of the Naval Postgraduate School or the U.S. Department of Defense.

When I was employed by the Justice Department, I had the opportunity to work on the defense of Commerce Secretary Mosbacher from lawsuits arising from his decision not to adjust the 1990 census. Although his decision was controversial at the time, there is less criticism of his decision today. Proponents of a sampling-based 2000 census are not eager to revive the memories of 1990. When they do, it is usually in reference to how they will do it better the next time.

Sampling in the 2000 census will be very different from that used in 1990. The integrated coverage measurement [ICM], will encompass more than 750,000 households, about five times as many as the post-enumeration survey. To accommodate the burden of a larger ICM sample, the plan for the 2000 census also calls for the use of non-response follow-up [NRFU] sampling. With NRFU sampling, close to 10 percent of the U.S. population will not be accounted for directly by the census. Instead, they will count only as estimates derived from other households' data.

The purpose of ICM sampling is to improve census accuracy by removing the undercount. In contrast, the purpose of NRFU sampling is to save time and money. NRFU sampling is a gambit to reduce the accuracy of the census up front, in order to free enough time and resources so that NRFU and ICM sampling together can increase census accuracy in the balance.

In my opinion, it is a risky gambit. Even without NRFU sampling, doubt has been raised about the accuracy of an ICM-based undercount adjustment. Statisticians have testified before this subcommittee that the ICM will not correct the fundamental flaws of the PES. If this occurs, the undercount adjustment will largely reflect errors in the ICM rather than the undercount itself. Thus, the only certainty in the gambit is that NRFU sampling will add error to the 2000 census. On this point, there is no disagreement.

Under NRFU sampling, the nearly 10 percent of the population with no census information will have other data substituted for them. The nearest sample household will donate its data to a nonsampled household. This is known as imputation. How much sampling error imputation will add to the 2000 census is uncertain. To my knowledge, the Census Bureau has not conducted an extensive evaluation to answer this question.

Like the ICM, imputation will rely on a form of matching. More than 10 million households will have to be correctly matched to their nearest donors. Errors in the master address file, in

geocoding, or in the matching algorithm could lead to large imputation errors. Experience with matching the 1990 PES suggests that the difficulties may be larger than anticipated.

NRFU sampling will also aggravate problems with the ICM. This is because the census blocks and the ICM will not be subject to NRFU sampling, but counted with 100 percent followup of non-respondents. As a result, the base census in ICM blocks will not be the same as the base census in non-ICM blocks. This poses the risk that the undercount may be fundamentally different in both types of census blocks. If this happens, an undercount adjustment based on the ICM will not be valid for the Nation in general.

NRFU sampling will complicate the treatment of movers in the ICM. People who move out of the ICM blocks after the base census will still be counted in the ICM. This is because NRFU sampling will make it too difficult to use in-movers, as was done in 1990. But there is a problem. If someone moves out of an ICM block and does not match to the base census, how will you know that they exist? This information will have to come from family members, neighbors, and other indirect sources which will lead to unreliable data, ICM matching errors, and possibly an inflated undercount adjustment.

A gambit is an active calculation. What is the evidence that the NRFU and ICM sampling used together will produce a more accurate 2000 census than traditional enumeration? I find it is disturbing that a credible comprehensive evaluation of the plan for the 2000 census has yet to be seen. In the absence of such evidence, I believe that the 2000 census is likely to repeat, if not magnify, the mistakes of the past. I thank you for this opportunity and I gladly yield to questions afterward.

[The prepared statement of Mr. Koyak follows:]

Statement of Robert A. Koyak, Ph.D.
Before the U.S. House Subcommittee on the Census
September 17, 1998

I want to thank Chairman Miller and the members of the Census Subcommittee for inviting me to testify today on issues facing the 2000 Census. The views that I express are my own, and do not reflect the official policy or position of the Naval Postgraduate School or the U.S. Department of Defense.

When I was employed by the Justice Department, I had the opportunity to work on the defense of Commerce Secretary Mosbacher from lawsuits arising from his decision not to adjust the 1990 Census. Although his decision was controversial at the time, there is less criticism of his decision today. Proponents of a sampling-based 2000 Census are not eager to revive the memories of 1990. When they do, it is usually in reference to how they will do better the next time.

Sampling in the 2000 Census will be very different from that used in 1990. The Integrated Coverage Measurement, or ICM, will encompass more than 750,000 households, about five times as many as the Post-Enumeration Survey. To accommodate the burden of a larger ICM sample, the plan for the 2000 Census also calls for the use of nonresponse follow-up, or NRFU[1], sampling. With NRFU sampling, close to ten percent of the U.S. population will not be accounted for directly by the census[2]. Instead, they will count only as estimates derived from other households' data.

The purpose of ICM sampling is to improve census accuracy by removing the undercount. In contrast, the purpose of NRFU sampling is to save time and money. NRFU sampling is a gambit, to reduce the accuracy of the census up-front, in order to free enough time and resources so that NRFU and ICM sampling together can increase census accuracy in the balance.

In my opinion, it is a risky gambit. Even without NRFU sampling, doubt has been raised about the accuracy of an ICM-based undercount adjustment. Statisticians have testified before this Subcommittee that the ICM will not correct the fundamental flaws of the PES[3]. If this occurs, the undercount adjustment will largely reflect errors in the ICM rather than the undercount itself. Thus, the only certainty in the gambit is that NRFU sampling will add error to the 2000 Census. On this point there is no disagreement.

Under NRFU sampling, the nearly ten percent of the population with no census information will have other data substituted for them. The nearest sampled household will "donate" its data to a nonsampled household. This is known as "imputation." How much sampling error imputation will add to the 2000 Census is uncertain. To my knowledge, the Census Bureau has not conducted an extensive evaluation to answer this question[4].

Like the ICM, imputation will rely on a form of matching. More than ten million

households will have to be correctly matched to their nearest donors. Errors in the Master Address File, in geocoding, or in the matching algorithm, could lead to large imputation errors. Experience with matching in the 1990 PES suggests that the difficulties may be larger than anticipated[5].

NRFU sampling will also aggravate problems with the ICM. This is because the census blocks in the ICM will not be subject to NRFU sampling, but counted with 100% follow-up of nonrespondents. As a result, the base census in ICM blocks will not be the same as the base census in non-ICM blocks. This poses the risk that the undercount may be fundamentally different in both types of census blocks[6]. If this happens, an undercount adjustment based on the ICM will not be valid for the nation in general.

NRFU sampling will complicate the treatment of movers in the ICM[7]. People who move out of the ICM blocks after the base census will still be counted in the ICM. This is because NRFU sampling will make it too difficult to use in-movers, as was done in 1990[8]. But there is a problem: if someone moves out of an ICM block and does not match to the base census, how will you know that they exist? This information will have to come from family members, neighbors and other indirect sources, which will lead to unreliable data, ICM matching errors, and possibly an inflated undercount adjustment.

A gambit is an act of calculation. What is the evidence that NRFU and ICM sampling, used together, will produce a more accurate 2000 Census than traditional enumeration? I find it disturbing that a credible, comprehensive evaluation of the plan for the 2000 Census has yet to be seen[9]. In the absence of such evidence, I believe that the 2000 Census is likely to repeat, if not magnify, the mistakes of the past.

---

1. In this context, a nonrespondent refers to a household that is not accounted for in the census through the mailback of a census form, or any other acceptable means (e.g. telephone) of household-initiated transmittal of census information. Follow-up refers to activities undertaken by the Census Bureau, such as direct contact by Census personnel, to obtain census information from nonrespondent households.

2. The plan for the 2000 Census calls for NRFU sampling to be conducted on one-third of nonrespondent households, or enough to bring the total response rate up to 90%, whichever is greater, in every census tract of the United States. A census tract is a unit of census geography averaging about 4,000 persons. If the anticipated 67% mailback rate is achieved, most census tracts will only reach the 90% target, and the response rate nationwide will be slightly higher than 90%.

3. Research conducted by Professor Leo Breiman (*Statistical Science*, **9**, 458–475, 1994) found that nearly 80% of the undercount adjustment based on the 1990 PES consisted of bias due to data problems. The Report of the Committee on Adjustment of Postcensal Estimates

(C.A.P.E Report, Bureau of the Census, August 7, 1992) reached a similar conclusion.

4. The method in reference is specifically referred to as "hot deck imputation." A Census Bureau memorandum dated December 10, 1997 presented accuracy estimates for the use of hot deck imputation and for three other imputation methods. The estimates came from a computer simulation study based on 1990 Census data from the West Sacramento, CA district office, which raises questions about the applicability of the results of the study to the entire nation. The memorandum stated that "future work will include an increased number of simulations to ensure a complete comparison of the estimation methods." On January 28, 1998, Census 2000 Decision Memorandum No. 41 announced the decision to use hot deck imputation in the 2000 Census.

5. Leo Breiman, *op. cit.*

6. One of the dangers is that the 100% nonresponse follow-up base census and ICM together will create severe time and resource management pressure. If the base census in ICM blocks adopts economizing measures as a result of this pressure, it is at risk of not being comparable to the base census in non-ICM blocks. Similarly, if the base census and the ICM are allowed to overlap in order to save time, the independence of the base census and the ICM may be violated. Independence of the base census and the ICM is a fundamental assumption underlying the undercount adjustment methodology.

7. Due to the mobility of the U.S. population, the treatment of movers will necessarily be a significant component of the ICM. According to the Census Bureau, about 16% of the U.S. population moved between March 1996 and March 1997 (1.3% per month). For groups of special interest to the undercount issue, the mobility rate is even higher. About 33% of renters moved during the same one-year period (2.8% per month), as did 32% of people 20-to-29 years old (2.7% per month) (U.S. Census Bureau, "About 42 Million Americans Moved Between 1996 and 1997," press release dated July 23, 1998). With the ICM following the base census one to three months afterward, a significant turnover in population will occur between the two counts.

8. In the 1990 PES, the problem of movers was handled with what is known as Procedure B. Procedure B used all PES-sampled persons as the base population for matching to the census: those that lived in a PES block on April 1, 1990, and those that moved in afterward (in-movers). Those that moved after April 1, 1990 (out-movers) were not included in the PES. In contrast, the plan for the 2000 Census currently calls for the use of what is known as Procedure C. Procedure C will use as its base population only those persons that had lived in the ICM blocks on April 1, 2000. In-movers will be excluded because many of them will enter from non-ICM blocks, where they may not be counted due to the use of NRFU sampling. Instead, out-movers will be included in the ICM under Procedure C.

In addition to the conceptual problem of identifying out-movers, Procedure C multiplies the match rate for out-movers by the ratio of total in-movers to total out-movers. This calculation relies on an assumption that the undercount rates for out-movers and in-movers are

comparable. For example, people who move out of an economically disadvantaged inner city census block in Philadelphia are assumed to be similar, with regard to undercount, as those who move in.

9. To date, the only large-scale evaluation of error rates that incorporates both NRFU and ICM sampling that has been published is a computer simulation based on the 2000 Census plan as of June 1997 (General Accounting Office, "2000 Census: Progress Made on Design, but Risks Remain," GAO/GGD-97-142, July 1997). The error rates were included in the Census Bureau's Report to Congress (revised and reissued August 1997). Since that time, the 2000 Census plan has undergone a number of significant changes (e.g., the decision to use hot-deck imputation for NRFU nonsampled households). Also, the error rates do not take into account the various sources of bias that affected the PES in 1990, and that are likely to affect the ICM in 2000. The Report to Congress acknowledged these sources of bias, also known as *nonsampling errors* (pp.41–43), but did not include bias in its error calculations for the 2000 Census plan (p. 44).

Mr. MILLER. Thank you. Dr. Brown.

Mr. BROWN. Thank you, Congressman Miller and other Members of Congress.

Mr. MILLER. Excuse me. I meant to ask each of you before to state where you are, and just a very brief statement of why you are here as far as your experience on census.

Would you like to do that before Dr. Brown begins?

Mr. KOYAK. Yes. I am assistant professor of operations research at the Naval Postgraduate School. My experience with the undercount issue was as an employee of the Anti-Trust Division of U.S. Department of Justice in the early 1990's, where I had the opportunity to assist the civil division in its defense of Secretary of Commerce Mosbacher and the lawsuits that arose from his decision not to adjust.

Mr. BROWN. I am professor of statistics at the Wharton School of Business at the University of Pennsylvania, former president of the Institute of Mathematical Statistics, a member of the National Academy of Sciences and several national research council committees.

There has been a lot of discussion here and elsewhere about the undercount. I think there's no question that the traditional enumeration methods leave an undercount. However, once that undercount is recognized, its existence has almost no political or economic consequences.

The political and economic consequences related to the census involve accuracy in the State and local proportions of the national total. Getting more accuracy in population shares of State and local shares of the national total then is provided by the traditional census enumeration methodology is an extremely difficult task. It's maybe beyond the limits of possible.

Let me try and pin this down a little bit. The 1990 PES, post enumeration survey that we have been hearing about, attempted to correct inaccuracies in the enumeration. The corrections are at the State and local levels in terms of proportions, but the median change in State share percentages was 0.008 percent. The average change was a little bit larger. That is very, very small on anybody's scale of accuracy. In other words, in terms of shares, the 1990 enumeration did a fantastically good job, in spite of missing some, in spite of its undercount, gross undercount or net undercount.

How well did that PES really do in terms of accuracy in shares? Well that's a very difficult issue which I have looked at and many other statisticians. It is very hard to judge.

I want to make only two essentially non-technical observations at this point. The first you have heard about. That is, that the actual analysis of the PES data had to be iterated several times in several variations, before acceptable results were derived. I find that an unacceptable state of affairs. If the census process is to be acceptable, the complete correct protocol really needs to be laid out before the data is gathered. That was not done before.

As another point which you have also heard about, there is an alternate independent type of analysis referred to as demographic analysis that provides widely accepted total figures for the overall United States, though they can't be broken down into State shares. But they can be divided into several other kinds of categories, for

example, sex and race. Although the race category is just black, non-black, and not the several categories of the current census.

The 1990 PES did not do a particularly good job of estimating the actual U.S. shares in these four categories, which are a relatively easier job than estimating State and local shares. Even more worrisome, correct figures for these four categories are an essential component of the entire PES process, so errors here are likely to be magnified elsewhere by the PES process.

I find it doubtful that the ICM in 2000 can perform better than the PES did in 1990. There is one change that in principle should improve the 2000 ICM, which is a similar process. That is the increase in sample size. But the 2000 ICM also involves a number of other changes which act against this improved accuracy. Some of these you have heard about already.

There is another major statistical component in the current plans for 2000. That's as you have just heard, the sampled non-response follow-up. I find this portion of the plan nearly as worrisome. Under this plan, households that don't return their initial questionnaires will be followed up on a sample basis rather than a 100 percent basis as in previous censuses. The sampling rule is designed so that approximately 90 percent of all households will eventually either respond to the original questionnaire or be contacted in the followup portion.

It could be that if nothing else impedes the accuracy of SNRFU, as it's acronymically called, it would do an acceptable job. However, that's a very big if. Many of these 3,000 elements of the process, and I had never done a count of elements of statistical process, but there are many small elements of this SNRFU process, all of which or most of which need to be essentially correct in order for this process to work satisfactorily. Problems in the SNRFU process, most of them would not be repaired by the presence of an ICM correction. On the contrary, most of them would be magnified by it.

What concerns me the most of all, however, is that as I see the planning for these statistical components, SNRFU and ICM, they are very tardy, very behind the schedule that it should be, for various reasons. Sort of theoretical calculations, textbook calculations don't answer the kinds of questions that need to be answered. What really matters is what happens in the field under actual census conditions. Methodologies which are to work under these conditions have to be created by looking at data from suitable field trials, and then should be separately validated. So the Census Bureau has what seems to be a process for doing that. In 1995, they had what were called test censuses. This year they carried out a dress rehearsal.

The term dress rehearsal makes it sound like you have got it all nailed down and you just need one more run-through to get out the final kinks and some last minute jitters. But in fact, this was a dress rehearsal after which large chunks of the script are still being written when it comes to both SNRFU and ICM, and there's no further opportunity for any sort of a real rehearsal or test before the public performance, before 2000. Under these conditions, you would need at best a very large amount of good luck for the process to work satisfactorily.

So at its best, planning for both SNRFU and ICM is well behind what it needs to be. As now contemplated, even a traditional enumeration will need to overcome new obstacles provided by more transient contemporary patterns of life and by such census innovations as the duplication and broad availability of census forms. Planning in the last 18 months, before the 2000 census, should concentrate on implementing a mostly traditional enumeration which will satisfactorily cope with the complexity inherent in any U.S. census in general and with these new additional obstacles in particular.

[The prepared statement of Mr. Brown follows:]

Testimony to be Presented to the House of Representatives
Subcommittee on the Census
Sept. 17, 1998

Lawrence D. Brown
Meiers Busch Professor of Statistics
Wharton School,
University of Pennsylvania,
Philadelphia, PA 19104-6302

CORRECT POPULATION SHARES ARE THE GOAL

First, there is no dispute that traditional census methodologies leave a noticeable undercount in the gross population totals, and this undercount is not distributed uniformly across the various geographic and demographic strata of the country. However, correct gross population totals are not the primary target of the decennial census. What is of primary importance is accuracy in the state and local population proportions of the national total as determined from the census.

IT'S AN EXTREMELY DIFFICULT TASK

Getting more accuracy in population shares than is provided by the traditional Census methodologies is a fiendishly difficult task which is at or beyond the limits of the possible. The 1990 PES (Post Enumeration Survey) attempted to accomplish this task. The corrections it proposed to population shares such as state shares were extremely small. The median change in state-share percentage was only $\pm$ 0.008%.[1]

HOW WELL DID THE 1990 PES REALLY DO?

This issue has been hotly debated in the statistical literature and elsewhere. No broadly accepted answer has emerged as to the accuracy of the 1990 PES corrections in terms of state and local population shares. It's probably impossible to definitively judge whether the proposed PES changes to the 1990 enumeration would have improved or degraded the overall accuracy of state and local population shares. This is a difficult and partly quite technical issue. Here I will make only two pertinant, non-technical observations:

(i) The first and most straightforward analysis of the PES data gathered in 1990 produced some implausible and hence unacceptable answers. Generally plausible answers were only produced after several reanalyses of the data using various technical data-analytic methodologies and options. Even then, the official version of the 1990 PES correction, as submitted for consideration by the Secretary of Commerce, contained at least one major error, and was open to reanalyses that arguably improved its results[2]. This state of affairs seems to me unacceptable; the complete protocol for both the sampling and the analysis of the resulting data must be laid out in advance of the census process.

(ii) An alternate independent type of analysis, referred to as Demographic Analysis provides widely accepted total figures for the overall U.S. population as divided into four categories by sex and race (Black or Non-Black). The 1990 PES did not do a particularly good job of estimating the actual U.S. shares of these four categories.[3] Even more worrisome, correct figures for these four categories (and several others) are an essential component in the entire PES process for correcting the basic Enumeration reults about state and local population shares. Errors in the share estimates for these categories have the potential to create much more drastic errors in the resulting figures for state and local shares.

IT'S DOUBTFUL THAT ICM IN 2000 CAN PERFORM BETTER THAN DID PES IN 1990:

The planned 2000 ICM involves a post enumeration sample size five times larger than that in the 1990 PES. Considered in isolation that increase could be expected to significantly improve its accuracy. However the 2000 ICM also involves a number of other changes which act against this improved accuracy. Most of these were adopted for legal, political, or logistic reasons in spite of technical statistical considerations showing that they tend to degrade accuracy in the estimates of population shares.[4]

THE PLANS FOR SNRFU IN 2000 ARE ALSO PROBLEMATIC

The statistical sampling plans for 2000 also include a plan for Sampled Non-Response Follow-Up. Under this plan those households that do not return their original questionnaires will be followed up on a sample basis, rather than on a 100% basis as in previous censuses. The sampling rule among those nonrespondents is designed so that approximately 90% of all households will eventually either respond to the original questionnaire or will be contacted in the FollowUp portion of the census. Textbook style calculations make it appear that this 90% rule could be sufficient to provide the desired accuracy in state share estimates IF nothing else impedes the accuracy of SNRFU. However, that's a big IF. There are many apparently small but actually important factors which could significantly degrade the accuracy of the SNRFU process. These problems would not be repaired by the presence of the ICM correction, and some of them could be significantly magnified by it.[5]

THE PLANNING FOR SNRFU AND ICM IN 2000 IS TARDY

The decennial census is an immensely large and complex task. Every small detail of every step of this complex process must be carefully and correctly planned and executed in order for the result to be acceptably accurate and fair in all respects. As you've heard, the plans for Census 2000 involve inserting two separate but interacting large scale statistical samples into the core of the traditional process. This adds additional orders of magnitude to the difficulty of planning and carrying out the analysis of the resulting data.

Textbook style calculations may help but they do not tell the whole story, no matter how advanced and sophisticated the textbook. What really matters is what happens in the field under actual census conditions. Methodologies which are to work under these conditions must be created

by looking at data from suitable field trials. They should then be validated -- and fine-tuned if necessary -- by testing them on additional, confirmatory field trials. Census Bureau terminology suggests just such a developmental and then confirmatory procedure. There were "Test Censuses" in 1995, and this year's field trial was called the "Dress Rehearsal". The term, "Dress Rehearsal" suggests that the performance has been completely rehearsed, and only one more run through is needed to get out the final kinks and overcome last day jitters.

In fact, this years trials were really designed as two separate field tests - one of them (in Sacramento, CA) was a test of the SNRFU-ICM combination and the other (in Colombia, SC) was primarily a rehearsal for a modernized version of the more traditional full enumeration methodology. Neither of these "rehearsals" was sufficiently large or broadly representative so as to be confirmatory trials at the necessary level of accuracy. So, this was a "Dress Rehearsal" after which large chunks of the script are still being written when it comes to SNRFU and ICM. There is no further opportunity for a real rehearsal before the public performance. Under these conditions a valid concept, cleverness, skill, hard work, and also a very large amount of good luck would be needed for such a complex enterprise as Census 2000 to succeed.

CONCLUSION: 2000 IS TOO SOON FOR SNRFU AND ICM

Statistics can be a valuable tool. Some limited statistical techniques will be useful, perhaps even essential, in obtaining acceptable enumeration results in 2000. However, improving on the accuracy in shares provided by the census in 1980 and 1990 is a complex task that may be beyond all capabilities of modern statistics. At best, planning for both SNRFU and ICM is well behind what it needs to be in order for these methodologies to stand a reasonable chance of success at the job of improving on the enumeration figures for state and local shares. As now contemplated, even a traditional enumeration will need to overcome new obstacles provided by transient contemporary patterns of life and by such census innovations as the duplication and broad availability of census forms. Planning in the last 18 months before the 2000 census should concentrate on implementing a mostly traditional enumeration which will satisfactorily cope with the complexity inherent in any U.S. Census in general, and with these aditional new obstacles in particular.

---

[1] For the data leading to this figure, see Wachter, K. "The Census Adjustment Trial: An Exchange" in *Jurimetrics* v. 34 (1993) p. 107 -115.

[2] Hogan, H. "The 1990 Post-Enumeration Survey:Operations and Results" *Jour. Amer. Statist. Assn.* v.88 (1993), p.1047-1060 contains such a reanalysis. This article suggests very significant changes to the state proportions promulgated in the July 1991 PES. In 1990 Hogan was Chief of the Undercount Research Staff of the Bureau of the Census. He is now the "Acting Chief of the Bureau's DSSD."

Other statisticians have proposed even more extreme reevaluations of the PES figures; see for example other articles in the same issue of the *Jour. of the Amer Statist Assn.*, and also "Adjusting the Census of 1990: The Smoothing Model" by D. A. Freedman, et. al. in *Evaluation Review* v.17 (1993) p. 371 - 443, and also Breiman, L. "The 1991 Census Adjustment: Undercount or Bad Data" in *Statistical Science* v.9 (1994), p. 458 - 475, with accompanying discussion.

[3] According to the 1991 PES correction, the U.S. population is 5.87% Black males. This compares to the Demographic Analysis figure of 6.10%. There is thus a disparity of +.23%. Other categories are, respectively, Black females: 6.51% vs 6.44% (disparity = -.07%); Other males:

43.05% vs 43.12% (disparity = +.07%); and Other females: 44.57% vs. 44.34% (disparity = -.23%). The average magnitude of these disparities is thus .15%. In the sense referred to in the first part of my comments, the PES is aiming to correct disparities in state shares of the order of .008%. But the error it seems to be making in the placement of population into these four basic race by sex categories is 20 times larger.

The figures here are from the Census Bureau and from the 1991 report of the U.S. Dep't of Commerce concerning the adjustment of the 1990 census.

[4] Among the key changes I have in mind is the decision to make "state" a primary stratifying variable rather than "region of the country." This has the effect of increasing the number of post-strata to be used in the ICM analysis and therefore increasing the variability of its population estimates. This decision may be necessary from Constitutional and other legal considerations, but has the apparent effect of approximately cancelling the benefit of increased sample size in the ICM.

It also has the effect of limiting flexibility in the choice of other stratifying variables, which may well increase problems related to heterogeneity within strata as discussed for 1990 in Freedman, D. A. and Wachter, K., "Heterogeneity and Census Adjustment for the Intercensal Base" in *Statistical Science* v.9 (1994), p. 476 - 485, with discussion. (For example current plans for post stratification in the ICM do not involve separation of large urban, small urban and rural areas, as was done in 1990.)

In the plans for 2000 there is an additional race-origin category used in the post stratification - namely, "Non-Hispanic Hawaiian/Pacific Islander". This may be desirable from a political or legal standpoint, but the inclusion of an additional category whose numbers are so small nationwide is inadvisable from the perspective of statistical accuracy.

The Census Bureau appears committed to avoiding extensive, sophisticated ad-hoc smoothing methods like those adopted in 1990. This should have the desirable effect of increasing perceived objectivity of the process, but -as in 1990 - should be expected to lead to some highly variable and erratic population and share estimates.

Logistic and cost considerations compel a variety of other modifications. To quote from a manuscript currently under preparation by myself and several co-authors, "The plan to conduct initial matching on site during ICM could make it easier to resolve [ambiguous or contradicting] responses by questioning the respondents. On the other hand, enumerators may prove less adept at matching than the specialized matchers of 1990. The ICM fieldwork will be done closer to census day which should reduce the number of bad census day addresses given by respondents. On the other hand, avoiding interactions between census and ICM operations will be harder. The capture recapture estimates used in ICM, of course, assume independence; that independence will be harder to maintain. .... In terms of its plan for handling "movers [those who change addresses between census day and the ICM] the ICM design is more problemmatic than that in the 1990 PES." (The 1990 design, which had problems of its own in this respect, cannot be used because of the presence of SNRFU in the current plans for 2000.)

[5] In 2000 census forms will be widely available (in banks, convenience stores, etc.) and census information can also be provided to a central telephone number. These innovations could potentially lead to difficulties under any census scheme in 2000, but the difficulties are likely to be more severe under SNRFU than under the traditional full NRFU. Under SNRFU the status of possibly duplicate returns must be accurately resolved at the time the NRFU population is identified. This leaves a very short time frame for such unduplication. Also, late returns to the census from people residing in households which were not designated to be sampled in SNRFU would create further problems. A scheme to handle such returns is proposed in a Census Bureau working paper by H. Hogan, "Sampling and Estimation in Census 2000 and the Dress Rehearsal". Validity of this scheme is based on an assumption which appears to me and some of my colleagues to be implausible.

Handling returns from housing units not in the Master Address File could be even more problematic. (According to a 1998 U.S. GAO report indications are that the MAF for 2000 is

incomplete and inaccurate, with up to 6% of the housing units missing in some counties and up to 10% duplications in others. This is quite troubling. If uncorrected it would cause significant inaccuracies in the census no matter what census methodology is used. It is unclear whether these inaccuracies would be reduced or further magnified under SNRFU and ICM.)

The SNRFU - ICM combination involves other questionable assumptions such as that census coverage will be the same whether non-response followup is done on a sample basis or a 100% basis and that residents of ICM blocks (where there will be 100% followup) do not change their behavior as a result of being interviewed more than once.

Mr. MILLER. Thank you.

Dr. Wells.

Mr. WELLS. I am Martin Wells. I am at Cornell University in the Department of Social Statistics. I have worked on various large data projects, the National Capital Jury Project and various projects involving large data sets, a lot of data collection and issues so involved.

But I thank Chairman Miller and other members of the subcommittee for inviting me to speak about the challenges facing the census 2000. In the past, there have been census undercounts that differ by race, ethnicity, sex, and age. The results of the integrated coverage measurement [ICM] adjustment create no new resources. It would only change the way existing resources are shared out.

Furthermore, resources are generally allocated to geographical areas rather than to racial or ethnic groups. For such reason, the population shares of places such as States and counties and cities matter more than census counts. In proving the accuracy of population shares for places, it's inherently more delicate than in proving national counts of demographic groups. Unless the adjustment method is quite exact, it can make estimated shares less accurate by putting people in the wrong places. For 1990, major errors in the proposed adjustment were documented using Census Bureau evaluation data. There was good evidence to show the adjustment could have easily made State population shares less accurate than more accurate.

One oft-stated goal for census 2000 is keep it simple. Sampling for non-response follow-up [SNRFU] and ICM adds layer upon layer of complexity to an already complex census. Complex systems from jet planes to space telescopes fail in unexpected ways. The results of the census are highly dependent on somewhat arbitrary technical decisions. Furthermore, mistakes are almost inevitable, very hard to detect, and have profound consequences.

Examples from the 1990 post enumeration survey [PES], were sobering. The lessons to be learned from 1990 is that the kinds of methods that will be used in the ICM are inherently vulnerable to error. Furthermore, the statistical assumptions behind the adjustment methodology are rather shaky. Senior technical staff at the Census Bureau candidly admit even the 1990 PES estimates, for the highest and most important unit states, rested on assumptions that could not satisfactorily be verified from the PES data themselves. If the ICM, like the PES of 1990 before it, puts in more error than it takes out, the census 2000 will be at considerable risk.

The census makes errors of various kinds. So does the enumeration survey. Some of the estimated undercount must result from census errors, but some is due to processing errors in the PES. In 1990, of the 5.3 estimated undercount, 3 to 4 million was contributed by processing errors in the PES. Processing errors are likely to be even more serious in 2000 than in 1990, as the sample sizes and a tighter time table make quality control more difficult.

Many of the statistical issues have been laid out by the Census Bureau. However, these have been addressed by textbook calculations, which depend on various assumptions and may be hard to put into practice in the field. Reading through the various statis-

tical plans, it is clear there are still a large number of issues that need to be resolved before the census 2000 plan is complete. It is the hope that the dress rehearsals will illuminate these issues. Neither of the dress rehearsals are large enough or sufficiently representative to serve as confirmatory trials at the necessary levels of accuracy for census 2000. Furthermore, there is no last chance to validate or to fine tune the decisions made as a consequence of the dress rehearsals.

Proponents of the Census Bureau make a number of arguments for sampling adjustment. One, everything is relative. The census is imperfect and data quality or survey data are better quality than census data. The imperfection of the census may be granted. However, even if the ICM is in some ways better than the census, the central question remains open. Are the ICM data good enough for their intended use? Will the proposed adjustment to the census take out more error than it puts in?

Two, sampling will save money and improve accuracy. Sampling three households in four within each tract cannot save very much money. Senior Bureau technical staff acknowledge this. On their figures, the savings will be in the order of $125 million out of a total budget of $4 billion. The case that sampling will improve accuracy rests upon the ICM and the corresponding adjustment. In my view, however, the adjustment failed to improve accuracy in 1990. As far as I can see in the papers outlining the statistical methodology in census 2000, there are no claims about relative accuracy for adjusted versus unadjusted census counts in the past or in 2000.

Three, an argument from authority. The arguments from authority don't impress me. The National Academy of Sciences and the American Statistical Association reports provide little analytic detail supporting adjustment. Furthermore, the technical arguments depend on false models which are subject to large biases that cannot be measured directly. Nor are the arguments quite as supportive of sampling as the Census Bureau makes out.

Sampling is scientific. This argument is diversionary, as well as murky. The real problem is non-sampling error. SNRFU and ICM will both suffer from non-sampling error, which will be as hard to quantify as the errors in the census. Thank you.

[The prepared statement of Mr. Wells follows:]

Statistical Controversies in Census 2000

Prepared for 17 September 1998 Hearing of the
United States of America House of Representatives
Subcommittee of the Census

Martin T. Wells
Cornell University
New York State School of Industrial and Labor Relations
Department of Social Statistics
Ithaca, NY 14853

I thank Chairman Miller and the other members of the Subcommittee for inviting me to speak about statistical challenges facing Census 2000.[1]

In the past, there have been census undercounts that differ by race, ethnicity, sex, and age. The results of the Integrated Coverage Measurement (ICM) adjustment create no new resources, but only change the way existing resources are shared out. Furthermore, resources are generally allocated to geographical areas, rather than to racial or ethnic groups. For such reasons, the population shares[2] of places, such as states, counties, and cities, matter more than census counts. [3]

Improving the accuracy of population shares for places is inherently more delicate than improving national counts of demographic groups.[4] Unless the adjustment method is quite exact, it can make estimated shares less accurate by putting people in the wrong places. For 1990, major errors in proposed adjustments were documented using Census Bureau evaluation data, and there was good evidence to show that adjustment could easily have made state population shares less accurate rather than more accurate. [5]

One of the oft-stated goals for Census 2000 is "Keep It Simple."[6] Sampling for non-response follow-up (SNRFU) and ICM adds layer upon layer of complexity to an already complex census. Complex systems, from jet planes to space telescopes, fail in unexpected ways. The results of the Census are highly dependent on somewhat arbitrary technical decisions. Furthermore, mistakes are almost inevitable, very hard to detect, and have profound consequences. Examples from the 1990 Post Enumeration Survey (PES) are sobering[7].

The lesson to be learned from 1990 is that the kinds of methods that will be used in ICM are inherently vulnerable to error. Furthermore, the statistical assumptions[8, 9] behind the adjustment methodology are rather shaky.[10] The senior technical staff at the Census Bureau candidly admit, "Even [1990 PES] estimates for the highest and most important units, states, rested on assumptions that could not be satisfactorily verified from the PES data themselves."[11] If the ICM, like the PES of 1990 before it, puts in more error than it takes out, Census 2000 will be at considerable risk.

The census makes errors of various kinds; so does the post enumeration survey. Some of the estimated undercount must result from census errors, but some is due to processing

errors in the PES.[12] In 1990, of the 5.3 million estimated undercount, 3 to 4 million was contributed by processing errors in the PES.[13, 14] Processing errors are likely to be even more serious in 2000 than they were in 1990, as the larger sample size and the tighter timetable make quality control more difficult.

Many of the statistical issues have been laid out by the Census Bureau; however, these have been addressed by a textbook calculation which depends on various assumptions that may be hard to put into practice in the field. Reading through the various statistical plans, it is clear that there are still a large number of issues that need to be resolved before the Census 2000 plan is complete, and it is the hope that the "Dress Rehearsals" will illuminate these issues.[15] Neither of the Dress Rehearsals are large enough or sufficiently representative to serve as confirmatory trials at the necessary level of accuracy for Census 2000.[16] Furthermore, there is no last chance to validate and to fine-tune the decisions made as a consequence of the Dress Rehearsals.

Proponents of the Census Bureau plans make a number of arguments for sampling and adjustment:

1. Everything is relative: the census is imperfect, and survey data are of better quality than census data. The imperfection of the census may be granted. However, even if the ICM is in some ways better than the census, the central question remains open. Are the ICM data good enough for their intended use? Will proposed adjustments to the census take out more error than they put in?

2. Sampling will save money and improve accuracy. Sampling 3 households in 4 within each tract cannot save very much money. Senior Bureau technical staff acknowledge this; on their figures, the savings may be on the order of ¼ x $500 = $125 million out of a total budget of about $4 billion.[17] The case that sampling will improve accuracy must rest on the ICM and its corresponding adjustment. In my view, however, the adjustment failed to improve accuracy in 1990. As far as I can see in the papers outlining the statistical methodology of Census 2000,[18] there are no claims about relative accuracy for adjusted versus unadjusted census counts, in the past[19] or in 2000. Furthermore, due to the non-sampling error in the ICM--like the PES of 1990 before it--puts in more error than it takes out, Census 2000 will be less accurate.

3. An argument from authority. The arguments from authority do not impress me. The National Academy of Sciences and the American Statistical Association reports provide little analytic detail supporting adjustment. Furthermore, the technical arguments depend on false models which are subject to large biases that cannot be measured directly. Nor are the arguments quite as supportive of sampling as the Bureau makes out.[20, 21]

4. Sampling is scientific. This argument is diversionary as well as murky. The real problem is non-sampling error. SNRFU and ICM will both suffer from non-sampling error, which will be as hard to quantify as the errors in the census.

[1] Much of this testimony is based on the paper "Statistical Controversies in Census 2000" by Lawrence D. Brown, Morris L. Eaton, David A. Freedman, Stephen P. Klein, Richard A. Olshen, Kenneth W. Wachter, and Martin T. Wells.

[2] For instance, according to the 1990 census, the total population of the U.S. is 248,709,873 and the population of New York is 17,990,455. So New York's share is 17,990,455/248,709,873 = 7.2335%. According to the adjustment, New York's share should have been 188,304,414 / 253,979,141 = 7.2071%. Although New York's total count would gain from an adjustment, other states would gain even more, so New York's population share would decrease by .000264 (less than 3 parts in 10,000). The share of California increases from .11966 to .12162 - that is, by .00196. California would have been the biggest gainer from adjustment. Pennsylvania's share would have decreased from .04777 to .04708, a decrease of .00069, which was the biggest loss. Oddly, Pennsylvania and New York supported adjustment in 1990. For one-third of the states, the adjustment would have been less than 30 parts per million. The median share change is 7.5×10^{-5} in absolute value. The Northeastern and Midwestern states with large central-city minority populations, like New York, Illinois, Michigan, Massachusetts, and Pennsylvania, would have lost shares from adjustment. If adjustment were accurately correcting for racial differentials in the undercount, these states would have been expected to gain population share.

[3] Farber, J.E., Fay, R.E., and Schindler, E. (1998), "The Statistical Methodology of Census 2000." pp.3-4, Prepared for Joint Statistical Meetings, August 13, 1998. Technical Report, Bureau of the Census, Washington, D.C.

[4] As in L.D. Brown "Chair's Remarks Following Scheduled Presentations", p.1-2, Prepared for Joint Statistical Meetings, August 13, 1998, since the median state share is .02 the observed the median share change is 7.5×10^{-5} which corresponds to an observed coefficient of error of .003 on a value of p=.02. One would need a simple random sample of 5.5 million to get state shares with this coefficient of variation. The ICM will be about half as large. Although the ICM is a complex cluster sample it could not provide usefully accurate state share estimates.

[5] For discussion from various perspectives, see *Statistical Science* 1994, Vol. 9, No. 4.

[6] U.S. Bureau of the Census (1998), *Census 2000 Operational Plan*, Bureau of the Census, Washington, D.C., p. I-3.

[7] A handfull of examples are: 1. A computer coding error added a million people to the adjusted count. 2. In total, about 3 to 4 million out of the estimated 5.3 million undercount resulted from errors in the PES rather than census errors. 3. The treatment of the Q-class by the imputation model subtracted 400,000 people from the adjusted count. 4. A decision to revise the weights of two block clusters out of the 5,300 in the 1990 PES subtracted 654,000 people from the adjusted count. 5. The decision to pre-smooth estimated variances added 2.5 million people to the adjusted count. The original adjustment would have shifted two congressional seats; the revised adjustment (with the estimated undercount reduced from 5.3 million to 5.3 - 1.3 = 4.0 million) would only have shifted one seat. Errors in the PES, the coding error being the dominant factor, would have moved a congressional seat from Pennsylvania to Arizona. The original adjustment went too far, and the revised adjustment may also, for it still incorporates processing errors of 1.7 to 2.7 million out of the 4 million estimated net undercount.

[8] The two major assumptions that are violated are the independence and homogeneity assumptions, leading to correlation bias and heterogeneity, respectively. Some persons are

missed both by the census and by the post enumeration survey. Their number is estimated by capture-recapture methods, under the assumption that within pre-defined demographic groups (the post strata) there is no correlation between being missed by the census and by the ICM. When this assumption is not satisfied, there is "correlation bias" in the estimated adjustments. People who are especially hard to reach by any survey, whether the census or ICM, are a prime source of correlation bias. Reachability varies from place to place around the country, and differential levels of correlation bias are a major threat to the accuracy of adjusted population shares. To illustrate the impact of correlation bias consider the shares paradox from Endnote #1. The northeastern and midwestern states with large central-city minority populations, like New York, Illinois, Michigan, Massachusetts, and Pennsylvania, would have lost share from adjustment. If adjustment were accurately correcting for racial differentials in the undercount, these states would have been expected to gain population share. One plausible explanation for the paradox is correlation bias. This bias is hard to measure at the national level, and almost impossible at the state level. However, "demographic analysis" supplies independent estimates of the total U.S. population by race, sex, and age. These estimates are derived from administrative records, including birth certificates, death certificates, and Medicare registration.

Table 1. Estimated net census undercounts, by race and sex, from the Post Enumeration Survey and from Demographic Analysis. As of 15 July 1991. "Other" is non-black, including whites and Asians.

	Post Enumeration Survey	Demographic Analysis	Difference
Black Males	800,000	1,340,000	-540,000
Black Females	720,000	500,000	+220,000
Other Males	2,210,000	2,140,000	+70,000
Other Females	1,540,000	710,000	+830,000

Table 1 compares the census undercounts as estimated by the post enumeration survey of 1990 and by demographic analysis. The data suggest there was substantial over-adjustment for females; moreover, a large number of the black males who were missed in the 1990 census were also missed by the proposed adjustment. For the latter group, there is evidence of substantial correlation bias at the national level. It is plausible that the missing black males were concentrated in states with large central-city minority populations, including some of the states with a share decrease. The post enumeration survey may well have been more successful at finding undercounted black males in California and Texas than in New York, Illinois, and Pennsylvania. Correlation bias is a serious and intractable problem. The bias is serious because it can result in adjustments that make state shares worse rather than better. The bias is intractable because it cannot be measured at sub-national levels. Efforts to model the bias at the state level in 1990 were quite unsuccessful, as acknowledged by senior Bureau technical staff (Fay, R.E and Thompson, J.H. (1993) "The 1990 Post Enumeration Survey: Statistical Lessons, in Hindsight", Proceedings, Bureau of the Census Annual Research Conference, Bureau of the Census, Washington, D.C., p. 76).

[9] Geographical shares are adjusted by assuming that undercount rates are constant within specific demographic groups called "post strata." This assumption is clearly wrong, which may have a substantial impact on the adjustment. The Bureau divides the population into post strata defined by demographic and geographic characteristics: one post stratum might be Hispanic male renters age 30-49 in California. Sample persons are assigned to post strata on the basis of the fieldwork. The rate of gross omissions, erroneous enumerations, and the net undercount rate are estimated

separately for each post stratum. It is the net undercount rate that matters for adjustment. To adjust small areas (counties, cities, ... , blocks), the undercount rate in a post stratum is assumed to be constant across geography. In our example, the number of Hispanic male renters age 30-49 in every single block in California, from the barrios of east Los Angeles to the affluent suburbs of Marin County and beyond, would be scaled up by the same adjustment factor, which is computed from sample data for the whole post stratum. This process, of course, is repeated for every post stratum. Ordinarily, samples are used to extrapolate upwards, from the part to the whole. Census adjustment extrapolates sideways, from 60,000 sample blocks to each and every one of 5 million inhabited blocks in the U.S. That is why the homogeneity assumption is needed.

After Census 1990, senior Bureau technical staff acknowledged that residual heterogeneity was appreciable- see Fay, R.E. and Thompson, J.H. (1993, p.81) or Thompson (1998, p. 7); compare Farber, Fay, and Schindler (1998, 14, pp. 17-18). The Bureau has not finalized its post stratification for 2000; current ideas are described by Farber, Fay, and Schindler (1998, pp.11-12); also see Waite, P.J. and Hogan, H. (1998), "Statistical Methodologies for Census 2000-Decisions, Issues, and Preliminary Results", Prepared for Joint Statistical Meetings, August 13, 1998. Technical Report, Bureau of the Census, Washington, D.C., p. 8-9.

According to current plans, post strata will not cross state lines; thus, each state would be adjusted only using data collected within that state. This is an improvement over 1990, because homogeneity of post strata that cross state lines is not assumed. There is a cross-classification within each state (and D.C.) by six race-ethnicity groups, seven age-sex groups, and two "tenure" groups - owners and renters. Post strata are then formed from these 6 x 7 x 2 = 84 categories by collapsing cells with small sample sizes, although that part of the process does not yet seem to be fully defined. These post strata do not take into account area of residence - whether respondents live in major metropolitan areas, suburbs, or rural areas. For that reason, heterogeneity within states may be even more of a problem in 2000 than it was in 1990.

[10] In 1990, the adjustment factors computed from the dual systems estimate had unacceptable levels of sampling error. The Bureau tried to reduce sampling error by means of smoothing. When model outputs were taken at face value, smoothing seemed to reduce variances by a factor of about 2. However, simulation studies and sensitivity analysis suggest that estimated variances were too small by a factor of 2 or 3: if anything, smoothing increased variance. [See Fay, R.E. (1992), "Inferences for Small Domain Estimates from the 1990 Post Enumeration Survey", Bureau of the Census, Washington, D.C. or Freedman et al. (1993), "Adjusting the Census of 1990: The Smoothing Model," *Evaluation Review*, 17, 371-443.] Farber, Fay, and Schindler (1998, pp. 13-14) agree that the 1990 smoothing model was problematic and describe plans for smoothing the adjustment factors in Census 2000 by means of a log linear model. The aim is again to reduce variance. (Although the post enumeration survey in 2000 will be 5 times larger than the one in 1990, there will be 2 or 3 times as many post strata; thus, sampling error may still be a problem.) There are strong similarities between the log linear model for 2000 and the linear model for 1990, although the one for 2000 may be simpler and therefore more robust. However, the basic issue remains the same: a reduction in variance is likely to be accompanied by some increase in bias, and the tradeoff is extraordinarily hard to assess. Farber, Fay, and Schindler (1998) do not attempt to assess the tradeoff. The Bureau may limit the number of post strata to 30 or so per state (Thompson, 1998, p.7); that should reduce sampling error, but is likely to worsen the problems created by heterogeneity.

[11] In Thompson, J.H. and Fay, R.E. "Census 2000, the Statistical Issues" , Prepared for Joint Statistical Meetings, August 13, 1998. Technical Report, Bureau of the Census, Washington, D.C.

[12] After the fieldwork is complete, however, some cases remain unresolved. Statistical models are then used to impute the missing match status. The number of unresolved cases may be relatively small, but it is likely to be large enough to have an appreciable influence on the final results. Imputation models have many somewhat arbitrary elements, and should therefore be scrutinized with great care. After Census 1990, senior Bureau technical staff acknowledge the "subjectivity" of model specification and the difficulty of quantifying the resulting uncertainties; they also agree that the missing data problem is quite intractable because the missingness mechanism is non-ignorable (see Fay and Thompson, 1993, p76).

Movers - people who change address between the time of the census and the time of the ICM interviews - are another complication. Unless persons are correctly identified as movers or non-movers, they cannot be matched. Identification depends on getting accurate information from respondents on where they were living at census time; this is not a trivial problem and is a large factor in the adjustment equation (Breiman, L., 1994, pp. 471-2, "The 1991 Census Adjustment: Undercount or Bad Data?" *Statistical Science*, 9, 458-75). In 2000, the ICM must identify both inmovers and outmovers, but it is the outmovers who will be matched to the census. Inmovers would have to be matched to the census at their census-day addresses, which are likely to be in non-ICM blocks where census follow-up was done on a sample basis. Consequently, matching inmovers would be quite troublesome, and that is why outmovers must be considered. Of course, outmovers will be hard to identify, and information about them will be hard to collect: apparently, such data will be obtained from "proxy interviews" with neighbors, current occupants of the household, and so forth; the alternative is to try and trace the outmovers to their current addresses.

[13] Some systematic error is inevitable in any large survey operation. The ICM is particularly vulnerable to such error for two reasons mentioned above: (i) the adjustments to state shares that need to be estimated are tiny, and (ii) relatively small errors in estimates of gross omissions and erroneous enumerations translate into relatively large errors for the net undercount. Moreover, the ICM will be conducted under extreme time pressure to meet the legal deadline for the transmission of census data to Congress.

[14] There are several operational changes from 1990 in the design for Census 2000 which may decrease either the correct enumeration rate and/or the match rate. A 3% decrease in both the correct enumeration rate and the match rate from 97% to 94% would not change the estimate much, but it could double the estimated variance, multiplying the estimated CV by about 1.4. These changes include: the easy availability of Be Counted Forms could increase the number of erroneous enumerations, decreasing the correct enumeration rate; the use of a five person form instead of a seven person form could increase the number of persons, especially children and non-relatives, missed in the initial phase, decreasing the match rate. The tight schedule and decreased public cooperation could increase the number of whole household imputations in the initial phase, decreasing the match rate - the rate was about 1% in 1990 but about 8% in the 1996 test in Chicago; not performing a surrounding block search for additional matches or performing a limited surrounding block search could decrease the match rate.

[15] See Farber, Fay, and Schindler (1998) and Waite and Hogan (1998).

[16] The goal of the Census 2000 Dress Rehearsal was to obtain a 1.5% coefficient of variation on the population estimates. The ICM is designed to yield state estimates with a 0.5% coefficient of variation on the population estimates.

[17] Farber, Fay, and Schindler (1998, p.6)

[18] Farber, Fay, and Schindler (1998), Waite and Hogan (1998), and Thompson and Fay (1998)

[19] According to the Committee on the Adjustment of Postcensal Estimates Report, 1992, Bureau of the Census (Cape Report), p.33 "... there is no intention to adjust the Census because research shows insufficient technical justification."

[20] "If sampling for NRFU frees resources for taking steps to reduce other sources of error in the final results, it may produce a more accurate census by some measures," in Steffey, D.L. and Bradburn, N.M., editors (1994), *Counting People in the Information Age*, National Academy Press, Washington, D.C., p. 101. This is hardly a ringing endorsement.

[21] On April 6, 1998, the American Statistical Association filed a Brief of *Amicus Curiae* in two lawsuits challenging the use of sampling and statistical methods to count the population in the 2000 census. The introduction to the Brief of *Amicus Curiae* makes it clear that ASA takes no position on the specific uses of sampling proposed by the Census Bureau. "ASA takes no position on the appropriate disposition of this case or on the legality or constitutionality of any aspect of the 2000 census. ASA also takes no position in this brief on the details of any proposed use of statistical sampling in the 2000 census." The sole concern of the ASA is to state to the court that "statistically designed sampling [is] a valid, important, and generally accepted scientific method." With this type of endorsement it is hard to say that all statisticians agree with the Bureau's sampling plan.

Mr. MILLER. Thank you all very much. I appreciate this. Being one who used to teach statistics years ago, it's a pleasure to be here today. I guess I am more comfortable here than over there.

Let me ask a question of each of you all. Basically you all are saying that based on your knowledge of the 1990 attempted use of sampling adjustment, it did not work. Is that right? All three of you are saying yes. Right?

Mr. KOYAK. Yes.

Mr. MILLER. Each of you have talked about accuracy a little bit. You just used that word, Dr. Wells, because it's scientific and it's going to be more accurate. Which is more accurate, do you know? In 1990, it was less accurate. Would you say that? If they would have adjusted? Dr. Brown.

Mr. BROWN. If the target, as I take it, the target is accuracy in shares, that is in national population proportions, then it's very hard, at best very hard to determine whether the PES statements were more accurate or less accurate for that matter than enumeration.

Mr. MILLER. So we don't really know which is more accurate?

Mr. BROWN. We don't really know.

Mr. MILLER. The purpose of the census is for apportionment of representatives and the redistricting effort that will take place in 2001 and 2002, which means most elected officials in America are affected by the census because they have to rely on this data. It relies on the issue of trust. You have got to have trust in this or you are threatening the entire system. In this country we have too much cynicism already. There's so much going on right now about this whole issue of trust. If we have a census that is not trusted, we are threatening a democratic thing. Which one do you think can be trusted to have—you are saying we don't know which one is more accurate?

Mr. BROWN. Right.

Mr. MILLER. Where does the word trust come into the choice between the sampling because it's scientific, versus enumeration? Do any of you want to answer that question?

Mr. BROWN. Well, I guess I would like to give it a try or throw it back to you in part. I raised this issue some year and a half ago roughly when I testified before the Senate. I think it is a question that I can raise and that you, as elected Representatives, are really more qualified to answer. That is, trust, public trust is important in the process, and which kind of process would the public trust more.

Mr. MILLER. You say you don't know which one is a more accurate method. Do you know which one could be—there is a risk that sampling could be even less accurate, because we don't know for sure. Is that a fair statement or not?

Mr. WELLS. Right.

Mr. MILLER. That's what you are saying, Dr. Wells, that—in fact, I think you made a statement that if the ICM, like the PES before it, puts in more error than it takes out, it will be less accurate.

Mr. WELLS. Well, you need to have the data to evaluate what happened so that there's ways of evaluating using loss function analysis. You can tell one story with one set of assumptions and another story with another set of assumptions. So it's very difficult

because you are carrying out this estimation procedure under some set of assumptions. If those assumptions are valid, then things will work out fine. If they are not valid, they won't.

Mr. MILLER. If all eight of you, all eminent statisticians, were in charge of the census and you had to make all of your own assumptions without talking to each other, what is the chance of all of you having the exact same numbers in each of the 60,000 census tracts in this country? You have to make assumptions, and if you all could argue the assumptions you make.

Mr. WELLS. We like to have assumptions that people will agree are close to being true.

Mr. MILLER. You all are members of the Statistical Association, right? I think it was asked earlier. There's a lot of people in the statistics profession that have doubts about sampling for the census. Is that a fair statement? There's a lot obviously that support it. The National Academy of Sciences, when they did their study, one of the critical factors was cost, accuracy and cost. When you put that as a parameter, what does that mean? They wondered and looked at the 1990 census. You want to change it, lower the cost, what are your choices?

Mr. BROWN. Well, that's a difficult question, how to make significant reductions in cost. But coming back to those studies, I think it is clear from their text that they were very concerned with the issue of cost and with attempts to find ways to dramatically drastically decrease costs, hopefully without terribly compromising accuracy.

Mr. MILLER. But when you put a parameter, you get a lower cost, you don't really have much choice of what direction you go. Any time you do experiments, when you set the parameters, you can't influence the result. So basically when you set cost as a factor or a parameter, you have already pre-determined the result.

Mr. BROWN. That's right. Especially when you are trying to drive cost below the levels at which you know you can do a good job.

Mr. MILLER. Dr. Koyak, would you answer, conclude with that?

Mr. KOYAK. Yes. Well I would certainly endorse what Professor Brown has said. One thing I think should be kept in mind. I don't want to be an advocate for wild spending programs, but the projection I have seen is that if we did traditional enumeration in the year 2000, I believe a cost figure on the order of $5 billion was cited. There has been great concern that the costs seem to be going up higher than the nominal rate of inflation from census to census.

But if you take the $5 billion figure and understand we do a census every 10 years, and understand we have a quarter of a billion people in this country, that works out to about $2 per person, per year. That is less than one half cent per day, per person or a little more than one half cent per day, per person. I guess if we want to do a good traditional enumeration, maybe we should look at that parameter and ask ourselves what are we really willing to spend to make sure that people are being properly represented in the population counts and then whatever fund allocation programs or political allocations are made based on the census.

Mr. MILLER. Thank you.

Mrs. Maloney.

Mrs. MALONEY. Thank you. I just want to go back to the element that was raised on cost. It was my understanding that the new majority is willing to spend whatever is needed. That was one statement that was made earlier. I hope it still is true. But if you go back to 1990, it was not a problem of cost. It really wasn't. The reports that I read are that they hired as many enumerators as were needed, that money was not a problem, that they were funded, that if cities came forward and said certain blocks were not counted, particularly public housing areas, they would send out additional enumerators. So it's not a matter of cost.

What we are coming down to is accuracy and getting an accurate count, assuming that the new majority will spend, as they said, no matter how much you need. It is my understanding, well I am not going to get into my understanding. I just would like to ask very briefly and go down the panel starting with Dr. Wells and go down. Excluding preparation for this hearing, please describe what research you have done to evaluate the 1990 post enumeration survey and the planned 2000 ICM. Please tell us when that research was done and where the results were published. Just quickly go down.

Mr. WELLS. Before the preparation of the testimony?

Mrs. MALONEY. Yes.

Mr. WELLS. I have a paper. Probably I'd say 6 months before concentrated research. I have looked at the issue throughout the years, but I didn't have anything to do with the evaluation of the 1990.

Mrs. MALONEY. You did not evaluate 1990.

Mr. WELLS. Did not have anything to do with the formal evaluation in press.

Mrs. MALONEY. You haven't published or done any extensive research except in preparing for this?

Mr. WELLS. Before preparing, I investigated it and just based on that——

Mrs. MALONEY. Did you investigate it in preparation for this hearing?

Mr. WELLS. Before the preparation. That is why I was asked to come.

Mrs. MALONEY. OK. Before. Did you publish anything on it? You didn't? You just personally looked into it?

Mr. WELLS. As an academic you are interested to study things.

Mrs. MALONEY. All right. Dr. Brown.

Mr. BROWN. I also had no direct relation to the studies of the 1990 census. It is an issue that I began to look at seriously, I guess, when I was president of the Institute of Mathematical Statistics as a public policy issue. That was 1994.

Mrs. MALONEY. Have you published? I know you testified in the Senate, but have you published?

Mr. BROWN. I testified in the Senate. I chaired a session on this issue at the recent American Statistical Association meeting and made some comments there.

Mrs. MALONEY. Have you published on it?

Mr. BROWN. I have not published a paper as yet, although I have a paper in process.

Mrs. MALONEY. Dr. Koyak.

Mr. KOYAK. I worked on the 1990 undercount issue as a Department of Justice employee. During that time I had worked extensively with data from the 1990 PES and the census.

My work was concerned with two different questions. One was what is known as the homogeneity assumption. When they defined their poststrata and they made the assumption that the undercount is the same——

Mrs. MALONEY. Have you published on it?

Mr. KOYAK. I have written a paper on it, but I have not published it.

Mrs. MALONEY. I'd love to see it, your paper, if you could get it to me.

Mr. KOYAK. I would certainly do that for you.

Mrs. MALONEY. Maybe the chairman would like to see it too.

Now I think that everybody realizes that there is an undercount. If you believe, and I believe from your testimony that the design for the 2000 ICM is not appropriate for correcting the racial bias in the census, what would you propose be done to correct the racial bias? What would you propose to correct it? I mean everyone who testified has testified that there is an undercount with a racial bias. We know that. What are we going to do to correct it? Just start down. Dr. Wells, Dr. Brown, Dr. Koyak. Again, using enumerators didn't solve it in 1990. It is not going to solve it in 2000. But do you have any suggestion besides more enumerators?

Mr. WELLS. No. I think that there is no easy answer. I think that we should re-evaluate the practices rather than sort of doing what we did before, which is re-evaluate and study it.

Mrs. MALONEY. OK. If you would like to re-evaluate them and study them and put your findings in this committee, I would appreciate it.

Dr. Brown.

Mr. BROWN. Well, it is actually—that's an issue that needs a several pronged approach. Some of the approaches are being investigated by the Census Bureau and are being implemented for the 2000 census, such as better community action, better publicity. How well they will work, I don't know. We don't have good field tests to really verify that.

I want to just, Congressman——

Mrs. MALONEY. So you have no suggestion on how to correct the undercount except for better publicity?

Mr. BROWN. I think better publicity, better field work.

Mrs. MALONEY. Better field work.

Mr. BROWN. I don't know whether those will suffice.

Mrs. MALONEY. You know what bothers me, and I don't mean to be argumentative on this, but some people have said in prior testimony and reports on the 1990 census that they had good field work. They had enumerators. They had the money. There was still an undercount.

And Dr. Koyak.

Mr. KOYAK. I can sympathize with your frustration. But I think that for some problems there just are not very good answers. We would like to have a cure for cancer. We have an enormous scientific investment in trying to find a cure but we don't have one. Maybe we should just look at the problems at this present point

in time with technology, with human resource devotion, with the very narrow timeframe that the census has to operate in, and understand we are going to have some errors in the census. The question is, are we going to try to count people as well as we possibly can. I think that is where the real answer lies.

Mrs. MALONEY. Mr. Chairman, I know Mr. Davis has—my time is up and I know Mr. Davis has questions. I have a series of questions. I would like to just put them in the record. I would just like to say for 2 seconds, I feel very deeply and strongly about this, and that we know the plan without using modern scientific methods will give us an undercount of at least 8 million people. So we either do what many scientists have told us to do to correct it or we accept an undercount of 8 million people. I am not hearing from this panel ideas of how to correct it, with the exception of more publicity and more enumerators. But it was my understanding that last time they had all the enumerators they wanted, all the field they wanted.

Mr. MILLER. Let me just respond because this hearing today and these panelists are statisticians. We are talking about the statistical issues. We are not talking about the outreach efforts and all that. That is going to have to be another issue, another subject, another panel. We need to work in outreach with these other groups, as you have mentioned before, the minority communities to come up with better techniques.

But as far as the scientific thing, that is what we are asking from these panelists, rather than how to develop better outreach. That is not their expertise. We'll have another panel where people can talk about outreach efforts and administrative records. We are talking about statistics here.

Dr. Brown, by the way—and we have I think a distinguished panel. Dr. Brown is a member of the National Academy of Sciences. Is that right?

Mr. BROWN. That's right.

Mr. MILLER. I hope Mrs. Maloney will respect that we have some capable people before us.

Mr. Snowbarger.

Mr. SNOWBARGER. Thank you, Mr. Chairman. Again, as a layman to all of this statistical stuff, I would like to pursue just one angle here briefly. That is, the issue of sample size. Dr. Koyak, if you could—maybe I could walk you through some questions and then maybe you can explain the significance of sample size to me.

It is my understanding that sample size for the year 2000 is about five times what the sample size was for 1990. Is that your understanding?

Mr. KOYAK. That is correct. We're speaking of the integrated coverage measurement, the ICM sample, comparing it to the PES. That's true.

Mr. SNOWBARGER. And it's your contention that the increase in sample size won't necessarily improve the accuracy of the adjustment process?

Mr. KOYAK. I don't believe that it will because the worst problem of the PES was its bias. That's what is known as a non-sampling error. As Professor Breiman detailed to you in earlier testimony, there are some serious problems in the details of using the ICM to

perform an undercount adjustment. It needs to be matched to the original census records. That's a process that is very vulnerable to error. Taking a larger sample does not get to that issue. In fact, it could make it worse because now that five times greater sample has to be done in an even shorter timeframe than in 1990, which means that there's going to be less time for going over the results and checking them carefully.

So I see no way that the larger sample is going to decrease the significant errors of the PES. They could even increase them.

Mr. SNOWBARGER. You mentioned the term bias. I presume that you are referring to the bias that was found in the PES 1991? You have studied that aspect of PES?

Mr. KOYAK. Yes. I was concerned with that when I worked on the 1990 issues. Bias should be understood in a technical sense.

Mr. SNOWBARGER. Why don't you explain it in a layman's terms as opposed to a technical sense, if I can catch it.

Mr. KOYAK. OK. When we speak of how much error is involved in an estimate, we break our error down into two types. One is called sampling error. That's the type of error that says that the larger the sample you take, the more you'll be rewarded for taking a larger sample by reducing a certain type of error. Bigger samples are more accurate than smaller samples because bigger samples give you a better snapshot of the thing that you are trying to address.

However, non-sampling errors or bias means that you have some sort of flaw in your estimation procedure itself, that you are somehow not being able to get your view of that population right on target. When I speak of bias, I am speaking specifically of estimates that really aren't on their target. So, for example, if we have matching errors between the ICM and the census, then that is going to leave the impression that more people are not matching than would really be the case if it were correctly done. That is going to move the adjustment factors upward when you try to correct for the undercount. That is a form of bias because it's not happening because of anything that's true. It is only happening because of errors in your data.

Mr. SNOWBARGER. Are those errors related to process?

Mr. KOYAK. Yes. I believe that they are primarily process errors.

Mr. SNOWBARGER. Does it appear to you that the proposed census for 2000 contains the same kinds of errors in process that were evident in 1991 or whenever the PES was taken?

Mr. KOYAK. That is a difficult question for me to answer in specifics because technology has changed in 10 years. I am sure that they are going to be doing certain things differently 10 years after what they know failed in the 1990 census. How those modifications, how new technology that didn't even exist 10 years ago will impact this equation, I really cannot say. But I think the fundamentals happen to be that you are taking five times as large a sample in half the amount of time. It is going to put a severe strain on the quality control of those data.

Mr. SNOWBARGER. What do you mean by quality control? How do you control for quality in a sampling census?

Mr. KOYAK. You control for quality through having good checking procedures, by having good—for example, for enumerators that

work in the field, to make sure they are properly supervised, to make sure that they do not fabricate data, that you have you know, just very careful checking, very careful looking over, re-matching certain sets of your data to make sure you did it right. The more time you have, the more that you can do. The less time you have, the less of that you can do.

Mr. SNOWBARGER. Do you know what the current plan proposed by the Census Bureau is? You have talked about the time, but have you seen their process for checking and do you have an opinion on what they are proposing?

Mr. KOYAK. I really have not seen the details of it so I really cannot give a detailed opinion on it.

Mr. SNOWBARGER. Thank you, Mr. Chairman.

Mr. MILLER. Mr. Davis.

Mr. DAVIS OF ILLINOIS. Thank you very much, Mr. Chairman.

Dr. Koyak, you say that the use of sampling for non-response is only to save time and money. However, we have been told by the GAO that the longer the Census Bureau is in the field collecting information from housing units that did not return the form by mail, the higher was the error rate. GAO also tells us that in 1990, the error rate at the end of the census approached 50 percent. Does not the use of sampling for non-response reduce the error identified by the Government Accounting Office?

Mr. KOYAK. I did not see that report. I wouldn't be qualified to comment on it.

Mr. BROWN. I guess I'll comment.

Mr. DAVIS OF ILLINOIS. Dr. Brown.

Mr. BROWN. There's some validity to what that report says, but the bottom line answer is no. That sampled non-response follow-up at a three out of four level can only save money at the possible cost of accuracy.

Mr. DAVIS OF ILLINOIS. So if money is to be saved, then the level of accuracy will more than likely increase?

Mr. BROWN. No, no, no. The level of accuracy can only decrease and the only question is how much will it decrease, and that is not clear.

Mr. DAVIS OF ILLINOIS. Though it will definitely decease?

Mr. BROWN. Definitely decrease.

Mr. DAVIS OF ILLINOIS. Last year, Dr. Brown, at a hearing before the Senate Governmental Affairs Committee, you said that every census contains error. Of course we all agree to that, and that head count errors tend to result in undercounts. Though the Census Bureau has proposed using an ICM plan to try and correct such an undercount, the plan is based on statistical sampling methods. In principle it should work satisfactorily, if the statistical sample is large enough and if detailed sampling procedures which are used are adequate to avoid measurement error from certain hidden biases. It went on to say that it is my opinion that the current plans are a very significant improvement over those in effect up to as recently as a month ago.

Then today you testified that the benefits of the increased sample size will be largely offset by other factors. A footnote states that adding State as a stratifying variable is the primary cause. I am just curious.

Mr. BROWN. I don't know if that is the primary cause, but it is certainly a contributing cause. The comments I made a year and a half ago in the Senate and the comments I am making today, I think, are actually very consistent with each other in the sense that a year and a half ago, the Senate had announced a new collection of plans, which are basically what you see before you now. That is a great improvement over what the Bureau had been talking about up until then. But I don't think it is enough. I think it's now too late to pin down all the details to avoid the suite of ifs that I was worried about then.

Mr. DAVIS OF ILLINOIS. Is it not true that the stratifying variable had been in place for quite some time?

Mr. BROWN. Well, the principle to use State as a stratifying variable has been in place. The issue of what the other stratifying variables will be is still an open issue, even at this date. That's one of the many problems that haven't been settled.

Mr. DAVIS OF ILLINOIS. Let me ask; are you suggesting in any way that perhaps a regional breakdown State by State would be a way of dealing with this?

Mr. BROWN. Well a regional breakdown might be preferable on statistical grounds. Might be. Runs into some homogeneity questions that were faced in 1990 and not sufficiently settled. It might be better but it's not clear that it satisfies the political constraints.

I want to just point out that 1990 PES correction didn't do a very good job, in particular with the black population. The demographic figure for the black population of the United States is 6.1 percent. The PES correction makes it 5.87 percent, which is a disparity of over two-tenths of a percent. It is also not clear that the PES put the people that it did put, that it put them in the right place. It is very peculiar. There's a very peculiar pattern of undercount correction that in essence penalizes, appears to penalize the whole northeast, Pennsylvania, New York, Massachusetts and so on, and much of the midwest, all of which areas one would think would gain in the PES correction, but instead in terms of national share are losers.

Mr. DAVIS OF ILLINOIS. Does this position put you in disagreement with other experts such as Marty Wells, Ken Wachter, and Dave Friedman?

Mr. BROWN. To make it very brief. No, the paper that Professor Wells mentioned and that I mentioned, a joint paper that we have put together, so it will be jointly by himself and David Friedman, Ken Wachter and others, combining various views on this issue.

Mr. DAVIS OF ILLINOIS. Thank you very much.

Mr. MILLER. Let me ask, Dr. Wells, I want to clarify an issue that is so often distorted in the reporting of the plans for 2000. Is this 10 percent that is not going to be counted, is this the 10 percent hardest to count? Or is the plan to count the 90 percent easiest and then the 10 percent, is that the plan?

Mr. WELLS. The goal is to count 90 percent.

Mr. MILLER. Then the 10 percent is not the hardest to count?

Mr. WELLS. That's the technical calculations. The goal would be to count the 90 percent.

Mr. MILLER. So is the 10 percent the hardest to count, or is it more random of the non-response?

Mr. WELLS. Well it seems that people don't—there are various reasons people may not reply. So it stands to reason that some of the people that don't reply are the ones that may not reply in the future. So it could happen that they may be the hardest to count.

Mr. MILLER. Do you consider this, when you only count 90 percent and then adjust with the ICM, a complement or is it an either/or type situation?

Mr. WELLS. Well it seems if they want to report one number, it can't be a complement because that's all, if they are just going to report the one number census, it isn't a complement if that is what they are going to report.

Mr. MILLER. Dr. Brown, you are a member of the National Academy of Sciences. I don't think you were there when they had three panels on this issue. Are you familiar with them at all?

Mr. BROWN. Well, I'm familiar with their reports, but not with the construction of how the panels are constructed. I have been active on several boards on the National Research Council, but these are boards devoted more to—well, I am currently a member of what is called the Commission on Physics, Mathematics, and Astronomy. The panels, the census panels were formed by a different subdivision of the National Research Council. So I was not involved in that decision.

Mr. MILLER. How do they select people for these panels? Do you know? They get a chairman and he selects or she selects?

Mr. BROWN. In some cases, although usually there's an oversight board or commission and staff, that the staff proposes names and the oversight board approves them.

Mr. MILLER. OK. They have had three panels on this issue. The issue of cost being the driving factor. I was asking about it at the very end. When you bring up cost, do you have any choice but to go to sampling? I mean if you really want to cut the cost, you would sample only 50 percent.

Mr. BROWN. Well, in fact some of those reports suggested some very drastic sampling rules, as little as 1 in 10, which would have seriously compromised accuracy of any sort.

Mr. MILLER. Is there some magical number to 90 percent, do you think?

Mr. BROWN. Not that I know of. I am not sure where that 90 percent number came from. It is not the result of any particular scientific calculation.

Mr. MILLER. Dr. Koyak or Dr. Wells, why the 90 percent, I heard it was a polling number. From a statistical standpoint, 90 percent is no magical number. Right?

Mr. BROWN. No. I've discussed this issue with people, some people at the Census Bureau, some of the scientists. They don't know where the number came from either. But it seems like a reasonable number if you are going to do any sampling. You might as well take a very large sample. Ninety percent is a very large sample.

Mr. MILLER. OK. Thank you.

Mrs. Maloney.

Mrs. MALONEY. I have further questions, but I would like to just put them in the record and ask for the response so that we can move to the other panel. We are at 1 now.

Mr. MILLER. We thank you all very much. We appreciate you being here today.

If the next panelists would step forward. If all three of you would please stand. We have to go through this little process.

[Witnesses sworn.]

Mr. MILLER. The record will show that all stated in the affirmative.

So we'll start with Dr. Barbara Bryant, if you would like to begin.

STATEMENTS OF BARBARA BRYANT, DEPARTMENT OF BUSINESS AND ECONOMICS, UNIVERSITY OF MICHIGAN; EUGENE ERICKSEN, DEPARTMENT OF STATISTICS, TEMPLE UNIVERSITY; AND STEPHEN FIENBERG, DEPARTMENT OF STATISTICS, CARNEGIE MELLON UNIVERSITY

Ms. BRYANT. I started to say good morning, but I'll now say good afternoon, Chairman Miller, Representative Maloney, members of the subcommittee, ladies and gentlemen. First of all, my name is Barbara Everitt Bryant. I am currently an adjunct research scientist at the University of Michigan, and managing director of the American Customer Satisfaction Index, which involves a large scale survey.

I start by emphasizing that I am currently not an official of the Census Bureau and that opinions I express therefore are my own. I have purposefully not talked with James Holmes, the present acting Director. Nor have I held any discussions with Dr. Kenneth Prewitt, the nominee for Director prior to this testimony.

What I do represent is one who, to use a currently much overworked slang expression, has "been there and done that." What I say today is based on the experience of having served as Director of the Census Bureau from 1989 to 1993. That means that I was Director throughout the 1990 census and through the following 2 years of research on its accuracy and the size and characteristic of its undercount.

Now like most of you in this room, I am pleased that the Supreme Court has expedited the schedule for hearing and issuing an opinion on the use of sampling and statistical estimation to augment enumeration in the 2000 census. At this point in the cycle 10 years ago, the major elements of the design for 1990 were far more locked up than they are now.

The Federal panel's interpretation of the Census Act with its ruling that statistical sampling cannot be used to supplement the 2000 census direct head count is in my opinion unfortunate. Unless overturned by the Supreme Court, that prohibition virtually guarantees undercount in the upcoming census.

In 1980 and 1990, the Census Bureau really pushed the envelope of what you can do by trying to contact every individual household unit and every individual. The addition of sampling for non-response and integrated coverage measurement are necessary, if the Census Bureau is to improve on 1990 or even achieve the accuracy of that year. Until this recent ruling, the weight of case law and Department of Justice reviews of that law and of the Constitution during both the present administration and the Bush administra-

tion were that neither the Census Act nor the Constitution prohibited the use of statistical sampling to improve accuracy.

We all wait now for a Supreme Court decision. In the meantime, we are here today to consider the plan for census 2000 that incorporates both sampling for non-response and integrated coverage measurement to correct for undercount. Today I want to make four points to dispel misunderstandings I think the public may have about this plan.

First, the plan for census 2000 is not a scheme conceived by the present administration. It is not a Clinton plan. Rather, it is an experience-driven plan based on the evaluation of the 1990 census to improve on the methods used. It's a research-driven plan that in mid-decade combined two streams of research. First, very applied practical research by Census Bureau professionals, and second, recommendations of a congressionally-mandated study by experts assembled by the National Academy of Sciences, that said that there would not be improvement without going to sampling.

Now that research all began in 1991. The first test of redesign of more user-friendly questionnaires and better mailing of strategies took place as of April 1, 1992. I testified on that later before the predecessor committee to this one. That was well before the election of the present administration later that year. The plan that has evolved is very close in outline, though with much more operational detail added, to the chapter on the 21st century census in the book I wrote in 1994, Moving Power and Money, the Politics of Census Taking.

The second point I want to make is that sampling for non-response and integrated coverage measurement does not replace direct enumeration. They merely top off the count with an estimated count of that small percent of people who either do not respond by mail or cannot be reached despite multiple efforts of census personnel. It is simply not possible to count everyone. Both common sense and census history prove this. Ours is a mobile society, as several of the other panelists have pointed out. In fact, in a 5-year period, 47 percent of us change our housing unit address.

The United States is made up of individuals with very diversified housing arrangements and lifestyles. There are persons without fixed addresses. There are households that return their questionnaire, but fail to mention the relatives who have been visiting for 2 years or the renter in the apartment created up in the attic or down in the basement. There are those who for reasons of their own do not want to be identified by the Government and avoid being so, even by an agency that guarantees confidentiality.

Now easier to understand questionnaires and multiple opportunities to receive them should improve the 2000 census. But these alone are not enough to approach the 100 percent count, and will never get there perfectly, of the residents of this Nation that the Congress, the mayors, the Governors, and representatives of constituency groups demand. Throwing more money at direct enumeration and the plan to use a professional marketing and advertising council won't close the gap.

The lesson learned from 1990 is that trying harder to directly enumerate doesn't work. When local governments complained of preliminary numbers and identified thousands of blocks they were

certain were undercounted at the local level, these local people who should have known, the Census Bureau mounted an enormous recanvass. Enumerators went back to the blocks upon which 20 percent of the housing units in this country are situated. This expensive effort added only one-tenth of 1 percent to what we had already counted. A net undercount remained.

Now the charge has been made that the Census Bureau will directly enumerate only 90 percent of the population. The fact is that the Census Bureau will be making an effort to enumerate at least that proportion in every census tract. If 100 percent or 95 percent or 98 percent of the housing units in a tract return their questionnaire, all of those questionnaires will be processed. The fact is that most census tracts don't get up to 90 percent without call after call after call by enumerators, with the last several percent impossible to find. At the time the mail return response ended in 1990, when we had to start sending the enumerators out for the followup, the best figures in the Nation were the city of Columbus, OH, that had returned 73 percent of questionnaires, where 150 miles north of them in Cleveland had only returned 51 percent.

The third of the four points I want to make is that integrated coverage measurement, the taking of a very large scale sample survey and matching names from it to those in the census to determine who was counted and who was not, is not an untested experiment. The Census Bureau is merely enlarging on survey methods that it knows how to do very well and has done before. The post enumeration survey was done in 1980. I'll only talk about 1990.

The 1990 PES, as we called it, completed interviews at a sample of 170,000 households and matched the approximately 400,000 names in them to the recently completed census. Now you have done pilot studies I'm sure, Mr. Miller, Chairman Miller, and that's a pretty good pilot test for a survey of 750,000 housing units. Like all research organizations or the good ones, and I think the Census Bureau is one of the good ones, the Census Bureau learns from experience. Integrated coverage measurement will benefit from the extensive studies of what went right and what went wrong with the 1990 post enumeration surveys, the studies of bias and other things. But I repeat, you couldn't have a better pilot test for 2000 than this 1990 survey was.

The fourth of my four points is that both opponents and proponents of the plan for census 2000 with its integrated coverage measurement over-estimate the political effects of adding those who will be uncounted to the count. They over-estimate for two reasons. Half, 52.2 percent to be exact, of the net undercount of 4 million, we've heard the 8 million undercount, the 4 million overcount. The net is the 4 million. Half of that 4 million are children. Now those under 18 can't vote.

The second reason is that uncounted adults are unlikely to be voters in the near-term future. These are persons who are disconnected from the society that you can't reach, you can't get to, not those who participate in civics ceremonies like voting and filling out census questionnaires.

Yes, adding 2 to 3 percent to the population, only half of whom will be adults, will shift one or two seats in the House of Representatives between the States. Most likely, the seat will shift to

a southern or southwestern State because those are the areas where our population is growing fast. Yes, distributing that 2 to 3 percent across 435 congressional districts and hundreds of State and local legislative districts will marginally affect the boundaries of some or many districts. But for the above two reasons, there will not be a change of 2 to 3 percent of the number of those who are eligible to vote and then do so. Besides, reapportionment of districts is a State activity, dependent in part on the party in power in each State. Reapportionment only starts with the count delivered by the Census Bureau.

Finally, there are those who are concerned about whether integrated coverage measurement is accurate at small area levels. Critics however, tend to overlook the fact that enumeration is not always accurate at very small levels either. My least favorite news story about the 1990 census was a picture and caption in USA Today. Unfortunately I did not retain the clipping because until today I really wanted to forget it. I talked to a former USA Today reporter 2 days ago who confirmed that the picture and caption did exist, but he didn't have the clipping either. The caption showed the official census count in big numbers of a very very tiny town. The picture showed the entire population turned out on the main street with more people in it than the census had counted by enumerating.

Fortunately, errors in both sampling and enumeration tend to cancel out when aggregated to the level at which apportionment is critical, the aggregates of census tracts that make up legislative and congressional districts.

I think taking a census is a lot like building a house of bricks. What is important is not that each brick is absolutely perfect, but whether all the bricks are available for construction of the whole. Without integrated coverage measurement to fill in the missing bricks and to remove a few extras that shouldn't be there, we will have an incomplete building. Unfortunately, it will not be a building that's missing a row of bricks so the whole house sits level. We will be missing a few bricks here, a few more there, 2 million children, 4 percent of those who live in rental housing, and have a slight oversupply of two bricks from those who have two homes.

For the 21st century, we need to build with modern methods. We need to use up-to-date statistical techniques to improve the accuracy of the national count, and know that our house is level, complete, and in order.

[The prepared statement of Ms. Bryant follows:]

SUBCOMMITTEE ON THE CENSUS

COMMITTEE ON GOVERNMENT REFORM AND OVERSIGHT

U.S. HOUSE OF REPRESENTATIVES

September 17, 1998 at 10 a.m.

"Oversight of the 2000 Census: Serious Problems with Statistical Adjustment Remain"

TESTIMONY OF

BARBARA EVERITT BRYANT

Adjunct Research Scientist
University of Michigan Business School
Ann Arbor, MI 48109-1234

Director, Bureau of the Census
1989-1993

Good morning, Chairman Miller, Representative Maloney, members of the Subcommittee, ladies and gentlemen. I am Dr. Barbara Everitt Bryant, currently an adjunct research scientist at the University of Michigan Business School, but from 1989 to early 1993 I was Director of the Bureau of the Census, U.S. Department of Commerce. My term in office included the taking of the 21st, and most recent, decennial census of the United States, the 1990 census.

I start by emphasizing to those in the audience that I am currently not an official of the Census Bureau and that opinions I express here are my own, not those of the Census Bureau. I have purposefully not talked with James Holmes, its present Acting

Director, nor have I held any discussions with Dr. Kenneth Prewitt, the nominee for Director, prior to this testimony.

What I do represent is one who—to use a currently much overworked slang expression—has "been there; done that!" What I say today is based on the experience of having served as Director throughout the 1990 census and throughout the following two years of research on its accuracy, and on the size and characteristics of its undercount.

Like most of you in this room, I am pleased that the Supreme Court has expedited the schedule for hearing and issuing an opinion on the use of sampling and statistical estimation to augment enumeration in the 2000 census. At this point in the cycle 10 years ago, the major elements of the design for 1990 were virtually locked up.

The federal court panel's interpretation of the Census Act, with its ruling that statistical sampling cannot be used to supplement the 2000 census direct headcount, is unfortunate. Unless overturned by the Supreme Court, that prohibition virtually guarantees undercount in the upcoming census. In 1980 and 1990, the Census Bureau pushed the envelope of the accuracy that can be achieved by trying to contact every individual and household unit. The addition of sampling for non-response and Integrated Coverage Measurement are necessary if the Census Bureau is to improve upon, or even match, the accuracy level of 1990. Until this recent ruling, the weight of caselaw, and Department of Justice reviews of that law and of the Constitution during both the present administration and the Bush administration were that neither the Census Act nor the

Constitution prohibited the use of statistical sampling to improve accuracy. We all wait now for a Supreme Court decision.

In the meantime, we are here this morning to consider the Plan for Census 2000 that incorporates both sampling for non-response and Integrated Coverage Measurement[1] to correct for any undercount. Today, I wish to make four points to dispel misunderstandings I think the public may have about the upcoming 2000 census.

1. The Plan for Census 2000 Is Not an Administration-Conceived Plan

First, The Plan for Census 2000[2] is not a scheme conceived by the present Administration. Rather it is:

- An *experience-driven plan* based on evaluation of the 1990 census to improve on the methods used then.

- A *research-driven plan* that in mid-decade combined two streams of research: First, 5-years of research by Census Bureau professionals and, second, recommendations of a Congressionally-mandated study by experts assembled by the National Academy of Sciences.

[1] Integrated Coverage Measurement (ICM) will use a large scale sample survey, stratified by demographic and geographic characteristics, to determine the number of people and housing units missed or counted more than once in the 2000 census. The names of those enumerated in the survey are matched to those counted in the census to determine those counted in both, or counted in one but not the other. From these an estimate can be made of those missed by both census and survey. (This is called capture/recapture methodology.) The numbers of those missed or counted twice in each strata can be estimated and the counts in each strata weighted to correct the census.

[2] Bureau of the Census, Economics and Statistics Administration, U.S. Department of Commerce, *Census 2000 Operational Plan* (July 1997) and *Report to Congress—The Plan for Census 2000* (Revised and Reissued August 1997).

Research began in 1991. The first test of redesigned, more user-friendly questionnaires and better mailing strategies took place as of April 1, 1992, well before the election of the present Administration later that year. The plan that has evolved is very close in outline—although with much operational detail added-- to the "21st Century Census" described in: *Moving Power and Money: The Politics of Census Taking* (Ithaca, NY: New Strategist Publications, Inc. 1995), a book that I wrote in 1994.

2. Sampling and Statistical Estimation Do Not Replace Enumeration; They Top It Off

The second point I want to make is that sampling for non-response and Integrated Coverage Measurement do not replace direct enumeration. They merely top off the count with an estimated count of that small percent who either do not respond by mail or cannot be reached despite multiple efforts of census personnel. *It is not possible to count everyone.* Both common sense and census history prove this. Ours is a mobile society in which 47% of households change their housing unit address within a 5-year period[3]. The United States is made up of individuals with diversified housing arrangements and life styles that did not exist a generation or two ago. There are persons without fixed addresses. There are households that return their questionnaire, but fail to include the relatives who have been "visiting" for two years, or the renter in the apartment created in the basement or attic. There are those who for reasons of their own do not want to be identified by the government—even by an agency that guarantees confidentiality to those enumerated.

Easier-to-understand questionnaires and multiple opportunities to receive them, should improve the 2000 census, but these alone are not enough to approach the 100 percent count of the residents of this nation that the Congress, governors, mayors, and representatives of constituency groups demand. Throwing more money at direct enumeration and the planned use of professional advertising counsel won't close the gap. I do applaud the plan to use paid advertising to promote the importance of answering the census. However, the lesson learned from 1990 is that just trying harder to directly enumerate doesn't work. When local governments complained of preliminary numbers, and identified thousands of blocks they were certain were undercounted, the Census Bureau mounted an enormous re-canvass. Enumerators went back to the blocks upon which 20% of the nation's housing units were situated. This expensive effort added only one-tenth of one percent to those already counted. A net undercount remained.

The charge has been made that the Census Bureau will directly enumerate only 90 percent of the population. The fact is that the Census Bureau will be making an effort to enumerate *at least that proportion in every census tract.* If 100 percent, or 95 or 98 percent, of the housing units in a tract return their questionnaires, all of these questionnaires will be processed. And how happy the Census Bureau will be! The fact is that most census tracts don't get up to 90% without call after call after call by enumerators, with the last several percent impossible to find.

---

[3] 1990 census figure on housing unit changes 1985-1990.

3. Integrated Coverage Measurement Is Not an Untried Experiment

The third of the four points I wish to make is that Integrated Coverage Measurement—the taking of a very large scale sample survey and matching names from it to those in the census to determine who was counted and who was not--is not an untested experiment. The Census Bureau is merely enlarging on survey methods it knows how to do very well and has done before. The 1990 Post Enumeration Survey completed interviews at a sample of 170,000 housing units and matched the approximately 400,000 names in them to the recently completed census. That's a pretty good pilot test for a survey of 750,000 housing units. And as far as turning large surveys around in short time, every single month of the year the Census Bureau completes interviews at a sample of 150,000 housing units for the Current Population Survey—and does it all within the third week of the month. The Bureau's been doing the Current Population Survey for 56 years to provide the nation's unemployment/employment data.

Like all research organizations, the Census Bureau learns from experience. Integrated Coverage Measurement will benefit from the extensive studies of both what went right and what went wrong with the 1990 Post-Enumeration Survey. But, I repeat, you couldn't have a better pilot test for 2000 than this 1990 survey was.

4. Overestimation of Political Effects

The fourth of my four points is that both opponents and proponents of the Plan for Census 2000, with its Integrated Coverage Measurement, overestimate the political

effects of adding those who will be uncounted to the count. They overestimate for two reasons:

(1) Half (52.2%) of the net undercount of four million in 1990 were children. Those under 18 can't vote.

(2) Uncounted adults are unlikely to be voters in the near-term future. These are persons disconnected from the society, not those who participate in civic ceremonies like voting and filling out census questionnaires.

Yes, adding a missed 2-3% to the population count (only half of whom are likely to be adults) will shift one or two seats in the House of Representatives between the states. Most likely the seat(s) will shift to a Southwestern or Southern state because those are the areas of growing population. And, yes, distributing that 2-3% across 435 Congressional districts and hundreds of state and local legislative districts will marginally affect the boundaries of some districts. But for the above two reasons, there will not be a change of 2-3% in the numbers of those who are eligible to vote and then do so. Besides, reapportionment of districts is a state activity, dependent in part on the party in power in each state. Reapportionment only starts with the count delivered by the Census Bureau.

Conclusions

Finally, there are those who are concerned about whether Integrated Coverage Measurement is accurate at small area levels. Critics, however, tend to overlook the fact

that enumeration is not accurate at very small areas either. My least favorite news story about the 1990 census was a picture and caption in *USA Today.* Unfortunately, I did not retain the clipping because until now I wanted to forget it. I talked to a former *USA Today* reporter a few days ago who confirmed that the picture and caption had been published, but he didn't have a copy either. The caption showed the official census count of a very tiny town. The picture showed the entire population of the town, turned out on the main square. There were three or four more people in the picture than the census had enumerated in this very tiny population.

Fortunately, errors in both sampling and enumeration tend to cancel out when aggregated to the levels at which apportionment is critical—aggregates of blocks and census tracts that make up legislative and Congressional districts.

Taking a census is like building a house of bricks. What's important is not that each brick is absolutely perfect, but whether all the bricks are available for construction of the whole. Without Integrated Coverage Measurement to fill in the missing bricks—and to remove a few extras that shouldn't be there-- we will have an incomplete building. Unfortunately, we will not be missing merely a row of bricks, so that the building remains level. We will be missing a few bricks here, a few more there, and have a slight oversupply at some points. For the 21st century we need to build with modern methods. We need to use up-to-date statistical techniques to improve the accuracy of the national count, and know that our house is level, complete, and in order.

Mr. MILLER. We will take a recess. A vote is taking place. I apologize. The beepers go off and we have to go.

[Recess.]

Mr. MILLER. The hearing will continue. We apologize for the necessity to go vote. I think we're not going to have any more votes for a while, so there won't be the interruption I don't believe again.

We'll proceed with Dr. Ericksen, if you'd like to make an opening statement.

Mr. ERICKSEN. In 1980, I was asked by the Census Bureau to give a talk in which I expanded upon the basic idea of my dissertation on the possibility of using a method or variation of that method——

Mr. MILLER. Can you move the microphone closer?

Mr. ERICKSEN. Sure. I'll start over. I did my doctoral dissertation. It was a statistical model used for estimating sizes of local populations through what is now called a smoothing model. In 1980, the Census Bureau asked me to come give a talk about the possibility of using that model or an expansion of that model to adjust the undercount of cities and States. After that, one thing led to another. I have now done 18 years of research on issues of census taking and adjustment. In the spirit of full disclosure, I should tell you that I was co-chair of a special advisory panel to Secretary Mosbacher appointed by plaintiffs from the 1990 census.

I am not going to read my testimony. I would simply like to hit a few of the high spots. The first thing, what is the problem to be solved. I think that it is important to introduce the concept of gross error as opposed to net error. If you say that the net undercount is 1.6 percent so that 98.4 percent of the population was counted, that is a performance that gets an A. Then surely if the census count equaled the demographic estimate, that percentage would be 100 percent, you would get an A+. However, in the first instance, if there were 12 million omissions and 8 million erroneous enumerations, there would be a net undercount of 4 millon. If the number of erroneous enumerations increased to 12 million, the net undercount would be zero, but the problem would be worse. So evaluations of PES data should actually focus on numbers of omissions and numbers of erroneous enumerations when evaluating the quality of data.

The second point I would like to make is that the problems of taking the census are getting more difficult. Taking a census in the year 2000 will be more difficult than it was in 1990, which in turn was more difficult than it was in 1980 and so forth. There were millions of erroneous enumerations and millions of omissions in the year 1990, and those quantities were greater than they had been previously.

This only begins to tell the story, however. The quality of the data collected by the Census Bureau was often times very poor. Some 8.5 million people were counted by last resort or closeout or were non-data defined. This information was often times collected from people not even living in a household where the counted person lived, such as from a neighbor, a building superintendent, or even a passerby. These cases only had zero in some cases, one, two, or three pieces of data collected about them. Moreover, the Census

Bureau found that 38 percent of last resort and 44 percent of closeout cases were erroneous enumerations.

These problems were especially frequent on long forms, on which crucial variables such as income and employment status are collected. For example, among blacks on long forms, some one out of seven were last resort, closeout or non-data defined cases. For this population, social and economic policies are poorly informed by 1990 census data. My testimony does not intend to be critical of the Bureau. It's just that the problem is very difficult.

The next point is that the longer the census collection takes, and the traditional enumeration in 1980 and 1990 lasted 4 to 6 months, depending on how you count it, the later you go in the year, the worse the data are. The rate of erroneous enumeration among people counted in April was 3 percent. It was 7 percent in May, 14 percent in June, 19 percent in July, and 28 percent in August or later. Sometimes trying harder only makes things worse.

When I hear people say the Bureau needs to do a better job of counting, I think of Yogi Berra, déjà vu all over again. Coverage improvement doesn't work. Coverage improvement has tremendously high rates of erroneous enumeration. In 1990, among those procedures put in place after NRFU, the rate of erroneous enumeration was 19.4 percent.

Statisticians will agree that if equality of measurement on a sample is better than the quality of measurement on the census, the sample is likely to give a more accurate estimate than the census. The key question is how accurate is the measurement on the sample. That is a technical decision that I believe the statisticians at the Census Bureau who are an extremely competent and politically neutral group are qualified to make. They are in a much better position to make that decision than any of us on the panel, with the possible exception of Dr. Bryant, who obviously has been there.

The idea of sampling for NRFU would be particularly focused on that part of NRFU, non-response follow-up, where the data collection would be late. So one of the true benefits of sampling is that you can improve the quality of data collection on a sample and reduce the rate of erroneous data collection in the latter part of the census.

Cutting to the final decision, I believe that there are three choices. The first choice, if we take an arbitrary date like June 1, which is approximately 2 months after census day, by June 1, by the traditional method, there will be many millions of people who have not yet been counted. One choice is to stop there and accept a very very substantial undercount, possibly of 20 million people.

The second choice is what I call the traditional census coverage improvement. You keep on counting. The problem with that alternative is, as previous people have testified, the rate of mobility in the American population is so great that you count many, many people in the wrong place. What you often wind up doing is collecting erroneous data, which gives the illusion of eliminating the undercount, but is being done so with erroneous data.

The third choice is statistical correction and the ICM. Based on my review of reports, I believe that that ICM would create a more accurate result than not taking the ICM. But it's important to understand what the two alternatives are. In 1990, I believe we

learned that the counting which was done in the last phase of the census was terribly erroneous. It's not the kind of data that American social policy or constitutional allocation should be based upon. Thank you.

[The prepared statement of Mr. Ericksen follows:]

I am a Professor of Sociology and Statistics at Temple University, where I have taught and conducted research since 1970. As part of my research I have written numerous articles on the conduct of the census, often on the subjects of census undercount and how to correct for it. I have also provided expert testimony on behalf of plaintiffs in lawsuits seeking to obtain a corrected census count for both the 1980 and the 1990 Censuses. In 1990 I served as Co-Chair on a Special Advisory Panel appointed to advise then-Secretary Mosbacher on the possibility of adjusting the 1990 Census.

In my testimony today, I have been asked to focus on the feasibility of (1) Integrated Coverage Measurement (ICM) and (2) Sampling for Non-Response Followup (SNRFU) to accurately measure the undercount of the 2000 Census. I will present my comments in five sections: (1) a statement of the problem to be solved, (2) a description of the problems of NRFU as observed in 1990, (3) a description of the anticipated improvements likely to be provided by SNRFU in 2000, (4) a discussion comparing the design of the 1990 Post Enumeration Survey (PES) with ICM as designed for 2000, and (5) my overall conclusions.

The Problem to be Solved

The 1990 Census had an unprecedented amount of error. According to data provided by the Census Bureau to the Special Advisory Panel in 1991, there were 20 million omissions and 16 million erroneous enumerations (including census substitutions) for a total gross error of 36 million. These quantities far exceed the comparable quantities for the 1980, or any preceding, census where evaluation data are available. The net undercount, or difference between rates of omission and erroneous enumeration, was disproportionate. The undercount for minority populations was greater than for Whites and the net undercounts for large cities typically exceeded those for small cities, suburbs, and rural areas. In general, where census-taking problems, as indicated by low mail-back rates, were greater, both omission and erroneous enumeration rates were higher. Comparing high and low mail-back rate areas, the differences in rates of omission were greater than for rates of erroneous enumeration. As a result, the net undercount was highest in those areas where census-taking problems were worst.

This only begins to tell the story, however. The quality of data collected by the Census Bureau was also very poor. Some 8.5 million people were counted by last resort or closeout or were "non-data defined." For these people, the census form recorded 3 or fewer variables, e.g., only race, or only age and sex. For these cases the information was often obtained from neighbors, building supervisors, or even passersby, and its validity is questionable. Some 38 percent of last resort and 44 percent of closeout cases were later found to be erroneous enumerations. These problems were especially frequent on long forms, on which crucial variables such as income and employment status are collected. For example, among Blacks on long forms, some 1 out of 7 (14.4 percent) were last resort, closeout, or non-data defined cases. For this population, social and economic policies are poorly informed by 1990 Census data.

My testimony is not intended to be critical of the Census Bureau. The problem is that the job of counting the nation has grown substantially more difficult as time has passed. Problems of counting in 1990 were compounded by the fact that it took so long to collect the data. Substantial numbers of people were counted during the summer, and others were not counted until the early fall. When people are counted so late, they are often counted inaccurately, or at the wrong address. Millions of people moved between April 1 and the date at which they were counted. In 1990, among those counted in April, the rate of erroneous enumeration was only 3 percent. It was 7 percent in May, 14 percent in June, 19 percent in July, and 28 percent in August or later.

To correct the problems of the 1990 Census, the Census Bureau recognized that it needed to improve the quality of the data it collected. The best way to do this is to collect fewer observations by sampling, but to improve the quality of the average observation. For example, a well-collected sample of 1,000 observations is likely to be more accurate than a poorly collected population count of 10,000 observations. This is because the errors due to sampling are very small compared to the errors due to a high nonresponse rate, missing data, or incorrect information written on census forms.

Problems of Non-Response Followup in 1990

Many of the problems of the 1990 Census occurred during Non-Response Followup (NRFU). First, the Bureau had difficulty hiring qualified enumerators, especially in areas where mail-back rates were low and the NRFU workloads were

greater. Second, due in part to the lack of qualified enumerators and in part to the greater workloads, NRFU took a very long time to complete, sometimes 3 or 4 months in difficult areas. Third, the quality of data was often poor, as I have just discussed. Finally, but very importantly, the cost of NRFU was immense. Any improvements in NRFU not involving sampling would cost even more money. They would involve somehow finding more and better enumerators, and paying them more.

In 1990, 11 percent of persons counted on NRFU were erroneous enumerations compared to 3 percent of persons counted on mail-back. This percentage was especially great in the latter stages. After June 30, the NRFU rate of erroneous enumeration was over 25 percent. After NRFU was completed there were millions of uncounted persons remaining. The Census Bureau attempted to count them through "coverage improvement programs" such as the Vacant-Delete Check and the Parolees-Probationers Check. Unfortunately, these programs were not only erroneous (19.4 percent rate of erroneous enumeration) but they left millions still uncounted.

Sampling as a Solution to the Problems of NRFU

The Census Bureau plan for the 2000 Census is to sample those households who do not mail their census forms back in. As I understand it, the Bureau plans to sample enough households so that at least 90 percent of all housing units in an area would have census forms turned in. Certain critics have derided this plan as one where the Bureau "intentionally decides not to count people." This characterization is misleading. The Bureau will send census forms to all households. It has made prodigious efforts to list all households on its address register, and the Bureau will send second copies of the census form to every household that does not quickly mail the first one back in. The Bureau has also made it possible to be counted by telephone.

The Bureau's sampling plan for the year 2000 will have several benefits. First, it is much cheaper than the non-sampling alternative. Second, it will shorten the time of census taking, so that a much greater proportion of persons counted will be counted near April 1. Third, the quality of data collected is likely to be substantially improved. This will be due in part to a smaller proportion of persons moving between April 1 and the date of data collection and in part to a higher quality enumeration. The use of sampling

for NRFU should reduce the rate of erroneous enumeration substantially, improving the overall quality of census data.

Integrated Coverage Measurement

The Census Bureau calculated corrections to the 1990 Census using a large-scale Post Enumeration Survey. These corrections caused substantial controversy concerning whether or not they should be incorporated into the official census counts. As a member of Secretary Mosbacher's advisory panel, I concluded that use of these corrections would improve the accuracy of the 1990 Census. As is well known, large numbers of statisticians both favored and opposed this position. There has been substantial debate and discussion in the scientific literature on the subject, and the disagreements continue to this day. Because the design of the 2000 ICM is very different from the design of the 1990 PES, though, the 1990 debate is not especially pertinent to the debate on the 2000 Census.

The design changes are important, and each is likely to lead to substantial improvement over the 1990 PES results. Among these design changes are:

1. The samples are much larger in 2000 than they were in 1990.

2. PES interviewing will occur at an earlier date in 2000, lessening problems of inaccurate matching.

3. Fewer people will have moved between Census Day (April 1) and the date of the PES interview, lessening problems of identifying the correct Census Day address.

4. In 1990, complex regression models were used to compute the corrections; in 2000 they will be based entirely on local samples without regression adjustments.

5. The regression models used in 1990 were intended to maximize the accuracy of total population estimates. They sometimes distorted the demographic distributions of local populations, e.g., adjusting females more than males. This problem will be lessened substantially by the use of local sample data without regression adjustments.

6. Reducing the amount of time between Census Day and PES interviewing will have a major effect on the accuracy of corrections, and use of larger, local samples will mean that information will not need to be "borrowed" from one state for use in another. In other words, for example, North Dakota data alone will be used for the North Dakota corrections.

The Census Bureau has worked continuously on the problem of census correction based on survey data for nearly 20 years. It has the benefits of its experiences on the 1990 Census as well as the extensive research conducted since then. Having read Census Bureau reports, as well as articles in the scientific literature and reports of panels appointed by the National Academy of Sciences, I am convinced that the Census Bureau has made substantial progress improving the quality of census corrections since 1990.

Conclusions

In thinking about whether use of SNRFU and ICM would improve the accuracy of census data, it is helpful to consider the dimensions of the problem they are intended to solve. In 1990, there were 20 million omissions and 16 million erroneous enumerations. While some of these occurred in the same places, by and large the distributions of omissions and erroneous enumerations differed, so there were substantial differences among areas in the size of the net undercount. In some areas, the undercount was close to 10 percentage points, while at the other extreme, some areas experienced overcounts, i.e., there were more erroneous enumerations than omissions. The total size of the net undercount, 4 million, is not the relevant statistic. Indeed, the national net undercount could be zero, and there would still be substantially differential undercounts among local areas. These could differ by 5 to 10 percentage points in many states. Any evaluation which compares the size of the possible ICM error to that of the total net undercount is therefore misdirected. Instead, statisticians should ask whether the estimated distribution of population among states, and within states, among local areas, has been improved.

Next, statistical errors can be divided into two types, systematic and random. Systematic errors, or biases, in the census would favor some areas over others. If the bias of the census was similar every time the census was taken, then some areas would consistently have an unfair advantage over others. Alternatively, if the bias was eliminated, but random errors remained, there might still be census errors, but different sets of local areas would be advantaged in different censuses.

Random error has a second relevant characteristic. While these may be large in small local areas, they tend to cancel out when the local areas are combined into larger ones. For example, let us assume an unrealistically high amount of error on the block level error due to sampling, 30 percent. This is a level of error almost certainly not to be

attained on the 2000 census. If a congressional district had 10,000 blocks, and the estimate for each block had a 30 percent random sampling error, then the expected random error for the congressional district would be 30 percent divided by 100 or 0.3 percent. This is almost surely smaller than the error that would be produced by an uncorrected census count.

Next, we should realize that for most local areas, even blocks, the size of the corrections would be small, or moderate. The differences between corrected and uncorrected counts will typically be just a few percentage points. While the effects of corrections will be noticeable, they will not be substantial. Fears of wholesale redistribution of population due to the ICM are therefore misplaced.

Fears that the corrections will be replete with error are also misplaced. The Census Bureau has a large group of able, well-respected statisticians who have designed SNRFU and ICM, and evaluated it. They have concluded that use of these methods would produce the most accurate census possible. Panels of experts appointed by the National Academy of Sciences have seconded this conclusion. These conclusions have been drawn without reference to political factors. In other words, the Census Bureau has designed the best census statistically possible at this time. Having reviewed their reports, I see no reason to contradict this decision. We need the Census Bureau to give us the best and most accurate census possible, and we should rely upon their judgements.

Mr. MILLER. Dr. Fienberg.

Mr. FIENBERG. Thank you, Mr. Chairman. My name is Stephen Fienberg. I am Maurice Falk, university professor at Carnegie Mellon University. I am pleased to have this opportunity to testify before the subcommittee today.

I have worked on practical sampling problems since the mid-1960's. For the past 20 years, I have been engaged in methodological research related to sample surveys and census taking. I have had numerous occasions to study and comment upon the work of statisticians at the Census Bureau. I've looked at the methods and the actual results in 1980 from the census, in 1990, and the plans for the year 2000. I have even looked at PES forms. Recently, I have personally checked the matching programs in my own work on record linkage.

In other professional settings, I am a former vice president of the American Statistical Association, and currently president of the Institute of Mathematical Statistics.

So, my testimony. My first message is very simple. The statisticians at the Census Bureau are professionals and they are scientists. I often disagree with them on specific details of methodology, priorities and so on, but their professional openness and the access to their work, which they provide to others, ensures that they would not knowingly and could not knowingly manipulate census results, whether through sampling or any other means. They have earned the respect of Congress and the American people, as well as that of their professional peers.

Other people this morning have noted that a few of our fellow statisticians, especially those in academia, have any sense of what it means to actually plan, conduct, or analyze the results of a census. There are a lot of myths about the census in American history and about the planning for the census in the year 2000 that are shared by statisticians and non-statisticians alike. I would like to address just a few of these, and share for the record a paper which documents more than I have written with my colleague Margo Anderson.

Here is an example. Myth. The census is just a physical head count. As Professor Ericksen has pointed out and Dr. Bryant in other ways, at no point in this Nation's history was there ever a physical head count of each person in the country. Those who call today for a return to traditional census taking seem blissfully unaware of both what the methods of the past have been and the extent to which errors have come to threaten the usefulness of the census.

Today Chairman Miller said that the 1990 census has accurately counted 98.4 percent of the population. This is another myth. In fact, this is the net national undercount subtracted from 100 percent. As Professor Ericksen just explained, that is netting out a large number of omissions, subtracting out an almost equal number of erroneous enumerations. Even if there were a net census error of zero so that the national total appeared to be correct, and I say appeared because it isn't, the census could be plagued by error that would have enormous consequences when carried down to look at accuracy at States and at lower levels of geography, which are used for apportionment.

My written testimony refers to the history and methodological improvements in the census. They involve sampling, in particular, post enumeration surveys, but other kinds of sampling, and imputation. Viewed from this perspective, the one number census proposal for the year 2000 is the natural culmination of improvements in methodology and a seeming decline in the accuracy of census counts.

It isn't a radical departure of ideas from the past. Nor is it something that was invented out of whole cloth by statisticians at the Bureau, or by people who are members of the National Academy of Sciences panels, like myself. In fact, if I were to complain about what the Census Bureau is proposing, it's that it isn't bold enough.

The plan for census 2000 when set in this context makes intuitive sense and it has a sound methodological basis. If the traditional census methods are fraught with error, which experience tells us they are, the GAO tells us they are, the Census Bureau reports on which everybody relies tells us they are, then we should try to replace the parts that just don't work. Using sampling to followup with those who don't return a questionnaire in the mail-out mail-back phase, is both efficient and systematic, and it offers the potential of correcting problems that we have encountered in the past. It is a way to get more accurate information from many of those who are admittedly hard to count. But even with sampling in its generalization to the other non-responders, there are still going to be omissions and still going to be erroneous enumerations. That is why the Census Bureau has proposed integrated coverage measurement. It will attempt to correct for these errors and in a reasonable way.

Despite some of the biases in the 1990 PES, I and many many others who have studied it with care believe that it was generally successful. The characterization I have heard today by many is wrong. The full plan for the census 2000 offers the prospect of a considerably more accurate census set of counts than anything that would come out of a plan based on the so-called traditional census methods.

I do not want to minimize all the problems and complexities. I wouldn't have counted the number that you had counted, but I can believe that that's how many steps there are in the process. But there are that many steps in virtually every census process, whether it is for some of their best surveys or for some of the things that I had the greatest problem with. But they must be judged in terms of the backdrop of haphazard and biased efforts in the past and what we know about the failings of traditional census methods.

If I can be allowed just a couple of seconds more, I want to end by returning briefly to sampling. Sampling techniques and methods for probability sampling represent one of the greatest statistical achievements of the 20th century. They are used throughout science and they are used by Government in particular. That is because they are both rooted in scientific theory and they work. The evidence shows that they work. The Census Bureau has an enviable record of careful design and implementation of large-scale surveys based on these methods. We, the Nation, rely upon the information that it produces from the surveys. Comparing their work to convenience samples, to market research, and poor political polling

is demeaning and unfair. The proposed methods for 2000 build on an attempt to reach everyone, and then augment that effort by using an approach that the Bureau is well positioned to do. That is, sampling.

There are lots of methodological issues that remain to be resolved. I don't disagree with that. But they shouldn't be interpreted as disagreements about the scientific nature about probability sampling or the fundamental approaches that the Bureau statisticians are taking.

Finally, no matter what the resolution is of lawsuits and the methods that are actually used to take the census, in the end we need a careful plan to evaluate what we have got when the process is done, and to evaluate the extent and the quality of coverage achieved. This will involve sampling once again in the year 2000 just as it has in the past.

Thank you for the opportunity to testify today.

[The prepared statement of Mr. Fienberg follows:]

PREPARED STATEMENT OF
STEPHEN E. FIENBERG
MAURICE FALK UNIVERSITY PROFESSOR
OF STATISTICS AND SOCIAL SCIENCE
CARNEGIE MELLON UNIVERSITY

Before the Subcommittee on the Census
Committee on Government Reform and Oversight
U.S. House of Representatives
September 17, 1998

Mr. Chairman and Members of the Subcommittee:

I am pleased to have this opportunity to testify today on plans for the 2000 decennial census. For the past twenty years I have been engaged in methodological research related to sample surveys and census taking, and I have had different occasions to study and comment upon the work of the statisticians at the U.S. Bureau of the Census. Beginning with my work for the National Commission on Employment and Unemployment Statistics, and continuing through research on the National Crime Survey and two National Academy of Science Panels on methodology for census taking, as well as in my own research activities including the preparation of a forthcoming book on the 1990 decennial census, I have interacted with my professional colleagues at the Bureau, critiqued their work, and gained a deep appreciation for the high quality of their methodological efforts and for their ability to collect high quality survey and census data in the real world. Having recently spent six months as a visiting researcher at Statistics Netherlands, I can also tell you of the great respect the professionals in the U.S. Census Bureau are accorded by their agency colleagues around the world.

The Integrity of the Census Bureau

My first message for the Subcommittee is simple. The statisticians at the Census Bureau are professionals and scientists. They are widely respected as such and work in an open fashion, sharing both their plans and their data with colleagues outside the Bureau, thus allowing others to examine, critique,

and reanalyze their efforts. The response rates of their surveys are the envy of every other statistical agency and survey organization in the world. While they often encounter difficulties by the sheer enormity of their task, and while I often disagree with them on specific details of methodology, priorities, and so on, it is my deeply held belief based on decades of direct observation that they never willfully manipulate or distort the data they collect or the results they report. Their professional openness and the access to their work which they provide to others, ensure that they could not knowingly manipulate census results, whether though sampling or any other means. They have earned the respect of Congress and the American people, as well as that of their professional peers.

Myths About the Census

The U.S. decennial census is our nation's most complex statistical activity, and planning for it is rooted in decades of statistical effort, both methodological and operational. Few of my fellow statisticians, especially those in academia. have any sense of what it means to actually plan, conduct, and then analyze the results of a census. And there are many myths about the census in American history and about the planning for the census in 2000 that are shared by statisticians and non-statisticians alike. Today I would like to address just a few of these, but I am happy to share with the Subcommittee a paper prepared with my collaborator Margo Anderson which addresses several additional myths and provides some detailed background drawn in part from our forthcoming book.

Here is an example. *Myth*: "The census is just a physical head count." *Fact*: At no point in this nation's history was there a physical head count of each person in the country. The first census involved federal marshals compiling lists without actually counting people per se, and from that census to the present day the federal officials who have overseen the census have always wrestled with the problems associated with the accuracy of the so-called census counts. Thus, those who call today for a return to traditional census taking seem blissfully unaware of both what the methods of the past have been and the extent to which errors have come to threaten the usefulness of census data. These errors impact the myriad of uses of census data, from apportionment and the redrawing of political boundaries, through their role

in federal funding formulas, to their role as benchmarks for assessing the status of the nation.

Error in the Census

A widely reported statement, repeated in courts of law and before this Subcommittee, is that the actual enumeration has been highly successful at counting the population and that in particular, "[t]he 1990 census accurately counted 98.4% of the population...." This is another myth. In fact, as Professor Eugene Ericksen, the GAO, and others have repeatedly pointed out, this figure represents 100% minus the net national undercount, and the net national undercount consists of the balancing of many millions of omissions against an almost equal number of erroneous enumerations. In 1990 the GAO reported to a predecessor of this Subcommittee that its comprehensive estimate of omissions was 15.5 million and of erroneous enumerations was 10.2 million. So, while the net amount of 15.5 -10.2 = 5.3 million represents only about 2% of our best estimate of the 1990 population, about 1 in 10 Americans was not accurately counted (15.5 +10.2 = 25.7 million). Because omissions and erroneous enumerations are spread across the country in different and unequal ways, we can't simply let them balance one another out wherever possible. Thus, even if there were a net census error of zero, so that the national total is correct, the census could be plagued by error that would have enormous consequences for census accuracy for states and at lower levels of geography.

Before someone starts to blame the Census Bureau for this level of error, we should all recognize that (i) things would in fact be much worse were it not for the efforts of the statisticians and field staff at the Bureau, and (ii) the only reason we know about the accuracy of the 1990 census is because of their coverage evaluation program, especially the 1990 Post Enumeration Survey. Most previous U.S. decennial censuses were also chock full of error; we just don't know how much error because we didn't have the tools to measure it. Further, given public attitudes towards government, general declines in response rates, and the fallibility of traditional census-taking tools, we can expect the level of error to be higher in the 2000 census, unless new tools and approaches are introduced to control error in a systematic fashion.

The Evolution of Census Methodology

The history of the changes in how we take the census, at least in the modern era beginning in 1940, is one of research, careful testing, and the slow evolution of methodology. In the U.S., we have actually had a post enumeration survey for coverage evaluation since 1950. But even though the Bureau worked with dual systems methodology in the 1950s, another 20 or more years passed before statisticians at the Bureau developed a version of the method which was potentially useful in correcting for the differential undercount. They published the core proposed methods in scientific journals and presented them at public meetings. Members of panels of the NRC/NAS, including those generally opposed to any form of adjustment, reviewed the dual systems approach for undercount estimation and commented upon it from both methodological and operational perspectives. The methodology was then successfully implemented in 1990, albeit with some problems, but the changes and improvements in the version planned for the 2000 census should make things work even better when the integrated coverage measurement survey is implemented this time around.

Similarly, statisticians at the Bureau have developed other applications of sampling and estimation in the census. For example, they implemented a sampling approach for vacancy checks during the operations of the 1970 census and, along with the results of a second sample survey, they used it to add over 1.5 million people to the 1970 count. The notion of using sampling for nonresponse followup appeared in an NRC/NAS panel report in the 1980s and drew in part on a proposal by a distinguished methodologist with extensive experience in sampling and knowledge of census taking. At the time, he was highly skeptical about the use of a post enumeration survey for adjustment, although he later changed his mind about that. At any rate, the Bureau briefly considered the idea of sampling for non-response followup, but precluded its use in 1990 because of the need for careful research and experimentation with implementation. Finally, I note that statistical estimation in census also has a long history especially in connection with imputation methodologies. All of these ideas of sampling and estimation for census taking have been considered and implemented in other countries as well.

Viewed from this perspective, the one-number census proposal for the

2000 census is the natural culmination of improvements in statistical methodology and a seeming decline in the accuracy of the census counts. It is neither a radical departure from the past, nor is it something invented out of whole cloth by statisticians at the Bureau or by members of panels at the National Research Council in the early years of this decade. In fact, the current Bureau plan is formally the outgrowth of a plan launched in 1991 by the members of the then-Republican administration, although its roots run deeply into Bureau research and other activities.

If I were to complain about the Census Bureau's approach to new methods, I would say that it is too slow to try new approaches and adopt statistically-based ideas, such as sampling and estimation. For example, the NRC/NAS panel on which I recently served urged the Bureau to "re-engineer the entire census process." What we as a nation got instead was an incremental change building on the traditional mail-out-mail-back approach.

The Census 2000 Plan

The methodological plan for Census 2000, when set in the context I have outlined, makes intuitive sense and has a sound methodological basis. If the traditional census methods are fraught with error and this error leads to the systematic undercounting of segments of society, we should try to replace the parts that don't work. Everyone seems to agree that the census must attempt to enumerate every household, and this is exactly what the mail-out-mail-back phase has done and will continue to do in 2000.

There is also considerable agreement that attempts to follow up with those who do not return a questionnaire in the mail-out-mail-back phase of the census have been particularly problematic, and unsystematic. Using sampling here is both efficient and systematic, and it offers a potentially more accurate way to elicit information from many of those who are admittedly hard to count. But even with sampling and the generalization from the sample results to all nonresponders, the new census approach will produce omissions and erroneous enumerations, although hopefully many fewer than what we found in 1990. Thus we still need a post enumeration survey to correct for these errors and thus complete the count.

I do not wish to minimize the operational complexities of the census 2000 plan and the fact that tricky methodological issues remain. But they must be judged in terms of the backdrop of the haphazard and biased efforts to complete the count which were used in the past.

Administrative Records as an Alternative

"There must be a better way to do an accurate census without sampling." Interested parties have made this or similar remarks over the past decade. In *Modernizing the U.S. Census*, the NRC/NAS panel on which I served described some radical alternatives to traditional approaches to census taking, including the use of administrative records.

Although administrative records have been used to support census taking for years. they are simply not a viable alternative to regular modes of census data collection or to the sampling components of the census 2000 plan. There are complex problems of *accuracy* and *quality* associated with all of the administrative record systems we reviewed, and data we currently collect as part of the census are simply unavailable in administrative form. But even more importantly, because of the way administrative records are stored and sometimes dispersed among the states, *accessibility* is a crucial problem and the *privacy* and *confidentiality* issues abound. Finally, even if data existed and were accessible, the *matching* of administrative records poses far greater problems than anything associated with the census 2000 plan.

Sampling is Scientific and Can Help Improve Census Accuracy

I wish to end this statement by returning briefly to the issue of the appropriateness of sampling for the collection of government information. Sampling techniques and methods for the selection of probability sampling represent one of the great statistical achievements of the twentieth century. These techniques are used throughout science and government because they are rooted in scientific theory and because they work. The U.S. Census Bureau has an enviable record of careful design and implementation of large-scale surveys based on these scientific methods, and the nation relies upon the information

it produces from these surveys. Comparing their work to convenience samples from market research and poor political polling is demeaning and unfair.

The proposal for census 2000 builds on an attempt to reach all residents of the U.S., and then augments that effort by using an approach that the Census Bureau is well positioned to implement, that is sampling. There are methodological issues about the details of the plan that remain to be resolved, but this should not be interpreted as disagreements about the scientific nature of probability sampling. Regardless of the methods that are ultimately used to take the decennial census in 2000, we need to have a careful plan to evaluate the extent and quality of coverage achieved, and this too will involve sampling, as it has in the past.

Thank you for the opportunity to come before the Subcommittee today.

References

Anderson, M. (1988). *The American Census: A Social History.* Yale University Press, New Haven, CT.

Anderson, M. and Fienberg, S.E. (1998). *Who Counts? Census Controversies for the Millennium.* Paper presented at the Joint Statistical Meetings, Dallas TX, August 13, 1998.

Citro, C.F., and Cohen, M.L. (1985). *The Bicentennial Census: New Directions for Methodology in 1990.* Report of the National Research Council Panel on Decennial Census Methodology. National Academy Press, Washington, DC.

Edmonston, B., and Schultze, C., eds. (1995). *Modernizing the U.S. Census.* Report of the National Research Council Panel on Census Requirements in the Year 2000 and Beyond. National Academy Press, Washington, DC.

Ericksen, E.P., and DeFonso, T.K. (1993). Beyond the net undercount: How to measure census error. *Chance*, Vol. 6, No. 4, 38–48, 14.

General Accounting Office (1991). *The 1990 Census. Reported Net Undercount Obscured Magnitude of Error.* GAO/GGD-91-113. U.S. Government Printing Office, Washington, DC.

Steffey, D.L., and Bradburn, N.M., eds. (1994). *Counting the People in the Information Age.* Report of the National Research Council Panel to Evaluate Alternative Census Methodologies. National Academy Press. Washington, DC.

Mr. MILLER. Thank you. We'll begin the questioning. Mr. Davis has time pressure, so I am going to give him my slot.

Mr. Davis.

Mr. DAVIS OF VIRGINIA. Let me start. I've just got a few questions. Dr. Ericksen, let me first of all thank you for your testimony. You were on the panel last time that reviewed the PES under Secretary Mosbacher. Is that correct?

Mr. ERICKSEN. That is correct.

Mr. DAVIS OF VIRGINIA. As I understand it, the recommendation, and I assume you supported it, would have allocated a House seat from Pennsylvania to Arizona had Mr. Secretary taken the recommendation of that panel at the time?

Mr. ERICKSEN. I don't recall what the actual States that were involved.

Mr. DAVIS OF VIRGINIA. But the secretary did not accept that recommendation. Correct?

Mr. ERICKSEN. That is correct.

Mr. DAVIS OF VIRGINIA. Do you still feel in retrospect that the recommendation you made was the appropriate one?

Mr. ERICKSEN. The recommendation I made was that the adjustment based on the PES would improve the shares of distribution across the United States. I still stand by that recommendation.

Mr. DAVIS OF VIRGINIA. I would note my understanding. That would have given Arizona an additional seat.

Mr. SHADEGG. I would have liked to see it.

Mr. DAVIS OF VIRGINIA. OK. I just wanted to get that on the record. Let me ask Dr. Bryant a few questions.

Your testimony up here notes on page 4. It says on your second point, it says they merely top off the count with an estimated count of that small percent who either do not respond by mail or cannot be reached, despite multiple efforts of census personnel. Is that what you mean?

Ms. BRYANT. My feeling is that the public is being given an impression that it's a sampling instead of enumeration. In actual fact, it will be all of the traditional methods of direct enumeration, even better actually, because there will be multiple chances to get the questionnaire and multiple mailings of the questionnaire. There will be the traditional calls on non-responding households. But then there comes a point at which instead of, as Eugene Ericksen has said, you go on just, you know, beating this dead horse to death and taking longer and longer and longer to do it, that you go out and sample those who in a sense have been so recalcitrant up to that point that you don't have whatever you don't have.

Mr. DAVIS OF VIRGINIA. Maybe you and I have a different understanding, and let me understand it. There is only one mailing. There is not multiple mailings, it is my understanding.

Ms. BRYANT. There is a second mailing if the person does not respond in the 2000 census. There was not in 1990. Coming from the private sector, I said we would never do a mail survey without——

Mr. DAVIS OF VIRGINIA. My understanding is they have canceled the second mailing. That's what the staff is telling me. Would that affect your decision?

Ms. BRYANT. I did not know that. As I said, I am the past Director.

Mr. DAVIS OF VIRGINIA. Your testimony is based on an understanding of the second. That's fine.

Ms. BRYANT. I supported a second mailing.

Mr. DAVIS OF VIRGINIA. Well I think they ought to do everything to enumerate everything they can, before they get to whatever else they are going to do.

Second, in merely topping off the count, topping off is, merely topping off the count would include do you know how many people?

Ms. BRYANT. It would include better methods for getting the people——

Mr. DAVIS OF VIRGINIA. Topping off the count is like 20 million people or something, 26 million people?

Ms. BRYANT. Well, we can argue about what topping off is. I am saying that it would—you know, you have baked this cake with enumeration and now you are going to ice it with something that gets you up closer to the 100 percent.

Mr. DAVIS OF VIRGINIA. Here is my concern. I just want to understand. They merely top off the count, which kind of is—merely top off the count with an estimated count of that small percent who either do not respond by mail or could not be reached despite multiple efforts of census personnel. It is not a small number. It is a very large number, it's my understanding. Maybe——

Ms. BRYANT. It could be as large as 10 percent, and if people cooperate, it will be less than that. The more people cooperate, the less it will be.

Mr. DAVIS OF VIRGINIA. Just for the record, my understanding is at this point they are not planning to followup with that second mailing. That gives me some cause of concern. But I appreciate your testimony being based on the fact of what might have been considered earlier with a second mailing.

Another question. I'm just going through your testimony. You talk at the end on page 8 about unfortunately we will not be missing merely a row of bricks, we'll be missing a few bricks here, a few more there, have a slight oversupply at some points. As I understand back in 1990, going back to 1990, the sampling would have deleted about 1.5 million people out of the count.

Ms. BRYANT. My row of bricks analogy is because the differential undercount is not evenly spread. If it were, none of us would be worrying about this because then the proportion of each State would come out the same.

Yes, there are people deleted who have been overcounted. They tend to be people like myself, a senior citizen woman, homeowner with possibly two households. I send in my questionnaire from both of them.

Mr. DAVIS OF VIRGINIA. I don't think they are going to miss you. I guess this goes to again the language, you know, in this case topping off actually could be as much as 10 percent. A few bricks, 1.5 million people out of a count possibly. Just it's all in the analogy. It doesn't affect the substance. But those aren't bricks, those are not even cinder blocks. Those are big concrete slabs when you start talking about numbers. It makes a lot of us a little bit nervous. We have had a lot of testimony on this. We want to get the best count that we can.

Ms. BRYANT. The thing that makes me nervous is the size of the undercount. I want to see the best possibility of fixing that differential undercount.

Mr. DAVIS OF VIRGINIA. Well we all want to do that, but we don't want to do it—we want to do it in a correct way. I think one of my concerns is that you improve a little bit here, but you disapprove—for example, eliminating that second mailing. I am not sure we are doing everything in the initial enumeration. I think too much emphasis is being put on the polling or sampling at this point. We could get a better upfront count. I think, I would hope you could agree with us on that. That's a great concern.

Ms. BRYANT. Well the better the enumeration, the better the census. But just extending it on, and trying to call and call and call on these recalcitrant households, we know does not improve the quality of the count. That is why there's many others that think that we have to do some changing.

Mr. DAVIS OF VIRGINIA. You have just testified that a second mailing you would support would be in order. We find that they have eliminated that. So I think they could do more. I think you agree with that. Thank you.

Mr. MILLER. Mrs. Maloney.

Mrs. MALONEY. I would really like to go back to a statement that my dear friend and colleague from the great State of Virginia just raised about the shifting of a seat from Pennsylvania to Arizona. Ever since I started working on this subcommittee, there has been a lot of talk about a computer, so-called computer error shifting a seat from Pennsylvania to Arizona. Is that correct, Dr. Fienberg? Can you explain what really happened? If anyone else would like to add to it, I would like it clarified.

Mr. FIENBERG. I can attempt to explain because I have been trying to sort out all of the materials and reports on this for the book that I am working on with Professor Anderson.

The computer error was described in several census documents, and it was discovered after the decision by Secretary Mosbacher in July 1991. The Census Bureau staff then did a number of things because they were planning, as Professor Ylvisaker explained a little earlier, for a report that would deal with inter-censal estimates. They did a series of changes along the way for a different purpose.

One of the things they did was they fixed up numbers that they thought they could fix up, including the so-called computer error. Then they changed the whole method by which they thought, for inter-censal purposes, they should be producing stratification structures. It wasn't the computer error that shifted the seat or would have shifted the seat. It was the combination of all of the things they did, including the restratification scheme. Those were all done after the fact. That is one of the reasons why specifying methods in advance and living with them, even when they do produce biases, is part of the plan for the year 2000 and it was the plan for 1990 as well.

Mrs. MALONEY. Anybody else like to comment on this?

Ms. BRYANT. Yes. I would like to comment on it because if you look at, we had an initial count. Then, we had the demographic analysis that showed the net undercount was 1.8 percent. One of the former panelists sort of agreed that maybe that was the best

estimate of where truth was. We did the post enumeration survey and came out with a 2.1 percent.

Later, as part of the study, both for the study going ahead to how do we improve for 2000, but more immediately to do with post censal estimates, went through and did a super super super evaluation of the post enumeration survey and did find this computer error. We also knew that we had to specify a year in advance exactly the techniques we would use, which I think the Census Bureau will do again in 2000, so that there can be no questions of manipulating, and made some other improvements, one of which is showing up in this 2000 plan. That is, to stratify the State level since you are going to be doing this for apportionment.

If you look at the 2.1 post-enumeration survey, the 1.8, the demographic analysis, the 1.6, which was the refined post-enumeration survey, any one of those three is closer to the real population of the United States than the census as enumerated and as left unadjusted.

Mrs. MALONEY. Thank you. Dr. Bryant, would you comment or would you agree with the characterization of the previous panels, that reducing costs drove the design of the 2000 census and that accuracy was secondary? The talk about the National Academy of Sciences writing so extensively on cost.

Ms. BRYANT. There was a congressional mandate in setting up that panel.

Mrs. MALONEY. By mandate, you mean a law? There was a law?

Ms. BRYANT. It was instructions. I don't know how formal or informal.

Mrs. MALONEY. It was a law, the one that we passed.

Ms. BRYANT. The two reasons for setting up that panel were given as improve accuracy and reduce cost. The Congress was very upset about the fact that we had a larger differential undercount than had been measured in 1980, and that costs had gone up more than the inflation of price and population. So the National Academy of Sciences was under that mandate to look at cost.

Mrs. MALONEY. So in other words, they were instructed to write and to take into account cost.

Ms. BRYANT. That was a particular concern of Hal Rogers, who was then the ranking minority member. He would now be the majority chairman of the Budget Subcommittee that oversees the census budget.

Mrs. MALONEY. Because we had continued to spend more and more money on the census, yet the undercount had grown and in fact gotten larger.

Ms. BRYANT. I think the panel in the end of course put their focus on how do you improve accuracy rather than cost alone.

Mrs. MALONEY. My time is up. So I hope we'll have a second round.

Mr. MILLER. I mentioned before this issue of trust. What we have to decide and make these decisions as we go through to the 2000 census, is so fundamental to our democratic system. We have to have census that is trusted by the American people. We all understand that. We're talking about statistics today, but the issue of trust has to be there. When we make such a dramatic change, as has been proposed this year, the burden of proof is on the new di-

rection they are going. That is where we have this issue of trust. How can we trust these numbers?

As I said, there's so much cynicism in this country about what goes on in Washington right now on all kinds of areas. You say well, we don't have any political appointees but one at the Census Bureau, correct? But you know, 1996, the Immigration and Naturalization Service only had one political appointee and thousands and thousands of people were illegally made citizens in order to vote, and there was only one political appointee there.

So we have every reason and the American people have every reason to say wait a minute, do we trust this. What you are saying is we have fancy computers in Washington and we have all these smart Ph.D.s up here. Trust us and we'll give you one number. We're not going to let you see all the details. We're only going to give you one number because we are afraid you will check it.

You talked about the issue, Dr. Fienberg and Dr. Bryant, about over the time there were three different adjusted set of numbers, as we worked on that. But it took a couple years to work through that process. You know, this proposed in the year 2000 is only 5 weeks to come up with adjustments. Are you aware of the 5 weeks? I mean GAO has got great concerns about the 5 week issue. Are you?

Ms. BRYANT. Well, anything to do with the census, you always have enormous timing problems because this is, as you have said yourself, the biggest operation outside of war.

I point out that the Census Bureau benefits from that, it has been working on this over a 10 year period and even going back further if you go back to 1980. So I think a lot of improvements have been made. The technology is much better, the computer programs for matching. The interviews will be done using laptop computers, which will make—I mean the ICM interviews, which will make everything go faster than before.

As far as trust, this is one reason why I wanted to point out that this is not a program that had been developed by the Clinton administration, which obviously has a trust problem right now. This is a program that has been developed by professional statisticians at a Bureau that is, I think, world class and that it is recognized in the rest of the statistical world as world class, and I think it is incumbent on this committee to show some trust in the Bureau.

Mr. MILLER. Dr. Fienberg, do you feel bothered by the 5 week constraint, that you think they can be so perfect in 5 weeks and it took years before in 1990? I mean there was a problem in the 1990, you would recognize?

Mr. FIENBERG. No, no, no.

Mr. MILLER. Oh, you think 1990 was a success?

Mr. FIENBERG. I think that 1990 was a success. Yes.

Mr. MILLER. Oh, the numbers were perfect the first time around?

Mr. FIENBERG. No.

Mr. MILLER. Well then it wasn't a success.

Mr. FIENBERG. Numbers are never perfect when they come out of statistical agencies for any purpose.

Mr. MILLER. Why should we trust these numbers?

Mr. FIENBERG. They are always constrained by time, by setting, and by the data collection context.

Mr. MILLER. Five weeks is enough time to get the best numbers possible, that we should trust them? I mean it took a couple years, 22 months in 1990, and they changed it three times.

Mr. FIENBERG. The 22 months represents a number of different efforts, not all of which were geared up to produce the counts for the Nation. We have now had a tenure learning effort. A computer error program in 1990 is not a computer error program for the year 2000. It was fixed. We don't have to spend a year trying to discover it.

Mr. MILLER. Let me go to another issue. Local review. They are not doing that this time around. Do you agree with that?

Ms. BRYANT. I agree completely. Local review was a public relations disaster and a statistical disaster.

Mr. MILLER. But isn't that something about trust? I mean the local review——

Ms. BRYANT. No. I think——

Mr. MILLER. Congressman Petri, didn't Congressman Petri's district find a mistake? I mean in the experience in Columbia, SC, they missed a whole zip code. I think Congressman Petri, there's something they missed in his district and local review determined that. And are you saying we are so good in Washington, we are not going to trust local people?

Ms. BRYANT. No. I think that they have got a better thing going than local review. That is this program in which thanks to the work of the Congress in changing the law so that the Post Office and the census can compare mail addresses and they can be compared with local. You know the local communities can look at the mail address in advance.

Local review was really to look at the mismatch between addresses and blocks.

Mr. MILLER. Didn't it turn up some mistakes in 1990? Congressman Petri, for example. I remember him testifying or stating that they found a——

Ms. BRYANT. Local review turned up all of these complaints that sent us back into the field with this very very expensive re-canvass that produced virtually nothing. Yes, there were errors in geographic coding, but remember, that was the first time we ever had the geographic mapping system. Those kind of improvements or corrections made in the field are now permanently in place.

Mr. MILLER. I am still scared of a rather risky plan.

Congressman Shadegg.

Mr. SHADEGG. Thank you, Mr. Chairman. Thank you members of the panel for your testimony. I appreciate it. I have got to tell you that in part, I don't really know where to begin. I have got so many questions.

Mrs. Maloney, I will give you multiple additional extensions as long as the chairman will agree to it because I have got a gazillion questions.

I do know one thing. Arizona would have liked that other seat.

Mr. FIENBERG. So would Pennsylvania.

Mr. SHADEGG. I also know, Doctor, you pronounce your name——

Mr. FIENBERG. It's Fienberg, and it is spelled F-I-E.

Mr. SHADEGG. OK. Fienberg, Dr. Fienberg. I know that I certainly agree with you on one point. That is, you have got to set a

set of rules and you have got to say those are the rules and they cannot be changed after the fact no matter what you discover, because I am deeply worried about the issue of integrity. I think the chairman hit upon this. I will tell you, I believe there is almost nothing more important that we can do in this system than to do this count accurately and to make the American people have confidence in it.

We have been talking about, the chairman has been talking about, errors at the congressional level. I worked in Arizona during reapportionment in the last cycle. I hired the statistician who did the computer work. I stood beside him as the lawyer. When I watched him doing those tiny little census blocks and making decisions that affect people's lives at a very finite level, tiny little areas, areas we have in Arizona. We have census tracts with very few people and census blocks with very very few people. I worry deeply about the consequences of that. I assume you agree with me?

Mr. FIENBERG. I do agree, but it is also the case that error for those tiny census tracts is not ultimately the issue because when you do apportionment, you take them and you piece them together. So it's errors at the boundary, when you take the collection and when you add one and subtract one out. So I believe that you have characterized the issue. The question is——

Mr. SHADEGG. Have you ever sat in front of the computer and watched the line that says you are in this district and you are out of that district?

Mr. FIENBERG. Yes.

Mr. SHADEGG. Well at that point, you may be at the boundary, but the errors in the census tract at that boundary matter.

Mr. FIENBERG. But for the collective that you produce at the end, what I would argue is that the numbers are quite different and much more stable.

Mr. SHADEGG. My friend Ed Pastor and I have radically different philosophies. I have constituents, people in Arizona, who are his constituents who intensely wish they were my constituents. I'll bet you he has constituents who feel intensely the reverse of that.

You testified, Dr. Fienberg, that in the 1990 census, which you believe was a mistake, there were 25 million errors. Is that a correct assessment?

Mr. FIENBERG. That number comes from the GAO report. It is it's comprehensive estimate.

Mr. SHADEGG. And you accept it?

Mr. FIENBERG. Yes. I actually believe, depending on how you count, you can get more. Indeed, if you look at Professor Ericksen's testimony, you will see an even higher estimate.

Mr. SHADEGG. Given what you know about the sampling plan for 2000 and the experiences in 1990 then, is it fair to say that mistakes will be made in the sampling census as well?

Mr. FIENBERG. In the sampling component of the census?

Mr. SHADEGG. Yes.

Mr. FIENBERG. Or in the entirety?

Mr. SHADEGG. Actually you can answer both.

Mr. FIENBERG. The answer is in every statistical data collection I know in practice there are errors. The only question is how big

they are and whether one has a mechanism for estimating them and perhaps correcting for them.

Mr. SHADEGG. Isn't it also true that errors in the survey portion, as a result of the fact the survey portion, to use Dr. Bryant's words, merely tops off, those errors actually have the potential of magnifying themselves. Don't they?

Mr. FIENBERG. No.

Mr. SHADEGG. You take a sample. You determine that that is the correct sample and you apply it to the larger effort. If there is a larger block that you are applying that sample to, that is going to magnify the overall error when the small sample is applied to the larger group. Isn't it?

Mr. FIENBERG. You are asking, if you attempt to generalize from a sample to a population and you have errors, the errors do magnify. Yes.

Mr. SHADEGG. And you just explained to me that that's exactly what is going to happen here. Right? You just told me that my concern about the error in a census block in Arizona, rural Arizona, that has 20 people in it, doesn't matter because you are going to assemble all those tiny census blocks into a district and so it really kind of doesn't matter.

Mr. FIENBERG. I didn't say that. You have conflated two different comments. They don't go together.

Mr. SHADEGG. All right. Is it not true that under the plan developed by the Census Bureau, they are going to take the sample, and as Dr. Bryant says, they are going to add it on the top? They are going to project from a small sample and say we have got to add this and we're going to say there are so many missed. We are going to add to that and we are going to base it on this little sample?

Mr. FIENBERG. It is not so little, but yes. We are going to generalize from the sample to the population of non-responders. Yes.

Mr. SHADEGG. So clearly there is going to be a magnification at that point?

Mr. FIENBERG. Yes.

Mr. SHADEGG. So can't one argue that by necessity, sampling is going to magnify existing errors in the system?

Mr. FIENBERG. Except that balancing the sampling error that one gets because you have sampled rather than gone to the entire population, is the elimination of two kinds of errors, omissions and erroneous enumerations. In particular, erroneous enumerations should, under the current plan, be drastically curtailed because of the timing issue and how it relates to the timing that Professor Ericksen explained a little while ago. That was roughly half of the overall error in the census.

Mr. SHADEGG. As I said, I have dozens of questions. Let me just, if I could comment on the end of that. My deep concern is that when you do that process and you magnify the one upon the other, you are going to in fact make more serious errors. I would cite as an example right now that we are told across America, that Americans believe by numbers in the 60 percent, nothing should be done about the President right now. But if you applied that to a small group of people, can we apply that broad generalization about the American people down to the State of Arizona, down to Maricopa County, down to my congressional district?

What about local characteristics that make it different? I am told, for example, national polling says the President's support is way up there at this high level. You go to the State of Utah and the same statistical broad poll gives you a dramatically different result because the people of Utah are looking at this issue differently. I have got more questions.

Mr. MILLER. There was an article I mentioned earlier, and I don't know if any of you all could be aware of this article in the Associated Press. It was a Tuesday article in Vermont about the unemployment rate dropped. It was the lowest rate since 1988. It dropped six-tenths of a percent in 1 month. The article in the paper says the department warned that the sharp decline may have been caused by a statistical sampling error by the U.S. Census Bureau. The numbers will be revised by the end of the year and then, more accurate data will be available.

I assume you don't know anything about the details of this and I'm not going to ask you. But this type of error happens. This is what creates doubts. You say the issue of polling—President Clinton used the word polling when he was in Houston with you talking about the census, at a very specific hearing, not a hearing but a program on the census. He just starts talking about polling because he obviously believes in polling a great deal. But does that, I mean this happens a lot I guess. I mean they are admitting a sampling error with this AP.

Mr. FIENBERG. Well, I would suggest that there's a misplaced article. I have a friend who says that newspapers are never accurate when they relate to something he knows about, but he always believes they are accurate for something that they don't.

I would guess that the original statement, because I have seen statements like this before, said may have been caused by sampling error, that is uncertainty. Therefore, we have to wait until the end of the year to have enough data and aggregate it in order to——

Mr. MILLER. That's the problem. You have to wait. We're only going to have 5 weeks to make these decisions in the year 2000 if it's proposed.

Does it bother you, by the way, that you know, we have three Federal judges in a unanimous decision. One was appointed by President Clinton, has said it's against the law. It ruled on the constitutionality of it and we're going to drag this out. I mean we have another three judge panel that will rule soon too. I mean we just need to move forward on this. I mean the sooner we make a decision, the better. Is that right?

Ms. BRYANT. The Census Bureau plan was set up with a review of all the rulings prior to this one. As I said in my testimony, the weight of case law was that sampling was not precluded for improving the accuracy. That goes back to all of the looking at whether Secretary Mosbacher would or would not have had the right to make a decision for adjustment. It goes back to all of the cases by cities against the Census Bureau.

Mr. MILLER. I'm not a lawyer.

Ms. BRYANT. I'm not a lawyer.

Mr. MILLER. I think it's wrong to use sampling, not because of the legal reasons, but my understanding is that the judges ruled

on the basis of law on this plan. In 1990 and in the other cases, rulings were about adjustment to a full enumeration, which is a different issue. That is the reason these three judges ruled not only on the standing issue, but as a unanimous.

Ms. BRYANT. The Department of Justice did give an opinion in 1994, I believe it was, that it was all right to go ahead with this sort of a plan.

Mr. MILLER. Well that's where you get into these trust issues. That is the reason we have three Federal judges and we're going to have more judges.

Ms. BRYANT. As I say, if this one goes to the Supreme Court, then I will have nothing to do with it.

Mr. MILLER. Well this is the first time they have ruled specifically on sampling and not adjusting the full enumeration. So there is a difference. To keep saying all these other courts have ruled that way, well that was the adjustment in 1990. That is a different issue than the fact that we're not counting, we're not going to do a full enumeration.

Let me clarify that, we are not going to count 10 percent. Is that right or wrong?

Ms. BRYANT. As I said, if more than 90 percent turn in their questionnaires, every questionnaire is going to be processed.

Mr. MILLER. But realistically, you are——

Ms. BRYANT. But the realistic thing is that last 10 percent you can't do a good job on. That is what Professor Ericksen testified.

Mr. MILLER. Is the last 10 percent the hardest to count 10 percent?

Ms. BRYANT. Yes. Because those are people who have not cooperated, not just with the mail but with the fact that the Census Bureau still goes out and makes followup calls to try and get them.

Mr. MILLER. My understanding, correct me if I'm wrong, is that let's say in an area, a census tract, 60 percent of the people respond. OK? By mail. We'll use easy numbers. My understanding is they will take 30 percent and they'll do the followup. The other 10 percent they won't. Is that your understanding? That is your understanding?

Ms. BRYANT. Yes.

Mr. MILLER. So you are saying——

Ms. BRYANT. A randomly selected 30 percent.

Mr. MILLER. Then if it is randomly selected of that 40 percent, it can't be the 10 percent hardest to count.

Ms. BRYANT. No. You are right. That would not be the hardest to count. Your question to me, as I heard it, was are the last 10 percent the hardest to count. If you are talking about direct enumeration, yes. The last 10 percent are the hardest to count.

Mr. MILLER. Earlier when we were talking, Dr. Brown talked about this matching problem, and that you really need to get a very high percentage of accuracy in the matching. That doesn't bother you, if you only get—he says we should get up to 97 or 98 percent of the matching. That's the key to the ICM. Right?

Mr. FIENBERG. No.

Mr. MILLER. Oh it's not? Oh, OK. If they don't match, it doesn't really matter?

Mr. FIENBERG. No.

Mr. MILLER. Because trust is in Washington. We have smart people and big computers and we can do it. Is that what you are saying?

Mr. FIENBERG. The answer is that if you could match perfectly you would get better numbers and better estimates. But the 98 percent number I have never—I have heard him say this before. I have heard others say it. I have never seen a calculation or a document that would explain why.

Mr. MILLER. What number do you think we should try to have a match for?

Mr. FIENBERG. I think we should aim to get them all read, but I don't think that we will.

Mr. MILLER. What we are doing is looking at this, and it's been stated before, looking at one set of errors you might like versus another set of errors you may not like. So there are going to be errors either way. We know there is going to be an undercount. But to say that we should trust this system whether it's the Census Bureau, which is going to be President Clinton's Census Bureau, assuming he is still in office, that will be doing that. We still have concerns about this trust issue.

Mrs. Maloney.

Mrs. MALONEY. First of all, I would like to allow Dr. Bryant an opportunity to respond to my colleague, Congressman Shadegg's question about the error on the census tracts. Would you like to respond to that set of questions from practical experience?

Ms. BRYANT. I have been in—I was with a company in 1980 that did reapportionment, so I have looked on the screen and watched reapportionment being done and the moving of a block here, a tract there and stuff like that. But always the test of how good each plan is that you aggregate those. That you aggregate those census tracts up and then look at did you come out with equal population. If you are on one side or the other, what does the past vote look like it would be.

So the tests are always done at this aggregated level of the legislative or congressional district. They are not done down at the block level. You will go through a number of alternatives.

Mrs. MALONEY. I would like to ask Dr. Fienberg, all the previous witnesses based their objections to the post-enumeration survey on Dr. Breiman's estimate that most of what was measured by the PES was error, not true undercount. Would you explain to me whether or not you agree with that assertion and why?

Mr. FIENBERG. Well I disagree with his characterization, in his use of the term bad data. I could go through his tables and explain why. I did in my comments that accompanied his. But there is a much more fundamental flaw in the reasoning he presented to this committee. It is wrong and it's potentially misleading. We should understand it. So let me try to give an example.

Let's suppose that this plan for the year 2000 takes place. I want to give a hypothetical. And that there are 30 million erroneous enumerations that occur in the census. They all happen to happen east of the Mississippi. Then there are 30 million plus one omissions. They all happen to happen west of the Mississippi, which includes Louisiana and California and Arizona. The net census undercount is one. The difference between those is one.

Then the census comes along with the ICM and actually measures the errors. By god, they get it right except for one omission. Professor Breiman would then tell you that 100 percent of the net census undercount is due to error, bad data. Yet Congressman Miller and Congressman Shadegg would be ecstatic to have all the omissions corrected and have all of the erroneous enumerations taken away from New York and Pennsylvania, where I come from, and elsewhere.

So the Census Bureau would get it almost perfect, unbelievable results. Yet Dr. Breiman's benchmark would tell us that 100 percent of the census estimate for undercount was error. That is just flat out wrong. That same thinking applies to a somewhat lesser degree to all the calculations and all the discussion in his paper.

Mrs. MALONEY. Dr. Fienberg, would you explain to us the difference between the 1990 post enumeration survey and the 2000 integrated coverage measurement, and how those changes affect the quality of the any adjustment to the census?

Mr. FIENBERG. Size. They go by different names. But the surveys have essentially similar characteristics. The new one is five times the size of the other. That is a big big difference. In all my re-analyses, in all the re-analyses that I examined from 1990, the biggest shortcoming of the post enumeration survey was its size. The National Academy of Science's panel said essentially the same thing.

So the biggest change is size. There's a different structure to the design as a result of the size. There is very different timing for this as a result of the plan to integrate this with sampling for non-response follow-up. The Census Bureau will be able to take it into the field in a different fashion, implement it with different care, and in light of the reduction or expected reduction of erroneous enumerations that would come about from sampling for non-response follow-up, the expectation is that all of the processes to check on quality will actually occur in a more direct and easy to understand way.

I can't promise it is going to be better. There are still details to resolve. But if ever there was anything that had a shot to improve things, I think this is it.

Mrs. MALONEY. My time is almost up before I start on a series of other questions.

Mr. MILLER. Mr. Shadegg.

Mr. SHADEGG. Thank you. Let me just start by saying I believe very strongly that as statisticians, you can have passionate views about this issue on one side, and as policymakers we can have passionate views about this issue on the opposite side. I don't think you have to consider, indeed, your profession doesn't require you to consider the policy implications that we have to consider. That is why I go in part to this issue of trust.

Before I go forward, I want to say, Dr. Bryant, I don't know where you did your redistricting, but I think you and I were on different planets. Because in point of fact, in Arizona where you have to live within the Voting Rights Act, it is simply not true that "everything is done in the aggregate." I watched and I know that we went down the street block by block, literally in rural Arizona. This family, this household in, that family, that household out. We did that all through the process. Quite frankly, under the Voting

Rights Act, which was my legal expertise, we said Hispanic families in, white families out. Native American families in, white families out. It gets right down to specific houses. So the notion that it's all just a big generalization is simply not true, at least not in a Voting Rights Act State like Arizona.

That takes me, in case I don't get a third round, to a huge issue for me, which I ask any one of you to explain to me because I can't understand it. The Census Bureau plans, specifically says point blank, that we are going to do enumeration for all of America except we are not going to do sampling—I'm sorry, we are going to do sampling for all of America, but we're not going to do it for American Indian reservations, Alaskan Native villages, Virgin Islanders, or in the Pacific Island Territories. Now A, on a policy basis, I do not know how you can justify doing it for all of America and not doing it for that group.

I have Indian reservations in my State, many, many, many of them. I think all of America knows that Native Americans was alleged the most undercounted category in America last time. I have even beyond the policy question of saying we're going to do it for white Americans, we are going to do it for black and Hispanic Americans, but we are not doing it for these categories, I have policy questions on that.

But beyond that, I have equal protection problems. Why in the world would not people in those areas have a due process right and an equal protection right to sue and say this is a completely bogus census because they never gave a reason which could be justified under the Constitution for applying a different set of rules. I offer all three of you a chance to respond.

Ms. BRYANT. I repeat of course that I am not the census director and do not know all the current details.

Mr. SHADEGG. Don't worry. They are not willing to answer either.

Ms. BRYANT. Indian reservations, as you know, are not Federal territory, not State territory. They are independent nations. Therefore, are always dealt with somewhat differently on the census, with the cooperation of the rulers of those independent nations. Alaska Native villages are a very unusual weather problem because most of them are fishing villages. You have got to get in there and get them enumerated before the ice melts. Then once the ice melts by April 1, when we do the rest of the country, a lot of the people are out fishing. So that's why there are actually special things done for them that are probably better and more complete just because they have to do them early, when everybody is locked in by the snow up there.

Mr. SHADEGG. Does that explain simply not applying the sampling techniques? If sampling techniques are good for everywhere else, why wouldn't they be good there?

Ms. BRYANT. It would be because you would be going back later after the ice had broken up and you would lose some of those people. But I don't want to——

Mr. SHADEGG. Reservations get snowed in.

Ms. BRYANT. I don't want to comment on specific operations of the 2000 census. That should be commented on by the Census Bureau or the incoming director.

Mr. SHADEGG. Dr. Ericksen or Dr.——

Mr. FIENBERG. I agree with you completely about the principles. Without seeing the detail of the plan, I can't say for sure, but it would be my expectation that under the plan as I have read it in the past, without these details there, that those populations would in fact be well in excess of the cut-offs that we've been talking about for sampling through these other special processes that are used for them. If they weren't, that the Census Bureau should be asked and indeed should be directed to do something about them.

Mr. SHADEGG. I guess I am confused by your answer. The Census Bureau is going to apply sampling techniques to every census block and tract in America, to the whole system, except for the areas that I have talked about. If you happen to live, by the way, in a Native village and you are not Native, you are an Anglo, well we're not going to apply sampling to you. So a Caucasian American in a Native village somewhere in Alaska gets treated differently than a Caucasian American somewhere else in the world?

Mr. FIENBERG. The people in those areas are enumerated in a different fashion than the mail-out mail-back enumeration that's applied to the rest of the Nation. So there are additional procedures. If they don't cover it, they should. So I agree.

Mr. SHADEGG. Just to followup on that point, if they are enumerated in a different fashion, then the only reason not to apply sampling to them must be that enumeration itself there is better than sampling. And if enumeration is better there, it's good enough, it doesn't need to be improved by sampling, why isn't enumeration good everywhere?

Mr. FIENBERG. No. I didn't say that. It is because of the special——

Mr. SHADEGG. That's the fact.

Mr. FIENBERG. No. It's a special set of techniques applied in special settings which provide a different yield because you start with a different base. You could apply sampling there and it may or may not depending on the timing, as Dr. Bryant has said, improve the count. But it may not.

Mr. ERICKSEN. Every census that I know of, going back for the last three or four censuses, there have always been special populations which require special methods of counting. Another example would be the population of homeless people. These are usually populations which are small in number with very particular characteristics. The method which is used for well over 99 percent of the remaining of the population in the statistical judgment of the Census Bureau doesn't work for them. Virtually every one of those special situations that I know of, the Bureau's alternative method has been a common sense, clearly superior way of doing it.

Were I able to review the details of the situation on the Indian reservations, it's quite likely that their method is better for that particular situation which is quite unique in a number of ways.

Mr. SHADEGG. If I could just conclude my comment. We have a principle in this Nation called equal protection of the law. There are people in rural Arizona whose living circumstances and therefore, whose ability to respond to an enumeration is identical, though they live off of a reservation and they are not Native American. How can I go back and tell them this is a fair system? Their living conditions, their rural lifestyle, their lack of mail coverage is

identical to the Native Americans living next to them, but a different rule applies. I don't think it can be justified.

Mr. MILLER. All right. We'll conclude the hearing now. One thing, let me thank all eight witnesses. I appreciate it. It was a longer day than I think any of us were expecting, but it was very important, I think, to have eight eminent statisticians here. We find there is a real division of opinion within this community, that is not all 100 percent of them think one way or another way. There are concerns and problems whichever way we go.

Let me have a couple of unanimous consents. In case there are additional questions that Members may have for our witnesses, I ask unanimous consent for the record to remain open for 2 weeks for Members to submit questions for the record. I'll let witnesses submit written answers as soon as practical. Without objection, so ordered.

I also ask unanimous consent that all Members' and witnesses' opening statements may be included in the record. Without objection, so ordered.

Meeting adjourned. Thank you.

[Whereupon, at 2:36 p.m., the subcommittee was adjourned, subject to the call of the Chair.]

○

ISBN 0-16-058357-8