**Proceedings, Second Workshop: Panel to Review the 2000 Census**

Committee on National Statistics, National Research Council

ISBN: 0-309-54182-4, 136 pages, 8.5 x 11, (2001)

**This free PDF was downloaded from:**
**http://www.nap.edu/catalog/10279.html**

**THE NATIONAL ACADEMIES**
*Advisers to the Nation on Science, Engineering, and Medicine*

# PROCEEDINGS, SECOND WORKSHOP

## Panel to Review the 2000 Census

Committee on National Statistics
Division of Behavioral and Social Sciences and Education
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C.

**NOTICE:** The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

Additional copies of this report are available from Committee on National Statistics, 2101 Constitution Avenue, NW, Washington, D.C. 20418; (202) 334-3096; Internet, `http://www.nap.edu`

Suggested Citation: National Research Council (2001). *Proceedings, Second Workshop. Panel to Review the 2000 Census.* Committee on National Statistics. Washington, DC: National Academy Press.

# THE NATIONAL ACADEMIES

National Academy of Sciences
National Academy of Engineering
Institute of Medicine
National Research Council

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

iii

## PANEL TO REVIEW THE 2000 CENSUS

JANET L. NORWOOD *(Chair)*, Chevy Chase, Maryland
ROBERT M. BELL, AT&T Labs–Research, Florham Park, New Jersey
NORMAN M. BRADBURN, National Science Foundation, Arlington, Virginia
LAWRENCE D. BROWN, Department of Statistics, The Wharton School, University of Pennsylvania
WILLIAM F. EDDY, Department of Statistics, Carnegie Mellon University
ROBERT M. HAUSER, Department of Sociology, University of Wisconsin
RODERICK J.A. LITTLE,[*] School of Public Health, University of Michigan
INGRAM OLKIN, Department of Statistics and School of Education, Stanford University
D. BRUCE PETRIE, Canadian Institute for Health Information, Ottawa, Ontario


CONSTANCE F. CITRO, *Study Director*
MICHAEL L. COHEN, *Senior Program Officer*
DANIEL L. CORK, *Program Officer*
AGNES GASKIN, *Senior Project Assistant*
MARISA GERSTEIN, *Research Assistant*
MICHELE VER PLOEG, *Program Officer*
ANDREW A. WHITE,[**] *Senior Program Officer*
MEYER ZITTER, *Consultant*


[*]Served until March 2000
[**]Served as study director until March 2000

v

## COMMITTEE ON NATIONAL STATISTICS
### 2001–2002

JOHN E. ROLPH *(Chair)*, Marshall School of Business, University of Southern California

JOSEPH G. ALTONJI, Department of Economics, Northwestern University

ROBERT M. BELL, AT&T Labs–Research, Florham Park, New Jersey

LAWRENCE D. BROWN, Department of Statistics, The Wharton School, University of Pennsylvania

ROBERT M. GROVES, Joint Program in Survey Methodology, University of Michigan

HERMANN HABERMANN, Statistics Division, United Nations, New York, New York

JOEL L. HOROWITZ, Department of Economics, University of Iowa

WILLIAM KALSBEEK, Survey Research Unit, Department of Biostatistics, University of North Carolina

ARLEEN LEIBOWITZ, School of Public Policy and Social Research, University of California at Los Angeles

RODERICK J.A. LITTLE, School of Public Health, University of Michigan

THOMAS A. LOUIS, RAND, Arlington, Virginia

DARYL PREGIBON, AT&T Labs–Research, Florham Park, New Jersey

NORA CATE SCHAEFFER, Department of Sociology, University of Wisconsin-Madison

MATTHEW D. SHAPIRO, Department of Economics, University of Michigan


ANDREW A. WHITE, *Director*

vi

# PREFACE

The Panel to Review the 2000 Census was established in 1998 by the Committee on National Statistics of the National Research Council at the request of the U.S. Census Bureau. The panel has a broad charge to provide an independent assessment of the quality of the 2000 census operations and results. It is charged in particular to review the statistical methods of the 2000 census, including the use of the Accuracy and Coverage Evaluation (A.C.E.) Program and dual-systems estimation (DSE), along with other census procedures that may affect the completeness and quality of the data.

As part of its work, the panel held three open workshops on topics related to the A.C.E. and possible adjustment of the census counts for population coverage errors. The panel has prepared proceedings of each workshop to be part of the public documentation of planning for the 2000 census.

The first workshop was held October 6, 1999. It considered issues of the A.C.E. design that had not yet been completely worked out by the Census Bureau staff. Topics discussed included methods and issues for determining post-strata for estimation, obtaining the final sample of block clusters from a larger initial sample, and imputing values for missing responses on characteristics needed to define post-strata.

The second workshop was held February 2–3, 2000. It covered the dual-systems estimation process from beginning to end. The third workshop was held October 2, 2000. It laid out the process the Census Bureau planned to follow in order to reach a decision by March 1 on whether to adjust the census counts for purposes of congressional redistricting.

This volume provides the edited transcript of the proceedings of the second workshop. (See National Research Council, 2001b and 2001c for proceedings of the first and third workshops.) The substance of the participants' remarks has not been altered. Some text has been added [in brackets] to spell out terms, clarify who is speaking, and make sentences complete. The only material deleted consists of introductions of speakers. A list of invited participants and their affiliations (at the time of the workshop) is provided at the end of the document.

Papers prepared by Census Bureau staff were distributed to participants before the meeting. They are often mentioned in the discussion, and appropriate citations have been added in the text to refer the reader to the relevant bibliographic entry in the references. All of the documents are available on the Census Bureau's web site: `http://www.census.gov`.

For explanations of terms and discussion of the design of the A.C.E. as it was implemented in 2000, please consult the interim report of the panel, *The 2000 Census: Interim Assessment* (National Research Council, 2001a).

I want to thank the Census Bureau staff who prepared materials and participated in the workshop. Particular thanks are due to Howard Hogan, who did a superb job in presenting a complex set of materials to the panel members and invited participants.

vii

I also want to thank the panel members and invited participants for their thoughtful comments and probing questions during the workshop.

Finally, I want to thank the panel staff who prepared this proceedings. They include Constance Citro, Maria Alejandro, Agnes Gaskin, and Daniel Cork.

Janet L. Norwood, *Chair*
Panel to Review the 2000 Census

# CONTENTS

# PROCEEDINGS

## OPENING REMARKS, FIRST DAY

**DR. NORWOOD**: I want to welcome all of you to this workshop on dual-systems estimation [DSE] that, as you know, is sponsored by the Committee on National Statistics Panel to Evaluate the 2000 Census.

This panel has been working for some months now attempting to learn as much as we can about this census and to gather background information about all aspects of it. We have had several meetings. We have, of course, read much of the literature. There is a lot of it, so I cannot say that we have read absolutely all of it, but we have done a pretty good job of that. We have listened to the Census Bureau describe its plans. We visited census offices during the dress rehearsal and visited several of the processing centers. We will be doing much more of that.

This workshop is another of our efforts to learn as much as we can. I want to emphasize that this panel has no position on any of these issues at this time. We remain completely objective—not an easy task, in this, one of the most politically charged census efforts in our history. But we are not interested in the politics of this effort. I spent my whole life staying out of politics. We are here to learn about the plans for 2000, not to reconsider what was done in 1990. Our purpose is not to provide specific advice to the Bureau on the issues that will be presented today, although, of course, it is possible that the discussion on some of the issues might prove useful to the Census Bureau. We are here doing the job that has been asked of us, to use our knowledge and experience to evaluate all aspects of this monumental effort of measurement—operational, preparation, collection, compilation, survey design, statistical methods—so that we can produce an informed evaluation.

We know that the issues to be discussed here today and tomorrow have been the subject of considerable controversy, in the Congress, the courts, and in the statistics profession. I expect that differences of opinion will surface among the speakers today. Indeed, I hope that will be the case. But let there be no mistake: this panel has no position on dual-systems estimation or on the issue of adjustment itself at this time. We have deliberately invited a number of professional statisticians who we know have different views on the issues to be discussed. We are sponsoring this conference today as a part of our effort to review the statistics of adjustment in a professional, non-political environment, so that we will have a fuller understanding of the issues we will need to consider in completing our evaluation.

Let me just take a moment to tell you how I plan to run this conference. I shall try my best to keep to what I think is almost an impossibly tight time schedule. Howard Hogan, from the Census Bureau, will present each set of issues. He will have a block of time in order to make those various presentations today and tomorrow. Members of the panel will, I am sure, raise questions or make comments. Of course, others of the specially invited guests will have an opportunity to do so as well.

The timing of all the breaks will be very carefully monitored. I ask that you return to your seats as quickly as possible after them.

The speakers will be expected to stick to the specific issues before us. Comments, of course, must be relatively brief and to the point. I know many of the

people here quite well, and that may be a little difficult. But I think we will manage it.

I hope that you will forgive me when I am forced to curtail the discussion. I fear that if I do not keep the time with an iron hand—and I have asked for a little monitor here—we will just not accomplish our goal today and tomorrow.

Once again, let me emphasize that we are getting at this, really, with no holds barred, and no comment, no question that is raised by members of the panel should be taken as an indication of any position of the whole panel. The panel has not discussed dual-systems estimation as yet. We are here, really—and rather eager, as a matter of fact—to engage in an intellectual discussion, not to take sides.

I would like to, first, introduce the special guests. We have a lot of guests here, and I am pleased that they are all here. We have eight participants who are sitting toward the front here, who will be participants: Barbara Bailar from NORC [National Opinion Research Center]; Stephen Fienberg from Carnegie Mellon; David Freedman, University of California at Berkeley; Bruce Spencer, from Northwestern; Philip Stark, University of California at Berkeley; Marty Wells from Cornell; Don Ylvisaker, University of California at Los Angeles; and Alan Zaslavsky, from Harvard.

We have some other invited guests as well, whom I would like to introduce: Lynne Billard, from the University of Georgia; Hermann Habermann, from the United Nations; Charlie Jones; Ben King, who is chairman of the Committee on National Statistics Panel on 2010;[1] Mary Mulry; John Rolph, from the University of Southern California, also chair of the Committee on National Statistics; and Kirk Wolter, from NORC.

I should say that we also have Ken Prewitt, the director of the Census Bureau. We also have Bill Barron, who is the deputy director of the Census Bureau.

Ken, would you like to say a few words?

**DR. PREWITT**: Thank you, Janet. It obviously goes without saying that the Census Bureau takes this meeting very seriously. We are much indebted to the Academy and to the panel for having organized it, for having given so many pro bono hours to struggling with this as we try to put Census 2000 in place.

A word or two on Census 2000, then I will turn more specifically to the agenda today. As I have said many times to the press—Census 2000 is on schedule and on budget, as of today. That does not mean that we will not have operational difficulties that have yet to be encountered. But right now we are in as good a shape as we could hope to be in terms of mounting it. Obviously, the census has started. Our operations in remote Alaska have been launched successfully, and not just on the PR front, but also operationally successfully. The hiring is slightly above target right now. We want a pool of 3 million. We wanted to be at 45 percent this week; we are at 48 percent. Our budget estimates look good. None of our major operations are not within the budget boundaries that we had set for ourselves. As you will hear in a moment, the A.C.E. listing is on schedule.

There is no major operation right now that is not where it needs to be for this

---

[1] The panel referenced here is the Panel on Research on Future Census Methods, which is charged to advise the Census Bureau during its early planning for the 2010 census.

moment in the launching and implementation of Census 2000—MAF [Master Address File] production, and I could go on down the line.

I just wanted to put that as part of the setting, because, as I will get to in a moment, obviously one of the issues for DSE has to do with operational capability, as well as some of the technical issues. Indeed, as I have said to this panel and I would like to say to this audience, if it [the census] does unfold as robustly as we hope it is on an operational front—I should also say, on the promotion front—it is going to be much, much bigger than any of us could have imagined. I do not mean the Census Bureau's promotional effort; I mean the country's promotional effort. I just left a meeting this morning with members of the California Assembly. The California Assembly has just voted $24.7 million to, quote/unquote, "help" us do the census. States and cities and counties all over this country are investing their own resources, their own labor effort into a successful census.

Needless to say, we welcome it, but it also creates some operational complexities at the edges. There will be many, many, many things connected to Census 2000 that will not be under the direction or control of the Census Bureau itself. In that sense, it is a very different kind of census. I do not want to underestimate that. It is a very big deal. There will be many, many little problems that will be created because we have between 40,000 and 50,000 different groups and local governments who, for their little area or their constituency, believe that they are doing the census—not just promoting it, but actually helping us do it. That raises lots and lots of complicated issues at the edges. I could give you endless anecdotes on that.

As a generic issue, as I say, it is to be welcomed, but it is not without its complications. I met yesterday with the GAO [U.S. General Accounting Office] and challenged them, in effect. When they actually evaluate this census, they are going to have to evaluate not just what the Census Bureau did, but what the country did. This is a metaphor that we have used, but it is within the realm of accuracy. It is not hyperbole. It really is a census being shared with the American people, with all the implications and complications that that creates.

So I just wanted to make a few comments of that sort to set the stage for today's meeting.

With respect to DSE, as I read and hear the criticisms, it seems to me that they fall into three broad categories. There is, of course, the concern or criticism that the Census Bureau, in its design, is engaged in partisan politics; secondly, that the Census Bureau is perhaps not sufficiently competent to manage an operation as complex as the dual-systems estimation operation—and sort of a sidebar on that one, at least as I read the literature, that even if it were, it maybe does not have the wherewithal (because it may not be doable) to get the public to comprehend it. That is, there is a sort of public comprehension problem, quite separate from the competency of the Bureau to actually do it. Third, of course, and by far the most important issue as far as we are concerned—because it is, fortunately, fact-based—are the serious differences of opinion and viewpoint about whether DSE will actually improve census coverage and counts.

Let me just talk about each of those very briefly. The first one, of course, is not of our doing—"our doing" including everyone in the room—that is, the issue of the

partisan concern. It has taken what is really a very important kind of technical, professional debate and captured it and used it for other purposes, well outside the boundaries, if you will, of the technical and professional debate. Nevertheless, it creates part of the environment in which we are all caught up. It is not good for the Census Bureau, it is not good for the national statistical system, and I think it is also not good for the statistical community, to have had its work be pulled into, in effect, a debate that is about partisan motivation and partisan attention. I think, as everyone in this room knows, no one has put on the table any evidence that the design was selected because it had a known outcome for partisan purposes. To this date, no one has put any evidence on the table that it would be implemented with that in mind. Nevertheless, it is the shadow, in effect, that hangs a bit over Census 2000. It is not the task of this meeting. Nevertheless, I think this meeting ought to be aware of the fact that anything that we can do to make it clear that this is a debate about a technical issue, that it is fact-based, that there are reasonable differences of viewpoint about whether DSE will improve the census coverage, and that it has nothing to do with anybody's partisan affiliations or interests—the more that the statistical community can make that statement, the better it is for all of us.

Let me turn to the second issue, which is the issue of the operational competency. I turn to this one obviously not in any defensive manner. I would hope we would be extremely frank. But I do think it is important to distinguish between those criticisms of DSE that are based upon technical judgments and those that are based upon whether the Census Bureau can do something of this complexity in a competent manner. The Census Bureau is remarkably non-defensive about conversations about its technical competence and operational competence. I have only been at the Census Bureau for a year, and I have been part of a lot of other institutions, a lot of good ones—Stanford University, the University of Chicago, Rockefeller University, SSRC [Social Science Research Council], and so forth. We should all be so lucky as to be associated with an institution that is as comfortable with criticism and self-criticism as is the Census Bureau. It is a remarkable culture in that sense—and I say this as an outsider—in which the more you beat us up, the better we like it. In some respects, it means we are probably going to do it better next time around.

I do not have to remind anyone in this audience (but I do have to remind from time to time our political and public commentators) that, when the country debates the undercount, after all, they are debating the report card the census gave to itself. There is no other way to talk about how well the Census Bureau did, except to look at the Census Bureau's own work on how well it did and its own publication of that effect, including its operational failures and flaws and so forth. So when I say that we want to talk seriously about operational competency, I do not mean that in any kind of defensive way. If that is the area that sort of organizes some of the criticisms of DSE, then I would much rather have it discussed that way, because I think it is actually better in the long run for DSE. If the problems are really the competency of the agency, then that is at least something we can work on.

So I urge us not to be hesitant, if that is one of the lead arguments, to make that a straightforward lead argument, and we will respond to it as best we can.

In that connection, I add a footnote. This is, to me, just as a reader of the literature, slightly disturbing—the debate about the coding error in 1990 and whether it did or did not move a seat.[2] Much of the conversation about competency uses that as one of its lead pieces of evidence. Since the issue of whether it did or did not move a seat (on which there are people in this room who have quite different positions) is itself a matter of fact, I would actually urge the statistical community to sort that one out, because it really is a matter of fact, having kind of a clean, clear position on whether it would or would not have moved a seat.

I would also urge those who use the coding error as an important piece of evidence about competency to recognize that, of course, a coding error can occur in any operation, not just in the DSE operation, but also in the enumeration operation.

I do not mean to back away from it, but I at least would like, insofar as that is the lead piece of evidence on competency—and not the only one; I appreciate that. There are very, very strong friends of the Census Bureau who were very concerned about the pre-Supreme Court design, from the point of view of its operational complexity and whether we could have pulled it off within the Integrated Coverage Measurement [ICM] within the nine months. So we are very comfortable having that conversation. But I see it as a somewhat different conversation from whether DSE itself will improve coverage the way that the Census Bureau hopes that it will.

On that front, there are many, many issues, and that is what will be discussed today. That is why we are so pleased by the day-and-a-half, as we turn to that. The issues of, obviously, correlation bias and heterogeneity assumptions and the fifth cell and the "wily trout" and all those kinds of issues, are extremely important, to see how wide the differences of viewpoint are within the statistical community. If we can actually focus on what really are the differences, then we can have a really quite important debate. That does not mean we erase them; no one is expecting to erase them. But at least let us frame them in the most precise terms that we possibly can. Obviously, the presentation that Howard will shortly be making moves in that direction.

I will just conclude with one final comment in the area of technical differences, which is, as I say, obviously what the meeting is to be about. Howard will talk about this. John will talk about it a bit as well. I just want to flag it as an issue. It would be very useful to make sure we understand the difference between distributional accuracy and numerical accuracy. They are different things. It is very difficult to maximize both of them, as we will try to present. It is important for the audience and the panel to know that the Census Bureau believes that it should

---

[2]Based on the 1990 Post-Enumeration Survey, the Census Bureau developed an initial set of adjusted census counts through dual-systems estimation. Subsequently—after Commerce Secretary Robert Mosbacher's decision not to adjust the 1990 census—a computer coding error was discovered by the Census Bureau, which necessitated revision of dual-systems estimates and undercount estimates.

favor numerical accuracy. That is its operating philosophy in the decennial [census]. If the criticisms focus primarily on distributional accuracy or proportionate shares, but not coverage, then we actually have a difference in viewpoint that is, in part, about the philosophy of the census, and not just the technical operations or technical components of DSE. If we actually are having an argument about the philosophy of the census, that is very important to sort out and make clear that that is, at least in part, the nature of the differences.

So we do hope that, before the day-and-a-half is over, we have come to some kind of understanding of why the Census Bureau believes it has to favor numerical accuracy as its core mission for the conduct of the decennial, and why that necessarily precludes us from maximizing in every instance the distributional accuracy. I urge us, as we have the conversation, to make sure that we know when we are talking about a difference in the philosophy of what the decennial ought to be as against the actual technical components of DSE.

Obviously, Janet and panel, if this meeting can make any kind of headway whatsoever—and I am sure it will—on trying to really identify cleanly what the differences are between the Bureau's recommended decennial and what some of the concerns have been, it will be extremely valuable. The more cleanly we have identified it, the more likely we are to be able to take some steps that would move us in the right direction.

Thanks very much.

**DR. NORWOOD**: Thank you. Before we go further, I would like to ask people who have not been introduced to very quickly stand up and tell us your names and whom you are associated with, if you have an association....

I guess we have the staff of the Committee on National Statistics to thank for these plush quarters. I would like to thank, especially, Joshua Dick and Jamie Casey for their hard work in getting all this put together.

We are a little bit behind, so I am going to take some time from John Thompson, who can come up and just give us a very brief statement on where the census is.

## CENSUS 2000 UPDATE

**MR. THOMPSON**: I will be very brief on where we are. I do want to talk just a few minutes about accuracy. As Ken said, the Alaska enumeration is under way. I hope you all saw this picture in some of the various pictures. You look really good. We are well under way to printing maps. We have to print over 20 million maps. We have just about finished printing the maps for the first major operation, which is where we go out and deliver questionnaires in non-city-style areas.

We have a program called "New Construction," where local governments in city-style address areas can review their address lists one final time before the census. We have delivered well over 90 percent of the materials to the local governments. They have until, basically, Census Day, April 1, to give us any new housing units that have been built.

As Ken said, recruiting is going very well right now, but the big challenge is yet to come. Our data-capture centers are all open. We have been doing what we call "operational tests" and "dry runs." We are finishing the operational tests and dry run in Phoenix, and we just started our last operational test and dry run in our Jeffersonville facility. Our Local Update of Census Addresses Program—this is where local governments can review the address list and challenge—is drawing to a close. We have completed our appeals process for the rural governments, for the most part. We have sent the materials out to the city-style governments, and they are in the process of starting their appeals. That should come to a close probably by the end of February.

Howard will talk to you about where we are on the A.C.E. [Accuracy and Coverage Evaluation Program]. Let me turn very quickly to a discussion of accuracy. As Ken said, there are several ways we view accuracy at the Census Bureau. One way we call numeric, and that is making a particular area as accurate as possible, or as close to truth as possible, as opposed to another kind of accuracy that we call either distributive or distributional, which is taking a set of areas or population groups and trying to bring the relative shares of these groups as close to a true distribution of relative shares as possible.

Let me first say that both of these concepts are very important in the uses of census data. I think a lot of attention has been focused on distributive accuracy, but not very much on numeric accuracy. As Ken said, as we plan the census— and, actually, those who give us a number of recommendations on how we should plan the census—we focus, by design, on numeric accuracy. That is, we design procedures that are designed specifically for unique areas or unique population groups. If these procedures are carried out to perfection, then we will, in fact, get distributive accuracy. But we do not use distributive accuracy as a criterion for our planning. If we were to do that, for example, we would not do our partnership program; we would not let the state of California, for example, spend $24 million to try to improve the count; we would not do coverage improvement programs that we thought would improve the count for one population group, because that population group is not uniformly distributed across states, which would hurt distributive accuracy. So we really focus on doing the best job we can for areas where we see problems.

That is also true of those who give us advice. For example, we get advice that we should apply more advertising to specific groups. We get advice that we should select particular tracts in selected cities and try to make these tracts as accurate as possible, without regard to the effect on the overall distribution of states or cities.

As we plan the census, we really focus on trying to achieve numeric accuracy, because that is really what I think we can focus on and plan for and try to develop solutions for.

When we evaluate the census, we evaluate the census in terms of both numeric accuracy and distributive accuracy—although I have to note that most of the evaluations to date have focused on the PES [Post-Enumeration Survey] in terms of distributive accuracy, and on numeric in terms of most of the rest of the census operations and coverage improvement programs. I think that is something we will

probably correct in our evaluations of 2000. We will look at distributive accuracy for all the census programs as we analyze it.

I hope this puts accuracy in a little bit of context as we go through the discussion of the dual-systems estimation and the assumptions that underlie it. With that, I will turn it back to Janet.

**DR. NORWOOD**: Thank you. We are now ready to start. Howard Hogan is going to have the hard job of being up here for most of the time. Howard is going to talk to us, first, about a series of topics that involve the Accuracy and Coverage Evaluation survey.

## A.C.E. OVERVIEW

**DR. HOGAN**: Thank you, and thank the panel. The only awkward thing is that I really feel I am supposed to be here talking to the panel, and I will be spending most of my day with my back to my primary audience. But the compensating thing is to look out and see so many friends from over the years, people who have been involved in this issue, some of them, even longer than I have. It is an amazing group that we have assembled, both on the platform and in the audience.

My time is completely flexible, because it is really the panel's time as far as I am concerned. I will talk a little bit about the model. I am going to go through the survey itself fairly briefly, talk a little bit about the changes from 1990, and then go into the assumptions of the model and how we relate that to the survey. Most of what I will be talking about is in the background materials and so I am not going to dwell on that. I assume you guys will interact.

**DR. NORWOOD**: Let me just say that we will save the discussion part for after Howard is finished, but if the panel or any of the invited guests have particular questions or if they want to get a little bit more information for understanding purposes, please raise your hand or speak up.

**DR. HOGAN**: The question is, do we have the transparencies in the folder? Most of them, with only a couple of exceptions, are blowups of what you have in your folder. The couple that are not I can give to Janet, and we can get copies made.

I want to introduce the subject of today's meeting. Let me assume that everybody here has probably seen it at least 10 times, so I will not go through the model. But I did want to put up the model explicitly, so that when we want to talk about [cell] $N_{21}$ versus $N_{12}$, we will have the rows and the columns the same.

This is the basic paradigm, breaking down the population of the four cells. The question, then, is how to estimate $N_{++}$ [total population], how we can estimate this.

The Census Bureau adopts the traditional Peterson or Chandrasekar and Deming model—this is in my paper, as well as several of the other papers—that says that our estimate, based on the independence assumption (I will not review that here for you all), is the cross-product divided by the number of matches, the number in both. The way I think about it and the way I am going to discuss it today is

that our estimate of the population is really made up, in my mind, of two components: the number of people in the census, the number of people correctly enumerated in the census—that is, first, the thing we have to measure—and second, the proportion of all people who are in the census. So, first, we have to measure the number of people in the census. That is just not given. Then we have to measure the proportion of the true population in the census.

Just so we can define some terms, there is another way we can look at this. We can take the standard independence assumption and apply it only to the fourth cell of those missed by both the census and the survey—I am completely abstracting from the question of sampling for the moment—and estimate the fourth cell as those in one and not the other, in the other, not the one, divided by those in both—again, the standard assumption there.

One term we will probably be discussing today—I thought I would at least give it a symbol we can discuss—is those left over. After we take the standard dual-systems assumption of the fourth cell, we have an estimate of those missed by both. Many people believe that there are other people missed by both, a residual in the fourth cell that is not picked up by the standard assumption. I am just going to define that residual as $N^*$, so we have something to talk about when it comes up.

In these kinds of terms, the true population, then, is those in both, those in the survey and not in the census, those in the census and not in the survey, plus, by definition, those not involved. This is the true decomposition of the true population. We can then decompose the final cell into those modeled by the standard Chandrasekar and Deming or Peterson assumption and this residual, those left over. It has a number of terms—the unreachable, the unmeasurable, the "wily trout," the fifth cell. There is really no standard term for that. But I believe all those terms are attempts to describe what I have defined here as $N^*$.

After you take the dual-systems estimate, you can then decompose it into the four cells that it is capable of measuring, plus sort of a fifth group here, which I have kind of defined as measurement error, matching error, balancing error, missing-data error—all the other components. Therefore, taking what is really in the operationalized dual-systems model, which lacks, by construction, this unreachable population but includes the measurement error, then the bias in the dual-systems estimate is sort of the difference between the unreachable and the measurement error. That is the bias in the dual-systems estimate.

The unadjusted census is made up of those captured by both and those captured only by the census. But it also includes two terms that we need to take account of. It includes the traditional statistical adjustments to the census, the traditional statistical additions to the census, through our traditional imputation process. We define that as $II$, or the census imputed population. It also includes enumerations that are not [correct] in the census. I will say much more about this. So the bias in the census is, of course, the difference between the true population and the census count. That is those picked up only by the survey, the modeled part of those missed by both, the unmodeled, the unreachable part missed by both. But that is balanced off in the unadjusted census by the traditional census imputation,

statistical adjustment process, and any erroneous inclusions in the census. So a lot of the discussion really is whether the bias here in the DSE is smaller or lower or distributed differently than the bias in the census. So that is kind of what I see as the model.

The question, then, is how to operationalize that model, how to get the numbers to put in it so that we can actually measure it. We do that through the accuracy and coverage evaluation. This is in the handout. It is at the end of the paper by Danny Childers and Debbie Fenstermaker [Childers and Fenstermaker, 2000], if you want to read along or take notes.

I am just going to walk through the operations, because when we discuss how the operations match into the model, we need to know what the operations are. The first operation, which we and the panel have discussed extensively in previous meetings, is the sample selection. I will not spend much time on that, because we have discussed it. Essentially, we take a sample of block clusters. The sample we took was about 2 million [housing units in a sample of] block clusters, because it was designed to support the original 750,000 ICM sample. That was done last summer. Then we did this in the fall. This is completed. It is not only on schedule, as Ken said, but it is actually completed. We go out and list all the housing units in those block clusters. We did this in all the housing units—the full 2 million [in the sampled] block clusters. That is indeed done, and I think was done well.

We then need to reduce the 2 million clusters—not clusters; I mean 2 million housing units. If I have been saying 2 million clusters, I apologize—clusters that include 2 million housing units, down to a number of clusters to support 300,000 housing units. We do our block-cluster reduction. That is also done.

This was done this time with sort of a change from 1990 that I think was quite an improvement. When we did our second stage of sampling, this block-cluster reduction, we had the most recent census housing unit counts, and we also had the number of housing units the A.C.E. listers listed. So we actually had a fairly good measure of the size from both the census and the A.C.E. We have the number in the block as listed by the census, the number in the block as listed by the A.C.E., and we actually have the difference. When we did our subsampling, we took account of both those differences, and I think drew a fairly good sample of large and medium blocks.

Small blocks are those blocks where the census said there were zero, one, or two housing units. Those who followed 1990 know that the small blocks were quite a problem. We have done a number of things to reduce the problem, reduce the sampling issues introduced by the small blocks. For example, we do have the differences—the number measured by the census, measured by the A.C.E., and the difference. That helps us control the blocks. Also we did a different clustering algorithm this time. Rather than having the whole universe of about 3 million small blocks to deal with, we were able to group about two-thirds of those with larger block clusters, and so the small-block sampling algorithm has to deal with a much smaller universe. We are using a two-stage small-block sampling scheme.

So I believe, in terms of simply the sampling issues, we are much better in handling the small-block clusters than we were in 1990.

[Transparency]

In an operation that is basically brand new, we now have a housing unit matching and follow-up—brand new since 1990. We did test this in our 1990 series of tests and the dress rehearsal. It is not brand new since the dress rehearsal. We did do it then, and it was fairly successful. We take, for the blocks that remain in sample, the A.C.E-listed housing units and the census-listed housing units, and we actually do a matching between them. This helps us in a number of ways. For one thing, when we do the block sample reduction, we will have the two linked. That helps us quite a bit in maintaining the overlap between the sample of people that we use to measure the proportion of people enumerated and the sample of enumerations that we use to measure the number of people counted in the census and are the proportion of the census enumerations that are correct. So we want to maintain that overlap. This matching helps us do that.

It also helps us identify any listing problem in the A.C.E., whether the A.C.E. is listed in a wrong block or whatever. It gives us a link that will help us in later interviewing and later matching. So it is operationally quite an important step, although it does not really feed into some of the equations directly.

One thing we do not do is, if we do find a housing unit that the census has and the A.C.E. has, we do not add that to the A.C.E. We do not violate independence there. We note that that is a leftover census unit and carry that along separately.

So that is large block subsampling. The housing unit matching is starting right about now. We are starting that on time. It will continue, together with the matching and the follow-up and the after-follow-up housing unit matching, until around April. This is something that happens before the census and gets us our files, our A.C.E. files and our census files, all lined up before the census takes place.

Based on that, if a block has more than 80 housing units, then we subsample it. Since we have already linked the census and the A.C.E. together, the subsampling is such that we are able, largely, to maintain the same housing units in both the census sample and the person sample.

That brings us to sort of the heart of the A.C.E., the person interviewing. This has two components, one of which we tested in the dress rehearsal but is new since1990. We use the telephone to interview some of the housing units from the A.C.E. How do we do that? We have linked from the housing unit matching which units in the census correspond to which units in the A.C.E., and so we know if the census person has completed the questionnaire and mailed it back, and, under fairly restricted circumstances, that people have completed their census questionnaires and mailed them back. It is a house-number/street-name kind of address, single family, very well controlled. We will then allow the A.C.E. interviewer to telephone the housing unit and conduct an interview over the phone. That can start fairly early. That can actually start in May, where we do the telephoning to the housing units to get the independent A.C.E. interview. They have the phone number, but they have none of the other census information.

This telephone interviewing really gets us two things. It gets us some interviews early, and I think that is important. But judging from dress rehearsal, it did

not get us that many—maybe 5 or 10 percent, if we are very lucky. But it gets the A.C.E. interviewers and crew leaders and the whole structure used to dealing with the interview. That brings up one of the other changes from 1990, a fairly important design criterion here. The A.C.E. interview is a computer-assisted personal interview [CAPI]. It is laptop. When they conduct the interview, it is done on a laptop. When they conduct it on the phone, they do it from their homes, but on the laptop. When they visit door-to-door, they will do it on the laptop. This helps us control some of the transcription errors, the keying errors. It lets us make sure that numerics are only filled in in numeric fields and gives us a much cleaner interview, which is then transmitted, ready for A.C.E. matching.

The person interviewing essentially starts with the telephone in May, really hits the ground in July, right after the end of the census nonresponse follow-up interviewing. Most of the interviewing is done in July. Then we have some nonresponse conversions and other things into late summer and early fall. At the end of that, we have completed the independent list of people for the A.C.E.

[Transparency]

Something else has been going on that I do not talk about very much. It is called "the rest of the census," the really hard work of going out and interviewing, not a sample of 300,000, but the universe of something like 120 million housing units. At this point, we have what the census has enumerated. By "this point," I mean September of this year. We have what the census enumerated. That comes to us, the A.C.E. folks, in what we call the unedited census person file, sometimes known as the census unedited file, or the CUF. That is the file we will match against. That is the file where we will say, these are the records that are in the census. That, of course, comes to us with all the people in all the clusters that we have sampled, together with some of the other surrounding clusters.

We have to identify which of those census records are in the sample—that is, are in the sample of census enumerations that we will use to measure the number of people correctly counted in the census. So we have to identify our enumeration sample. Again, we have done the housing unit linking earlier, so we have a real head start on that. But we actually, at this point, have the final census of what the census has counted. We take a sample of that, we take it from the same block clusters as the A.C.E. person interview sample, and we bring them together and do the person matching. Under certain circumstances, if there are unmatched people, we do follow-up, and after follow-up, person matching. At this point, we also do the person duplicate search and other codings.

So coming out of this step, everybody in the independent person sample, what I will call the P-sample, is determined to have been correctly enumerated, to have been missed (at least missed where the person should have been counted), or some unresolved or non-interview cases. All the records from the census side [E-sample] are determined to be correctly in the census, erroneously in the census, or will have some missing data, at that point. We will also have missing characteristics, missing age, missing race, whatever. So we have our missing data procedure, which we will talk about later.

Then, after all of this work, we bring [in] the census edited file. They have now imputed age and race, as well as everything else.

[Transparency]

At that point, then, we can begin discussing what we are here today to discuss, the dual-systems model. We can now compute that. We can try to get those ratios.

[Transparency]

Let me just go very quickly over the changes from 1990, so that we are all together. The sample size, as we all know, is bigger. The sampling probabilities—I think we have done a much better job of controlling the weight variation in the sampling. I have mentioned the small-block sampling. I think we have that better controlled. One of the differences that needs to be noted is, in 1990, our universe, our PES universe, included group quarters, very importantly—say, college dorms. That is no longer part of the A.C.E. universe. That is a difference.

I have mentioned the initial housing unit matching and the subsampling method for the E-sample identification, the automated CAPI instrument for the A.C.E. We are overlapping the A.C.E. and the nonresponse follow-up a little bit, through the telephone interviewing that I mentioned before.

One of the differences we will talk about as the afternoon goes on is, in 1990, the heart of the PES interview was, who is living here now; where did you live on Census Day; what were the cross streets, and what were the neighbors? We searched for them where they were. For 2000, we are using what is called a PES-C interview. That is much more "tell me who was living here on Census Day." The interviewing is more difficult, but the matching is much simpler.

The matching is about the same. A couple of the big differences—the search area, the area around which we look to see if the people are correctly enumerated, is different between 1990 and 2000. Now we only have one processing office. All the matching will be done in our permanent Jeffersonville matching office. In 1990, we had seven offices, seven staffs trying to do a uniform job of matching. I think that is actually an important change.

The universe of cases we are following up is slightly different. We are following up unmatched cases in the P-sample, where it is a partial household match. We did not follow those up in 1990. But some whole-household non-matches that were followed up in 1990 will not be followed up in 2000.

Post-stratification has changed somewhat. We are now looking at a mail-return variable, adding that to our post-stratification variables. Also we will be replacing the urbanization variable with a combination of metropolitan statistical area (MSA) and type of enumeration area, whether it is mailout/mailback or list/enumerate or list/leave. Our initial 1990 estimates included smoothing, this whole complex smoothing model that was greatly discussed in the literature and in the halls of Suitland, which we actually did not use in the estimates that we put out on our Web site. We gave that up in about 1990. It was probably not a bad model at all, but it was very hard to explain, and the discussion was focusing on the smoothing model rather than on the undercount. So we decided to simplify at least the explanation process, and we will not be smoothing. That is out.

So that is sort of an overview.

[Transparency]

The first step in doing a dual-systems estimate is figuring out the number of people counted in the census. We have the number of census records. That is easy. But we have to see if there is a correspondence between a census record and a person who should have been counted in the census. That is the purpose of the enumeration sample [E-sample], where we take a sample of census enumerations and verify that they are correctly in the census.

[Transparency]

In doing this, we look at essentially four criteria, some of which are not too controversial and pretty easy to carry out. Appropriateness: is it a real person or is it Donald Duck? Is it a tourist or is it someone who resides in the United States? Or is it someone who is made up by this enumerator? Is that enumeration an appropriate enumeration?

Is it unique, or is it duplicated? One of the things is completeness. The issue is, is that record sufficiently identified that we know it refers to a unique person? Essentially, for the A.C.E., does it have a name and two characteristics? There are other census enumerations that are counted in the census, which have far less. I think three letters in the name and one characteristic is what the census requires. Is it complete enough that we can say, yes, this is a unique person, and, yes, this person is appropriate?

Finally, geographic correctness. That has essentially two dimensions, one of which is, is this enumeration in the area where the person thinks he should have been counted? Is this enumeration where the person is correct as he considers his residence is? If someone moves into the census block on May 3rd and is counted there, but says, "No, on April 1, I lived at this other address," then that is not a correct enumeration, because it is not where it should have been. There are a number of things—two residences, and whatever else. So we need to determine the correctness.

That is done in sort of two stages. If we go out on the A.C.E. and we find the person independently and he says, "Yes, this is where I lived," and the census finds him, we say we have two consistent reports there; this is probably where the person should have been counted. In fact, there may have been some balancing errors, but that is our assumption, our statistical assumption, in the A.C.E. If we do not find him in the A.C.E., then we have to go out and do the follow-up. We are doing this even later. The A.C.E. interview, by and large, is in July—we are asking them where they consider their April 1 address—but when we are going out on the follow-up, which is in the fall, we have to ask them even further back. In both of these interviews, we sometimes take proxies. So one of the issues is how accurately we can measure that aspect.

The other aspect is sort of geographical correctness. We do not say you are correct only if you are enumerated in the housing unit where you should have been enumerated, but anywhere in the block. We have what is called a search area, an area where you are correctly enumerated. If you are enumerated anywhere in the block where you should have been enumerated—that is, where your residence was—we will say that is a correct enumeration.

However, we have found that, because of census geocoding errors, sometimes the person is counted across the street. So we have expanded this since dress rehearsal to include the surrounding blocks in certain circumstances, but not in all circumstances. This is one of the changes from both 1990 and dress rehearsal. You are correctly enumerated if your housing unit was counted in a surrounding block, but not in the sample block. But it has to be a housing-unit geocoding problem. It cannot simply be that you were living with your uncle across the street.

So the geographic correctness, then, is a more complex issue this time than it was last time. It is, first, that you are counted where you should have been counted, where your residence was, but the area includes the block, the block cluster, and mis-geocoded housing units in the surrounding block. If we go out and we find a housing unit in the census that is not in the A.C.E., and we go out and find that it is not in this block but it should have been in the surrounding block—it should have been in a block touching that—we say that is a correct enumeration and count that as correct.

[Question from Participant]

**DR. HOGAN**: Yes, if the person was anywhere in the right block or block cluster; in addition, if the housing unit was in one ring over. We will be doing that on a sample basis. That is a big change.

**DR. BRADBURN**: They both have to be true, the household and the individual?

**DR. HOGAN**: Yes, within the block cluster, it is completely person matching. If within the block cluster you should have been counted at your mother's house but you were counted at your aunt's house, you are correct. However, when we go to the surrounding block search, then the housing unit has to be in the surrounding block and you have to be there. To be matched in the surrounding block, we have to find you in a mis-geocoded housing unit.

There are a number of reasons for that. The general reason for doing a surrounding block search is that it reduces the variances tremendously. The reason for targeting it is that we found that doing it 100 percent was exceedingly tedious. Mostly, there are no geocoding errors. So by targeting it on geocoding errors, we think we can do a more consistent job.

[Transparency]

At the end of this process, we have, essentially, our estimator of the number of people correctly enumerated in the census—that is, $N_{+1}$. This measure is made up of essentially two components (in our backup material, I think we have been consistent): the number of census records—and that is the number of census records on the census edited file—less the statistical additions to the census, the people we know were added through the census statistical hot-deck, whole-person imputation process, records that we know did not uniquely correspond to a person. What is in parentheses, then, measures sort of what the census considers a data-defined person, what the census considers a person with sufficient information for census purposes. That is, of course, the difference between total census records and those added through the traditional imputation process.

The E-sample is a sample of census data-defined records. The A.C.E., then, applies an even more stringent criterion to that, subtracting out records that the census may consider sufficiently data-defined—after all, it had two letters of the last name and race—but the A.C.E. says we need a complete name and characteristics. So that is one thing that gets subtracted out, duplicates, people enumerated in the wrong place in terms of residence, people enumerated outside the search area, fictitious, Donald Duck, whatever. That is subtracted out through the E-sample process. What is left is the measure of the correct enumerations, and the ratio between total and correct enumerations is sort of the point of the E-sample. Multiplying this together gives us the number of people counted in the census.

Then we need to turn to the second part of the process, which is this: of all the people who should have been counted, what proportion were counted?

[Transparency]

So we have measured that. We now need to measure this ratio, the ratio of people in the A.C.E. to people who are in the A.C.E. and the census. What we are trying to measure is the ratio of the true population counted in the census.

[Transparency]

That brings us to sort of the DSE model, as applied operationally. The things we have to maintain in doing this are operational independence, consistent reporting, accurate matching, and homogeneity within post-strata. Operational independence has essentially two components that we have to pay strict attention to. To the extent we do not, it is a bias in the model. First, in terms of what is going on in the field, we have to make sure that the field interviewers are not colluding with the census. That is one of the main reasons that we are not using the same interviewers—the traditional stuff that we have talked about a lot in the past. In addition, we need to, in our office processing, make sure that our rules for whom to accept as being good A.C.E. interviews are independent of whether they match to the census or not. It is one thing to keep field interviewers separate, but once you bring it back, if you let the clerk say, "I can't find this person in the census. Maybe it wasn't a good A.C.E. interview," you can introduce operational dependence then.

**DR. NORWOOD**: Howard, what kind of process do you have to decide how good the A.C.E. is?

**DR. HOGAN**: How good the interview is?

**DR. NORWOOD**: You do the A.C.E., and you are going to use it. How do I know that you have completed it in a valid way and so on? Do you have some specific things that you are going to be looking at?

**DR. HOGAN**: In general, we will be looking at the sorts of things that in a traditional survey process are looked at—the interview rates, the non-interview rates, breakout by sampling and post-strata. We will have various quality control measures, through the interviewing, the listing, the matching.

**DR. NORWOOD**: You will have all those available before you do the....

**DR. HOGAN**: I believe we will have that available, yes. In addition, we will also have—and this is fairly important—the estimates and the variances on those

estimates. One of the things we look at in all survey processes is—usually, if something goes wrong, it will show up as a very strange variance, because the squared terms of the variances are much more sensitive to differential errors that we will be looking at. Of course, we will be reviewing our results by stratum and post-stratum, and comparing that to what we expected. We will have the demographic projections from the previous census, together with demographic analysis [DA], to give us sort of an independent reading of what the total population might be. We will have various readings of how the census went. So we will have the traditional survey measures, the traditional census measures, that we will be looking at throughout this process to see if what we are doing is under control, if what we are doing is what we are intending to do.

In 1990, for a number of reasons, we had a huge separate process of P studies and E studies, designed to support the Secretary of Commerce's decision process. As most everybody here probably remembers, the Secretary of Commerce in 1990 laid out criteria for whether or not we should correct the census. Those criteria seemed to call for various hypothesis types of things, whatever else. We had a fairly elaborate structure in 1990 designed to support the secretary's criteria that he had laid out to make that decision.

The current secretary has not laid out any such criteria, nor do I believe that the current secretary plans to be the one making the decision. The whole process we had laid out in 1990 to support the secretary's decision we do not see being repeated.

Does that answer your question?

**DR. NORWOOD**: Not completely. Later on I will ask it again, a different way.

**PARTICIPANT**: Just a follow-up to that. My understanding is that the methodology assumes that any discrepancy between the census and the sample is the fault of the census. If you have someone in the sample but not in the census, that is a gross omission. If you have someone in the census but not in the sample, that is an erroneous enumeration. Tacitly, you are assuming that the results of the survey, the A.C.E. bit, are completely correct. If you do not have any idea what the error rate is on that side, I do not know how you can judge the accuracy of the adjustment.

**DR. HOGAN**: Depending on how you are using your words, I think you are incorrect. We do not assume that someone who is in the census and not in the survey is incorrect. That is, we would not assume that this, "in the census" [and] "out of the survey," is zero. Our assumption is that, after our initial matching, we will have people left over who were in the census and not in the survey, and we will then send census employees to the field to verify whether those people are real people—

[Transparency]

—and to verify the criteria, that those enumerations are appropriate, that they do not refer to Donald Duck, that they are unique—well, uniqueness is an office operation, and completeness is an office operation, but also geographic correctness.

[Transparency]

So on the leftover census folks, we have a process that attempts to verify that the leftover people in the census, not in the survey, are correctly in the census. It is not true that we assume that all of these are incorrect and are zero.

The flip side is that at least some of the non-matched A.C.E. cases are verified. When we have reason to question the accuracy of the A.C.E., we will again verify it. So the premise of your question, I think, is incorrect.

**PARTICIPANT**: I want to be sure I understood the differences in the implications of matching, which I think went back to what you just had up there about the geographic completeness. The difference between 1990 and 2000—that is, the change—is focused now on the geography of the housing unit in the adjacent blocks as opposed to simply matching. Do you have some sense of what the implications of that are for numbers of matches or the proportion of matches out of block?

**DR. HOGAN**: There are two differences in terms of the surrounding block search from 1990, one of which is that the universe is defined differently. Second, we will only be doing it on a sample basis. So in terms of defining the universe, we look back at our 1990 records. Most often, when we searched the surrounding blocks, we found nothing. That is, by and large, what you find, because by and large the census gets it in the right block. Of what we found in the surrounding block, most of those were because the housing unit had been mis-geocoded. That is the component we will be picking up. I believe it was the bulk of what we found in 1990. Occasionally in 1990, we would find the kid, who should have been counted with the mother, who was counted with the aunt—very small. We will not be picking that up as part of either the surrounding block search for duplicates or the surrounding block search for enumerations.

We are going to be doing this on a sample basis. It is going to have a certainty component that we will talk about later, for cases where we think there is a lot of geocoding error and a sample just to make sure that it is a reasonably unbiased estimator.

I do not think that is true. I have heard that said, but it does not correspond with anything I recall from 1990. Let me answer two ways. First, if you continue to expand the search area, you would get more matches. That is true. I certainly do not think it would be anywhere near the majority of the non-matches that would be converted in—I think it would be a fairly small proportion.

But, in addition, the more matches you convert by expanding the search area, the more erroneous enumerations you convert to correct enumerations by expanding the search area. So I believe that had the 1990 search area been expanded from—the 1990, in urban areas, was one ring, the sample block, the sample cluster, and any block that touched that. I believe, based on my recollection, had that been expanded two, three, four, or five rings, the number of non-matches that would have been changed by adding additional rings would have been fairly small. We did some research on that to see, for 2000, whether to have one ring or two rings or three rings. Certainly, we did not convert the majority of the non-matches to matches. That simply did not happen.

But as you expand the rings, you are doing two things. You are converting some non-matches to matches, some people you thought had not been correctly enumerated to saying they had been correctly enumerated. You are also changing some census records that you thought were incorrectly enumerated, but now you are giving a looser definition of what you mean by "correctly enumerated," changing those to correctly enumerated. At least as a first-order approximation, those are balanced. It does not change the expected value. There are some second-order effects that are fairly subtle. But I do not think that expanding the search area would have had any overall net effect on the 1990 estimates.

**PARTICIPANT**: I think I am citing the case report.

**MR. THOMPSON**: I think what he might be referring to are some studies where there were two blocks in the PES, where it was a geocoding error and if you had expanded the search area for those two blocks, you would have found some individuals who were there. It is a very limited analysis of two blocks. It may have been the two outlier blocks.

**DR. HOGAN**: In 1990, of over 5000 blocks, we had two blocks that had many coverage errors. On the one side, there were many missed people; on the other side, many erroneously enumerated people. On those particular two, out of the 5000 sampled blocks, had we expanded the search area, we would have gotten a lot more matches. So in a couple of cases, that was true, but as an overall proposition, it is not consistent with my recollection from 1990 at all.

The 1986 test in Los Angeles? What about it? My recollection—and this is 14 years ago, so my recollection may not be 100 percent—is that we tried to search the entire site, not just for geocoding errors, but to see if anybody had moved from one side to the other. Our definition of being correctly in the census was having been counted in the census anywhere in the test site, and you should have been counted in the census anywhere in the test site. On the P-sample side, our definition was that a person was correctly enumerated if he[/she] was counted anywhere in the test site. If you remember the proportions in and outside the clusters in 1986—I sure do not—I do not think it was the majority of matches, by any means, that were outside the sample clusters. But I have not reviewed the 1986 pretest in many years.

**DR. NORWOOD**: Would you review why you think this different treatment of movers is better? It is important, I think.

**DR. HOGAN**: It is an important difference.

[Transparency]

Our goal is to get a sample of everyone who should have been counted on April 1, 2000. Essentially, there are three ways of doing it. For any location, you have the number of people who moved out of that location to somewhere else, and you have the number of people who moved in from somewhere else. By and large, the people who move out, at least nationwide, are the same people who move in nationwide. There are essentially three exceptions. There are people who move to Canada or Mexico or wherever, and the people who move back from there. Also there is a small exception where, if someone moves in or out of the A.C.E. universe—that is, moves into group quarters or moves out of group quarters—then there is an

exception to the inmovers not being the same as the outmovers. For most of the nation, most of the population that moves between the time of the census, April 1, and the time of the PES/A.C.E. interview, July—it is the same people who move out or move in.

You can sort of define the population as those who do not move, who are in the same place in July as they were in April, and those who move. The question is, what is the best way to measure the undercount of those who move? In 1990—and, as a matter of fact, in 1980—we went to the housing units in July and we said, "Who is living here now?" Then we said, "Where did you live on Census Day?"—more than just, "Where did live on Census Day," but, "Can you give me the names of the nearest neighbors? Can you give me the names of the nearest streets? Are you near any Piggly-Wiggly's?"

The advantage of that approach—and it is called the PES-B in the jargon—is that you are talking to the person. The person is there and you are talking to him. The advantage is that you are not getting a proxy interview.

The disadvantage of that is that the matching is fairly difficult and can lead to matching errors and can lead to nonresponse on the matching. If I ask you where you were living several months ago, you might give me the right address and I could write it down wrong—just a transcription error—or you could say you were on Rural Route 3, just outside of Buffalo, North Dakota, and I have to try to figure out where Rural Route 3 just outside of Buffalo, North Dakota, is. So the matching is fairly difficult. It takes a lot of time and effort, but it can also lead to errors.

There are a couple of alternative approaches, one of which we tried in the dress rehearsal. All these go back in the census evaluation literature at least 20 years. The alternative approach is to show up and say, "Tell me who lived here on Census Day," to try to reconstruct the Census Day housing unit. That can be difficult. You are asking for several months before. But to the extent you have reconstructed the April 1 household, then the matching is easy. You know where you should be looking. You should be looking here. It goes into the matching just like any other nonmover, and so the matching errors are much, much fewer with this approach.

**DR. NORWOOD**: Geographic matching errors.

**DR. HOGAN**: Geographic matching errors, yes.

**DR. NORWOOD**: What about the two characteristics that you want to have?

**DR. HOGAN**: The interview can be less complete. You have to take more proxies. Even if you are talking to a household member, you are asking about who was living there several months ago. So the interviewer is probably going to have a higher non-interview rate, and there may be more people down in the fourth cell that are missed. You probably will not get as many movers using this reconstruction approach as you do by saying, "Who is living here now, and have you moved?" If you go to a household and you try to do both—if you say, "Who was living here who has left? Who is here now who has moved in"—traditionally, you get more people reported to have moved in than you get reported to have moved out. You get more reported inmovers than reported outmovers. But we know that, since they are the same people, the number should be the same.

But there are two components to matching movers, to getting the right number of matched movers. One component is getting the right number of movers, and the other is getting the correct enumeration rate for movers. The technique we are using for 2000 tries to compromise between the two ways of reconstructing movers, the PES-B, which was used in 1980 and 1990, and the PES-A, which was used in dress rehearsal.

**DR. NORWOOD**: It is really a tradeoff.

**DR. HOGAN**: It is very much a tradeoff.

**DR. NORWOOD**: With some plus, you think, on this side.

**DR. HOGAN**: It is trading off. The B, in terms of whom you initially get in the interview, is probably more complete. You are going to lose some of them because they do not give you a complete address to match. But to the extent you can hold them in the survey, you have more complete coverage of movers directly by the PES/A.C.E. The flip side of that is that the matching is more difficult and perhaps could lead to more errors. You have to say, "Where were you living," get cross streets, get out maps—and I have done matching in Alaska where the maps were as big as this stage, literally.

So what we are doing in 2000 is a compromise between PES-A and PES-B, which we call PES-C.

[Transparency]

What PES-C says is that the best way to get the number of movers is to say, who is here now, who has moved in. We get the number of movers as the number of inmovers. The "$i$" there means inmover.

However, as I said, the matching of inmovers is fairly difficult. We get the coverage rate of movers from those who were here on Census Day and who have left [outmovers]. We get the coverage rate here. It is designed to get the best of both worlds. It gets the best number of movers from the best source, and it gets the best coverage rate of movers from the best source, and multiplies them together.

We did some research in dress rehearsal. The research had some problems. We could spend hours discussing the problems. We tried to trace outmovers. We tried to ask people who had left, find them and interview them, and see whether the proxy interview was a good interview. In general terms, what we found was, when you actually followed up, you got a different number of movers. But at least among those we were able to contact and follow up, the match rate was about the same as what we had before. You can say that the ones we did not follow up obviously—depending on your predilection—had a much higher or much lower whatever. But at least among the ones we could follow up and trace coming out of dress rehearsal—and I do not want to put too much on this one study—it indicated that this was a fairly good estimator and one that we are comfortable with.

**PARTICIPANT**: I wonder if I could follow up on a point you made about the group quarters. College students are in group quarters. It is my understanding—I could be wrong—that A.C.E. is not really providing an adjustment for the group quarters population. But the college students who leave their dormitories at the end of school, which would typically be June—so they are in a possible P-sample housing unit after that point—would be inmovers. Is this something to worry about?

**DR. HOGAN**: I do not think it is something to worry about; it is something that is there. It is certainly here. Some of the inmovers were people who moved in from, essentially, out of the universe, and so will raise the number of inmovers inappropriately. To the extent that overcompensates to the number of just general movers—I think A.C.E. probably does not get all the movers, but, at least in this universe, we could slightly overcompensate for the number of young mobile adults somewhat. So it is there. It could happen.

**DR. BROWN**: Bruce raised one issue having to do with this static assumption on the population. I have another question relating to that. The way you have this map laid out, with people going out and coming in, does not allow just for seasonal movements, like from the South to the North. Do you have any way of judging whether that is a problem?

**DR. HOGAN**: I think it is part of a general issue, certainly in terms of the nation, if you ignore international migration. In group quarters, the number of inmovers and the number of outmovers are the same. However, that may not be exactly true, and probably is not exactly true, post-stratum by post-stratum. If between April and July there was a large net movement—and "net" is the important word here—between large MSAs [metropolitan statistical areas] and rural areas, then we would measure a lot of inmovers in the rural areas, which is more than the number of movers, and we would measure too few [in MSAs]. So to the extent that there are large net flows, migration flows, between April and July, crossing post-strata, then that will either artificially inflate or artificially deflate our measure of the number of movers within each post-stratum. It should balance, obviously, over the nation, but post-stratum by post-stratum it may not balance perfectly.

I do not think that the level of net migration in those two months between post-strata is significant. A lot of people move, but a lot of people move within the same block, and a lot of people move within the same neighborhood. We can check with our demographers and run this down. The numbers we could expect within a two-month period—the net difference of migration between post-strata in a two-month period we have not explicitly looked at. My guess is that when we look at it, we will find it is quite controllable.

**DR. LITTLE**: When you have somebody who has moved out after April 1, and you are then interviewing a proxy, I would have thought they would know something about the people who were in April 1, but I would not have thought they would know detailed characteristics of the household. My question is whether you actually try to follow up. Presumably, a lot of people might leave a telephone number or something. Do you actually try to get a real interview for the people who have moved out?

**DR. HOGAN**: We tried that in dress rehearsal. It is a fairly complex process. It requires a highly skilled interviewing task force. In the dress rehearsal, we were able to deploy such a staff, and they were able to follow up and interview, very roughly, about half of them. The result was, the rate was about the same; the number was different. We carefully looked at this and decided that, given the complexity of this kind of tracing, it really was not realistic with the sample size that we are talking about. So, no, we are not going to be doing that as part of the production A.C.E.

**DR. NORWOOD**: I think it is time for a break. We will come back to further discussion.

## DISCUSSION OF TREATMENT OF MOVERS

**DR. NORWOOD**: I hope everyone is in place, because we have to really move on. I have been asked by our recorder to ask people to use microphones when they speak. Second, will you please identify yourself so that when we go back over this record, we will know who was here and who said what. We are also going to be passing around a pad for people in the audience to please identify themselves, so that we will know who attended. The one thing, of course, that I forgot to do was to give the panel a chance to identify itself. I would like very much to do that....

We have had the first set of issues presented, and we had some discussion. My sense is that there were some more questions about the issues that Howard raised. Phil?

**DR. STARK**: Regarding the difference between PES-B and PES-C—correct me if I am wrong—it seems that in PES-C not only are you using inmovers as a proxy for outmovers in estimating numbers, but you are also using people's recollections about other people instead of people's recollections about themselves, both of which seem as if they would introduce possibly some biases.

**DR. HOGAN**: On the first issue, the way I think of it is, I am trying to measure movers. To me, inmovers and outmovers are, by and large, the same people. At least in the way I conceive of it, we are not substituting inmovers for outmovers; we are measuring the number of movers by asking about the number of inmovers, and we are measuring the coverage rate of movers by trying to match the number of outmovers. So substituting one for the other is, I do not think, a big issue, except, as Larry pointed out earlier, if we have a net flow across post-strata. Then it could be a problem.

You are right on the second issue. We are, in the PES-C, at least for outmovers, taking the reports about them from not themselves, but either members of their households, because they moved individually, or their neighbors or the people who moved in, who might happen to know. So we are relying on the recollections of proxies to find out about the outmovers. This has two consequences, maybe three, one of which is that you probably do not get the right number of movers this way. That is why we go to the C. Second, the number of people who move around that you get reported by proxies may be a differential subset of all movers. In other words, they may be somewhat more likely to report the number of stable, easy-to-count movers. So even among the movers, you may be more likely to get easy-to-count movers than all movers. So that is a problem with the C. You are getting the proxies.

Contrasting that with the B, you are trading that off against the matching problems and the reporting problems of someone telling you with sufficient detail where they lived several months ago, detail enough not just to find it, but when you do not find it, to be absolutely certain that you are looking in the right block—that is, that you did not find it because they were not enumerated (pardon the double

negative), and you did not find it because you were not looking in the right area. When you do not find a corresponding enumeration, you want that to be because the person was not enumerated where he[/she] should have been enumerated, not because you were looking in the wrong place.

**DR. MULRY**: On this topic of movers, what is the match rate that you have been observing for the outmovers? How does that compare with the match rates that were observed in 1990, before imputation and after imputation, both ways?

The other topic is, I am a little confused as to whether you really have a consistent definition of the P-sample population, particularly in the area of college students. What is the definition on Census Day residency versus where they have to reside at the time of the P-sample interview? There seems to be some inconsistency there as to people in and out of group quarters.

**DR. HOGAN**: On your first question, I do not have those numbers. We can get the numbers. The problem in all of these comparisons is that we used PES-A in the dress rehearsal, which was only in two sites, Sacramento and South Carolina. Not only are those not particularly representative of the nation, but the census procedures that we were measuring the undercounts of were different from 1990. So comparisons between that and 1990 are probably problematic. But I can find the match rates.

**DR. MULRY**: Your match rates ought to be somewhat in the same ballpark, going back to your premise that everybody who is a mover is a mover, and it is just how you capture them. The concern is, with PES-A for the match rate, the people that you can find with proxies are going to probably be more stable people, who are normal people who participate in government and are more likely to be enumerated than the ones that the proxies do not know about.

**DR. HOGAN**: I can get those numbers. But in terms of comparing, the dress rehearsal specifically, especially in South Carolina, the census part of it, suffered from some big deficiencies. We had some block canvassing and other things that we had done in 1990 and will do in 2000 that were not done in the dress rehearsal. So just in general, the non-match rates, the coverage rates for South Carolina, in the dress rehearsal were very different from anything we have ever seen or probably will see again. I still do not think you can make a 1990-2000 comparison in numbers. But in terms of how the movers or nonmovers look like each other in the dress rehearsal, we have that, and I think we can get it. I just do not have it.

**DR. MULRY**: What about my question about the definition of the P-sample population with regard to movers? Is that really consistent? If you have a college student, it is not where he lives on Census Day. It is not in the P-sample population if he is in a dorm, but where he lives at the time of the P-sample interview he is in.

**DR. HOGAN**: I think our definition is consistent, although, as pointed out by a couple of the other questions, there is a measurement error in measuring that concept. The concept throughout the P-sample is that the P-sample should be a sample of everybody who resided in housing units as of census reference date. That is the concept that we are trying to apply. When we get the rate and we try to reconstruct the housing units for the outmovers, we are able to apply that very well.

It is true that when we try to measure the number of movers—we are trying to get, actually, the total number of people. The number of people who did not move, plus the number who moved, together, is everybody who lived in that post-stratum. That is what we are trying to measure. When we try to measure that second component, we have a couple of measurement problems, one of which is that the outmovers that you are able to find information about probably are not a random sample of all outmovers or outmovers in general. So there is a measurement bias there.

In addition, you are trying to get the number of people who live there and who have left by the people who have moved in within that post-stratum. To the extent that there are net flows between post-strata or net flows between group quarters and non-group quarters, or indeed net flows internationally, there will be a measurement problem. I think it is not a big measurement problem, but there will be a measurement problem on that second number.

But our concept of what we are trying to measure, our concept of our universe, is fairly clear.

**DR. ZASLAVSKY**: Another question, which you also probably cannot answer, would be the other piece of this. How many of these outmovers are you actually able to get matchable information for? Do you have a number? Do you have some basis for making a guess?

The other question is, more broadly, do you have plans for some kind of evaluation of this aspect, since it seems as if you are coming in without a great deal of experience in counting through an operation of this type?

**DR. HOGAN**: I do not have it. We can get the following numbers. Of the people who are reported as outmovers, we will know, and we can tabulate, those that we said were data-defined in terms of PES criteria and those that we said lacked sufficient information for PES criteria. So those two components we get.

Obviously, to the extent that the PES has a coverage deficiency itself—and so there is a group you are never told about—we can infer that size, in some way or the other, but we obviously did not observe that.

**DR. ZASLAVSKY**: If you have a lot fewer outmovers than inmovers—it seems you do think so; that is why you are using the other number.

**DR. HOGAN**: Yes, we do.

**DR. ZASLAVSKY**: Are you getting 80 percent matchable or are you getting 10 percent matchable? There is a question of how much room there is there for the difference to be important.

**DR. HOGAN**: I think it is closer to 80 percent than 10. The differences between the inmovers and the outmovers—we get more outmovers, but not by orders of magnitude. We will get that for you. We do not have it here.

On the evaluations, we did an evaluation by trying to trace the outmovers in dress rehearsal. It has been written up and is publicly available. I do not believe we are planning to repeat that. I saw Ruth Ann Killion here. She can verify—we are not planning to repeat that. We did it in dress rehearsal and wrote it up. If it has not been provided to you, we would be more than happy to do that.

**DR. LITTLE**: Isn't a more serious problem not so much that you have missing data, so you do not have enough information to match, but that you have data from the proxies that are subject to response error, and therefore you are going to get bias from underestimating the match rate? It seems to me the fundamental concern of this whole process is the response error issue, not inability to match.

**DR. HOGAN**: Right, exactly. There is the response error, and specifically, since we are getting the number from this other component, it is whether the proxies, when they tell us about the movers—are those movers more likely to have been counted in the census than all movers? That is probably true to some extent.

**DR. LITTLE**: No, that is not the point. You have a list of members of the family, for example, and you have the ages of the kids wrong and you have the names of the kids wrong. If that is treated as true information and then you try to match against the census information, then you get more non-matches, and that goes into the match rate.

**DR. HOGAN**: Oh, okay. That can and does happen. Any non-match case from a proxy is followed up. If we have gotten it from the neighbor or the landlord, a proxy respondent, and we do not match it to the census, that does go to our follow-up.

**DR. LITTLE**: Is the proxy information sort of taken into account when you try to do the match?

**DR. HOGAN**: Not explicitly. It is taken into account explicitly in what we send to follow-up, but I do not think our interviewers are told to reference that. David [Whitford], are they? They are. I stand corrected. Yes, they are told about the proxy information and told to take that into account.

**DR. BELL**: I want to ask about another group that is sort of coming in, and that is newborns. If someone reports a newborn in a house, is that person basically ignored?

**DR. HOGAN**: Yes.

**DR. BELL**: By analogy, ideally, one would want to treat somebody who had been in group quarters and comes out of group quarters into a sample population in the same way. But you are not doing that, presumably because of measurement difficulties.

**DR. HOGAN**: It is the net flows that count in terms of measuring. We are just measuring the numbers, the net flows. For newborns, the net flow is, of course, deaths. But they are almost certainly in different post-strata. Saying I am balancing newborns against old people who are dying makes no sense whatsoever, so we do not do that. But with, I think, the exception of college, the flow in and out of institutions—it is balanced in net, and so I think we do a better job by including those people.

**DR. BRADBURN**: Isn't it conceptually different? The newborns are not in the census. The newborns are not, conceptually, to be counted in the census, whereas the group quarters people conceptually were counted in the census.

**DR. HOGAN**: Yes, but [not] in the A.C.E. universe.

**DR. BRADBURN**: Right. But with the newborns, they should not be in the census. It is asymmetric.

May I just follow that up with a more general question about this whole set of—not just the group quarters. This is not moving. In the dual-systems estimate, what do you do with that whole group of people who are not in the A.C.E.—that is, group quarters, homeless, various kinds of people who are not going to be in the universe for the A.C.E.? Where do they fall in those cells?

**DR. HOGAN**: For those, we just take the census count as reported. For college students, for remote Alaska, and for military group quarters, homeless, the census count is sort of the final word. Implicitly, you could say they get a coverage factor of one. We do not apply the one universe to the other.

**DR. STARK**: I am concerned about the change in the definition of the correct geography for A.C.E. My understanding from your paper [Hogan, 2000] is that what is determined to be the correct location to have counted someone is where the person recalls having lived at the time of the interview, as opposed to where the person, in fact, was living at the time of the census. It would seem that, to the extent that people do not recall correctly where they were living, that would inflate the mismatch rate and introduce bias that way.

**PARTICIPANT**: Your paper says where the person believes, not recalls. It is actually subjective.

**DR. HOGAN**: It is. Here is the issue. Let me try to explain it. It is a fairly subtle one, but it is an important issue. We want to define on the census side, in our whole dual-systems model, a set of enumerations that are correct, and want to allow matching only to that set of enumerations. So the question, then, is, when is a particular enumeration correct? If someone has two addresses, for example—to take an easy example—it really, in some sense, does not matter which of those two addresses, at least as a first order of approximation, we assume is the correct one, as long as we consistently apply our rules in defining the number of correct enumerations and the proportion of people who are correctly enumerated. As long as we say this person should have been counted in Florida, if we apply it consistently and count him as correctly matched only if he was counted in Florida, and erroneously enumerated if he was counted anywhere else, then the model works, in that we are matching to the right universe.

Then the question is, how do you define what you mean by "correctly enumerated"? There are a number of ways you can think about this. I think, operationally, they probably do not have a lot of difference. But it is important to think about these things—at least I do.

Let us say you are correctly enumerated where you should have been counted on April 1. That is a rule we could try to apply. The thing is, most of our interviewers are not going to conduct it on April 1. Most of our interviews are conducted in July. So to say it is really sort of a theological truth as of April 1—you can, if you like, conceive of the model that way. I think it does no big harm. But the fact is, we only have the information about what you believed in July. If you had two addresses, Florida and Maine, and you came to conceive of your usual residence as being one or the other, because your intentions changed or whatever, we only have your response.

As long as we have a consistent reporting in the E-sample and the P-sample, the model will work. It is the consistent reporting that is very important.

Just having said that it is correct based on what you believe in April does not necessarily mean I am going to get consistent reporting. We accept proxies, and where they believe you lived may not be where you believed you lived. We have follow-up interviews months later; therefore, the response you gave us in September may not have been the response that you would have given us if we had conducted the full interview in July. But at least as a way of thinking, we are saying, we are here in July, this is our reference date; you get to choose an address, and once you choose an address, that is where you should have been counted, and we are going to apply it consistently.

But, as I said, if you want to think about some sort of overarching truth, you can think that way. But then you have to somehow figure out how you are going to measure that.

I do not know if that answers your question or not.

**DR. NORWOOD**: Are there further questions on the subjects that we have talked about?

[No response]

Let us move on now to post-stratification, on which I know many of us have a lot of questions.

I can just tell you that I remember when my husband was doing some genealogical work, he went to the 1920 or the 1900 census and came home and said, "You know, in the census, I found that I had another uncle who had never been heard of before, and who, upon further examination, really never existed."

I think that was in the days when interviewers were paid on the basis of the number of people they found.

## ESTIMATION DOMAINS (POST-STRATIFICATION) FOR A.C.E.

**DR. HOGAN**: Okay, post-stratification: Let me sort of remind you of the roles the post-strata play in what we do, and the roles they do not play. I think, for other groups, the third thing is very important.

[Transparency]

First, the post-stratification determines the cells for the dual-systems model, and therefore we are trying to get sort of the homogeneity for the dual-systems model to promote independence. Second—and it is conceptually a different question, but we use the same post-strata—we use the post-strata for carrying down the adjustment to the small area as part of our synthetic model. Those are really two uses, the dual-systems model itself and the synthetic model. We use the same post-strata for both because I think the same criteria lead us to the same conclusions in both. But you need to think of the two uses when we think through this problem.

Secondly—and I do not need to say it in this group, but I say it in every other group—post-stratification has nothing to do, or very, very little to do, with how we tabulate. So some of the issues in terms of how we group together people for DSE

post-stratification and synthetic adjustment post-stratification, the A.C.E. post-stratification, have nothing really to do with how we tabulate. When we group two states together in a post-stratum, that does not mean we are not going to tabulate two states separately, or when we group a couple of small towns together in a post-stratum, it does not mean we are not going to tabulate two separate towns. Similarly, when we get into the issue of multiple-race people and we start grouping people together for post-stratification, that does not mean the Census Bureau will not tabulate the separate groups separately.

I know some groups I have discussed this with, less learned than this group, when I talk about grouping people together for DSE, they leap to the conclusion that the census is ignoring the important differences. So I wanted that on the record.

[Transparency]

In choosing post-strata, we explicitly wrote down a number of criteria that we would like to take account of in forming the post-strata. First is the familiar one: we want to put in together people with as similar as possible capture probabilities. That is, for the dual-systems model to work, we want within a post-stratum people who are missed by the census at, as nearly as possible, uniform rates.

Secondly, we want similar net undercounts. That is a different criterion. It is the criterion that is used for the synthetic assumption. When we do the synthetic assumption, we are assuming that the net undercount rate is uniform for the post-stratum. That is different. The first is really a gross omissions rate; the second is the net rate, including the erroneous enumerations, and indeed the traditional census adjustments.

The third criterion takes a little explaining, the differences in geographic areas. It has very little to do with the DSE model and it has something to do with the synthetic. We would like criteria that kind of differentiate areas geographically, to the extent we can. For example, age/sex—except maybe for some college towns and military bases—tends to be fairly homogeneous. Every county has young and old, men and women; every city has young and old, men and women. So age/sex does not really help us a lot in differentiating areas for the synthetic adjustment, although it might be very important for the dual-systems model in terms of separating out different capture probabilities.

The fourth criterion that we use is similar to what we have done in ratio estimates. We want an adequate sample size within post-strata. We want the expected sample size within each post-stratum to be about 100, being fairly conservative. That helps control the bias, the ratio bias, as well as keeping the variance under control.

Next is operationally feasible. I will come back to this. It turns out to be fairly important. We want the post-strata definitions to be something we can carry out, something the Census Bureau can carry out, not based on things that are not available, not available in time. For example, I was in Los Angeles and people said, "Boy, you really should post-stratify [by] immigrant status. That is really the thing to explain undercount." Immigrant status, of course, is only on the long form. The long form is not going to be tabulated until much later. It is not operationally feasible.

The next criterion is that we want to minimize classification error. This has a couple of components. The dual-systems model places people in that $2 \times 2$ table, but it is getting them from two different data sources. It is getting them on the census side from how they reported in the census, and it is getting them on the A.C.E. side from how they reported in the A.C.E. If they report differently, that model is going to be slightly off.

Also if, say, the group of people whose net undercount we measured in the A.C.E. is different from the group of people that we are applying it to synthetically in the census, we will have a synthetic bias there. So we want to choose categories that are robust to response variance.

We want to account for changes since 1990. The census has changed; the A.C.E. has changed. What a variable might have meant in 1990 is quite different. How we are conducting the census door to door is different. We need to take that into account in coming up with our variables.

Finally, explainable—variables that are explainable. Some of the things we looked at we could barely explain to ourselves. They were a lot of fun. We have gifted people at the Census Bureau, with clever minds, and they came up with some really interesting stuff, but it was not stuff that I thought we could explain to anybody. So that was one of our criteria.

Those were what we were trying to optimize in forming our post-strata. We sort of had a rule of thumb that if we had a group left over and we did not know what to do with it, we would group it with the largest group. The reason for that is, for the largest group, you can break it down on other variables as well.

So these were our criteria. We did a lot of research. I will tell you where we are, and then we can discuss various aspects of it.

[Transparency]

By way of history, in 1990, for the original estimates, we had 1,392 post-strata, based on a number of things. But then that was refined by 1991 into the famous 357. The 357 is sort of where we started here. These post-strata are similar. I will talk about the differences between what we have here and the 357.

First, we have seven race/origin groups: American Indians living on reservations being one group, a separate group, which is why we get seven; American Indians and Alaska Natives not living on reservations, including those living on tribal or trust lands, but not formally recognized reservations; Asian; Hawaiian and Pacific Islander; Hispanic; non-Hispanic black; and non-Hispanic white and other races.

Then we have seven age/sex groups. Those are exactly the same that we had in 1990, the 357, one group for children under age 18, boys and girls together; 18 to 29, males and females separately; 30 to 49; and 50-plus. So these are the same groups we had in 1990.

Tenure, owner and nonowner—this is a variable we first identified coming out of the 1980 census. We used it quite successfully, I think, in the post-stratification in 1990, and we are continuing that through. A lot of our results have to do with that.

This comes to our first change, metropolitan statistical area [MSA] and type of enumeration area. First, let me say what we did in 1990, which we based on census urbanized area. The census, after the census, classifies areas based on density and a fairly complex set of criteria, including boundary changes and everything else, whether they are urbanized or not. Then we can sort of group the urbanized areas based on the size. That is a process that happens after the census. It is based on the census data.

In the second 1990 PES, the 357, the one we came up with in 1991, this had already been done, and we used it. In 2000, the census will be based on the results of the A.C.E. So, presumably, how urbanization is defined, how density is defined, will be based on the results of the survey. Therefore, it is a little tricky to post-stratify on that variable. We did look at it. We did work with our people in geography to see if they could do something preliminary for us that we might use. After a lot of discussion, we decided that it really would not work, that it probably was not operationally feasible to use urbanization.

So we have looked at other measures of the same kind of criterion—density, urbanization. What we have come up with is, essentially, MSA classified by type of enumeration area. The type of enumeration area in this discussion really breaks into two parts, the part of the census done mailout/mailback and everything else. Everything else includes list/enumerate, where we simply show up at the door, list the housing unit, and count it right then; update/leave, where the census interviewer has a questionnaire, goes out, and updates the address and leaves the questionnaire to be mailed back; and there are a bunch of other little things left over.

But we use these other types of enumeration procedures in more rural areas, in areas where mail delivery is not adequate. So we think that doing MSA and type of enumeration area as one kind of combined variable gives us a lot of what we had before in urbanization. Essentially, what we will do is take the mailout/mailback universe, classify it into large MSAs, medium MSAs, small MSAs, and everything else that is done mailout/mailback—towns and whatever else—and a fourth category, which is why I have four there, which is all non-mailout/mailback types of questionnaire delivery/enumeration methods.

**DR. SPENCER**: Could you explain about the urbanization and why you need the A.C.E. results to define an urbanization variable for individuals?

**DR. HOGAN**: Let me answer in two parts. The official urbanization variable will be defined based on the official census counts, which will include the A.C.E. So it is not possible to use the official one to construct the A.C.E. So that is clearly out.

Then the next question is, could you construct another one that might be fairly similar, but constructed based on the unadjusted census numbers? It could theoretically be constructed, but the people who would have to do it looked at their workload and said, "We really cannot do this twice, and so we will only do the official ones."

**DR. SPENCER**: When I was reading one of the reports, it looked as if you had considered a housing density variable, which you could even get off the MAF and

how many units per block. It would vary locally. In a city like Chicago, you could have lots of different urbanization levels, and you could distinguish high-density housing from lower-density housing.

**DR. HOGAN**: Yes, we looked at a number of things, including that. It came down to the people to do this. Our geography support staff—it would have to be done block by block—simply said they do not have the resources to do that until they have already been scheduled to do it. They basically said it was not operationally feasible.

We could have, to be quite honest, if we thought that that was miles ahead of what we came up with, we could have gone back and said, "Look, you are going to do it," and they would have done it. But we looked at it, and we said, given the overall goals of the census, this is good enough so as not to jeopardize the other things that geography has to be doing. I think the A.C.E. is the most important thing in the census, obviously, but getting the maps of the local areas to the enumerators and the other things that the Geography Division has to do, balancing the needs of the rest of the census, the other 90-whatever percent of the census, against how much better urbanization would have been over this, we said this is probably what is best for the census as a whole.

**DR. SPENCER**: But you did not actually construct this kind of very local urbanization or housing density variable and try post-stratifying on it to see what effect it might have?

**DR. HOGAN**: No, we did not. All of our research that is reported in the background material is obviously based on 1990 [see Schindler, 2000]. Some of the decisions we have made are based on our reading of how the census might be different. One of the things we were not able to test on 1990 data was the impact of the type of enumeration. The 1990 census defined the areas that were list/enumerate quite differently than will 2000, and the methods are somewhat different.

To a certain extent, we could not have tested this combined TEA urbanization variable on 1990 against the combined MSA/TEA variable on 1990. It came down to a judgment of how we thought 2000 was most likely to come out.

One big change from 1990—and I think a very good change—is the mail return rate. This is measured, essentially, at the tract level. For every tract in the country, we know how many questionnaires we delivered and how many were returned by mail. We can compute for every tract in the country how cooperative they were. So for every person within that tract—it is an environmental kind of variable—we can ascribe the tract's characteristic. This person lived in a tract that was highly cooperative; this person lived in a tract that was ot very cooperative.

Having transcribed that to people, we can look at the people and find out, say, among our major groups, our major race/origin groups, the median mailback rate of the kinds of areas they live in and post-stratify the people based on whether they lived in a high-mailback area or a lower-mailback area.

This is fairly important in two aspects. It helps us, we believe, on the dual-systems model itself. But it also helps us in the second goal of the post-stratification,

the synthetic estimation. It is, within some of these metropolitan areas or what-ever, what kinds of people are cooperative and what people are not. We think it helps a lot there. We have been quite pleased with how that has come out.

So that is one variable we have added. It is quite new.

The final variable was in the 357, the four regions, but we are really moving back from it. In 1990, it was sort of a major variable that was used often. With the mail return variable, we think that is less important. A lot of what we believe we are picking up in region, we believe is now being picked up by the mail return variable. In terms of undercount, the Midwest looks different, because they mail back their questionnaires. So we think this is less important than it has been. We are keeping it only for non-Hispanic white and other owners, just for our largest group, where we have plenty of sample, where we believe we are going to have enough sample to support region. But it is only for that group that we are keeping region. Otherwise, that is not part of our post-stratification.

That is it on the overall. I can take some questions on the general, and then come back to the—if you want to create a discussion, talk about census adjustment and multiple race in the same sentence. It works every time.

**DR. NORWOOD**: How are you going to get to seven in race/origin? I keep hearing that the Census Bureau is going to be compiling something like 62.

**DR. HOGAN**: The Census Bureau will be tabulating, for many purposes, every combination and permutation of the major race groups, crossed by Hispanic/non-Hispanic. So I believe it is actually 124. We looked at that, and we decided to—let me say how we looked at it.

[Transparency]

We looked at the goals of the post-stratification. These were explicitly our goals. We said, what groupings of the people who choose a multiple-race option meet our criteria the best? For example, it was suggested that we could put all multiple-race people together in a category, and that will be the multiple-race category. For some tabulation purposes, for some uses of the data, that may make a lot of sense. But we could see no reason to believe that someone who marks white and American Indian is going to react to the census in the same way as someone who marks black and Asian. Just because someone identifies with two or more races does not mean that he reacts to the census differently.

So we thought, what is race measuring in the census, in the adjustment in the A.C.E., post-stratification? Race is measuring cultural, linguistic, housing iso-lation, economic-social isolation. That impacts on how the people perceive the census, how the Census Bureau is able to find the housing unit, how they and the census interact. So using those criteria, we suggested a fairly complex method of handling people who mark more than one race. I will walk through it. I hope I do not make too many mistakes.

First, we said that American Indians who live on the reservation or any other tribal or federally recognized Indian area and who mark American Indian as whatever—one race, one of two races, American Indian and Hispanic—if someone lived on a reservation and marked American Indian, we said that person has the cultural, the tribal, the community affiliations that make him[/her] part of that group. So we

said that everyone who marks American Indian and lives in Indian country goes into that one category.

I am going to simplify a little. Many people who mark more than one race and many people who choose to mark the "other" category, some other race, which is one of the categories, also mark Hispanic. We know that from our tests. Based on what we could infer from our tests, based on what we could infer from dress rehearsal, we decided that, with essentially the exception of the American Indians, anyone who marks Hispanic is Hispanic. If you mark Hispanic and five other races, you are still Hispanic.

This is somewhat of a change from 1990, when Hispanic blacks were put in with blacks. Now Hispanic blacks are put in with Hispanics.

We believe many of the people who mark multiple race will fall into this category. Remember, the other criterion we are using, which is very important to us, is classification error. Whatever grouping we do of the multiple-race people, we want to have them in the same group in the census and in the A.C.E. So looking at how Hispanic and race were reported by Hispanics, we figured that Hispanic was a more stable variable to measure.

For some of the other groups, we have come up with what look like inconsistent rules, but I think—and we have talked to at least some of our advisory committees, our minority advisory committees, and they agree with our logic—people who mark American Indian and white or American Indian and anything else and do not live on a reservation or in Indian country will be put in with white. So multiple-race, American Indian and white, will be put in with white, not American Indian. The reason is, we fear—we do not know, but we fear—that many people will discover their Cherokee grandmothers or Cherokee great-grandmothers, that many people with no tribal, cultural, community affiliation with the American Indian community will somehow mark something American Indian. If we put all of them into the American Indian stratum, we would not be able to measure anything that reflected the true undercount of the culturally, community-affiliated American Indians.

So, except on those living in Indian country, anybody who marks two races, one of which is American Indian, will be put in the other race.

On the other hand, for African Americans, we are using the opposite rule. Our reading of this society—and I think this is reflected by what our advisory committee is telling us—is that people do not lightly mark African American and something else, that the people who mark those tend to have a connection to that community. So African American and white would be put in with African American.

Finally, anybody who marks three or more races, with the exception of American Indians living in Indian country—actually, there is another exception, for Hawaiians living in Hawaii—anybody who marks three or more races is put in the "White and Other" category. We cannot come up with a rule that says where they should go. We do not understand this group at all, and we think that, at least, if they are put in with the largest group, we will be able to post-stratify them based on all the other variables—the size of place, the tenure, the mailback rate. The

place to keep this much information about this individual who marks three or more races is to put them in with the largest group.

So those are the rules that we have laid out. They are based on our reading of pretests, of the tests, of reporting of race on birth certificates, of sociology. They are not based on any previous census, because we have never done this before in a census.

## DISCUSSION OF POST-STRATIFICATION

**DR. NORWOOD**: So you are essentially using a behavioral aspect for determining it, how they behave, really.

**DR. HOGAN**: Exactly, yes, how they behave in the census, and also the response variance. We would like to put the same person in the same stratum twice. It is how they behave in terms of mailing back their census questionnaires, interacting with the person at the door, and the response. If you take Hispanic, we think that that—say it is someone from Cuba, who is Hispanic, who marks white or marks white and black or who marks only black. They probably react to the census because of their linguistic and other background, rather than because of which particular race they identify with.

**DR. STARK**: If you use this post-stratification for the 1990 data, what are the biggest differences in the adjustments that you would see? Can you characterize how much difference that makes versus the original post-stratification or the 357?

**DR. HOGAN**: I cannot, but I will see if Donna can summarize it. But before I do, let me point out that some of these do not translate back into 1990 very well. The type of enumeration area, as I said, we cannot really apply to 1990, because the areas included were quite different. So it is not possible to directly copy this back to 1990 and compare it. But we have done that to a limited extent. I do not know if Donna wants to share anything with us.

**MS. KOSTANICH**: There are not massive, big changes, but I cannot remember exactly what they are.

**DR. HOGAN**: In some of the background documents, we have some of these runs. But, as I said, with the MSA/TEA, you are not going to find it there, because it really was not there in 1990. But we do have runs where we compared them. I believe, at the higher levels of geography, it did not make a whole lot of difference. As you get down to smaller geography—for example, congressional districts—obviously, it starts making more difference.

But for the big ones that we have kept the same between 1990 and 2000, the race/origin and the tenure, in both categories, the differences are not huge.

**DR. BROWN**: This is kind of a technical question. What mail return rate are you going to use? There are different ways to measure. I remember we discussed earlier this issue of UAAs [returns marked "Undeliverable as Addressed" by the U.S. Postal Service] and whether they go in the numerator and the denominator or in one and not the other, and so on.

**DR. NORWOOD**: What is a UAA?

**DR. BROWN**: "Undeliverable as addressed." The Post Office returns the questionnaire. In one mail return rate, those got put in both the numerator and the denominator.

**DR. HOGAN**: This is like *Do You Want To Be a Millionaire?*, except that I have lots of people in the audience I can call on. Can either Donna or Eric answer that?

**MS. KOSTANICH**: I think the Census Bureau is coming up with an official mail return rate for each of the tracts that will be available. We are going to go with the official census.

**DR. BROWN**: Do you know how that is created?

**MS. KOSTANICH**: I do not know. I do not think that has been completely defined.

**MR. WAITE**: The return rate would not include in the denominator housing units that did not exist or "undeliverable as addressed" addresses. We have two kinds of rates that we measure during the census. There is the response rate, which is really intended to predict the nonresponse follow-up workload, in which case we do not know whether particular addresses are existent or not at the time they are going through. At the end of the census, we have a measure of cooperation, which is the return rate. We do not expect nonexistent units to return a form, and so we do not include that in our calculations—or vacant units either, for that matter.

**DR. NORWOOD**: So do you use different terminology for those two?

**MR. WAITE**: We do. The term that we measure in real time to try to predict what our nonresponse follow-up is going to be, and the rates that are going to be published on the Internet about how well communities are participating, we call a response rate. After the fact, after we know who is vacant and who is nonexistent units—so that we have a final picture of what the universe looks like—we call that a return rate. The return rate is always higher than the response rate.

**DR. NORWOOD**: I did not know that. Ben King?

**DR. KING**: Howard, you were talking about how reconsideration of 1990 has caused you to make some changes in 2000. This question has two parts. Are you doing anything especially interesting or innovative to make it easier to go through this same exercise for 2010? The proposed master trace sample will deliberately cover A.C.E. block clusters. Have you had time to think about the sorts of things you might do with the information from that to improve the A.C.E. for 2010?

**DR. HOGAN**: I can answer the second question very easily. No, I have not had time to think about it.

**DR. KING**: You are forgiven.

**DR. HOGAN**: I do not know if Ruth Ann Killion, who is more in charge of this kind of thing than I am, wants to add anything on what we are planning.

**MS. KILLION**: I think that clearly, as we look toward 2010 and start to plan the research for 2010, whatever happens in 2000 is going to be key. We have been working closely with the folks who are doing the A.C.E. for 2000, and the trace sample will, in fact, help us, in terms of keeping track of things and giving us answers to questions that we did not have this time because we did not have something like the trace sample.

So the answer to your question is, we are going to use every bit of information we have. Yes, there are some things that are being done that should make it a little bit easier for 2010.

**DR. BROWN**: Let me just ask another question. I think I know at least part of this answer. The way I work this out, you have something like 400 post-strata.

**DR. HOGAN**: Four hundred forty-eight.

**DR. BROWN**: Which means that, on average, the size of post-strata is somewhere between 1500 and 2000, something like that. You said a minimum size of 100, and your average size is 1500. That means there must be a tremendous variability in post-stratum size. How does that play out? Is that an efficient way to allocate resources or people in this sense?

**DR. HOGAN**: If you look at the groups we are worried about, they are groups that are both small and geographically dispersed. [One of] the two groups that I think we are most worried about having sufficient sample size for are the American Indians who are not on the reservation—American Indians and Alaska natives on other tribal and trust territories and other places—they are diffuse. We were not able to control them for sampling. We hope we will get enough in. We are pretty confident that we will, but we really do not know until we go out there. It is not something we could control our sampling stratification on.

The other group is Hawaiians and Pacific Islanders. We have tremendously oversampled Hawaii in the hopes of making sure we have an adequate sample of Hawaiians and Pacific Islanders. But even in Hawaii, it is not easy to target a sample to make sure you have enough to support a sample.

So it is these fairly small groups that, in terms of local estimation, are fairly important, in terms of, certainly, getting some of the race categories correct. Those are the ones we are quite worried about. Most of them, I think, were comfortably above 100. Of course, all this is divided by age/sex. So 448 includes the age/sex.

**DR. BROWN**: I guess my question, in part, goes the other way around. Could you do more post-stratification among the larger groups? Would that be advantageous or not?

**DR. HOGAN**: We have spent a lot of time thinking about that, arguing it, debating it, in a very lively way. We want to make sure they are small enough to minimize or reduce the synthetic bias, but also large enough to have a decent variance. We obviously do not have a really good grasp on the synthetic bias. After all, our data sets do not have the absolute undercount from 1990. But we have looked at how to balance those two, and we are pretty comfortable here.

Let me add this. In 1990, I think we did not allocate our sample very well, given the purposes the sample was ultimately used for. The sampling in 1990 was targeted on the smoothing model, which eventually we did not use. We really did not have enough sample in some of the big population groups, like non-Hispanic whites, that when we dropped the smoothing model—the areas we had a big problem with were the areas where we had a lot of non-Hispanic whites, with high sampling variance. Since non-Hispanic white owners are the biggest group, we really want to make sure we control the sample variance of that very well, because it is going to dominate almost all of your synthetic estimates for almost all areas.

I do not know if that answers your question. It probably answers it as best as I can today.

**DR. FREEDMAN**: I think this question is directed to Larry. I cannot get 448. It is 56 times 7, and then there is some collapsing.

**DR. BROWN**: No. The way I read the report, it is 56 after the collapsing.

**DR. HOGAN**: It has been somewhat changed since the report [Griffin and Haines, 2000], I am sorry.

**DR. FREEDMAN**: I am lost on this point.

**DR. NORWOOD**: Why don't you find him? Tell us where we are.

**DR. HOGAN**: I believe there are now 448 post-strata, including age/sex, and there are 64 groups summing over age/sex.

I will grant that what I put up here is not what is in the briefing book. We have done research since then. What is in the briefing book, for example, gives great emphasis to the urbanization variable, for which we substituted the MSA. In addition, what is in the briefing book, I believe, did not include at all a breakdown by region. Adding region, at least for the non-Hispanic white owners, increased the number of post-strata.

Yes, it is not what is in the briefing book.

**DR. FREEDMAN**: So is the current estimate of the post-stratification publicly available?

**DR. HOGAN**: The current number? As of now, yes. You heard it first.

**DR. NORWOOD**: What I think we are asking is, could you clarify what your position is now on the post-stratification?

**DR. SPENCER**: I have a question and an observation about the criteria for choosing the post-strata. Maybe you are going to tell us later how you decide whether one set of post-strata is superior to another, in terms of what you expect to be better accuracy. It looks to me as if the criteria you set out were used to rule out certain post-stratifications. I would just observe that I understand the first one, similar capture probabilities, because you want to minimize correlation bias. The similar net undercount you are using because you are requiring that the same post-stratification be used for adjusting for undercoverage as for erroneous enumerations.

**DR. HOGAN**: That is correct.

**DR. SPENCER**: That is a choice that you made. It is not necessary. Could you tell us how, given some alternative sets of post-stratifications, you would decide which one was better?

**DR. HOGAN**: In terms of ruling out variables, we went through a process of—everybody, I think, in the Census Bureau had his favorite post-stratum variable, and we put it into some logistic regression modeling and saw which were the greatest predictors and played around, data mining 1990. Some of those we ruled out based on these kinds of criteria. For example, in early 1990, we had many age/sex post-strata. We have collapsed them to only seven, because age/sex really does not buy you anything geographically. We talked about collapsing it a little bit more. But collapsing more, you might be collapsing across the young adults, and we thought it was important to maintain that for the DSE purposes. So that is how the third criterion came in.

The sample size of 100 was mainly important in trying to convince ourselves we could support a separate thing for the American Indian [population].

**DR. SPENCER**: I am sorry, I understood the additional criteria that you set up. But I am saying, given that you are sitting around the table and you have five alternative post-stratifications that all meet these criteria, how do you decide which one looks best?

**DR. HOGAN**: We ran the post-strata, as best we could simulate, on 1990, and then we summed up synthetic estimates for the 357, as domain, for the states, for congressional districts, and the group of 16 largest cities. After we summed those up, we then looked at the kinds of variances we would get, given the 2000 design, the 2000 sample size, so we could see what kind of variances we could get. We also—and this was fairly difficult, and we discussed this at a previous panel meeting—tried to come up with some targets to measure the bias, to see if the various ways of cutting it reduced the bias. None of them did a tremendously bigger job in reducing overall bias. Some of them seemed to be better at picking out, say, differences between congressional districts, while maintaining what we felt was adequate variance.

**DR. SPENCER**: If you have already gone through with the panel how you got these targets and how you estimate bias, I will not ask you to go through it again.

**DR. NORWOOD**: We reserve that. We can come back to that if you like. I want to say one thing. The staff and I insisted that the Census Bureau provide us with information at least two weeks in advance. They did that, even though they made it clear that some of this was still in process. My guess is that a lot more will be in process—I hope, perhaps on the basis of some of the discussion today or other things that you will think about. At some point, you will have to finalize what you are actually going to do. I think we all would like to know at what point that would be.

I think we are going to give you a rest right now, and also some lunch.

## CONTINUED DISCUSSION OF POST-STRATIFICATION

**DR. NORWOOD**: A major part of the discussion, the last part, was on post-stratification. Don Ylvisaker, do you have some questions or comments on that?

**DR. YLVISAKER**: I am still in the uncomfortable position that I am not totally in favor of adjustments, and you are giving me the question of how I like the tools for doing it. I guess my problem, generally speaking, with what is going on at the moment is timing. I got something in the mail that says January 12th. We are in the middle of doing research on the question of post-stratification. I think in August, in Baltimore, we were in the middle of doing research on post-stratification. I have had a hard time understanding why it was not being done in about 1995 or so. We could not have foreseen multiple races, for example, but a lot of what is here is based on, as I understand it, 1990 data, and so much could have been done.

Unfortunately, I will bring in history. The CAPE [Committee on Adjustment of Post-Censal Estimates] report suggested that it was not clear that this was a viable

strategy at all, and somewhere in between 1992—post-stratification in 1990 really was 1,392, and by, I think, 1992, it was 357. That report said, effectively, we are not quite sure what is going on here, whether this is viable or not. Somewhere between then and now, something has transpired, so we are doing it again.

Here are post-strata which I guess are built on pretty much judgment calls. Let me start at the bottom and just make a few remarks.

I like mail response rate. The idea of what they call the MSA/TEA kind of regional base I have no particular objection to. I get very uncomfortable when one begins to talk about race. I get very uncomfortable when we decide arbitrarily that we think white and black is black, that Asian and white is white, that Asian and black is black, that white and Asian and black is white, and a variety of things of that nature. It seems to me that, if we were in another forum, this would be very embarrassing. But we are suggesting that these are similar attributes and so on.

I will only bring out one fact. When you talk about similar characteristics, similar linguistics, similar economic isolation, things of that nature, you are talking pretty much about the definition of a census tract. The census tract is given out in pretty much that kind of language.

So I think the real problem is that post-strata should be local in character, and we cannot really do that, so we have national post-strata, and I am afraid a lot of heterogeneity. I do not know how that is addressed in just choosing.

**DR. FREEDMAN**: I would like, actually, to return a little bit to the question of movers. Now that geography has come back into the post-stratification, it does not seem feasible to me to claim that there is even approximate balance of inmovers and outmovers within post-strata. Forgive me for this brief geography lesson, but, to my recollection, people have been for a long time moving from the Northeast and Midwest to the South and West. So the notion that there will be approximate balance within post-strata does not seem viable to me.

**DR. NORWOOD**: Howard, do you want to say something about that? I am sure there will be other comments.

**DR. HOGAN**: I would like to address some of the issues. First, in 1990, we did, as Don said, start with 1,392 post-strata in a fairly complex paradigm. Those post-strata were designed specifically to be supported and to support a regression smoothing model that we used, which we have since decided no longer to use. Once the decision was made not to use this smoothing model, then it was clear that the 1,392 were not supportable. We never thought that they could stand alone. They were never designed to stand alone. That is how we came up with the 357, which, with a couple of changes that I have mentioned, are similar to what we now have.

Why are we doing the research now? Why did we not do it in 1995? That is a fairly easy question to answer. Up until last January, when we were going to use the integrated coverage measurement, the ICM, for apportionment, we operated under the assumption that we could not share information across state boundaries. Therefore, if you have to post-stratify and run 51 separate sets, the set of post-stratum variables that you can even consider is fairly restricted. Therefore, we had a fairly simple set, and that is what we were testing. Once the Supreme Court

made its decision, in January, we started a fairly intensive process of looking at other variables—and we have talked to the panel several times—to come up with a set that takes into account more information, more of what we have nationally. Of course, we made the explicit decision that post-strata could cross state boundaries, a decision that not everyone agrees with, but certainly one we were up-front about.

So this research really only got started in the last 12 months. In the last few weeks, as I said, we were working with our geography folks trying to operationalize some of the variables and ran into a glitch. They thought it would probably be wise not to pursue the urbanization, so we have gone back and relooked at it. What I am presenting to you now has slightly changed, but certainly the spirit is quite similar to what we had two weeks ago. We did not head off in another direction.

I am glad you like the mail variable and the MSA. Let me sort of agree with you, but disagree with your wording on the combining of racial groups into post-strata. The fact is, we have very little knowledge about the particular characteristics of people who might mark white and American Indian, or white, Asian, and American Indian, or whatever else. We have had some tests in the mid-decade, but with the publicity surrounding the census, with advocacy groups going out and saying, "Do mark more than one race," "don't mark more than one race," we have very little information about how these people are going to react. We are making decisions that we think are prudent, but they are based on judgment. They are not based on facts. I have to state that right away.

My disagreement with you is in saying what we are doing is making whites and blacks, blacks; we are making blacks and Indians, whites—whatever. We are combining groups together to form post-strata. We are no more making people who mark white and American Indian, white, than we are making New York and Philadelphia, New York. We are putting together groups into post-strata. We are not changing the fundamental nature of what people marked. If someone marks four different races in the census, that is how he[/she] will be tabulated, the PES notwithstanding.

The discussion about whether the post-strata should be local, whether they should be national—that is a continuing dialogue we have had, with the panel and with others. I feel fairly strongly that those variables that determine how people react to the census—variables such as attitude, housing arrangements, whatever else—are only very loosely connected with state boundaries, with city boundaries. They are to some extent, but only loosely. We get more homogeneous post-strata by grouping, say, Hispanics in Alabama with Hispanics in Mississippi than having to group Hispanics in Alabama with whites in Alabama, to take an example. So I think we do get better, but that is sort of a judgment call we have made.

On David's point about moving across post-strata, there are indeed people who move from the Northeast to the South or to the West. There is no doubt about it. The question in my mind is, is the net migration of non-Hispanic whites and others—which is the group that we are splitting up by region—over a two-month period, a three-month period, so large as to make this assumption problematic? It is not exact, but I think that the period we are talking about is fairly small, and we are talking about the net differences, not the gross differences. It is true that

about a third of the people move every year, but most of those people move in the same neighborhood, the same city, in my experience. So I think our assumption is quite viable. Maybe we can get our people in the Population Division to try to approximate what they think the net migration between post-strata might be over a two-month period, based on whatever information they have.

We have not done that yet. We can ask them to do it.

**DR. FIENBERG**: I have a comment on post-stratification and a question. The comment relates to, actually, Don's comments on the post-stratification structure. I think he was suggesting that it was not a good thing for the Bureau to bring to bear its subjective knowledge in the construction of the post-strata. It struck me that that was not a comment that either I or, perhaps, others would agree with. It may be subjective judgment, but it is also rooted in lots of background and other information. If it is the information we have, then you should bring it to bear. So I do support that.

I think it is also important to remember that race is—at least in part, if not largely—what adjustment is about. That is, we are discussing adjustment now for two reasons. One is to get better counts. But the second is to correct what we now know and have measured over a number of years to be a disparity in the net undercount, at a variety of levels, which manifests itself by racial categories, even if it is not something that we attribute directly to race. So thinking of putting racial variables in somehow, in the stratification, is, to my mind, inevitable, and it would probably be a mistake not to have something there, either directly or as a proxy.

The other thing is that race labels track to 1990. In 1990, it may have been the case that people could only say one thing, but lots of people said that they were multiple things, and then they got recoded. So you could think of this as simply a recoding of the variables that is very much like what the coding rules were in 1990.

But the labels are important. I think Don's point about that is absolutely crucial. I think it would be a colossal error to use the racial labels per se. I think Howard's response sort of indicated that that is not what is meant. But I think that for public use of the stratification scheme, some other set of labels is critical.

Now the question, which goes back to where we were before the break, actually, but propagates through: As the A.C.E. will be conducted, it starts as early as a telephone call and extends through July. We are asking a set of questions that relate to projection back to April 1. If you actually took a given household and moved them through that period and could ask the same questions at different points in time, you would get, conceivably, different answers. Do we know anything about the added variability associated with the estimate that comes from spreading the interviews out across time in this way, as opposed to the very concentrated way that the Bureau typically does a survey? For example, in two weeks' time, it will go out and, in one week, do measurement for the previous week for the CPS. This is a different kind of thing. If we do not know about that, is there a way we can get that information, or something about that information, so we can learn about that component of variability? It just strikes me that it is time we knew about it.

**DR. HOGAN**: We do not know very much. We know a little bit. Certainly, in some of the characteristics, when we match people between the census and the A.C.E., we know to the extent, say, their race changes. We have that.

In dress rehearsal, we had a small study that attempted to take cases that would have gone into the early telephoning and hold them out and put them in the personal interviewing, to try to compare the two. It was a very small study. It had some difficulties, given the timing. It found very few differences. But I would not hang a big hat on that hook. But we did try.

Short of that kind of controlled study, the people we get in May are going to be different from the people we get in the nonresponse follow-up later, so the only way to explicitly address your concern would be a study within a study. I do not think we are planning one. It is something we need to, I guess, think about, and maybe think about how, through our data, we can address that.

But, no, we have no data to answer that.

**DR. NORWOOD**: Don't you have data on other surveys that look at that particular kind of issue?

**DR. HOGAN**: There are lots of data on recall. That is a big issue that we have spent a lot of time on, and why my earlier conversation about what the reference point for the A.C.E. really is, whether the reference point is April 1 versus July 1— there is a lot of research that says it would be better to be closer to the reference date. I envy my English colleagues, who conduct the census in two weeks and start the PES not later than four weeks after Census Day.

We know that is an issue. We try to address that through the way we interview, the probes and the CAPI instrument, and the follow-up.

But I interpreted the question slightly differently, not just how far we are from Census Day, but given a mean or modal distance from Census Day, to the extent you could concentrate all the interviews on that day and gather all the information—I think that would be better. We have no data on that, but it would, intuitively, make a lot of sense.

**DR. NORWOOD**: But the further you get from the day, the more recall error there is. I remember that from the Consumer Expenditure Survey. It cost us millions of dollars to—

**DR. FIENBERG**: But the question is, what does the curve look like as you move away, and how can you separate out that decay from the difference in the people? That is, people whom you do not get to until July may be intrinsically different than the people you get in April.

**DR. HOGAN**: Within post-strata.

**DR. FIENBERG**: Within post-strata. So it is very tricky to measure, and you cannot simply bring to bear what we knew from the surveys in the '60s.

**DR. NORWOOD**: I see. I am just glad to know we did not waste money.

**DR. MULRY**: This relates to the telephone calls versus the personal interviews. I am not sure how it is going to work. Is the same interviewer going to do the phone calls and the personal interviews? I am wondering about sorting out mail mis-deliveries. This does have some post-stratification implications. We have had problems with mail mis-deliveries in the past. It seems as if that would

be hard to straighten out over the phone. But if the same interviewer was doing the phone calls and doing the on-the-ground interviews, that might have a better chance of coming together when it occurred.

**DR. HOGAN**: The universe eligible to be telephoned is a fairly narrow universe. It is single-family housing units, not multi, with good house number/street names, that mailed back their census questionnaire and included a telephone number when they mailed it back. So it is not a big universe. In dress rehearsal, we did about 5 percent of all the interviews, roughly, on the telephone. We kind of planned in 2000 for not more than about 10 or 15 percent.

We are very cognizant of this mis-delivery problem, and knowing that on the telephone you are never going to sort that out. So we restricted it to a universe where we think that problem is going to be minimized.

Essentially, it is the same workforce doing the telephone and the personal visit. It is probably more likely to be the crew leaders and people who later graduate to more leadership positions. But it is essentially the same workforce. They have the same instrument. They are in their homes doing the telephoning. However, what we do is, we set up the telephone universe, they do the telephoning, and then we bring the workload back, and then we redistribute it—including everybody—to personal visits. So it is not the same interviewer who goes out and he or she cannot get an interview.

**DR. BROWN**: Howard, doesn't that telephone protocol violate the independence assumption between the E- and P-samples, to the extent of the 5 or 10 percent of people who are contacted?

**DR. HOGAN**: I do not believe so. First, they have already mailed back their questionnaires. We have them in hand. They are in our processing office. Whatever goes on on the phone should not change the questionnaire. Second, we do not have someone in the field who might bump into the census nonresponse interviewer. So at least in terms of the big things going on, it should not—

**DR. BROWN**: It is neither of those. It increases the chance that those people will be included in the P-sample.

**DR. HOGAN**: No. It is a good question. Let me go back to how we get the universe we will be telephoning. We have, based on the independent A.C.E. listing that we did this past fall, a list of housing units that are in the A.C.E. That list is sort of closed. However, we match that list to the census list, and if a unit is on both, then we take the telephone from the census unit. But we do not add census units to our list. The list of A.C.E. housing units that are in the sample to be interviewed is never augmented based on the matching to the census. So I do not believe, in any fundamental way, that independence is compromised.

You look as if I did not answer your question.

**DR. BROWN**: I still—among other things, those people are contacted earlier, so they are less likely to move. They are treated differently from the rest of the P-sample in ways that increase their capture probability, increase the accuracy of the data they give, so it makes the matching more likely.

**DR. FIENBERG**: I think Larry's point is an important one. The distinction here is between the operational independence and the actual dependence of probability. Howard is arguing that there is operational independence, and I think that is correct. You are arguing that that can induce a dependence of the model, and I think you are correct. They are not inconsistent. That would argue, therefore, for a separate analysis that at least took a look at that to see if there was something to worry about.

**DR. HOGAN**: Certainly your issue about how it could affect the number of movers is something that we do need to think about. That is an important issue. I had not put those two together in my mind at once.

**DR. NORWOOD**: I would like to go back to Don's point about race categories. The way we generally look at the whole question of the undercount is not whether there is an undercount that is evenly distributed, but rather that it is disproportionately demographic. Are you saying that you do not think we will be able to use these data to know that? Or am I misunderstanding?

**DR. YLVISAKER**: I am saying that if one were not trying to make a smaller number of boxes to reduce variance, we would not be in this position.

**DR. NORWOOD**: But if you just look at this at a national level, and not have any of the boxes that you are objecting to, can you use it for that?

**DR. YLVISAKER**: My problem is with methodology, and now you are saying, given that you are going to use this methodology, what would you do? If I were interested in correcting demographic imbalances that I understood, I have other ways of doing it, and I would not be in the business of constructing 448 post-strata, which had to invoke race in ways in order to come down to a small number.

**DR. NORWOOD**: I understand that. My question, though, is a little bit different one. I have been thinking about this a good bit. We use race categories in order to have some idea of whether there are groups that are undercounted more than others. In general, that is a racial category. I guess I am asking a more general question: if we forget about adjustment and we forget about the localization of this, but we want to look at the country as a whole, how do we use the categories of race to determine that? What I thought I heard you saying—and I am not arguing your basic points about the localization of it—do you think that those data can be used, at least at the national level, to compare where we are, in demographic analysis?

**DR. YLVISAKER**: The post-strata, according to the criterion up here, were to capture, at least as one reads the documents, people of similar character leading to similar undercount rates and similar responses to census and so on. We are agreeing amongst ourselves that this is heavily racially motivated. We are into the idea now that we have to understand race, and we have to make comparisons of the type that Howard is talking about—if someone says American Indian, that is pretty convincing to me that it is American Indian; if he says white and American Indian and he is not on a reservation, then it is likely it is somebody—we are making up stories. I say, fine, but I wish I was not in this business.

**DR. NORWOOD**: I am sure there are people who wish they were not in the census business. Thank you. Lynne?

**DR. BILLARD**: The way I understand your question, it is if, instead of there being 442—whatever the number is—strata, the country is divided into one stratum itself. Is that what you are saying?

**DR. NORWOOD**: Don't you have the same problem?  That is what I am saying.

**DR. YLVISAKER**: If we had only one post-stratum?

**DR. NORWOOD**: If we are just looking at the country as a whole. That is how, in many ways, in the past, we have looked at the whole question of differential undercount. We can grind this into the ground. I did not want to focus too much on it.  But I do think that there are a number of complications for the Census Bureau, in general, because of the changes in the race and ethnicity categories.

**DR. YLVISAKER**: I certainly agree with that.  Certainly, demographics does not attack this multiracial thing yet, as far as I understand it.  Our knowledge of the undercount is still based on the old days, as nearly as I understand it.  That is, we believe it is a 5 percent undercount for blacks, and things of that nature, for demographic reasons.  This is handled demographically.  What they will do now I do not really know.  I do not think the census should be leading the charge, so to speak.

**DR. NORWOOD**: Are there other comments from the floor?  How about the panel?

**DR. BRADBURN**: May we pursue for a moment this line of argument? What I think people are saying is that there is some kind of correlation between the time at which people will be interviewed in the PES and their probability of moving—and not just the temporal. The longer it is, the higher the probability that people will not remember it. I will not pass myself off as expert about memory curves, but I do know a fair amount about it. The thing is that they are different. Some are linear; some are curvilinear of various sorts. But you really have to establish, in the particular content area, the nature of both the slope and the shape of the distribution.

But apart from that, what I thought Larry was saying, and Steve a little bit, too, is that, regardless of whether you do the telephone or not, there is a kind of correlation between when you get to somebody—because of just the difficulty of locating him, finding him at home and getting him to respond, and so forth—such that the people you get to later are, (a) more likely to have moved, and (b) probably more likely to have response error problems, because you may have to rely on proxies—that whole sort of thing, because they will be the more difficult households. So there is a kind of underlying correlation, which is not operational, between the characteristics of the household and the thing you are worried about—the accuracy of their reports about the moves.

I do not know how you get—I mean, obviously, operational independence does not—unless you do—I mean, as Howard was saying, you could do a study.  Actually, that was one of the things I wanted to raise.  I thought last time there was some reference to some experiments that were being built into the census for learning some things that would be useful for 2010, and so on and so forth.  I have

forgotten what that was called. Anyway, would it be possible to build an experiment into the A.C.E. in which you took a subset of the households—since you already know and the A.C.E. sample is already drawn—take a subset of those and control the time in which they are sent out to be interviewed, so that you have, essentially, random replication? You control the time at which you start to give assignments for this controlled sample and say, this has to be done in this two-week period, this has to be done in this two-week period, and so forth.

It might be that that would have to be too big to make the estimates that you want. But, in principle, I do not think it is a terribly difficult thing to design. It may be difficult, operationally, to bring out. In principle, it seems to me, it is doable.

**DR. ZASLAVSKY**: I have been sitting here trying to figure out why this is an issue. I can think of two reasons why we might care about this, aside from the one that Norman brought up, which is that it might help in planning for the future.

One of them is that every time you look at things more closely, you find some heterogeneity that you had not accounted for. So we look a little more closely at early and late respondents. We have not actually analyzed this, but it is fairly easy to anticipate that they would have different coverage rates and might actually be substantially different.

So what do we do about that? I guess, if we are in the world of thinking that the post-stratification is supposed to divide the population into post-strata of people who are exactly alike with respect to coverage, then we get very upset about that. If we think that post-stratification is supposed to form groups that are more similar within the group than between groups, and within the constraints of various other criteria that have to be satisfied for some of the things to be practical to be implemented as post-stratification, then we just have to accept that as being part of what is left over after we have done what we can do by dividing the population into groups.

The other reason I think we might be concerned about this is if the differences between the early and the late people, in some way, are causing us to mess up the estimation. For example, if we just threw away the very late people when we knew that they were actually different, then when we see that the early people are better covered, we would know that we had biased estimates. We do not actually do that. We go out and do a bunch of follow-ups, and eventually we do imputation from a missing data model. If that missing data model takes into account the fact that the late people might be more like the ones you never actually got, and not so much like the early ones, who are different, then you have probably gone about as far as you can in taking care of the difference along that direction.

I guess where I come out on this is that this sounds interesting, but I have not yet heard a specific reason, a specific model, or a specific story for why this would really be a big problem for the estimation.

## MISSING DATA

**DR. NORWOOD**: Howard, how would you like to give us the remainder of this discussion, the next block, which is: dual-systems estimation, missing data, synthetic estimation?

**DR. HOGAN**: There are essentially three components of missing data that we have to deal with. I will go through them fairly rapidly. The first component is whole household non-interviews. This really has two subcomponents now, because we are using the PES-C methodology. We have the non-interview of the current residents and the non-interview of the household that was here on April 1. So, really, we have two interviews and two non-interviews for each of those two conceptually different households. Thus, we have two different missing data whole household adjustment models going on. Both of them are very simple weight distribution within-sample-cluster kinds of approaches, very similar to what we have used in the past, similar to what is used in other sample surveys.

One thing I do want to point out is, as you can guess, the non-interview rate of who was living here a couple months ago is higher than the non-interview rate of who is living here now. Comparing the levels of missing data, at least on that important variable, that we will have in 2000 to 1990, they will probably be higher, because, as I said, it is easier to get an interview for who is here now. So we suspect that the level of whole household missing data, at least with respect to who is living here on Census Day, might be higher in 2000 than it was in 1990. So this kind of missing data household non-interview adjustment probably is somewhat more important.

But the approach, as I said, is a simple weight distribution within cluster. I do not know if that is controversial or not.

The next step of missing data that we have to deal with is missing characteristics. These are the characteristics we need for post-stratification, to place people into the cells. They are race, broad age groups, owner/renter, sex. Here we use either a nearest-neighbor kind of hot deck, based on missing tenure—we will go to the last household that reported tenure and borrow their tenure status—or, for things like sex, if the spouse of the head of the household [has sex missing] and the head of the household is female, then we assume that the spouse's sex is male. For age, we have distributions based on relationship to head of household—pretty standard missing data, missing item, missing characteristic, nonresponse adjustment here. It allows us to place people into post-strata.

Finally, we have missing enumeration status. This has two components: missing on the P-sample side and missing on the E-sample side—that is, people who are in the survey with enough information to at least know the person was there, but where we could not determine whether the person was enumerated or not, and census enumerations, but we could not determine whether the person was correctly enumerated or not. Here we have a change from 1990. In 1990, we used a fairly complex mixed model that was developed by some staff members at that time, who have gone on. We are not going to try to reconstruct that. It was an interesting model, but we decided, given where we are in planning for Census 2000, to go to a simpler model, an estimation cell approach. In some of our literature,

it is described as a ratio estimate. Rod reminded us that this is really just an estimation cell. The cells take into account, essentially, at what stage in processing the particular case became missing data. For example, the cells we are doing—this is in one of the handouts, the Q-19, on page 9 [Cantwell, 2000]—matching cases that needed follow-up, that failed follow-up, possible matches, non-matches from partial household matching households, missing people (this is done on an individual basis, where nobody in the household matched, and it needed follow-up), non-matches from households that had conflicting information. Let me explain that to those who are not complete aficionados here.

We have the situation that is fairly important in how we design the PES, the A.C.E., where we have the same housing units in sample on the P-sample side [and the E-sample side]. As I have said earlier, one of the most important things is to make sure that we have consistent reporting. We may have errors, but if they are consistent we have a lot of balancing out. When we get a situation where one family on the P-sample side said, "I lived here," and on the E-sample side, another family was living there, then it is possible that two families were living in the same housing unit. That does happen. I am not saying it does not. But it is also possible that we are getting conflicting information on the two halves of our surveys. One of the most important things we do in follow-up is sort that out to make sure we have one consistent response for that address, that housing unit.

That is a very important class of cases that need follow-up. It is a class that will go into the missing data model, the estimation cell model, separately. Then there is another set of cases where we did not have sufficient information to even put them into the matching. Either the A.C.E. interviewer did not get enough information—did not get name, characteristics—or the census got name and characteristics—actually, the census side does not go into this. The A.C.E. did not get sufficient information to go into matching.

Each one of those cells will be weighted up, based on the resolved cases. We have not spelled out the extent [to which] these groups will be segmented based on race or any other variable besides these operational variables. That is something we do need to spec out and are in the process of spec'ing out, but it is not laid out yet. But it is a simple missing data kind of model.

I know various people have comments.

**DR. LITTLE**: On the unit nonresponse adjustment weight, it strikes me that you sort of established the fact that the time from the census to the time of the PES interview might have a considerable effect on nonresponse, and is potentially also related to the outcome, which would suggest that you might want to use that variable as a stratifier for computing a unit nonresponse adjustment, if you actually have it available. That might be something to think about as an alternative.

In general, I am not particularly happy about the move from a logistic regression model for dealing with imputation to an adjustment cell style of model. This will probably not be a surprise to Howard. Particularly when you have a lot of potential information for predicting values of items, I view an adjustment cell method as being kind of limited, because you are basically in the same problem as doing post-stratification adjustments, that you can only form so many boxes based on the covariate information you have.

Also, you can view the adjustment cell method as basically a special case of a logistic regression, where you are including all the interactions between the variables that you have available, whereas I think it would probably make more sense in a lot of settings to have a logistic model that has main effects of these variables and allows you to choose more main effects in setting some of the 10-way interactions equal to zero.

I understand why the Bureau has done that. This is part of a general move towards simplicity and things that are easy to understand. But, as a statistician, I think there is a danger in trying to make things as simple as you possibly can, and going, as I view it as a methodologist, in the wrong direction, because hopefully the statistics profession is developing better methods rather than worse methods. So I rather regret this move backwards in terms of that general methodology.

The final comment I might make is that I would like to see a detailed explanation of what the imputation model actually is. All I have seen so far are sort of generic descriptions—"I am going to use an adjustment cell method" But a lot of the details of actually what goes into these adjustment cells really matter, and I would like to see that.

**DR. FREEDMAN**: David Freedman, Department of Statistics, UC-Berkeley.

It gives me great pleasure—a rare moment here—I am going to stick up for the Bureau. I am very happy they gave up that logistic regression model and have moved to what they are doing now. The issue, Rod, in my opinion—and I think many others would be found to agree with me—is the bias/variance tradeoff. When you do the modeling, you are trading off bias against variance and forgetting about the bias, which I personally consider unsporting.

But I do have an information question for Howard, if I may. Where do you get the resolve [resolved status] from for those people who are so weak that they do not go to follow-up?

**DR. HOGAN**: The way we think about that is that those cases are pulled out right at the beginning of processing, before matching, and therefore they are imputed from any case that either is resolved before follow-up or resolved after follow-up. The way I think of it is as sort of a branching process, and you go down the trunk to the point where this case became unresolved, and anything that happened after that goes into a cell to impute that. These cases are essentially very similar to the whole-household non-adjustment. They are imputed based on everything that is resolved after the point where they are screened out, and they are screened out very early in the process.

May I say one thing to Rod? Simplicity is important, but it is not just simplicity in either discussion. One of the things we are giving a lot of attention to, and rightfully so, is verifying our programming and testing our programming. We think that going to the cell model, while it has some properties that Rod does not like—I love having something in agreement with David. I think we should end the meeting right here and call it a victory.

But the simpler model lets us program it more easily and verify our programs, and lets other people verify them as well. So there are other advantages.

**DR. NORWOOD**: Steve Fienberg.

**DR. FIENBERG**: Department of Statistics, Carnegie Mellon.

People will not be surprised that I agree rather strongly with Rod on this. I just want to remind David, when you talk about bias-variance tradeoffs, if you use the cell model and you do not put all the variables in, then you have bad bias there, too. You do not have any sense of what the variance is usually, anyway, except for the wrong estimate that comes out of the wrong survey structure. So doing the modeling seems to me to make lots of sense. There is actually something somewhere in-between putting all the interactions in for a subset of variables and putting in only main effects for a larger number. That is why we have all those models that some of us have worked on for the last 30 years.

**DR. MULRY**: I was sort of surprised that you were not going to use the modeling anymore for the enumeration status estimation, because in 1990 the evaluations that were done showed that it was surprisingly effective. I was impressed at how effective it was, with what we could determine.

My question about this is—what I have looked at, I have not seen. It is simpler. From the dress rehearsal, there are these phrases like "had limited effect," to do the modeling versus not doing the modeling. I saw no evaluation of what would have happened with the 1990 data if there had been cells instead of the modeling. Also in the evaluation that I looked at with the dress rehearsal, it was a 1 percent difference, at most, on each post-stratum. There was no discussion that I saw of aggregate effects—for example, what would happen in the whole site or with demographic groups.

I would like to know what those were—a larger discussion, what happened in 1990, what would happen in the aggregate each way. What are you really losing versus what you are gaining?

**DR. HOGAN**: I do not have that discussion here. The one thing that has been asked for, and I think we do owe the audience (but not today), is a very detailed discussion, a very detailed outline, of exactly what we planned to do, which we do not have. When we have that, we certainly will provide it.

As part of that process, to the extent we can reconstruct 1990, we can try to run that. Part of our plan is to go back and try this on 1990. We are working on it now. We just do not have that.

**DR. ZASLAVSKY**: I find myself in the really surprising position of not having to say anything about the model, in defending it. I want to go back to the non-interviews. I think that also deserves some scrutiny. Again, the weighting adjustment may have this somewhat deceptive simplicity to it, but there may actually be a fair amount we know about a case being a non-interview, which can enable us to do something more sensible than saying that it is just like the average of every other case that you have in the post-stratum. For example, if there is a lot of geographical variation, a non-interview located in a block or in a small area that has, let us say, an unusually high or an unusually low coverage rate may also be more likely to have a corresponding rate. So you may want to do—staying within the range of simple methods that you can explain a nearest-neighbor imputation rather than simply spread the weight over everything else in the post-stratum.

Another possibility is just the continuation of the comment I made in the last discussion about the timing of the interview. If the non-interview looks like late interviews— for example, it is a non-interview because the form came in very late or did not come in at all, and it was held up for a long time—if it looks more like those, that is a place where a difference between the early and the late, if there is such a systematic difference, on the enumeration or coverage variables you are looking at could cause a bias in the results, if the non-interviews are more like the late ones than they are like the early ones.

**DR. HOGAN**: Let me correct something if I misspoke. The non-interview adjustment spreading of weights is done within cluster, not post-stratum. We very explicitly take advantage of the local geography. If I said it the other way, I apologize. It is within the cluster.

**DR. STARK**: I know this is not a straw poll, but I am also relieved that we are away from the logistic regression model. I wanted to take the opportunity, having spoken about computer programming, to be responsive to one of the things that Ken Prewitt said this morning about the computer bug from 1990. I, at least, and I believe some of my colleagues on the con side of the adjustment issue, do not feel that that is any kind of indictment of the competence of the Bureau, but merely illustrates how incredibly complex the whole procedure is. Bugs do occur in all kinds of contexts. The whole adjustment procedure is extremely complicated. Rockets crash. Space telescopes do not get ground correctly. There are bugs all over. This is not anything against Bureau personnel.

**DR. NORWOOD**: Howard, I have not heard anything yet about correlation and heterogeneity.

**DR. HOGAN**: I think there are two aspects to correlation and heterogeneity, one of which is within the dual-systems model and one of which is in the carrying-down. To a certain extent, a lot of our discussion on the movers is addressed to minimizing the correlation among one subset.

[Transparency]

In general, the model is an approximation. The post-strata are somewhat heterogeneous. Even to the extent that we work very hard at maintaining operational independence, there are indeed groups of people out there where the regular dual-systems assumption of independence and homogeneity fails. I have labeled them $N^*$. We do not observe them directly. We do not even observe them indirectly. We have some studies from ethnographic whatever that give us maybe some indications of where these kinds of people might be. But certainly from the survey itself, we do not observe them.

A couple of things. They are there. I do not pretend they are not. To the extent that A.C.E. and the PES do not correct for all the undercoverage—we only measure some of what is out there; we only measure $N_{12}$ and $N^*$, we are moving closer, and not all the way. Certainly, numerically, I believe that even with the $N^*$, we are doing better.

In any case, $N^*$ is neither, by construction, by its very nature, in the census nor in the A.C.E. So whatever problems $N^*$ is causing, if all these people are living in rural Minnesota, then they are being missed equally by both systems.

Another way of looking at it—and that is certainly, numerically, the easiest way to look at it—in terms of the distribution, there are a couple of ways of thinking about this problem. One can be defeatist about this and say there are some people we will never observe, and therefore we will never be able to prove that we are making things distributionally better. After all, if there is a group out there, $N^*$, that you can never observe, then you can never prove distributional improvement. You can always argue that these people could be somewhere else.

On the other hand, we believe that these residual unreached, unmatchable people are probably residing in areas where the $N_{12}$ and the $N^*$ are living. In other words, the unreachable people probably are living in areas that have high undercounts. They are probably living in places where we measured high undercounts in 1990. The highest undercount we measured in 1990 was among rural Hispanics. I would think that many of these unreachable people are living among rural Hispanics. I cannot prove that, but I believe that that is not an unlikely hypothesis.

So we think that if these unreachable people are distributed, at least in the kinds of areas where the reached people also live, we are able to make an improvement, both numerically and distributionally. We worked very hard in terms of improving the coverage of PES, but there are people out there who do not like the government, do not like surveys. We do have a rule that the Field Division insists on. If people have guns and they say, "We'll shoot anybody from the government that shows up at our cabin in rural Idaho," we take a non-interview. That is one of those things.

**PARTICIPANT**: You could send the director.

## GENERAL DISCUSSION

**DR. NORWOOD**: Comments, questions?

**DR. FIENBERG**: We have now accumulated a large number of issues. If you went around the room, maybe my list is different from David's. It usually has been. But it will overlap. We actually share some topics where we think there are issues, and that you could actually collect data on, to learn about them in the process. They go to the heart of either assumptions or the quality of the data that you are gathering, or both.

We heard this morning at the outset that we are not going to do the set of evaluations, the so-called P-study evaluations and related sequelae evaluations, that were done in 1990. The question is, what will you do to know whether you have done a quality job? Which evaluations are going to be done? How will we, in the end, be in a position to know if the decision to (a) release the data and then, later, perhaps to say that the data were good (or bad)—how will we be able to support that? What studies will be done, and how will they be done? Will they be articulated in advance?

**DR. HOGAN**: I think there are two aspects of your question. What I said was that the P studies that we did in 1990 in order to help the secretary make his decision, done in advance of the decision, will not be done until after the data

are released. We do have a program coming up with a set of evaluations of the whole census, including the A.C.E. These A.C.E. evaluations are part of a larger program. I think at some future meeting we are going to roll that out with this group. I think that is true.

**DR. NORWOOD**: We have asked for it.

**DR. HOGAN**: Good. Just like for any other census process, census operation, after the census is over, we plan to evaluate it and plan to share our results. I believe the current thing includes a matching error study, evaluation follow-up, maybe not all 27 studies we had in 1990, but the important ones. We are also working to develop some error models. We are working with Bruce and other people on some error models that we will bring to this committee at a future time.

So I do not want to say that we are not going to do that. We are going to do it, but just like we do every other census evaluation. We run the census the best way we can. We look at it step by step and say, is this operation doing what we expected this operation to do? Is this operation in control? If, when it comes time to release the A.C.E., we believe the A.C.E. did what it was meant to do, it was in control of the variances or in control of whatever, we will release the numbers. Then we will evaluate it, as any other census operation, and share our results widely.

**DR. NORWOOD**: But you are really looking toward 2010 for that. Most of its use is for 2010.

**DR. HOGAN**: The evaluation program is an evaluation program; it is not a program designed to support a decision process. It is very different from how it was conceived and carried out in 1990.

**DR. YLVISAKER**: If in 1991 the secretary had made the opposite decision and then you had done the evaluation studies, as in fact took place, how would you have couched your openness to the public at that juncture? That is, you have a computer error, you change post-strata, and you come up with a conclusion that these things are probably better now that we have revised it, at perhaps the state level, perhaps the national level, and so on.

**DR. HOGAN**: I think in two ways. First, with more time, one can do a better job. This is something that is true for all surveys and censuses, and we freely admit that. I would assume that, had we come up with a better estimate, with more time and more effort, we would have been forthcoming on that. The re-post-stratification was post-stratification designed to support a different purpose. After the post-stratification for the political, regular census, we were then asked by the Secretary of Commerce to come up with new estimates for a new purpose. We did. Part of the new estimates for a new purpose was a new post-stratification scheme. I am not sure we would have pursued that, necessarily. We might have. The purpose that was done was one the secretary laid on us when he chose not to adjust the regular census.

**DR. YLVISAKER**: If you were responding in 1992 or so, talking about what happened in the 1990 census, would you have considered that a success?

**DR. HOGAN**: I think, even with its flaws, even with its errors, the proposed census adjustment, the original census adjustment, was more accurate numerically and distributionally than the unadjusted census. It was not as good as what

we did when we had more time and corrected some errors, but I think it was numerically more accurate, and I believe distributionally more accurate, for the use of the census. Obviously, the secretary who was in charge at the time made a different decision than I would have. I voted to adjust. In retrospect, I still think that would have been the most accurate.

**DR. YLVISAKER**: I guess I did read in the CAPE committee report a sentence to the effect that it may be the case that this methodology will never be effective. There is a point that is made in the CAPE report to that effect.

**DR. HOGAN**: If that is in the report, it is a sentiment that I did not share at the time and do not share now.

**MR. THOMPSON**: This is John Thompson, one of the people who helped write the CAPE report. A lot has been said about the CAPE report; a lot of quotes have come out of the CAPE report, and various points. Let me try to summarize what I believe the CAPE report said. The CAPE report was an analysis of the data for the purpose of adjusting postcensal estimates. As part of the CAPE report, we found and corrected some errors in the PES, and we did a significant amount of analysis on shares, proportionate shares, for states, for large cities, for counties, and for small towns. What the CAPE report found and stated was that we believed that, for states and large cities, the proportionate-shares distribution was improved by the adjustment. We also found that for areas below that we could find no difference between the adjusted and the unadjusted.

We at that time felt bound by the guidelines imposed by the Secretary of Commerce, that, in order to adopt something, it had to be shown to be better at all levels. For example, showing that something was better at some levels and no different at other levels, which one might consider to be a basis for improvement, was not in the cards. So at that point, we said we could not show that, and therefore the unadjusted census should stay in place, if one restricted oneself to guidelines that said you had to demonstrate something is better at every level for which it is used.

That is what the CAPE report, in a nutshell, said, just to put a little clarity on that.

**DR. NORWOOD**: Other comments?

**DR. STARK**: May I just quote something from the CAPE report regarding heterogeneity?

"A first analysis showed similar homogeneity for the 1,392 design, as well as the 357 design, as well as for a design with only two strata." I do not know; I do not see this, somehow, as being in favor of the methodology.

"The level of bias in the PES was close to the point where artificial population analysis shows that the homogeneity assumption fails to hold."

"The panel cautioned that artificial population analysis was inconclusive about whether the homogeneity assumption held."

"The fourth cell in the DSE is an estimate of the number of people missed in both the PES and census. Both the committee and the panel of experts were very concerned about the negative values in the fourth cell. Correlation bias should be a component of total error. However, there is concern about our method of estimating it and very serious concern about the method of allocating it."

**MR. THOMPSON**: Briefly, the report had a number of discussions of the analysis, some of which were extremely difficult and extremely technical. There were reservations about various portions of the analysis. The heterogeneity, I think, is explained in much greater detail in an addendum that was written. It is also described in a paper Bob Fay and I wrote, which showed some pros and some minuses for it. So those data are out there and available.

I could find individual statements in virtually anyone's report that, if I were just to quote those, would say things. What I am trying to do is give you the broad essence of what the report said. If we want to really go into it in detail, it would take an all-day session.

**DR. NORWOOD**: We do not.

We are not here to reinvent 1990, and I hope the Census Bureau does not want to reinvent 1990. I think it is very important for us to understand what differences there are in the Census Bureau plan, not necessarily to agree with them, but to understand what is different now from 1990. But that is really the only purpose for bringing 1990 into the discussion.

**DR. FIENBERG**: I want to follow up on Howard's response on the two issues of heterogeneity and correlation bias. It strikes me that they have been among the most contentious. Is there a document planned that swept back through all of the reports you have now constructed on A.C.E. and explained how the different features of A.C.E., as opposed to PES, were going to deal or not deal with those particular issues? At several points during the day, you said, "Oh, this change in matching will have a particular kind of impact on the estimates, and therefore on things like heterogeneity and correlation bias." The exchange that Larry and I had about that change may yet be another, which goes in the other direction.

It would be really useful, it strikes me, to, if nothing else, have a catalogue of those, and perhaps a summing up of what the expectation is. If we are not going to evaluate until 2002, then perhaps we should know what you expect to see, as opposed to what you did see in 1990.

**DR. HOGAN**: In the document we distributed, the one I wrote, I do not quantify, but I do try to lay out, at least on a theoretical basis, changes and what [effect] those changes might have. I do think we currently plan to try to quantify what 1990 might have looked like. It is a very difficult thing. We did not do a PES-C in 1990, so we do not have that. Once you have done that, of course, Census 2000 might be quite different. Given our other responsibilities and our staff workload— it will go on the list, but it will not be item number two or three.

**PARTICIPANT**: This question was asked before, I think. Maybe it was answered, and I missed it. What is the deadline for announcing the actual algorithm or procedures for calculating the adjustments, the A.C.E. adjustments? When will the final announcement be made on post-strata?

**DR. HOGAN**: On post-strata, it has to get through all of our decision and sign-off process, but I think it is going to be in the next couple or three weeks— very soon. With things like missing data, where we still have to completely specify and test our model, it is going to be a little bit further. It is like any other census process. We have nailed down most of what we are doing, but there are some

details that need to be specified, specifications written and circulated. Each of those will be well in advance of the operation. But I am not acquainted with a deadline by which all of our specifications have to be done. If there is one, my boss has not communicated it to me.

**MR. WAITE**: It has to be done prior to execution.

**DR. HOGAN**: Right, and prior to programming.

**MR. WAITE**: And if they are not, there will be an execution.

**PARTICIPANT**: Is it prior to April 1, or is it when you start the telephone calls for the A.C.E.?

**DR. NORWOOD**: May I just say, you have had the question, and you can answer it later. It is my understanding—let me rephrase that. From the perspective of a statistical agency, you have to define what your procedures are going to be before you use them. That does not mean that you have to define them six months ahead or a year ahead, but it does mean you have to define them and have them very specific. Otherwise, people will think you keep changing them. You cannot be in that situation.

So I am sure that that will come out, but I think it is not quite our purpose today to pin the Census Bureau down on a particular date.

Ken, you have heard this. You and Bill know, certainly, that you need to do something about that. We will just see what happens. But you cannot issue a major set of numbers without having stipulated in advance what the procedures are going to be. I think everybody understands that.

I am surprised, frankly, that this last discussion has elicited so little comment and discussion. This panel has talked about this quite a lot. Did you have something?

**DR. BROWN**: I was going to raise something, but I think it is a question that I am not going to get answered now, because the answer does not exist. I really wanted to hear a lot more about characteristic imputation, what the algorithms are. I have a kind of suspicion that there is—I guess it is just not clear to me what is going on in some of those imputations.

**DR. HOGAN**: I think on characteristics—we do not have it here today—we can give you some pretty good detail, because that has been fairly stable in terms of what we plan.

Pat [Cantwell], on imputation of characteristics, do we have something?

**DR. BROWN**: Yes, it is open in front of me [Cantwell, 2000]. There are some questions I would have about that. That is the sort of detailed description, then, of—

**DR. HOGAN**: We can make Pat available or other background and more detailed documents available. We certainly have exactly what we did in dress rehearsal, which I do not think we are changing very much. In terms of characteristics, I think we can do a pretty good job of answering your question.

**DR. BROWN**: Does this memo refer to characteristic imputation in both the E-Sample and the P-sample or just the P-sample?

**DR. HOGAN**: For the E-sample, we are using the census editing.

**DR. BROWN**: So that is in some other report.

**DR. HOGAN**: That is in a different specification. The missing data in the census, whether it is in the A.C.E. block or not, go through the regular census edit and imputation process, which is documented and we can share with you. We use whatever the census imputes. If someone does not report his race to the census, and the census imputes the person as being whatever, we accept that for A.C.E. purposes.

**DR. BROWN**: If you otherwise match that person—

**DR. HOGAN**: Yes. There is, I think, a misconception out there that maybe you do not share, but other people do—for example, that race is a very important matching variable, or whatever. In fact, most of our matching, since we are localizing it to just the block or the search area, has to do with address, has to do with name, family relationships, relative ages within the family. Something like race gets almost nothing in the matching. It is because most blocks in this country, for better or for worse, tend to be predominantly Hispanic, predominantly white, predominantly black. Once you are in a particular block, the fact that the person is missing race affects the matching very, very little. But it does affect which post-stratum cell the person goes into for the estimation. For that decision, we follow whatever the census imputes, on the census side.

**DR. BROWN**: But age is a matching variable.

**DR. HOGAN**: Age is a matching variable, yes. But again, it is done in the context of the whole household. If we have husband, wife, and a person with a name and without an age, but it is their son, it is used, but it is not positive. But it is used. Missing age does make matching harder, which is why on the A.C.E., both on the census side and on the PES side, we do have these rules that the response has to be relatively complete. That is why, on the census side, first, what the census says is data-defined, which is fairly minimal—three letters of the name and one characteristic, I believe—but for them to go into the A.C.E. matching, they have to have a name and two characteristics. So we have to have reasonably complete data on the census side. The same is applied to the PES side.

Some of these cases where a lot is going to be imputed are subtracted out as—I use the phrase loosely—erroneous enumerations, or certainly not correct enumerations.

**DR. BROWN**: That discussion helps me a little bit.

**DR. HOGAN**: I would be more than happy to continue this, but perhaps not this afternoon.

**DR. STARK**: To what extent is matching going to be done on census data that have been obtained by OCR [optical character recognition] from census forms?

**DR. HOGAN**: That is a good question. As people know, the census now is going to be essentially image-filmed, and then the information is going to be read off of it using optical character recognition. After various edits and whatever else, the census, as data captured by OCR, is what goes into our computer matching process. So the initial computer match on the A.C.E. will be against OCR data.

However, once it goes to the clerical matching, the clerks will have available to them the image. According to our rules, it is the image that matters. If the census garbles the name, they have the image they can pick up, and they can read the

name and match against the name as the person wrote it down on the questionnaire. So the matching is against, in our minds, the image of the questionnaire, rather than the OCR, although we do have this intermediate step of the computer matching, which is against the OCR ASCII file.

Does that answer your question?

**DR. STARK**: Kind of. Two more questions along that line. Will "Be Counted" forms also be OCR data captured? I did not quite understand how the ASCII is used as a first pass. You turn up possible matches that are then refined by looking at the image, or you rule some things out?

**DR. HOGAN**: The way our matching process goes, in general, is, first, we have the census records and we have the PES/A.C.E. records, and we do a computer match. Coming out of that, there are four categories of records: matched records—and we have set it up to say, if the computer matches it, it is very likely that we would think it is a match; possible, probable matches, where the computer says, these look similar, but we are not quite sure; and then leftover census people, leftover A.C.E. people.

The clerks are required to look at all leftover people, as well as all possible matches. They are not required to look at matches, which is why it is fairly definitive. We believe that coming out of the ASCII, OCR capture problems, where the OCR garbled the name or whatever, will fall out either as an unlinked census record, which will go to the clerks, who will have access to the images, or as, at worse, a possible match, where the clerks will have access to the images. We do not think these will result—we have no reason, based on the dress rehearsal, to think they would result in many false matches coming out of the computer processing. So we are pretty comfortable on that.

In terms of our matching, we will be matching against the "Be Counted" universe, as well as the Internet universe. We will have a questionnaire on the Internet.

**MR. WAITE**: The "Be Counted" forms do not have the bar code ID on them which would allow the data capture system to check them in and read them and categorize them. When they first come in, they do not have any geocoding information on them. They have to be sent to a place where we are able to identify exactly what block this address is that came in. Basically, the data file is created by keying from the "Be Counted" form. The "Be Counted" form still exists, but there is never an image out of the data capture center of the "Be Counted" form.

**DR. STARK**: One more question. Where does the computer unduplication fit into the matching process? Are you matching things that have already been unduplicated, or are you matching against the records before having done that, in case there are some differences?

**DR. HOGAN**: There are two kinds of duplicates going on here. First, there is the census process, if you have two returns for the same ID number, of selecting the ID that is going to be used, as well as the merging of the "Be Counted" and whatever else. The first one we call the primary selection algorithm, or PSA. The second is just a merging of the "Be Counted." On those, whatever the census does—right, wrong, or indifferent—whatever the census says—here are the records

I am counting for this housing unit, all the right ones or all the wrong ones—that is the universe that we sample from and match to, in terms of the PES.

In terms of measuring duplicates, I think my paper incorrectly states that we do that as part of the A.C.E., a computer-matching process. Actually, that is, I believe, entirely clerical. Whenever, as I said, coming out of the computer match, we have these leftover people, leftover census people, leftover A.C.E. people, one of the things the clerks do is say, is this person left over because the census duplicated it; is this left over because it is paired with a match case up here? In measuring the duplicates as part of the A.C.E., our clerks do that. But in terms of sort of creating the census file, the primary selection algorithm, whatever else—whatever the census does is our universe that we take.

**MR. WAITE**: And all that is done prior to the creation of the CUF [census unedited file]. So when the CUF is presented, the census at least believes that it has unduplicated. Basically, it is the last chance the census has to say, this is what we think our universe is. That happens in July, August.

**DR. EDDY**: I was in the Phoenix data capture center a few weeks ago, and I thought that I was told that 80 percent of the OCR forms are actually sent to keypunch for correction of one or more fields, and that no such thing as "?M?" is likely to slip through. I wonder if you could confirm that for me—that is, over 80 percent of the forms go to keypunch for visual confirmation by a human in keypunch.

The second thing I wanted to ask—I also recall that they said they threw away the forms when they were done, the digitized forms that go to OCR. I am wondering if you could explain the procedure by which they are selected out for the A.C.E.

**DR. HOGAN**: Let me answer the second and I will defer to Jay for the first.

Because we have identified our blocks, for the whole universe, the listing universe, we know which blocks and surrounding blocks are in our A.C.E. universe. We have asked them to save those for us. They actually copy it out on CD-ROM and give it to us. We have a contractor who is setting up our own data base to access the images for the A.C.E.

In terms of the percentage of questionnaires that go to key-from-image, Jay would know that off the top of his head.

**MR. WAITE**: I would know it off the top of my head if I just could find the top of my head. I do not know the exact—80 percent is extremely high.

**DR. EDDY**: That is what I thought.

**MR. WAITE**: I would be very surprised if it was that high. Whatever the percent is, I think it is important that you realize that if a questionnaire is decided to have some problem, and therefore is sent for purposes of key-from-image, they do not look at the entire form and go through all the things on the form and redo everything. All that the key-from-image person sees is the particular field that the computer had a question about or that the image had a question about. Just because a form went there does not mean that everything on that form has been rechecked by a human being and he has verified all of those answers.

We do have very high levels of quality on this data capture system. We have less than 2 percent error on any of the fields. The computer has to be 98 percent sure that it is correct before it accepts it. There is a very small proportion of things that are going to get through. But 80 percent going to keying, I think, is way too high. I would have to check where you got that information, and perhaps I get the information back to the committee—or fire somebody, I think, if 80 percent are going there.

**DR. YLVISAKER**: I was along with Bill on the same journey, and I think 80 percent is, in fact, low. I think it was 86 percent.

**DR. EDDY**: I said over 80 percent.

**DR. YLVISAKER**: Yes, I am agreeing with you. I think it was something on the order of 86 percent. But this I found encouraging, basically. They are being careful. I say fine. If they are not sure, then they are going to check. Fine.

**DR. EDDY**: I would just like to add the comment that I did watch the process, and I was pretty impressed. Yes, they were only doing specific fields that needed help, but the people who were doing it were doing a very, very good job. I was really impressed.

**DR. WOLTER**: I just have three short questions for Howard. First, why did we drop group quarters from the universe?

Secondly, you mentioned a technique used in 1990 called smoothing, which I believe you introduced to reduce sampling variability. What was the final report card on that for 1990? How did that bear on your choices for the year 2000?

Finally, one of the key evaluations that you will be doing as you go forward is to compare demographic analysis estimates versus PES estimates. In light of the introduction of multiple races, how are the demographic analysis estimates being assembled? Do those map on to your racial categories?

Those are my three questions.

**DR. HOGAN**: On GQ [group quarters], I believe the decision was made that when we went to PES-A or PES-C, the reconstruction of the Census Day residents— we felt that doing that in a college dorm or doing that in a fraternity house or wherever else is not going to work. So we decided to be conservative and not try to make it work. That was the real process that led to the exclusion of the GQ. It flowed from the decision of going to either PES-A or C.

I believe the smoothing model—I think there are two ways of looking at it. One, perhaps, is less controversial than the other. First, did it work statistically? That is the controversial one. Many people in this room believe it did not work statistically. I think it did. But what we felt was that it was becoming a sideshow. The attention was increasingly being focused on, could you pre-smooth the covariances, and should you pre-smooth the covariances, that go in the pre-smoothing of the covariances? We felt, whether it is the right thing to do or the wrong thing to do, let us refocus this on the undercount. So a lot of our decision, back in 1991-92, as well as now, is—not that we think there is anything wrong with smoothing, statistically, but it really was becoming a sideshow and distracting from what we think is the important thing.

In terms of demographic analysis, it really depends on how many people change. Remember, demographic analysis, traditionally—Kirk knows this; I will say "remember" for the rest of the group— really has only been able successfully in the past to do black [and] nonblack. It has never been able to do Asians or Hispanics, or whatever else. My feeling—and this is borne out from our dress rehearsal in South Carolina—is that the reporting of black, or of African American, is probably still pretty stable. In our society, that marker is still pretty well understood. So, although a lot of the multiple races we got in the dress rehearsal were white/Asian, Pacific Islander/white, whatever, reporting of African American is still pretty darned stable in this society and is likely to be in 2000.

So I would think demographic analysis will have the same limitations and the same advantages in 2000 as it has had in the past.

**DR. NORWOOD**: So does that suggest that the Hispanic is not very good in the demographic analysis?

**DR. HOGAN**: Demographic analysis, traditionally, has only published black [and] nonblack. In order to do any demographic analysis—

**DR. NORWOOD**: I know that, but in this whole process, don't you look at Hispanic? Won't you be forced to, given the tremendous increase in them?

**DR. HOGAN**: We will be forced to look at Hispanic in terms of the A.C.E., but I do not know—Greg is here, if you want to address how Hispanic might be addressed in DA.

**DR. ROBINSON**: First, I will agree with Howard. Demographic analysis relies on historical data, as consistent as possible. You go back to 1935, where we have births tabulated by race, carry that forward. We have historical data on blacks and whites and other. Those are the three categories. But when you compare census data for other races, Asians and American Indians, a classification error occurs, and we really cannot produce meaningful estimates. The estimates, as Howard knows, we produce from demographic analysis are for black and all other races combined. That is our plan for 2000.

What we have done on an experimental basis—one reason we do not do estimates for Hispanics is that we do not have the historical time series of births classified by Hispanic origin. We did an experimental report, by Ed Fernandez on my staff, looking at the estimates of the undercount of Hispanics under age 10 in the 1990 census. We can do some things for limited age groups, but not at the level of comprehensiveness of our total population. We can do some rough benchmark evaluation for some groups, but the main flagship estimates will be for black and nonblack, by sex and by age, like we did in 1990.

**DR. NORWOOD**: So I should take that into account when I look at the population projections for Hispanics?

**DR. ROBINSON**: The population projections are not—when I say demographic analysis, I am talking about adjusted for undercount. Population projections are a basic Bureau product, which are—

**DR. HOGAN**: We will have the basic projections from the 1990 census using components of change by Hispanic and these other things, so that in terms

of comparisons to the adjusted 2000 census (assuming we did something tremendously wrong or did something, prima facie, silly), we would have the projections from 1990 available to say we have estimated something very different from what demography expects to be there, even though they are not the proper demographic analysis estimates.

**DR. ROBINSON**: We take the 1990 census, we project it forward to 2000, we have an estimate of the number of Hispanics at the coverage of 1990, and there are big differences.

**DR. NORWOOD**: Okay, that is what I thought.

I think we are getting close to time for a break. But before that, let me say that we will provide tomorrow for each of the invited guests to have a few minutes to tell us anything they want to tell us. We have two people, however, who will not be able to be with us tomorrow. David Freedman, I think, has to leave a little after this break. Then Alan will be here after the break. So I would like to ask David to take a few minutes and tell us anything he wants to tell us now.

## COMMENTS OF DAVID FREEDMAN

**DR. FREEDMAN**: I want to thank the panel for the invitation to be here.

[Transparency]

Before leaving San Francisco, I saw a quote from Samuel Johnson that may be relevant to our considerations here. What he says is, "The use of traveling is to regulate imagination by reality, and instead of thinking how things might be, to see them as they are." The chair has been very strict that we should not be re-fighting the battles of 1990. If you will permit me, however, I will violate your instructions. But before you rule, let me try to tell you why I am going to do that.

To me, the census of 2000 seems to be similar in many respects to the census of 1990. The adjustment of 2000 seems to me to be similar in many respects to the adjustment that was proposed and not used in 1990. Therefore, in order to get a handle on the various magnitudes that are involved here, to see what things like correlation bias and measurement error might be, I think it would be advantageous to look at 1990 and see if we can learn from that as we plan for the future. So it is my thought that, by taking a brief look at 1990, we might learn something that would be relevant in thinking about 2000.

With that, if you will permit me to proceed.

**DR. NORWOOD**: I so permit.

**DR. FREEDMAN**: This is a topic on which I can argue with certain members of the audience, and they with me, for days, weeks, months. It has been in the journals, in the courts. I promise not to do that. Just one chart.

[Transparency]

On July 15, 1991, Secretary Mosbacher faced a decision whether or not to adjust. Some of you in this room may, down the road, towards the end of this year or perhaps the beginning of next year, face a similar question, and so it might be of interest to put yourself in Mosbacher's position. The adjustment that was proposed at the time would have added 5.3 million people, net, nationwide. There

are various measured errors in the adjustment process, due, for example, to the coding error that was mentioned. There are rematching studies that measure matching error. People go back out into the field and measure other kinds of errors. When Mosbacher was making his decision, the Bureau's estimate for the measured errors was something like 1.7 or 1.8 million people.

Later the coding error was discovered that was another bias on the order of slightly over 1 million people in the 5.3 million. A certain group of blocks was rematched. There was discovered another error of about 200,000 to 300,000 people, which brought the Bureau's estimate up for measured error from 1.7 million or 1.8 million to right around 3 million.

That, I think, is already interesting as you think about the decision to adjust.

My colleague Leo Breiman at Berkeley estimates measured error at about 4.2 million. I do give this talk once in a while. I did it one time in Dallas when Mary Mulry was in the audience. I am not going to hold you to it, Mary, but at the time, as I understood you, you were willing to go up to 3.6 million. Ken Wachter and I will certainly come down to 3.6 million. I cannot commit Leo, so if you want Leo's vote, you just have to get him to come.

I am going to use that intermediate number of 3.6 million. If you take that out of the 5.3 million, the corrected adjustment is 1.7 million people. So something like two-thirds of that original proposed adjustment is measurement error.

Of course, the story does not end there, because there is an error that works in the opposite direction. Generally, these measured errors inflate the adjustment, but there is a compensating kind of error, which has been mentioned several times today, correlation bias. Ken Wachter and I have taken to calling that the fifth cell. There are the four cells in the $2 \times 2$ table that Howard is presenting, and then the fifth cell is the people who are not on that chart.

You cannot estimate that from the census and the evaluation data. But the demographic analysis would have added 4.7 million people to the census of 1990. You can now estimate correlation bias as 3.0 million. That is what you would have to add to this corrected 1.7 million to get up to 4.7 million.

So there is a snapshot of the principal sorts of error in the production adjustment of 1991.

Some people have objected to this on the grounds that I am comparing net errors and you should not be comparing net errors. It is net because the 5.3 million—actually, we would have added slightly over 6 million person records. We would have subtracted slightly under 1 million. So that is a difference. I myself am not so resistant to comparing nets.

But I will tell you, at the state level anyway, this argument is somehow irrelevant. Every single state would have been adjusted upward in 1991. Rhode Island got the smallest adjustment, several thousand, but every single state would have gotten adjusted upward.

So the thing to think about is, we are putting down numbers of people through the census operation in the 50 states and in D.C.—248.7 million people. Adjustment would have added 5.3 million people, with positive additions in every single state. It is just that some states are adjusted up more and others less than average. I will come back to that in a second.

If two-thirds of the people you are adding reflect measurement error rather than good data, the addition of people to these various states seems to me to be a very insecure business. That is one point.

The other point is that even the 1.7 million people who are left on the board by this calculation—we actually really do not know what states they belong in. We know what post-strata they belong in, but they have been allocated to states not on the basis of data, but on the basis of the synthetic assumption. Therefore, even if we were to follow Ken Prewitt's idea and talk about numeric accuracy—we are trying to move the state populations in the right direction, and not go too far—it is not entirely obvious from these data that the adjustment of 1991 would have been successful at a numerical level.

When we look at shares, it is perhaps even more problematic. Let me try to make that a little more vivid. The state that would have been adjusted upward the most in terms of shares in 1990 was my home state of California, two-tenths of 1 percent. Texas was about a third of that. That was the next biggest, then Florida. The state that would have been adjusted downward the most is Pennsylvania. Everybody gets adjusted up, but they cannot all go up in share, because there is a constraint of 100 percent. So some shares go down. Pennsylvania, New York, Illinois, the sort of large industrial states of the Northeast and Midwest, would generally have been adjusted downward.

I ask you to imagine an alternative adjustment which simply takes the production adjustment of shares and reverses every single sign, so that California goes down by 2/10, Texas goes down by about 8/100, Pennsylvania goes up by 7/100 of 1 percent. How would we defend the production adjustment against the alternative adjustment? I think that is very hard to do when the adjustment is driven not by the data, not by the 5.3 million of solid additions, but by measurement error and by the homogeneity assumption.

That is, I think, the real problem with the production adjustment of 1991. I am afraid that that will be the problem with the adjustment of 2000.

The arguments I have presented here are due largely to Ken Wachter and Leo Breiman. I have been more interested in things like the smoothing model, loss function analysis, synthetic adjustment, missing data. The smoothing model, I am happy to say, is off the agenda. But I think the other problems will be there in the year 2000, will need to be dealt with. I think a good way to deal with them is by looking at the record of 1990. The people who favor adjustment, the people who oppose adjustment have argued it out in court, at professional meetings, in journals. There is a rich set of data and analyses. I would urge the committee to take advantage of this. I am afraid to say that I do not think the previous panels of the Academy have. I would hope this one does.

Ken Prewitt asked two questions at the beginning. I guess I want to try to answer them. Do we prefer numeric accuracy, or do we prefer distributive accuracy? I personally was in favor of distributive accuracy in 1990. I have not changed my opinion so much. Ken, I never thought I would get to say this, so thank you very much. I rather thought there had emerged something of a professional consensus in the argument that distributive accuracy was the way to go. So I am happy to be here representing the consensus and seeing you as the rebel.

**DR. PREWITT**: You might at some point tell us how to design the census cheaper.

**DR. FREEDMAN**: You cannot really be asking me that about Census 2000, since it is well under way, and I cannot imagine that you want any advice from anybody about how you want to do it. If you want to talk about 2010, I would be very happy to do that.

The second question that you asked me I think is more of a side point. Although the issues of 1990 I regard as salient, the question of how you judge the adjustment I regard as salient, the other question is more a matter of the historical record. Let me try to clarify.

You asked about moving congressional seats. Let me try to make a short claim. The 1991 adjustment, with the 5.3 million people, would have shifted two congressional seats. Sometime after that—and John says it was for other purposes—the Bureau proposed another adjustment. They took out the 1 million coding error and they took out the 200,000 to 300,000 false non-matches from the bad blocks and they re-post-stratified with the 357 post-strata. To me, the salient change is taking out the coding error; the next biggest thing is taking out the false non-matches from the bad blocks; and number three is the 357.

Those adjustments, the one from 1991 and the one from, I think, 1992, are public-domain numbers. You can, yourself, adjust the House of the Congress based on those two data sets. One of the few thrills in this business, I can tell you, was when I apportioned the House of the Congress, all by myself, on my computer at home, two different ways, using those two different data sets. In the first one, you move two seats, and in the second one, you move one seat.

I think I will stop there.

**DR. NORWOOD**: Thank you very much. We appreciate your having come and your contribution. As I said earlier this morning, I think it is very important for us to hear all points of view. I think we have heard a number of them. We will hear more later today and tomorrow.

Now I would like to take a 20-minute break.

## CONGRESSIONAL MONITORING BOARD REPORT TO CONGRESS

**DR. NORWOOD**: I have asked Charlie Jones, who I think needs no introduction to this group, to present the research that came out in a paper from the congressional side of the Monitoring Board [U.S. Census Monitoring Board, Congressional Members, 1999].

**MR. JONES**: Thank you, Janet.

Director Prewitt has stated that the requirements of the census are twofold. One is to enumerate the total population, and the second is to allocate them to the addresses where they lived on April 1. Similarly, we can look at the A.C.E. as addressing these two features.

[Transparency]

The re-interview of the housing unit and the people matching and dual-systems estimation yield an estimate of the total population.

[Transparency]

Synthetic estimation, on the other hand, attempts to allocate the missed persons derived from that estimate back to their April 1 addresses.

[Transparency]

The focus of the analysis we have undertaken is to see how well synthetic estimation works. So it is not the whole thing. It is a very narrow focus, but I think an important focus.

[Transparency]

After the 1990 census, we had two counts for each census block: the census as enumerated, and the census count as adjusted by synthetic estimation. To evaluate that adjustment, we needed a standard that we could compare these to. For this standard, we selected the 5,180 populated PES blocks in the 1990 census. We chose what we call the direct dual-systems estimate.

[Transparency]

That is, for each of these block clusters, we combined the number of persons found in both, the number of persons in the PES and not in the census, and the number in the census and not in the PES, and formed an estimate for that block of the total population. We call that a direct dual-systems estimate, as opposed to a dual-systems estimate based on synthetic estimation.

John Thompson and I had some discussions about this. He had recommended that we use Census Plus for a standard—that is, combine the three cells, but do not include an estimate for the fourth cell.[3] We actually ran the analysis both ways. No matter which estimator you choose, the analysis essentially presented the same results.

I am going to talk about one of the graphs in the report.

[Transparency]

This is on page 25 of the report. We first computed the census coverage rate for each block. That is, we took the census count and divided it by the direct dual-systems estimate. This allowed us to sort all these 5,180 blocks by the severity of the undercount or overcount in the original census. This is displayed along the *x*-axis. The worst undercount group we have there is those that have less than 50 percent coverage. It goes up on the other end of the scale to where we have over 120 percent coverage in the census for the block. Right in the middle, we chose a group that is plus or minus 2 percent, 98 percent to 102 percent, which we felt was good coverage. So we have overcoverage, undercoverage, and the good-coverage groups.

We next computed the average undercount or overcount in each of these groups before the adjustment is made. This is shown by the red bars on the graph. We next computed the average adjustment in these groups—that is, how much was added by synthetic estimation—and this is shown by the yellow bars.

[Transparency]

The amount of adjustment by synthetic estimation is almost invariant across the various undercount levels. Rather than the adjustment being proportional to

---

[3]Census Plus is a method for estimating population coverage, which uses field reconciliation of discrepancies between the PES household roster and the census household roster to estimate the "true" count for each household.

the size of the undercount, as we might have liked, the adjustment is about the same regardless of what the original error is.

The correlation between the amount of undercount and the amount of adjustment on this was about 0.09.

Without weights to apply to these sample data, which we do not have yet, we cannot carry out some of the important analyses, like looking at estimates at higher geographic levels. But the same data themselves provide a basis for some conclusions. For example, if these various undercounted blocks that we saw in this graph were distributed randomly, we might expect the errors to cancel out when we aggregate these to some higher geographic levels.

[Transparency]

In my experience, however, there are a number of neighborhoods where these high-undercount blocks will tend to cluster. That is, the blocks are not distributed at random, and the errors will not cancel out. Let me just name a few of these kinds of places: Robert Taylor Homes in Chicago, the Watts area in California, the Colonias in South Texas, Liberty City in Miami, Piedmont Courts in Charlotte, Bedford-Stuyvesant in New York, the Hough area in East Cleveland, and Anacostia right here in D.C. In these neighborhoods and others like them, with concentrations of high-undercount blocks, it is clear that the adjusted count as currently planned will give a low count.

Our concern in undertaking this analysis was local-area accuracy. It seemed to me that in these local neighborhoods, if you do not have a good census count before adjustment, you will not have a good count after adjustment. I know the assumption, as I read it, in the paper is that— there is an assumption of randomness underlying the distribution on which the synthetic adjustment is made. I suppose, if we did some research, we could actually find out at what geographic levels these do balance out when we add the blocks together. But at least for the neighborhoods that I am familiar with, where I think there is substantial undercount in most of the blocks in those neighborhoods, I do not think it will average out.

[Transparency]

I had a few recommendations I wanted to make for the committee's consideration. One of these was to ask the Census Bureau to release detailed 2000 A.C.E. data so that researchers outside the Census Bureau can carry out independent analysis. This seems to me to be consistent with Director Prewitt's desires to have an open and transparent process.

The second recommendation I wanted you to consider was to include an estimate of a nonsampling error introduced by synthetic estimation with other nonsampling error estimates in your loss function analysis. I know you do not plan to do a loss function analysis before you publish the data, but, as a part of your evaluation process, I think you could do that.

Third, now that you are no longer going to do all this evaluation before the census, I would suggest that you only publish the census estimates down to the levels for which the Bureau can demonstrate improvements in the data. The reason I say this is that I think these local areas—and there are a lot more than the eight that I mentioned—these people need good data as well as anyone else. The fact is, the census is in the business of producing local-area data.

That is all I have, Janet.

**DR. NORWOOD**: Thank you very much. Don?

## DISCUSSION OF MONITORING BOARD REPORT

**DR. YLVISAKER**: I apologize for the poor appearance of the handout. In 1990-91 I did a study of the TARO [Test of Adjustment-Related Operations] test site data, from 1986 in Los Angeles. In the course of doing that—I did basically the same argument here—I computed direct DSEs for the PES blocks, smoothed estimates and so on, for comparison's sake, and took these as truth and looked at undercoverage and what happened.

I was doing a lot of different things, and so I just sort of left them there. But after hearing about Charlie's work, I decided I should maybe go back and do something with them.

This goes to, I think, the first point of Charlie's. I discovered a memo, which was in my file somewhere, August of 1991. It says, there needs to be better information on the synthetic assumption. Charlie suggested an example of an option to compare the direct estimates for the 5,000 PES blocks with the actual adjusted counts and to a synthetic model, and make estimates for those blocks. This was in 1991. I guess, eight years later, Charlie has the data and has done this.

At any rate, I went back and looked at my own stuff. This was a heavily minority district in East Los Angeles, 73 percent Hispanic, 15 percent Asian, 12 percent other, I guess. Generally speaking, what you look at is, there were 169 block clusters that had population. I find the correlation between the undercount and the adjustment to be about 0.038. That is, adjustment is almost independent of perceived undercount. You will notice one point way to the far end, beyond 0.6. I put a correlation star to delete that point and consider the correlation subsequently.

I aggregated up. Block clusters were in tracts—one block cluster, perhaps, up to five or six block clusters, in tracts. I aggregated and looked at the same exercise when you did tracts as opposed to blocks. Doing the straight DSE on the tracts was just about the same as aggregating the block estimates up. What you see here is that the correlation does go up. There is a somewhat better relationship between adjustment and size of undercount. It does not have to do with small sizes, such as three people in the block or something like that. When you look on the reverse side, I made sure that these things were of sufficient size to do some sort of decent estimation.

I think this exercise of getting these data out very rapidly is a very useful one. I have only had a few minutes to look at Charlie's file. But when you do that and you look at some of the adjustments and some of the undercounts that are in that file, you say, boy, this is certainly a strange-looking block; somebody made an error here somewhere. Sometimes I think that is probably in the PES. There is an instance where, I think, in the enumeration phase, there were something like 70 owners in a block, and by the time the PES showed up, there were no owners. Of course, these were not matched people. Something had happened. My guess is that it was an error of some sort. But this is a very useful device for finding errors, I think, to look at the block level and see what is going on.

There are many, many tests that I think can be done, and should have been done back around 1991 and on. You have these PES data. You have three cells for many, many blocks. Lots of homogeneity/heterogeneity kinds of things should have been done. Given what happens in 2000, they should be done.

**PARTICIPANT**: What changes do you think could occur in the design of the 2000 census that will give different results in this? Or do you think the same thing would happen?

**DR. LERNER**: I am afraid, actually, that we have seen that the A.C.E. is very similar to the PES. While there are some differences in post-strata definitions, I think we might get the same thing. I hope we do not, but I have a feeling that we will.

Let me say something else about the sensitivity of this analysis. I did the runs for it and crunched a lot of the numbers. We tried all sorts of ways of changing this thing. I dropped out, for example, the block clusters with the bottom 10 percent, that is, with count under 30, and at the same time, eliminated the upper 10 percent. The PES, as you guys probably know, was subsampled, and we dropped those, too, to see if that would make any difference in this kind of result. It did not make any difference. As Charlie mentioned, we also used the Census Plus as our standard of, quote/unquote, "truth," to see if that would make any difference. It really did not make any difference. Finally, we originally had done it without imputations, and then we added imputations. That also did not make any difference.

I want to suggest that this is a fairly robust result at this level of analysis. What really needs to be done, I think, is to probe the synthetic assumption, especially in light of the fact that I guess we are going to have new post-stratum definitions, as we heard today.

**MR. JONES**: I might say, I do not think there is anything we can do at this late date to change the 2000 census, except maybe what we decide to publish. I think that is an issue that should be looked at: at what geographic levels we are going to publish these synthetic estimates.

**PARTICIPANT**: I echo Don's remarks. Getting the data is wonderful, because you can actually do something. I pooled the blocks by county, because we had county identifiers, and found a correlation of 0.18. If you pool it up, you get a correlation of 0.45. You really do see a change as levels of aggregation change.

**DR. LERNER**: I did some of those analyses myself, and I got similar results. Unfortunately, without the weights, it is really difficult, because at the smaller levels, the tract level and the county level, which is what I was looking at primarily, you do not have all that many block clusters, and therefore all that many people. I did a version of that at the county level by simply dropping out all of the counties that had only one block cluster. Obviously, if you have one block cluster, you are not aggregating anything. I got about the same result. I got about a 0.16 correlation. It comes out about the same.

Clearly, the higher you get, as you said, from an obvious statistical principle, the better it does. Really, the issue again, as far as the synthetic assumption is concerned, is, at what level does it really work, and how far down can you go? That is my sense of it.

**DR. FIENBERG**: Bill Eddy and I have spent some time analyzing the numbers, too. I have a prefatory remark, but then I want to tell you a few of the things that we have been looking at. We have actually just gotten partway in, so this is not in any sense definitive.

I agree with everybody who said that looking at numbers helps a lot. It sort of focuses the mind about what you are talking about.

I want to remind people that the structure of the PES and the crucial role of post-strata are essential to understanding what the synthetic DSEs are in this analysis. The comparison here is with something that directly used that, as well as a sample-based structure, to combine them with weights that are, I think, essential to any analysis. In the absence of the weights, you can only look at restricted numbers of things.

In the absence of what made up the post-strata, you can also only look at a certain number of things. Therefore, everybody should understand that what are in the data file that were analyzed are not the numbers that were used to create the synthetic estimates. Those aggregate people up by demographic groups. The post-strata consist of people grouped together in a demographic sense. These are counts, raw counts, for the block. They are also raw counts for the PES. We do not know how they break down. That is an interesting question that I will come back to.

Nonetheless, I like the idea of looking at blocks—they are the units—and unweighted, too, because at least you get a sense of what the real numbers are, rather than one number representing some larger number, but also because it lets us map back conceptually to where dual-systems estimation came from.

The other reminder I wanted to make is— the preface is that when Howard began his presentation this morning, he began with a $2 \times 2$ table. For me, every time I think about a block that is in the PES, I think about how to construct a $2 \times 2$ table for that block. Of course, the rest of Howard's presentation told us that that is not what he was doing. In fact, the whole process leads through a variety of paths to estimate things that the Bureau is good at estimating. We are never quite able to take things back to the $2 \times 2$ table.

So one of the things we actually tried to do was to, in effect, reconstruct, wherever we could, the $2 \times 2$ tables, to understand what was going on in those blocks, to clarify the differences.

Unfortunately, at this level of analysis—that is, at the level of the block—the data do not cohere for $2 \times 2$ table construction. It is not so easy to create a $2 \times 2$ table, with three cells that you get to observe and one that you get to fill in, that makes any sense. That is a real problem.

We began by looking at the data in a variety of ways, slicing them up into groups. Very early on, when we got one subset of the data, we discovered that for many of the blocks we were looking at—I will not even tell you what the subset was, because it turned out not to matter—for many of the blocks we were looking at, it turned out that the number of matches exceeded the number of people who were in the census. We said, oh, that means the $N_{11}$ cell is bigger than the margin. That violates everything I teach my students. Is something wrong with the arithmetic? The answer is, of course not. It has to do with the nature of the matching

procedure and, of course, sometimes the size of the blocks.

Maybe this is an aberration. It happened only a few times. Let us find out how many times. In 25 percent of the blocks, the number of matches exceeds the adjusted census count, where you adjust by removing erroneous enumerations. If you take out the people who do not belong in the block—assuming that that number is correct—then, in fact, in one out of four blocks, the match number is bigger than the census count.

What is the implication of that? It turns out that if you sort the data according to match being bigger than census count—one group where you do that and another one where you look at the blocks where it is bigger than this adjusted version, and then the residual—you get three very different pictures about how direct dual-systems estimation matches against synthetic estimation, with match rates and coherence differences of 10, 20 percent—radically different stuff.

I do not quite understand, but one of the things that is pretty clear is that in many of these cases we have very small blocks, something that has been mentioned. What we know [is] there is something about the dual-systems estimate. That is, without doing the fancy sample-based weighting version, the variance associated with the dual-systems estimator is proportional to the sum of 1 over the count. So it is 1 over the count in the first cell, plus 1 over the count in the second cell, plus 1 over the count in the third cell, plus 1 over the count in the fourth cell. If those numbers are really small, then the variance is gigantic. We are looking at blocks where the variance dominates absolutely everything in the analysis.

So if you look block by block, the variances are just enormous, except for a few very big blocks. I want to mention a couple of them.

A second thing is, this is a wonderful picture to have up there. It really suggests that a lot is going on out in the tails, and we are missing enormous amounts of stuff. In the tails, things are really bad. How many blocks are in the left-hand tail bar? The answer is 45. That is out of 5,180. How many are in the next bar? Another 45. How many are in the next bar? Another 100. About 200 blocks compose most of the left-hand side of the graph.

What is the characteristic? They tend to have very small and erratic behavior. So they are ones with very small counts and very erratic behavior.

What about the other tail? The other tail is bigger than 120 percent coverage. That is gigantic error. It turns out they are very different. If you look item by item, there is even one where that factor is a factor of 14, 1400 percent. That suggests something really screwy is going on in that block. But for that bar there are only 100.

So if you add up the two tails, and you have about 5 percent of the sample of blocks, what is left is not much going on at all and things are tracking, actually, remarkably well the dual-systems estimator, even if you believe that it was the smart thing to compute, which I do not.

Finally, I should note—and this is something that is in one of the documents that was distributed today—that there are 2,000 blocks with zero household non-matches. These turn out to be right in the middle of the distribution.

So what is the number of blocks you should be looking at? How should you be normalizing it? What analyses should you be doing? It is pretty clear to me that

you want to peel out the two tails, because you cannot look at them in that way. You need to peel out about two-thirds of the middle. I do not know what is left after you do that, because we only had a little bit of time to do the analysis.

I began by noting that the census real way of doing this, through synthetic estimation, actually aggregates up to post-strata by using demographic characteristics. You get that in the file that was analyzed. We spent a lot of time actually seeing if we could verify the numbers. We went into the published data on the P.L. 94-something file [1990 P.L. 94-171 file for redistricting]. We took the largest block in the PES, and, because it actually had some labels on it, we tried to identify what it was.

It turns out to be a very interesting block, as best we can discern. It is Fort Polk in Louisiana. Fort Polk is one of the most interesting integrated blocks in the nation. It has virtually every minority group, in substantial numbers. It also is a case where the synthetic estimator goes in one direction from the census count and the direct DSE goes in another direction. What is one to make of that? I think that this analysis offers us no clues whatsoever.

**MR. JONES**: I do not know, Steve. You apparently have a lot more data than I have. These estimates were given me by the Census Bureau. I certainly did not get the records by race or anything.

**DR. FIENBERG**: We took your data. We chose to analyze nothing but your data. But the question was, which blocks are they? You were able to identify the Robert Taylor Homes in Chicago, right?

**MR. JONES**: There were no samples in Robert Taylor Homes.

**DR. FIENBERG**: It turns out that you have a state number and a tract number. The P.L. [94-171] files also do. We have matched block by block, for several blocks in the PES, into the published, public data that are available on the Census Bureau Web page, with very high probability of a match.

**MR. JONES**: Well, good for you.

**DR. EDDY**: I would like to just clarify that a tiny bit. In the data that we got from the Monitoring Board Web site,[4] there is a state, a county, a tract number, and a PES cluster number. The state and county nail things down pretty well, and the tract number actually tells you the census block numbering area that is under consideration. Then you can just go through and select out those blocks. That is what we did. In some cases, they match exactly. That is, you can just add up the block counts and you get the block cluster. In other cases, there seem to be additional things, such as for Fort Polk, as Stephen mentioned, in Louisiana, where there appears to be a huge number of additional people, I assume from group quarters or something like that. (I used to be in the military; I think they still do it that way.)

So we actually were able to achieve quite detailed information from the P.L. 94-171 file by matching it with the PES.

**DR. YLVISAKER**: I guess I totally missed the variance component of this. I used the production estimate for the given block, and I used the block figures. A principle I normally use is, if I cannot predict what I see, then I am going to be

---

[4]See http://www.cmbc.gov, under "Reports to Congress."

hard-pressed to predict what I cannot see. In this case, I believe I was looking at the block business, and I was looking at what the estimate for the block was. I do not know whether variance is supposed to dominate all of this. I had no weights. I had nothing is what I was told. This is what this is supposed to be for this block; here is what the block looked like when we looked at it.

**DR. ZASLAVSKY**: I was going to say some more general things, but I think, actually, it ends up answering the comment that Don made.

In the report from the CMB [Census Monitoring Board, Congressional Members], the explicit conclusions of that—for example, that there is still a strong incentive for people in local areas to work hard to improve their count, because it is not going to be adjusted away, or the fact that the Census Bureau should use the best possible methods to improve the count in every local area, because they cannot all be corrected by adjustment—those are perfectly true. I really hope no one here disagrees with that.

It is the sort of implicit conclusions, which have to do with the view that failure of the synthetic assumption invalidates the usefulness of adjustment or the usefulness of estimation using dual-systems estimation, that I think are what need to be addressed.

First of all, synthetic estimation was never intended to be able to predict what is going on in extremely small areas. Taking it down to the most extreme level, we could do a synthetic estimate for my house. In my house, either all eight of us who reside there will be counted or one person will be missed, which gives you an undercount rate of 12.5 percent. I really doubt that any kind of synthetic estimation can get that.

When you get up to the block level, we are dealing with essentially the same thing. Blocks are very small units. Even if you believe the most extreme synthetic assumption, that at the process level, what is going on in every block in a post-stratum is the same, you are still going to see something like binomial variation, because not every household coordinates within the block, to miss exactly the right number of people that was predicted for the block.

So you expect to see this kind of variation. It is actually more variation than binomial, because there is clustering, people are missed within households, and so forth.

Beyond that, I do not think anyone really even believes the synthetic assumption the level above that. We know that there are very few post-strata relative to the kind of diversity that exists in factors that affect undercount in this country. With any reasonable-size PES, unless you had some really fabulous way of drawing post-strata, which would not only pick out all those areas that Charlie listed on the slide, and the other thousands of areas across the country that are like that, from all the neighboring areas that are not as extreme—would not only do that, but also divide those up ccording to the ones that have 10 percent undercount, 20 percent undercount, whatever—the synthetic assumption cannot really be true. Maybe the term "synthetic assumption" is a little misleading. It is an assumption for developing a procedure, not something you have to believe is true in order to use the procedure, any more than every time someone does linear regression he believes that all errors are perfectly normally distributed.

So synthetic assumption is wrong. Not using the PES also has a synthetic assumption, which is that the undercoverage rate is the same all across the country and is zero. That is the same assumption in a stronger form. If the first assumption is wrong within post-strata, then it has to be wrong to treat the whole country as the same thing.

What is going on here in these charts? There is this chart, and there is a different kind of chart that Don was showing us. Let us look at this one first.

Basically, this is a regression-to-the-mean problem. You have a bunch of numbers that are estimated in these blocks, as Steve pointed out, with large amounts of variance in most cases. There are other things going on that contribute to variance. I believe some of these blocks were subsampled for the PES, and not subsampled for the E-sample. So there are a lot of different things going on.

Anytime you have random variation and you choose a group on the basis of the outcome of the random variation, you are not going to be able to predict it. If you take any distribution, take the top 10 percent, that is going to look extreme, even if you have a very good regression model. If you choose on the *y* variable and take the extreme 10 percent, you are not going to predict how high they were or how low they were, even from a very good set of predictors.

So the question is, are we getting anything useful out of this? It is not whether you can predict every block that is defined by the extremeness of the group. I would argue that, as long as the model predicts some part of what is going on, then you are getting something useful out of it. So in this case, if we think, not of trying to predict individual blocks, but predicting the kinds of things that we actually deal with in uses of census data—which are typically political divisions or congressional districts, legislative districts, things like this, which are much more aggregated—a lot of this random variation at the individual block level, which is reducing the correlation in the plots that Don was showing us, gets averaged out.

There is lack of homogeneity in the post-strata, and there is also lack of homogeneity in the coverage of the original census, the same lack of homogeneity. All we have done by doing dual-systems estimation is to take out a part of it. People fit models with low $R^2$'s all the time and find them useful. There may be a lot of variation in individual observations, and yet getting broad patterns, especially when the things you are most interested in are aggregates of large numbers of observations, is still very useful.

As a sign of the limitations, which I think we all know, of dual-systems estimation for predicting down to very detailed levels and correcting all of the undercount at those levels—that is clearly the case. But to say that this shows that you are not doing anything useful I think is really wrong.

**DR. STARK**: I think this is a very sensible thing to look at. The fact that the pattern persists even when you look at larger blocks I find fairly compelling. But I can understand proposing a model where you look at the adjustment factor or the population in each block as being a sum of a correct adjustment, which the direct DSE is probably more likely to give than anything else (simply because if the homogeneity assumption holds on any level, it ought to hold for blocks or block groups), plus some random error that has to do with people's binomial probabilities of responding in various ways.

I am repeating hearsay here, but I will do my best to be accurate. David Freedman did some simulations recently looking at something like a binomial response probability. You do not get anything like the variability in the adjustment factors that you would need to explain the difference between the synthetic and the direct DSE. In order to get variation on that level, you would need an enormous amount of covariance among responses. That is not reflected when you look at the covariance matrix of the adjustment factors as reported by the Census Bureau. For the $357 \times 357$ covariance matrix, the diagonal is dominant by a multiple of about 100 over the off-diagonal terms. So there is very little covariance among what is going on with the adjustment factors for different post-strata.

**DR. BAILAR**: I am thinking back to a workshop we had on synthetic estimation almost 20 years ago, where we all came to the conclusion that it was a very conservative tool to be using. I think we see this happening again today, and I think the Census Bureau has been well aware of it and, in fact, has been counting on it for many reasons. It would be better, in many senses, to try to go in the direction of not overadjusting the population.

Coming back to thinking about the use of synthetic estimation, how can you make it a somewhat stronger tool? I think the charge that the A.C.E. and the 1990 PES are not true sort of negates what we have been sitting through a lot of today. Looking at the post-stratification variables, the fact that the mail return rate at the tract level will come in is probably going to sharpen that up considerably. There are other things that are being looked at as well.

I think a controversy that has been going on for at least 20 years is this whole thing about whether the synthetic estimate has to improve things at every level of geography, including the block. The feelings about the block-level counts—anybody who has worked in the Census Bureau knows that the block-level counts themselves are just subject to a lot of error. They are just not very stable. You know that there are error problems. Now you are looking at something that comes from these direct estimates, from the dual-systems estimates, and they have very large variances, so they are error-prone. Then you are using those as the truth to judge this other thing that is error-prone. I cannot accept using that as the truth in a situation where I know that there is just a lot of error.

Another thing that was said in the report is that—it sort of sets up a straw man—you have to have the best census, and then you do not need this adjustment. But everybody who has worked on adjustment also agrees that the best underpinning to an adjustment is a really good census. Nobody is arguing that it is one or the other.

I found the case of the Robert Taylor Homes in Chicago of interest. First of all, I notice that the estimate of the undercount came from the Chicago Housing Authority. I wondered if the Census Bureau had accepted that estimate of an undercount of 29 percent. Usually the Census Bureau does not accept estimates of undercount that come from local areas.

According to the Monitoring Board report, statistical adjustment would have added 5 percent. So the choice was between zero, which is no adjustment, and 5 percent, in 1990. The politicians decided on zero.

There is nothing to say that in 2000 the use of new census tools and census publicity would not reduce that 29 percent—if it was 29 percent—somewhat, and the sharpening of the adjustment tools would move that 5 percent up to some larger number. But the real choice is again between no adjustment and an adjustment of at least whatever that 2000 estimate will be for that area, and then the adjustments for that area joined to others as statistical building blocks. That is what is important.

One thing that has not been mentioned in the discussion this afternoon is that the Monitoring Board also in the report recommended the reinstatement of the Parolee/Probationer Coverage Improvement Program. In every analysis of the results of the 1990 census, that program was shown to lead to gross overcounting. It was put in place without the usual careful testing that the Census Bureau does with most of its practices. I would suggest that that is not the way to get a good census, through a poorly designed program that relies on overcounting.

**DR. SPENCER**: First an observation, then a question. If you think about two other sets of population estimates produced by the Census Bureau, not for Census Day, but postcensal estimates (where you estimate change since the last census, but still current, or a retrospective) and population forecasts (where you are forecasting the future), you see qualitatively similar patterns with the errors. States, for instance, that grow very rapidly are under-predicted; states that decline markedly are over-predicted. The change is estimated too close to zero. It is not a regression effect. But you find a similar pattern in other demographic estimates produced by the Census Bureau.

I do not know if that is relevant, but it is interesting.

My question is—I did not see it in this report—did you compute the average squared error for the two sets of estimates? You have your best shot at the true value, so you look at the deviation between the census and between the synthetic estimate. You can square those and then you can average them out. I realize that you said you did not have the block weights. But even taking an unweighted average of the squared errors—did you do that? If so, what did you find?

**MR. JONES**: No, we did not. We were waiting to get the weights that we thought were coming so that we could actually look at this at higher geographic levels. We were not just interested in blocks. We were interested in neighborhoods and other kinds of local-area groups. What we have are the block data. That is it.

**DR. SPENCER**: I understand, but it is still interesting to know at the block level, warts and all, whether the synthetic estimate gives you, on average, a more accurate number than the raw census count.

**MR. JONES**: That is certainly something we should look at, Bruce.

**DR. BROWN**: I think that follows from the positive correlation. It has to. You can sort of unwind it on the unit-by-unit basis just from that.

**DR. YLVISAKER**: In the TARO data, roughly, the adjustment—and most blocks came out to 8 percent. Therefore, all blocks that were undercounted between whatever it is are going to be closer to the truth. Everybody who is undercounted a fair amount, unless you are undercounted by less than 4 percent, is going to be closer to the truth, by the time we lift everybody 8 percent. This is a

very simple fact. The only thing is, you are also taking people who are overcounted and you are lifting them 8 percent at the other end.

**DR. ZASLAVSKY**: On the TARO, I do not know the exact boundaries of it, but I think pretty much all of it was in a very undercounted area. If we had done a TARO which had included also Sherman Oaks, say, as part of the area and calculated adjustment factors for South Central L.A. and Sherman Oaks, I think we probably would have found that, looking at those two areas, we would have come closer.

I cannot understand the point that Phil was making. Your main argument is that the blocks differ systematically more than you would expect due to binomial variation. If so, that is affecting all the post-strata in the blocks. Since it is the between-PSU [primary sampling unit] variance that drives the covariance matrix, then you should have strong off-diagonal correlations in the covariance matrix, at least for the blocks within related post-strata.

**DR. FIENBERG**: I just wanted to come back to the point Barbara was making about the assumption of the analysis—that is, that we know truth, and truth is the direct estimate. Just think about Fort Polk for a minute, which has every minority group in substantial numbers, and ask yourself about the heterogeneity assumption associated with the dual-systems estimator for Fort Polk. Clearly, it is violated. I give that as a good example, but there are other such blocks. Unfortunately, we do not have the breakdown, so we cannot look at them all.

But that means that truth in this study is something with not only high variance, but also potentially high bias, and bias that flops around depending on the block. I just do not know how you can interpret this graph. The base line is a moving base line. We have ignored both the bias and the variance. I cannot deal with the variance-bias tradeoff, or any other tradeoff in the analysis, because I cannot use any of my statistical tools to understand what is going on.

**DR. NORWOOD**: David Murray is sitting in the back, from the Monitoring Board [Congressional Members].

**DR. MURRAY**: Thank you. I guess what gives me the most difficulty with regard to the problems of synthetic estimation, when applied here, from the perspective of a Monitoring Board person looking at the process unfolding, is the use to which the data are put. That presents the greatest difficulty—not just a theoretical exercise about the numbers. The effect of this non-correlation and the potential difficulty of the actual need of undercount not being satisfied by the impact of the adjustment can best be seen most transparently at the very small level of resolution of the neighborhood, of the local block, of someone's individual house. But the effect is continuous.

What gives me the most difficulty is that there is a potential tension between the claims of numerical overall accuracy and distributive accuracy, which the synthetic estimate puts into a certain kind of stronger shape. That is, sometimes one of them is achieved at the expense of the other. If distributional accuracy is the most consequential politically, as we move back in our resolution, as we step up to higher levels of aggregation, indeed we do begin to gain something with regard to the variances of synthetic estimation, and we are moving towards a goal of helping the count.

But the difficulty for me is, processes are initiated within states, at levels that are intermediary between an individual house and the state—processes of apportionment, processes of resource allocation. There is nothing in this method that can tell me at what point the synthetic-estimate aggregations become reliable enough for those processes to be initiated without an expense at the cost of distributive accuracy, where those decisions have to be made. It is a continuous difficulty; as the aggregation increases, we do get better. But there is nothing internal to this that can tell me what my criterion is or what my threshold is. We have shown that the correlation is likewise not particularly striking even at the state level, and we are using between-state regional data. This gives me a difficulty about the initiation of processes that do depend upon local-level areas of accuracy for distributional consequences. I cannot tell when the effect is sufficiently ironed out that I now can move forward with confidence for those political processes.

**DR. FIENBERG**: All the analysis from 1990 showed that at state levels and at high levels of geography within state, there is no tradeoff, in the sense that both distributive accuracy and numerical accuracy can be improved by use of the synthetic adjusted counts. That was the result of the analysis. That is what Howard referred to before. There is not a single study that shows a reversal of that.

At the block level, we just had a discussion that said you cannot do much. But as you aggregate up towards those levels where we have fairly definitive evidence from 1990, you know that you are going to get toward that and cross a threshold. Where the threshold is, is an interesting one, but I do not think that what it says is that there is a problem with synthetic estimation for those within-state uses. I think we need to know more about what levels of aggregation you do well at, but I do not think there is that tradeoff at the multi-multi-block level, where most of the action is, where you tack on one block and take off another one, and are interested only in the total.

**DR. NORWOOD**: That is not our job to discuss, really. I really do not want to get into that. Let us stick to the statistical issues. That is an issue of what the director decides he is going to do, and I certainly do not want to advise him on that.

**DR. WOLTER**: I apologize, I am just coming back to this issue after 10 years. I thought these issues were settled.

In particular, I would like to urge for your reading list probably no fewer than a dozen articles written by people at the Census Bureau, like Cary Isaki and Bev Causey, who did extensive studies over roughly a 10-year period, analyzing the relationship between the geographic level of aggregation and accuracy. If you go through those papers, they answer all these questions very clearly, I think. At least in my opinion it did, 10 years ago.

The other comment I wanted to make is that I am not sure I quite understand this analysis or the analysis done of TARO, in light of the fact of erroneous enumerations. I do not see how one gets a direct dual-systems estimate at this level. Again, I am just coming back to this. But I do not think it works at that level. In TARO, for example, erroneous enumerations were measured at the citywide level.

In this analysis, erroneous enumerations are measured at the cluster level. You are going to have to explain to me (not now, but some other time) how in the world one even gets to a dual-systems estimate at this level. I do not think it makes sense.

**DR. NORWOOD**: Are there any comments or questions from the panel members? Oh, I am sorry. I forgot about you, John. I know you have been wanting to speak for some time, and now we will let you.

**MR. THOMPSON**: We talked about this report a great deal at the Census Bureau. Howard was interested, and I was interested. I was actually kind of encouraged by the report, for three reasons.

First, as everybody has noted, at the block level there is a lot of noise in the data. What that chart really shows—it takes the noisy data, and the synthetic adjustment makes an improvement. I found that to be fairly encouraging.

The second thing is that, if you look at the synthetic adjustment, the noise in the synthetic adjustment will come from the noise in the unadjusted census. This sort of illustrates a point that the Census Bureau has been making. That is, we are not relying on the adjustment to be perfect. We try to do a good census, and then we use the adjustment as a final quality check to make an improvement. This, again, illustrates that the better the census, the better the adjustment will be. In that sense, I found it encouraging.

The final point is, as near as I could tell from reading the report, it looked as if it was basically an analysis based on numeric accuracy as opposed to distributive accuracy. Again, I was encouraged. It seems that when one starts analyzing the data, one starts to look at the concept of numeric accuracy. I found that to be encouraging.

**MR. JONES**: Just one comment. One of those three things, John, I agree with you on, which was the middle point. That was really the purpose of what we set out to do, to say to people, do not just depend on the adjustment to take care of your problems. You have to get out there and get a good count, and then if adjustment comes or does not come, you are still in good shape. But if it comes, you will be in even better shape. I believe you have to have a good census in order to do the adjustment.

**DR. NORWOOD**: I think that is a great note to end this part of the discussion on. Thank you very much.

## COMMENTS OF ALAN ZASLAVSKY

**DR. NORWOOD**: I told you earlier that two people could not stay for tomorrow. We have heard from David Freedman. Alan is the other statistician who has to leave, unfortunately. We are glad that we had both of them here today. Alan Zaslavsky, do you have some comments? Everyone else will have a chance to speak tomorrow.

**DR. ZASLAVSKY**: I was panicking a little bit about what I was going to say, because I do not have any slides, like David. I will just speak very briefly to a couple of things that are on my mind at this point.

I have to make one brief statement of a couple of pet technical peeves, since Howard has to listen to me. It might help some of these discussions if the dual-systems estimator were presented the way it is actually calculated. To my mind, the $2 \times 2$ table is often a distraction from describing what the estimator really is in practice. There is something that is much closer to it on pages 3 and 6 of Rick Griffin's report on the calculation methods [Griffin, 2000]. Even there, there is some ambiguity about which things are coming from the sample and which things are coming from the population, the imputed cases, inadequate-data cases. I think it would probably be helpful if we really described it the way it works. Again, the $N^*$, the unreachables—I do not want to let the concept go unchallenged. We do not really know that there is such a thing as "unreachables." We just know that a number we estimate is smaller than we think that number should be. Whether it is one group that is unreachable, or it is just that there is a bigger group and you only happen to get half of it anytime you do things, is really something we do not know about what the process is.

My substantive point—and it is really just one big point—is to emphasize to the Census Bureau the need for evaluation. In 1990, all of us who care about the census, mainly the Bureau—I did not do this work, except a little bit—were forced by the fact that there was this very specific timeline for a decision to create something that at least looked like it was an attempt to address all possible sources of error through evaluations. There were a lot of components to it. It was a lot of work. Some might say it was somewhat unwieldy. Yet at the end of it, you were able to say that at least an attempt had been made to measure everything you would need to measure in order to answer questions like the relative accuracy of the raw census versus the census with coverage measurement incorporated, or even to take that down to some lower levels and think about how the differential impact of certain kinds of error might have affected different areas or different groups in the population. Of course, the overall framework for this was the total error model.

Almost every piece of that evaluation has been contested. But it was there to contest.

In this situation, as Howard emphasized this morning, even though some of these things will be done, it is not so integrated into the process, because there is not that kind of a formal decision track that we are working on. But I think that what we got out of 1990—not necessarily every detail of what the evaluations were (because I am sure, with the benefit of hindsight, we could think of ways of improving them), but the basic concept of thinking about the elements of error, the work that Bruce and Mary did, for example, on the total error model, which, I gather, is being carried forward—setting in place the evaluations needed to get some quantitative estimates for each of those pieces is going to be really essential. Otherwise, whatever comes out of 2000 will hang under kind of a cloud of uncertainty, a cloud of uncertainty seeded by the dust that is still flying around in the air from 1990. I think we can do much better than that, and it will do a lot for the stature of the Bureau and the acceptance of the census.

## COMMENTS FROM THE FLOOR

**DR. NORWOOD**: It is our custom at the end of each day of a public meeting to ask if anyone would like to make a brief comment. Ed Spar [Council of Professional Associations on Federal Statistics].

**MR. SPAR**: This goes back to the issue of telephone interviewing. We know that about 50 percent of those missed were children. We also know that many of those who were missed were in households that were reached. So if I may, I will make the assumption that many of the children missed were in households where the questionnaire was returned.

The question I have is, would it make more sense for those households to have personal interviews rather than telephone interviews? For whatever reason, they left the children off. I am wondering, if you do a telephone interview—and I assume there is going to be probing—would you still be prone to miss the children? If the interviewer got into the household, would there be a better chance to actually pick up those children who had not been counted?

**DR. NORWOOD**: Isn't that a lot of households?

**MR. SPAR**: It is for the Census Bureau, I think, to answer that part of it.

**DR. BRADBURN**: May I just comment on that? This is not definitive, and so forth, but I think the literature on household screening on the telephone does not reveal any big differences in composition of various sorts. There is no presumptive evidence that the mode would affect the accuracy of the reporting.

**DR. BROWN**: Once the issue of the missing children has been raised, let me ask—you know about that, Howard—what, if anything, is being done to improve that, besides crossing your fingers and hoping they get counted? It is not necessarily an adjustment issue.

**DR. HOGAN**: In general, there are the general changes or improvements to the census—the user-friendly questionnaire, the work we have done earlier in the decade to try to make the questionnaire friendlier, just that general approach. I do not know if advertising would motivate people to add children who might not otherwise add children, so I guess I cannot point to that.

I do not know of anything specific, except for the fact that the questionnaire is a little bit easier for people to understand, and the questionnaire for the interviewers is a little bit easier. John might be able to think of a couple more things.

**MR. THOMPSON**: We also have a Census in the Schools Program, where we are trying—actually, quite successfully—to go into classrooms and teach lessons on the census, to sensitize the children to the importance of being counted, so that when they go home, they will be counted. We have been going at it that way, too.

**DR. BROWN**: Unfortunately, the ones who are missed are even younger than that, right?

**MR. THOMPSON**: No. Our misses are zero through 18. That is the category of children.

**MR. WAITE**: This is clearly one of the "wily trout" issues. One of the things that we are doing that we hope will give us a handle on missed individuals is that, when you get your census form in the mail at the appropriate time, you will get an

opportunity to write on the first page of the questionnaire the number of people in your family, and you will list the individuals in the census. If you should get tired of listing people and send in a form with four people even though you had written five on the front, you will get a friendly call from the census giving you one more chance to 'fess up to that child that you have forgotten. It gives us an additional hook and an additional chance to call back on it, and we are going to call back on those cases. We have a process, in fact, anytime there is a discrepancy between the number of people filled out and the number on the front.

That does not mean that it is perfect, but that is another hook into those families, where perhaps they might have not mentioned the younger child.

**DR. EDDY**: I just want to follow up on that. There is room for six people on the form, if I remember correctly. It would seem to me the most likely case is that someone is going to put seven in that first box and then only fill out six people. I can imagine your calling them and they go, "Well, six. I don't want to deal with this." What is the actual process going to be? I am very worried that seven-person families are going to be undercounted.

**MR. WAITE**: We do not call them and say, "Hey, we noticed you only filled out six on the form and you put seven on the front." We would call and try to understand how many people there are in the family, without regard to divulging to them whether we saw two on the front and six inside, or whatever, and go through from scratch how many people are in the family, the names of those people, to try to figure out who is missing and who is not.

This is not a perfect world. I did not mean to imply that this is going to completely solve the child problem. But in answer to Larry's question, this is another hook or another probe into trying to get at that issue. You can list on the roster up to 12 people. What I am talking about is a comparison of the box on the front and how many people you actually filled something out for on the census.

The children issue is a very complicated process. We have not completely cracked it, but we are trying harder to do some of those things.

**DR. MULRY**: Some of you may know better than I do about the national immunization survey. My impression is that that survey was a telephone survey that had trouble finding two-year-olds.

**DR. NORWOOD**: It had a lot of trouble.

**DR. MULRY**: That is the reason I thought Ed's comment about the telephone and children is worth considering. Would that same phenomenon happen with a telephone call from the Census Bureau?

**DR. NORWOOD**: Let me hasten to say that it is not done by the Census Bureau.

**DR. MULRY**: That one is not done by the Census Bureau. It is an RDD [random digit dialing survey]. It is done by government contract, private firm.

**DR. BRADBURN**: It is a whole different kind of problem, I think.

**DR. MULRY**: There are a lot of differences. However, it is a very large government-sponsored survey that had trouble finding children on the phone. I can see where there is some cause for concern about it.

**DR. NORWOOD**: Is there anyone else in the audience who wishes to say something?

**DR. BRADBURN**: I do have a question, if I may revert to an issue from early on. I wanted to ask Jay about the return rate. Something you said later raised a question in my mind about the calculation of the return rate. How does a household that does not return its bar-coded form but returns a "Be Counted" form get counted in the return rate?

**MR. WAITE**: First of all, if a household did not mail back a form, number one, it would almost certainly be visited by a nonresponse follow-up interviewer. But let us say they filled in a "Be Counted" form. Your question is, does that "Be Counted" form represent a mail response, in a sense?

**DR. BRADBURN**: It was really in terms of the return rate. Would it get counted as a household that returned a form?

**MR. WAITE**: I do not know. I will have to check on that.

**DR. BRADBURN**: Because if you are going to calculate a return rate, you need to know.

**MR. WAITE**: I do not know. Danny?

**PARTICIPANT**: It does.

**MR. WAITE**: It does. It gets counted as a return.

**DR. BRADBURN**: That would be my sense, that it ought to be.

**DR. NORWOOD**: Ken, would you have any word of wisdom at the end of this day, before tomorrow, to give us?

**DR. PREWITT**: No. I am absolutely certain that by noon tomorrow I will be so much wiser than I am now that I will wait.

**DR. NORWOOD**: Andy?

**DR. WHITE**: I would like to thank everyone for coming and sticking with us today. I hope all of you will come back tomorrow to see the end of this story, and Ken's wisdom. We will be meeting here early. The meeting will start at 8:30.

I would like to thank you all for coming and giving us your time and your interest.

## OPENING REMARKS, SECOND DAY

**DR. NORWOOD**: I would like to call this meeting to order.

I spent a good bit of time after dinner last night thinking about what we had accomplished or not accomplished yesterday. What I would like to do is to exercise my prerogative as chairman this morning to change the agenda a little bit. I want to ask a question—actually, three questions, by the time I am through—and I would like to get, especially, our invited guests (but others as well) and the panel—the panel both to listen and to help to interrogate and to raise other questions. Let me explain.

We have to look at what the purpose of this panel is. As I interpret it, in any case, it is not to decide whether the numbers should be adjusted or not be adjusted. That is really the role of the Census Bureau and its director, who will have to face

that issue and make a decision, decide how to do it. Our role is a very different one. Our role is to evaluate what the Census Bureau has done.

There are obviously, on adjustment, two possibilities. One possibility is that the director can decide to issue adjusted counts, as well as unadjusted counts. Then, if that is the case, the role of this panel would be to use all the information that we have at our disposal and that we will get from the Census Bureau, as well as our understanding of all the material that was presented yesterday, plus all the discussion from yesterday, to try to look at what has been done—look at variance and look at correlation, and look at all the sorts of things that one can look at. That is one way to do it. There is another way, but that is one way.

There is another possibility. That could be that the Census Bureau decides, when it is completed, that it has done such a great job that it does not want to adjust, and the decision is against adjustment because you have such confidence in the numbers.

Even if the director decides to adjust the numbers and we try to evaluate the specifics of the adjustment, we also have to look at what would happen in the absence of adjustment. I know we have a lot of statisticians here, but I can tell you that, at least when I was in a position of decision making, I could not get away with just saying I was going to look at these things that we were going to do. I always had to look at the alternative and the strengths and the weaknesses of both, and then try to make the best decision that I could possibly make. I am sure that is what Ken Prewitt is going to do.

The question that I have, that I would like to get some discussion on, is, if you were all sitting on this panel, what is it that you would want to look at? How would you evaluate the accuracy of the census without adjustment so that you could compare it to the census with adjustment?

The focus yesterday was all on adjustment. We know there are lots of problems. I do not think that is particularly new. We know more about what the problems are here, and there are many. I am not pushing those aside. I think they are very important. But, as I said yesterday, I do not know where we are going to come out.

But we have, I think, another responsibility. This panel has discussed this a little bit, but not really very much, because we really need help. The question that I would like to put out to you is, how do we look at these two issues together?

Alan Zaslavsky touched on this a little bit yesterday when he said, if you do this, you have this kind of problem; if you do not do it, you are making an assumption of zero. I do not know how to do that in a way that—the other problem we have is that anything we say is going to have to go through clearance with the [National] Academy of Sciences, and I am sure that we will have people selected who will not be people that we would select, but people selected who are on both sides of this issue. I think we would have to have scientific evidence that we have looked at both sides.

So I throw that question out to you. I would love to have some discussion of it.

## EVALUATION OF THE CENSUS VIS-A-VIS ADJUSTMENT

**DR. FIENBERG**: I have actually reflected on this question a couple of times. There is a curious feature about what is going on, which poses a problem and has an advantage. I do not know where what I am about to say leads.

If what you are interested in is how well the enumeration did, how do you know? The answer is, we get reports from the field and we have a few—

**DR. NORWOOD**: We know about operations.

**DR. FIENBERG**: —a few numbers on operations, and we learn from the post-enumeration survey, now A.C.E. A.C.E., in this particular setting, as the PES in 1990, is actually designed to do two different things. It is an evaluation program, and it produces adjustments. The fact that we know something about the level of gross error in the 1990 census is due primarily to what we learned from the post-enumeration survey in 1990. I do not know where we are going to get any other metric for that kind of error, unless there is an attempt to set it in motion very quickly. And I do not know what the tools are.

So the fact that the two are intertwined, I think, presents complications. I do not have a solution for you, but we do have a procedure in place to understand aspects. The question is, what, besides the things that have traditionally come out of that program—for example, the things that came out in 1990—could you glean from the post-enumeration survey that would tell you more about what is going on? How can that information be combined with information from the other reports?

It is very hard to think about combining information when the other things are primarily operational as opposed to evaluational.

As I say, I do not have an answer, but it is pretty clear that you have one tool for looking. The question is, what other ones are there? I am not sure what they are.

**DR. BAILAR**: I am remembering back to some past censuses, when we were trying to decide whether we were in the middle of a good census or one that had some gaps here and there. It seems to me that one of the things is that you do get an early reading from demographic analysis. But this time I would be very loath to say, on the basis of that, that the coverage was going to be what it should, because I think there is a lot of fear about overcounting in this census. There is a lot of opportunity for people to respond in multiple ways. What that is going to depend on is that the Bureau really can unduplicate all these forms. I do not know that process, but I know that there can be a lot of error in it.

So you could get a good reading, that you are doing fine, and it could also show up later on in the PES that you still have these gaps, you still have the differential bias, and so forth.

I agree with Steve. When it comes right down to it, it is going to be the PES that tells you how well you did.

**DR. SPENCER**: The answer to your question depends partly on the timing. If you can wait until the evaluation studies are done—

**DR. NORWOOD**: We cannot.

**DR. SPENCER**: I understand. If you could do that, then you could do what the Bureau did in 1990, which was to do a very careful look at the accuracy of the PES and the DSE. The DSE is a read on the accuracy of the census, but going into it, you do not know how good the dual-systems estimator is. Down the road, there should be very good information about that. The question is, in the short term, trying to decide what numbers you release, how do you know how good A.C.E. was; how do you know how accurate the dual-systems estimator is?

I do not have an answer for that right now, but I think people need to do a lot of thinking about what indicators of quality there are, operational things like missing data rates, what information you will have about matching errors, what diagnostic information you will have. They will only be indicators about how well things worked, and then you can draw on information from past experiences—if you think that correlation bias is likely to be what it was in the past, or synthetic estimation error—and try to borrow information from other settings.

I do not know at this point what indicators you want to look at. But I think some work needs to be done to try to put together a list of what the indicators are, what they would indicate, what their fallibility is—just start thinking about how bad things have to be to tip the balance one way or another. You cannot totally specify things ahead of time. We can expect to be surprised by something or other. But you do not want to leave it totally open to judgment, either. I do not think you can eliminate judgment from the process, but you do not want to make it totally judgmental. I do not know how you draw the line.

But I think those are the issues.

**DR. NORWOOD**: So if there is adjustment, there is subjective judgment. You are suggesting that if there is not adjustment, there is subjective judgment about that decision.

**DR. SPENCER**: For instance, one of the things you would want to look at it is sex ratios, look at sex ratios from the census and see if they roughly conform to what you would expect. As Barbara Bailar pointed out, you can have misleading information there, because of overcounting, that could mask undercounting and could hide a very bad distribution of error geographically. We also have unknowns about the race/ethnicity question, what that will do to sex ratios from demographic analysis. But even so, that is something you want to look at.

Where judgment comes in is, given that you are going to look at sex ratios—and you can look at them from the census, you can look at them from A.C.E—how far off do they have to be for you to say you are very uncomfortable with this, and we should not do it? I think some judgment comes in there. I do not think you can form a fixed threshold ahead of time. You can have a guideline, but ultimately—thank goodness I am not having to make the decision—I do not see how you could be comfortable in saying, if it is to the left of this point, we do X, and if it is to the right, we do Y.

**DR. NORWOOD**: So if this panel were to evaluate what the Census Bureau has done, whichever side it goes on, and you were one of the reviewers of the panel report, you would expect us just to look at a bunch of indicators, and that would be enough.

**DR. SPENCER**: I would hope that you would have input into what indicators the Bureau should look at. I would think that is an interesting project for this panel.

**DR. NORWOOD**: We are evaluating; we are not telling the Bureau what to do. It is a conflict of interest.

**DR. SPENCER**: Okay, then they can present their plans for what indicators they are going to look at, and you can evaluate their evaluation plans.

**DR. BILLARD**: I think the bottom line is that you want to get as accurate a count as you possibly can. The question, as I understand it, is, to get that most accurate count, do we adjust or do we not adjust? If we do adjust, how do we adjust? That gets us into the statistical scientific question on how one does it.

I am not an expert in survey sampling, not by any means. But the impression I have is that we have these different methodologies out there, and each of them makes various assumptions; each of them has certain goals. My reading of some of the literature on this is that either the assumptions hold or they do not hold, or there are certain goals that might just be "this is what I would like" as opposed to any scientific support for wanting that particular goal one way or the other. Maybe there is scientific support.

So what I would like to see is that the assumptions for the different types of methods—the models, I guess—what are the weaknesses and the strengths of each model? Are some of the assumptions upon which they are based reasonable ones that should be there, or it would be nice if they were there, but it does not really matter? Basically, evaluate it scientifically each way.

That is, as I said, my reading of some of this battle, that it comes down to subjective opinions on assumptions. It is not as simple as that, but I would like to see that sort of looked at—and maybe even to focus on one particular one, although I do not think you should focus on any one in particular. I have the impression that there is this balance between bias and variance. I would have liked to see a discussion yesterday on how the different methods help one at the expense of the other, and to what extent.

**DR. NORWOOD**: I have to tell you, I have never heard a statistician tell me that there is a balance between variance and bias. Bias, I was always told by all the statisticians at the BLS [Bureau of Labor Statistics], was something we could not have. That is a different kind of question.

**DR. BAILAR**: Anyway, I have one other point. Maybe in addition to all of that—again, the impression I have is that if you are looking at the block level, the variances are wild, but as you move up in geographical area—

**DR. NORWOOD**: That is a different kind of question.

**DR. BAILAR**: That is right, but I think this is a question worth looking at: how are the data used? It seems to me, if they are never used at the block level, focusing the conclusions and the discussion on the block level as the end is the wrong focus. I think we need to look at that issue as well: what is used, for what purpose?

**DR. LITTLE**: I agree very much with Lynne's second point about the uses to which the census is being put. I do not think we have talked enough about that. I do not feel that I have enough of a grasp of the different objectives.

On the first point, though, I think, in some sense, I view it as being a little bit more complicated than that. It is not a question of the relative merits of alternative models, really. If you are just viewing it from that point of view, it seems to me the model that says that you do not do any adjustment is clearly an inferior model to a model that says you have to do some adjustment, because there is no doubt that there are differential undercounts across these groups.

The issue is that, to fit this more complicated model, you need additional information, the information from the PES. The question is, is the quality of the information you are getting from the PES sufficient to push so that the estimate from that model ends up being better than the model without adjustment? So it is not a question of which is the better model. The question really relates to the quality of the A.C.E. data. There are basically two assumptions that are central: the zero-correlation bias assumption (which is secondary from my point of view, because even if it is violated, you are still going in the right direction by doing an adjustment); then the matching-error issue, the response-error problem, which is the real key, because that is the thing that is very hard to evaluate, and potentially could push the adjustment in a direction that makes it worse than the unadjusted estimate.

So I do not think it is a question of just comparing models; I think it is a question of whether you have quality of data to fit the better model.

**DR. BILLARD**: We are not disagreeing. Your comments just fit in with how you might do what I was saying.

**DR. YLVISAKER**: It is all right, I think, to think about the census as one model within many models. But, as near as I understand the Supreme Court, it is a very distinguished model because it will be used, and people know that it will be used, at least for certain circumstances, so we are talking about the other.

Far be it from me to bring up loss functions and things like that, with the people up on the stage present already. But getting away from the technical issues, what always assaults me is the "what if?" Suppose that we say we will do adjustment, then what are the problems? The problems I have looked at more from that direction, because I think there have been plenty of people speaking from the other direction. So I will only address that particular kind of thing.

What worries me there is that the adjustment will show up ultimately to have defects that are not really what one would like to see.

**DR. NORWOOD**: If I might interrupt you, on the other side, there may be defects; there may not be. I hope that John Thompson does a superb job.

**DR. YLVISAKER**: I understand the census is imperfect.

**DR. NORWOOD**: We just do not know about them. But then how do we justify whatever it is that we decide? That is what is bugging me, really. I do not know how to explain and justify to these reviewers, or even to myself, which way to go.

**DR. YLVISAKER**: If I may just pursue one point or so in the direction that worries me, from a technical standpoint, what could happen with adjustment. I think it did happen in the case of 1990, although one does not know. But there is more information this time around. That is the problem of something like

Englewood. We were told yesterday that, really, we are interested in numeric, and not distributive, accuracy. But I think most people are not going to look at absolute numbers and stop. We have 30,000 people in our community who are going to compare their community with the next community. At the present time, to my understanding, they will also know response rates for these communities, which complicates the problem. In 1992 or so, the mayor of Englewood said, "We are the most undercounted city in the country. My gosh, I told everybody to go out and be counted, and still we are the most undercounted city in the United States." There was one PES block there. It was not at all clear that they were, but adjustment said that they were because they were minority.

If this happens again this year, and Englewood has done a very fine job with their community program and so on, and has a very high response rate, and they are still the most undercounted—these things can happen. I am not saying that this is any kind of decision. It is a "what if?" As I say, I concentrate on this.

The other one is rather more benign in some sense. That is, the Census Bureau has not solved your problem for you because we have not adjusted. Other people can speak to that. I have only spoken in one direction here.

But it seems to me that these are the real issues. Technical issues go into this makeup, how likely these things are to come about. But these are rather more global and more policy-oriented than, I think, scientific.

**DR. NORWOOD**: But suppose we know that Englewood or someplace like that has quite a low response rate—or Alaska (in the absence of the director's visit) has one of the lowest in the country. Suppose we know that, for that area, even if it is a broad area, like a state. Don't we have to look at what that means, really, in the census, and maybe decide that that is okay, that that is the way it is, because if you did something else you would make it worse? I do not know. The more I learn about this, the more confused I become, I have to tell you, about sorting these things out.

**DR. YLVISAKER**: I am confused as well. But there are two new things going on this year. The multiple-race problem is a very large issue that is going on. When we go to compare demographic estimates, it is no longer a simple story, as it has been in the past. That is going on, certainly.

The fact that response rates are going to be well known is another feature that I understand is going to be available this year. I do not know what the response rate in Englewood was—I have not the slightest idea—from 1990. Apparently, I will know in 2000 much more about this kind of thing. There will be much more inspection of what is going on than there has been.

**DR. NORWOOD**: I know. So what worries me is that people will say, what is the matter with the panel? All this material was available. How did you evaluate it?

**DR. SPENCER**: You want to know the relation between the indicators and the accuracy. When you talk about the response rate, if it is the response rate to the census mailout, if that is what you are talking about, I do not know what that implies for accuracy. If it is a low response rate, it means it costs more, but I do not know what it means in terms of accuracy.

**DR. NORWOOD**: How about if we substitute "return rate"?

**DR. YLVISAKER**: It is not going to change.

**DR. SPENCER**: Then there are personal visits to the households. When I talked about the need for judgment, I think there is inevitably going to be a gray area, and we hope that we can make that as small as possible. There are certain things where almost anybody would say, if it looks like this, we are in trouble.

Is Greg Robinson here? Greg, the sex ratio for blacks in some age groups dropped as low as 0.7. Do I remember it right?

**DR. ROBINSON**: Yes, the sex ratios for ages 25 to 50 or so were between 0.8 and 0.85.

**DR. SPENCER**: And demographic analysis shows?

**DR. ROBINSON**: About 0.9 to 0.95.

**DR. SPENCER**: Okay. So you could imagine, if the census sex ratios for blacks in these age groups dropped to something like 0.5 in the census, you would say, we have a disaster here; this looks very strange. You have to worry about the race reporting. You are really suspicious. Then you look and see what happens with A.C.E. If that is also 0.5, you say, now we are really in trouble.

But these are the sorts of things that you want to look at. I think you can imagine scenarios ahead of time where you would feel extremely uncomfortable using the census. You can imagine if you have missing data rates in A.C.E. that are 20 percent or 30 percent—just terrible. There is no reason to expect these. But if you had them, you would say, I do not see how we can use this. You can try to imagine scenarios where you would go one way, where you would go the other way, and you see where the indicators come out. Again, you want the indicators because you have some sense of—if the indicator is in a certain region, it tells you something about the accuracy.

**DR. MULRY**: One thing that you will have this year that you have not had in the past—and you will have it early—is the address list matching. Whenever they have done housing unit matching, it has always been a couple of years later. I think that will give you an early indicator of whether the MAF is very good or not. If the housing unit match finds a lot of duplication, you know that there are a lot of extra forms being mailed out. If there is a lot of undercoverage of housing units, that would tell you a little bit about what to expect.

**DR. BROWN**: Do we have a comparative standard from 1990 with the housing match similar enough so that we can learn from that what we would hope to find?

**DR. MULRY**: There was a housing unit coverage study done. It was done in 1992 or something. Howard would know.

**DR. HOGAN**: There was a housing unit study done in 1990, very similar to what we plan in 2000. The bigdifference in terms of this panel is that it was done after the fact, based on the final—which we will do again, but the study that Mary is referring to will be done based on the early census file. So to the extent that the census process makes major improvements in the list, that will not be picked up by the early readings. But, otherwise, methodologically, I think they are quite similar.

**DR. MULRY**: That would tell you about the foundation, whether you have a good foundation, to begin with, to build on. If the foundation has problems, then the rest of your operation can only compensate so much.

**DR. FIENBERG**: Bruce was referring to something like sex ratio, which takes returns from the census which reflect net error. The thing that you want to get a handle on is the two components of error. You want to know about both omissions and erroneous enumerations. Erroneous enumerations have a number of different characteristics. Mary was pointing to the geographic issue. There is actually a wonderful table in one of the reports that we got in the agenda book that points out how concentrated the geographic errors were [Navarro, 2000]. You can see how they propagate through the PES.

That tells you something about the census. That is real information about the census. It is not subject to the kinds of biases, in many senses, that we spent a lot of time discussing yesterday. So I think that is a targeted issue.

In 1990, there was an interesting set of analyses done that looked at the quality of the data over the course of the year, as you moved away from April 1. You can find it buried in the secretary's report as a set of analyses that was an appendix to one of the secretary's advisory group subsets. It is an empirical analysis. Actually, you could try to get that information. It strikes me as being very important. What it and auxiliary information tell you about is the other features of erroneous enumerations that are different from the geographic errors. You get information on duplicates, on the successively poor quality of things that spreads out over time. That is one of the ways to get a serious check on the quality of the census itself. If you do not get that deterioration, then you have eliminated a major source of problems. Then, if the net tells you that things are in good shape, then suddenly you have very strong information about the census itself.

So I think you have to combine a number of these. My guess is that there are some other ones. But that one is a clear indicator of quality of the census, and it targets one of the components of error, as opposed to omissions, which is what we talk about but do not mean.

**DR. EDDY**: I want to identify the table that Steve referred to, because I found it very instructive myself. It really does go to what Mary was talking about. It is in the memo Q-18 by Alfredo Navarro. They are Tables 2 and 3, showing that the geocoding errors are extremely concentrated—there are fewer than 700 block clusters with problems—whereas the household non-matches are spread out substantially more. Sixty percent of the block clusters contain household non-matches. So there are two sides to that issue. It is really very interesting and instructive.

**DR. NORWOOD**: But you had a very small PES. Any other comments on this?

**DR. STARK**: I agree with a lot of what has been said and disagree with some of it. I certainly do not think we would trust A.C.E. if it did not agree, at least roughly, with demographic analysis. I think even the believers would say that. So I think demographic analysis is a very important touchstone, with deference to what Barbara Bailar said.

**DR. NORWOOD**: If I may say so, I think there are believers all over. The question is, how do they get together?

**DR. STARK**: The other thing, and what I think is really crucial, is that what allowed people to discover that at least 60 percent, and possibly as much as 80 percent, of the 1990 adjustment was bias was, first of all, time. It took something on the order of 18 months. It was in 1992 when this was eventually turned up.

Secondly, there was a very complete set of evaluation studies that were done. I am concerned that if a comparably complete set of evaluation studies are not done this time around, then if the decision to adjust is made and an adjustment is released, we will never have the tools to evaluate whether the adjustment improved or decreased the accuracy.

**DR. NORWOOD**: Clearly, evaluation studies are very important. But from my perspective, since I do not expect to spend the rest of my life working on this issue—and I think the rest of our panel members feel that we need to finish this within a reasonable amount of time—it seems to me that our panel was appointed in order to develop an evaluation sooner rather than later. I do not know quite what "sooner" means. That depends, really, on how much information is available, and so on. Obviously, we cannot report until the data are issued and until we have sufficient time to consider them. But I certainly do not see this in terms of a matter of years.

I think the evaluation studies are going to take some time. I do not see that we, at least— other people can, and certainly the profession can go on for years discussing these things—we do not have the luxury of waiting to get all of the evaluation reports. We really should be reporting, hopefully, before that. I think everyone on the panel would agree with that.

So what I am really searching for are ways to help us to look at both sides of this issue. There are some people who say, you have to look at the use. I have had enough experience with putting out data to know that, of course, that is important. There is no question about that. The problem that the whole statistical system has is that there are multiple uses for almost every statistical indicator. I used to go up to the Hill to talk to people, to congressmen and senators, about the use of the CPI [Consumer Price Index] and say to them, the CPI was not constituted for this purpose; you are making a mistake using it. But they used it anyhow. That is all there was. It had a kind of aura about it, as the census does.

One person might say that it is the seats. But the apportionment at the state level—that is no longer an issue at the moment on this thing. So then it gets down to political districts maybe. But then there are some people who say they want to know about this housing project. They are down here; the political district is up here somewhere, where you are taking bunches of things and moving them around. I might tell you that I am much more interested in the population counts that go into things like the Current Population Survey and a series of other surveys, which are going to carry this country through for economic policy for the next decade. For that, you would probably use something still different. You would certainly want to have the best possible demographic data you could have, and you would want to have the best possible count you could have for the survey designs.

I do not know how to juggle these various things. I guess I just dump it in Ken's lap.

**DR. SPENCER**: I have spent a lot of time studying this problem, and you cannot do it. The problem is not only multiple uses; it is that any single use—what you are interested in is not just the use; you want to know what the consequences are of errors in the data, and then what the losses are, in some sense (money or otherwise), associated with changes in the uses because you have errors in the data. That is just too tough to do, even for a single use.

**DR. NORWOOD**: But, on the other hand—I should not be dominating this, and I will turn to other people—I have always felt that, if you are a statistical agency producing something, you should not look at the effects of its use, because sometimes it is used incorrectly or it is used for things that you did not really create it for. You have to do the best possible job you can. So you want to know about errors.

**DR. SPENCER**: Of course you do. I think what you have to do is set up a conventional loss function. It is the kind of thing we study in statistical theory. The loss is minimized when there is no error. Otherwise, we are in grave trouble, even though the uses might be better with certain patterns of error in some cases. We have all come across those situations.

So we set up a criterion that says we seek to minimize error. Then there are questions about errors, consequential and others. At what geographic level do we worry about very large errors being catastrophic? Do we go to a quadratic or do we go to something that does not grow as fast? Ultimately, there is some choice there. That is where I think you have to go.

**DR. BRADBURN:** Let me see if I can summarize what I have heard in this discussion. We all start, I think, with a very simple model that our observations are made up of. There is the true value, and there is some measurement error of various sorts. In the census, we have whatever the true value is, plus the various measurement errors that occur. In the PES, we have two sources of error, the sampling error plus the measurement error. We assume the truth is the same in both cases. The argument that is being proposed by the PES is that the improvement in the reduction of measurement error offsets whatever sampling error there is. Obviously, as you aggregate up, you get better things.

Our problem, the way Janet phrased it at the beginning, is, how do you decide that the measurement error in the census for a particular—at what level? You can talk about the whole census. You can talk about tracts. You can talk about blocks. You can talk about individual housing units. How do you decide that the measurement error is, in fact, reduced in the PES sufficiently to offset the sampling error, as compared with the census?

The problem, of course, is that in many situations of this sort that we face in other measurements, we have something we take as the gold standard. Eventually, the census, adjusted or unadjusted, for a lot of other uses, like those of us who draw samples to do other things and so forth, is taken as the gold standard. There is a whole world out there that adjusts their things to the census, whatever they are. That is one use that we have not sort of talked about.

So how do we decide? What I have been hearing is that you look at the process by which—one thing is, you look at how it was done. You look at all kinds of indicators of how well the process was carried out. Some of the things are evaluations of the actual census process. Some of the things, of which I think yesterday there was some more pressure on than is usual in cases, perhaps pay more attention to, actually, the measurement error in the PES, really demonstrating that that is reduced sufficiently to offset the measurement errors in the census.

The other one is appeal to some third standard, which is demographic analysis—a third element that I think I have heard here.

So we essentially have two kinds of things to be looking at. One is a third standard, estimate of the truth, which would be demographic analysis. It has the kinds of problems that—that is fine at the big levels, but you get down to the lower levels and it does not do too well, although perhaps sex ratios and things like that might be one way to get it at least [at] somewhat lower levels.

The other is all of these things that we have talked about, what Bruce called indicators of quality. It is basically putting our money on saying, this is the way to do it, and then how well it was actually carried out. I think, in terms of the realities of when the Bureau has to report numbers, the big challenge is to get those indicators or evaluations out soon enough—whether out or in and so forth, at least for the Bureau people who have to make the decision—fast enough that one can make that decision. Was the PES carried out sufficiently well in terms of the processes that we can say—of course, we will not know whether we are really reducing the measurement error. Since we do not know the truth, we cannot subtract off the measurement error to get it, or vice versa. So we are putting our money, essentially, on the process. It is the documentation of the process and the evaluation of the process internally that we have to use, and then externally is the demographic analysis.

That seems to me the sum total of the kind of thing. That is different from the—I think the model issues that we were hearing a lot about yesterday, and so forth, and disagreements, to my mind, are more directed to the question of having decided—if the decision was that, oh, yes, the measurement error really is reduced in the PES, now what do we do with it? That is, how do we take those numbers and combine in a way, then, to adjust the—at what level, block, whatever, and so forth. That is a different sort of issue, it seems to me, than how you come to decide which set of numbers you are going to go with as the best estimate for this block or unit, assuming you are building it up from blocks.

The other question, which I do not think we have addressed—and I do not know whether the panel is in any position to address it—obviously, what I get out of this is that a lot of the trouble or the disagreement comes from the fact that we want to apply the same thing to everything. One of the things that is quite clear is that the quality with which different blocks are measured is quite different—or different tracts, or whatever. You can talk about the quality of the census as a whole, but that is an aggregate of a lot of things, some of which are perfect, some of which are overcounted, some of which are undercounted, and so on. Ideally, in my mind, what you would like is, for every statistic, be it a block or a tract, an

estimate and some estimate of what the error is. It would be different—we think the error is pretty big on this one; it is not very big on this one. That is probably not practical, at least in the short run.

One of the things that always comes back to me is that after the 1990 census I was asked by—the Chicago mayor's office did not like the numbers they thought were going to come out, and so they were asking us, what can you do? So I began giving them some lecture like this, and they cut through this and said, "We don't care about all that. The population of Chicago is what the Census Bureau says it is. That's the truth, from their point of view. We want that number to be as big as possible." They did not care what the truth was; they just wanted it to be big, and bigger than they thought that—but the reality for a lot of the world is, ultimately, the truth is whatever the Census Bureau says it is. That is a very weighty responsibility, it seems to me. I am sure the Bureau knows that, and they take it enormously seriously. But in our evaluations, we have to take that very seriously.

**DR. NORWOOD**: Bruce, how did Canada get out of all this?

**MR. PETRIE**: I am not sure we have gotten out of it. We are just not quite there yet. We have simply taken the position that we are going to try to do the best census we can, and the census results are the census results. We will do an adjustment of the census for purposes of producing our population estimates. Those data are the ones that are taken down to provincial and, in some cases, some large sub-provincial areas. We have convinced the people who use the data that the census information is fine for some purposes. We cannot do a good adjustment based on the kind of evaluation and coverage studies that we do to make low-level adjustments. We do not know enough about the kinds of people we miss or where they are located within the urban areas that are part of the enumeration program. We simply do not have, based on our coverage studies, good enough information to carry it down. We think it is good enough to adjust at the larger levels, and we do that. Those estimates, we have convinced our users, are the best ones to use for distributing funds and the other major uses based on census data.

It does create an awkward problem. I remember, for the 1996 census, standing in front of the national news TV cameras trying to clearly say why our population estimates, which we had just released, showed that the country had finally gotten to 30 million population, and we were releasing the latest census results that showed 29 million. It took a bit of explaining. That was the most awkward part of releasing the census results. From there on, it was fairly easy.

But I think the important issue is dealing with the uses of the census data. We focus much more on the census content in terms of the information, other than the count information. That gets a lot less emphasis here, I see, for reasons that are obvious.

**DR. NORWOOD**: I do not know whether they are so obvious. I think one of the important parts of the work of this panel should be to evaluate the quality of the data that are collected, especially on the long form. But if you are going to take a long time getting it out, I am not sure we will still be in business at that point.

I do remember having discussions with the Congress, not on the census, but on local-area unemployment rates, which were not very good. They were full of various kinds of variance, but also other kinds of problems, because they were, in those days, based upon unemployment insurance data, and when we did a survey of the UI [unemployment insurance] offices, we found that they were—these are administrative data, which are, of course, supposed to be perfect, but the administrative data were all over the lot in terms of what they were collecting.

I went up to the Hill and talked to congressmen about error in local data. For a while, we brought up with us variance information that we had produced for each of these areas. Unfortunately, they could not have cared less. They knew what they wanted. They wanted either higher or lower—it did not matter what it was; that is what they wanted, and they did not want to know anything about any of the rest of it.

It is interesting to me that the U.K. [United Kingdom] is moving toward a so-called one-number census, and the Australians seem to adjust their numbers without having a lot of controversy. But we have had controversy, according to our historian statisticians, Stephen [Fienberg] and Margo Anderson, since George Washington and Thomas Jefferson, who wanted a number, and the number they wanted should be high enough so that they could prove to the French that this country was really growing and was going to be much more important. I think, almost since then, we have had this kind of controversy. Some days, I think it is because we have a fairly low level of statistical intelligence in the population. But that is a different kind of question.

**DR. HABERMANN**: Not from a statistics department.

Let me try to follow up a little bit on what Janet just said. It seems to me that the Census Bureau has three tasks, from listening to this: one, to provide the best possible set of numbers; two, to be perceived by reasonable people, who may disagree, that they have done a professional job in doing that; and three, to bullet-proof as much as possible against people who are unreasonable. The last category is the one Janet just talked about. You cannot change their minds, but you do not want to provide any crevices, nooks, or crannies for them to pry into to give aid and comfort when it really should not be there.

The best the Census Bureau can do is an approximate answer. I do not know what the gold standard is. The best the Census Bureau can ever do is an approximate answer to this very important question.

What I think the panel can do is help in all three areas—not in a political way; I do not mean that. But it can, through its expert advice, help the Census Bureau produce that best possible set of numbers, that approximate answer. It gets back, I think, to what Rod was saying, and Alan, the other day. We are not looking at a model that is perfect. By "model" here, I mean the whole process that is used to produce the approximate answer in the end. There is no perfect model. The question the panel has to wrestle with is, are there things we can do that are better than other things? What are the criteria for that? What is the professional judgment that can be used? What is our evaluation of what the Census Bureau is doing? How can we help them in all those three areas—and then help, of course,

the whole federal system. It is not just the Census Bureau; it is everybody in the federal system.

**DR. NORWOOD**: I agree with all that.

**DR. LITTLE**: I agree with that. I think there is this tendency to focus on the right answer. But from a scientific point of view, clearly there is no right answer here. I think part of what we could do should be viewed more as kind of a sensitivity analysis, where you present a range of answers for a range of sensible models and assess the size of that range. In some sense, you can get a lot of information from the PES or A.C.E., just in terms of comparing the unadjusted versus the adjusted answers, and how different they are. I think that is a pretty good indicator of the quality of the census, and does not necessarily require us to say that this is the best, rather than the other, particularly when we do not have a scientific basis, necessarily, for making that decision.

That may not be what people want, but it seems to me that that is the scientific thing to do.

**DR. BELL**: I would like to pose a question to the invited experts. We need to figure out what to look at. Based upon what you have heard about the A.C.E. and the DSE, what do you feel are the major sources of bias that we and the Census Bureau need to be cognizant of? I guess there is a corresponding question that should also be asked, which is, what are the major sources of bias for the basic census that those need to be balanced against?

**DR. BAILAR**: I look at the census and I am worried about a possible overcount this time, because of the multiple ways people can respond. It seems to me that is one thing people have to be very careful about. In spite of all the increased emphasis on publicity and trying new things to get more publicity out—just because people know there is a census does not mean they cooperate with it. So I do think that the mail return rates are going to be a good indicator of whether you are going to have cooperation. It does not matter if you go out knocking on doors in those areas. It is still a good indication of where there is an undercount.

On the PES side, I found yesterday that there has been a lot of research at the Census Bureau. A lot of things that we used to worry about are not as worrisome anymore. I think Howard and his staff have done a lot of work to shore a lot of these things up. Twenty years ago, it was the matching, and we were so worried about the matching. The Bureau itself decided that we did not want to make the adjustment. I think the matching is under much better control, except for a few groups, where I think we still have some questions. There I am concerned about the movers.

Howard, I listened to you and Eli Marks talk for what seemed like years about all these different ways of dealing with movers. Now you are trying a third way. I am still concerned about it, because I do think that movers are a lot of the problem. We know that they get undercounted.

What I heard yesterday about sharpening the definitions for the post-strata is very heartening. I like the idea of using the mail return rate, and I do think that will make some differences there. I would be very interested in seeing your analyses where you run these things with the 1990 results.

So I guess it is those major things that worry me.

**DR. STARK:** I am going to try, off the top of my head, to catalogue some of the biases in the PES, or in A.C.E. There are fundamental assumptions of the method, like the synthetic assumption, which essentially is assuming that response rate is a function of which post-stratum you are in, regardless of geography, except to the extent that the post-strata label geography. There is correlation bias, fifth cell, whatever you want to call that.

But what seemed in 1990 to be the largest source of bias were bad data and things in the matching that have to do with ad hoc decisions that you have to make to implement the algorithm. I will give you a couple of examples of that.

One is the search region that you use for matching. You have to decide when to stop. My understanding is that for the 1986 L.A. PES rehearsal, of those individuals who were not matched in their immediate block, 38 percent matched six or more blocks away. That is a large fraction. My understanding also is that in 1990, had there not been a search of a ring of blocks around the target block, the estimated undercount would have been double what it was estimated to be. So there is a fair amount of sensitivity to the search region. There are indeed the tradeoffs that Howard Hogan mentioned yesterday, but I do not really know how one calibrates this in a way that is not, in some sense, ad hoc.

The next piece where there is an ad hoc decision that needs to be made to implement the method is, what do you do with the cases where there is not enough information to determine whether or not there is a match? Last time there was this hierarchical logistic regression model. This time we have boxes, which I think are an improvement. But to give you an idea of the sensitivity to what you are doing, the estimate from 1990 is that, if you had distributed these cases where there was insufficient information to determine match status as adversely as possible, you could have gotten an estimated undercount of 9 million, and if you had distributed them as favorably as possible, you could have gotten an estimated overcount of 1 million. So there is a lot of sensitivity to exactly how you treat these cases where there is an extremely limited amount of information.

Some of these things, I believe, are going to be exacerbated in 2000 by some of the new things in the census. I am feeling more comfortable about OCR, having listened to Howard yesterday. But I am still kind of concerned about what that does. I am concerned about the multiple modes of propagating forms and the impact of the computer unduplication and its accuracy on the accuracy of adjustment. All of these things are going to affect the ability to match or the accuracy of matches.

Those are the things that come to my head immediately as biases in the adjustment procedure.

**DR. YLVISAKER**: I guess, if it is online things in time for a timely report, only so many statistics will show up. I believe in response rates. There is going to be a lot more information about that around. That ought to be taken into account. It does raise the duplication issue a lot. I think excellent points were made yesterday about movers, movers across post-stratum lines. I wonder to what degree this can be made manifest in a timely way.

One thing that I noticed from yesterday's discussion—Bill Eddy pointed out that in 25 percent of the 5,180 blocks from the PES in 1990, the matches exceeded the enumerations, which said that the search area, presumably, had a lot to say about things.

In going through that file and similar files, which will be available, presumably, as soon as A.C.E. is completed, and long before A.C.E. is processed, one can inspect blocks. If you go through the file, you see an awful lot of strange blocks. I think 140 blocks were subsequently redone and reexamined by 1992. I think, to some degree, one can find blocks like this by looking at local estimates and noticing strange patterns. I think some of that ought to be done online, to make sure that blocks that look totally out-of-whack, and which will have large influences on particular post-stratum estimates, are in fact firmed up and understood.

**DR. FIENBERG**: I want to begin with the census. I guess I share a number of the concerns that Barbara raised. I am very much concerned by what I would characterize as differential overcounting. It is not simply that it is overcounting, but we now have, as a consequence of the visibility of the issues surrounding Census 2000, committees in localities where they have never existed before, every one of which is designing an attempt to make its number bigger than everybody else's, in a relative sense.

While we know that they all cannot succeed—there is a theorem about that— what the implications of all those programs are I do not think we have the foggiest notion, except that they are going to lead to overcounts. We are going to get erroneous enumerations as a consequence. We do not know how many. We do not know how they are going to manifest themselves. We do not know that we have procedures in place to catch them early enough to understand the implications.

So if I had to express any major concern about the census, other than the traditional ones— like very low mail returns that will be lower than they were in 1990, and the only question is how much lower—then I think a focus in that direction is my major concern.

I come to the PES. It is true that we have a catalogue of things where the assumptions are not satisfied; none of the assumptions in any of the basic methodology are satisfied, including the "unbiasedness" of the sample estimates, of the sampling quantities, when you weight them in the appropriate way. Howard and his staff have told us how to weight them.

Then the question is, how serious are the departures? Phil has catalogued a number of concerns. He keeps on making reference to the net census error. I simply want to remind you—and I will again before we are done—that it is not net census error that the comparison for errors should be to, but to gross error. If you add the omissions and erroneous enumerations together, then the fraction of errors that occurred in the PES relative to that total is relatively small.

That does not mean that there are not concerns there. But I took the evidence that we found in [the] existence of a large number of blocks where the number of matches exceeded either the enumeration count or the enumeration adjusted downward for erroneous enumerations, as indicating that perhaps the search was already too far, rather than there being something wrong with the search and you

had to go further. So I think that evidence points in a different direction than the interpretation that we just heard. I was actually struck by the revision as addressing some of what I thought were fundamental concerns there, in a way that I was not sure happened in 1990.

**DR. MULRY**: One thing I would look at is the sampling weights. We have heard about all these different operations that are trying to reduce the range or keep the weights within a reasonable range, but then again, there are all these other— for example, the missing-data thing is now weight adjustment, and subsampling weight adjustments, and so on. So that will probably drive them back up. I always thought that some of the odd things that made people uncomfortable about 1990 could have been caused by the huge range of the sampling weights. When we were looking at the evaluation data, those had very large weights. Some of the time, the design effect and the error measurements were 22 [reference not clear]. It would always be, not a clustering of errors, but one error in a block that had a huge weight.

If the sampling weights are within a not-too-huge range, then maybe some of these things do not happen. Or if you see odd things, look at the weights.

I just have one final thing about the search area. Keep in mind when you examine that, that this housing unit-list matching at the beginning—they are searching in surrounding blocks then—that is going to find a lot of geocoding error. Previously when they searched surrounding blocks, they did not have the housing unit matching. They were just looking for people. They looked at all the names. They had to look at that, as opposed to looking at whether the housing unit was geocoding it. It just was not as fine a system. With computerization of everything, you can be more precise on that. Just keep that in mind when you look at the search areas.

**DR. EDDY**: To follow up on Mary's point, I would like Howard to tell us a bit more about the targeted extended search that is going to be done in A.C.E. If I understand correctly, there still is going to be a search in neighboring blocks, despite the fact that the housing units have been matched previously at the individual level.

**DR. HOGAN**: You have to think of this in two steps, one of which is, what is the universe that we are going to be trying to find duplicates for, erroneous enumerations for, or find matches for? Then we are going to be doing this on a sample basis.

In essence, we have the sample block and the cluster of which it is part, which is all we search. If we have either whole-household non-matches or whole-households in the cluster that we cannot find in the PES, then, at least in the universe, we are going to search for those in one ring of surrounding blocks. In fact, we will be doing this on a sample basis. Where there are a lot of them, it will be a certainty sample. Where there are none of them, there will not be a sample. In the middle, there will be a weighted sample with what you find weighted up based on the sampling.

So in terms of expected value, it is quite similar to 1990, with the exception that if just one person moved across the street and happened to get counted across

the street, as opposed to the whole household, we are not going to look for that, either as an erroneous enumeration or as a miss. But if the whole household is mis-geocoded, we are going to look for it. In 1990, that was the majority of it. Let me say, some of the uses of the 1990 data to study search-area imbalances are misleading. It is a fairly technical thing, but in terms of the production use of how things were coded in 1990, I think it worked. There was not a bias in terms of the intended use of those data. We did this ourselves. When we tried to use those same data for other uses, specifically studying how the estimate might change with changes in the search area, there is a flaw in the data—it is a fairly technical one, and I can explain it to anyone at the break—that makes those data inappropriate for that use, or at least potentially misleading.

**DR. FAY**: I am going to add a footnote to what Howard said. I can be addressed at the break, too. There is an additional question about trying to use the 1990 data on search area. We would get, in a sense, different results in a PES-B context compared to PES-A, because movers often provide only an approximate indication of their Census Day address under PES-B. In 1990, there was a real need to make an extended search just because of imprecise addresses. That also affects, however, interpretation of the data, which I do not think I have seen broken out by mover/non-mover.

**DR. NORWOOD**: I think it is time for a break.

## COUNTS VERSUS SHARES

**DR. NORWOOD**: I think we have had suggestions that will be very helpful to us in helping us to plan where we go from here. If you will all bear with me, I have two more questions.

There is a lot of discussion in the literature, and the director commented at the beginning, about the whole issue of counts and shares. In a way, that gets to part of the discussion that we had earlier: what are the uses of the census? As I think of the history of all this, I think people were thinking clearly about counts. Certainly, as I said earlier, Washington and Jefferson wanted to have the largest number of people in the country possible. If I put on my old hat as commissioner of labor statistics, if I were to look at this from the perspective of the Bureau of Labor Statistics, I certainly would want to have population counts that were as good as they possibly could be, especially for black men in the labor force, and, to the extent possible, Hispanics and a lot of other people.

On the other hand, we do have the allocation of seats and of districts. I think the allocation of seats for this census is off the table, because the Supreme Court has ruled—my reading of that, by the way, is that they are really more interested in counts than in shares, but there can be room for disagreement on that.

But I would like to know how people here feel about that issue. Clearly, you have to have at least some counts in order to have shares. So it is not a question of exclusivity. I guess it is a question of emphasis. I open that issue up for discussion.

**DR. FIENBERG**: If you think about how we have constructed the dual-systems estimation process and actually look at the derivation of the original formulas

as estimators, they are formulas for counts. It is a simple fact, going back 105 years. The goal is to get an estimator of a count. What we are doing in the DSE activity is, in fact, adapting that with sampling-based weights and a variety of bells and whistles. So the basic methodology has as its estimators things that were individually designed to be counts.

If you take a theoretical perspective about what to do, there are ways to get different estimators. Larry is as good at thinking about that as anybody in the world that I know. However, nobody has ever done that. The formulas that everybody has worked from are formulas for counts.

I think implicitly there has always been an expectation that if you actually improved counts a lot and made the data better, you would probably do better for distributive accuracy. The empirical evidence that we had from 1990, I think, supported that. But the formulas were not designed for that.

So I think it is a great question, but I hasten to add that nobody has ever sat and proved a theorem for what a good estimator is when the goal is distribution across multiple units. There is no theorem that I know, and I think I know the literature.

**DR. BROWN**: All right, I will weigh in. Of course, you are right. The formulas are formulas for counts. But that is, I think, not determinative of anything. That is a judgment about what statisticians can do. We can improve counts. But the real question is, are we improving counts in order to improve estimates of shares, or are we improving counts in order to improve estimates of counts?

**DR. FIENBERG**: I just wanted to be explicit about what the methods were designed to do.

**DR. BROWN**: Right. I know of no method that attacks directly shares without working through counts. But I do not think that that is definitive of any—

**DR. FIENBERG**: I also agree.

**DR. BROWN**: Distribution is an evaluation criterion, and it differs from counts as an evaluation criterion. So we are agreed there. Then it is a question of uses of the census, and emphasis, perhaps, on evaluation of methodology rather than—

**DR. FIENBERG**: But let me go back to the starting point. If I was sitting addressing not this panel, but Ben King's panel and looking ahead to 2010, then it seems to me that there is actually a set of methodological things that could be done that would produce different estimators using the same kinds of data for the purposes of improving distributional accuracy. I do not know how they differ from the ones that we are using. If they differed a lot, then you would have a hard choice. It may turn out that they differ hardly at all. They may differ in a term and an expansion. If that were true, that would actually tell you a lot about the way in which the goals come together. We do not have that. What we have are empirical efforts at looking at this after the fact.

**DR. BROWN**: Let me make two separate kinds of comments. The more direct one has to do—I think, to some extent, we cannot really separate the issue as an issue of counts versus shares. Think of subgroups of the population. Janet has mentioned blacks. Indians are another very obvious case in terms of the point I want to make. We spend a lot of effort, in a sense, to get their count right,

but it is really more nearly, if you think about it in statistical terms, getting their share right. If we were really only interested in counts, we would spread the effort differently and improve the count of whites, who are the biggest group, more. We would get a better overall count. So, in a sense, the emphasis is on counts, but in another sense, it is on shares. We cannot really separate the issue out that way.

The other comment is a completely separate comment. It has been mentioned. I have been hoping we would have someone here who was much more authoritative than I, because I am really just passing on secondhand statements. It has been mentioned that one other way of getting count accuracy and also share accuracy would be to take the shares from the census and upweight them across the board to match demographic analysis. That is a different scheme. I have only heard about that scheme. I do not want to try to evaluate it.

**DR. FIENBERG**: May I react to that? We have no numbers that would let you evaluate that relative to the proposal on the table here. Why? Because there are lots of sources of error. Despite all of the valiant efforts of Greg and his group, they cannot even give you a variance, let alone a bias. There are big biases. They know about them. They documented them. They are the source of all my knowledge. The notion that you could do that and you could actually assess it requires that you have both bias and variance. That is what we spent two days arguing about vis-a-vis the dual-systems estimator and A.C.E. You need comparable information if you want to do that, or you are not acting as a statistician.

**DR. BROWN**: I pretty much agree with those statements, that you certainly need that. I was hoping to have someone in the audience who has done it or who would claim that it is doable. But I also need to point out, on the other side, that although we have spent a lot of time talking about dual-systems estimation, we have not yet talked about the issue of estimating the variance that is in there, which may be more tractable than the demographic problem, but is still a difficult and very much unsettled kind of issue.

I do not know if the one or two demographers we have here want to make any comments on the other side.

**DR. YLVISAKER**: [Technical problems with the room sound system; inaudible phrases indicated.] I was a little surprised yesterday to hear the emphasis on numeric versus [not clear]. I have heard two examples, I think, in the last day. Bruce's 29 million versus 30 million was very important to Canadians, and I guess Jefferson's and Washington's size of the country was of some merit. But we have to deal with scales all the time. As a statistician, I have never seen very many problems that did not involve comparison. I do not say I ran a survey and I found 40 people who favored Clinton, and stop. I have to also talk about the people who did not favor Clinton and talk about proportion. So almost everything is automatically a comparison [not clear]. The 30 million versus 29 million they are already comparing what they were last time [not clear].

The difficulty that I see here is that we have been able to attack via loss functions and various things of that nature very specific problems—that is to say, let us look at shares of states, let us look at shares of regions, things of that nature, and let us suppose that we understand the truth for these few things. But they are

very specific problems. We are talking about users who are coming from all possible directions, and my contention is that almost all of them are concerned with shares—what is my share versus this guy over here, with whom I am competing.

**DR. HOGAN**: A couple of comments on different things. First, on the demographic analysis idea, which has been floating around since even before I came to the Census Bureau, I think that might have worked very well in 1960, when the U.S. population could be conceived as sort of white and black. In the year 2000, when the undercount is white, Hispanic, Asian, American Indian, as well as black, to propose a method that has never been able to do more than black/nonblack, I think, is not addressing the undercount of 2000, but addressing the undercount of 1950, really. I think that is not a viable alternative.

On the general question of count versus shares, after the census has been run, the Census Bureau has looked at both. Everybody looks at both. The counts are important. To give you an example, it has always been very important to the mayor of Detroit. But anybody who uses the data for planning or other purposes uses counts. Shares are important. The Census Bureau has said that. The differential undercount, the fact that we count lots and lots of blacks and Hispanics, but we undercount them differentially, is a share problem, which, after the fact, we talk about. We do not pretend it is not important.

But as a planning tool—I would like people to tell me how you use shares as a planning tool. After the census has been run, you can say everything else is now fixed; did the PES or did the Parolee/Probationer or did whatever improve shares? But as a planning tool, when everything else is up in the air, how do you apply shares? How do you say, I could do this in North Dakota, but I do not know what the person in South Dakota might do? As a planning tool, that needs to be addressed.

The thing I raised yesterday let me raise again: If one really believes—which I do not—that there are large numbers of these unreached people, large numbers of the "wily trout" or fifth-cell people, hiding in unknown locations, then one can never show anything about shares. You can always say the "wily trout" could be someplace else.

**MR. BARRON**: Just to add to what Howard has said, Janet—and I do not know the literature, but this is sort of from the experience of 30 years at BLS and a couple at the Department of Commerce now—it is a question I have asked internally, and I do not mind asking this publicly, even if it reveals a profound ignorance on my part.

The question is something like this: if you make shares your ultimate goal—and the only background I have on this is a very long document issued by the Commerce Department (not by statisticians, but by others) some time ago—if you make it your ultimate goal, does that not mean that if you find a way to measure a group, like American Indians, with absolute precision, but if that count is not distributed proportionally across all geopolitical jurisdictions, you would then walk away from that improvement, if you make shares your ultimate goal?

I would submit, based on 32 years in the federal statistical system, that a government statistician simply cannot do that. Maybe you can in some other

environment, but I just do not see—and maybe I misunderstand what "shares" means. What I am talking about is not ignoring it. I am saying, if you make it your ultimate goal, the absolute goal, what does that mean operationally? At least as described in one Commerce document of a few years ago, it looks as if you would actually walk away from an improvement if it did not distribute itself across the whole nation in some proportionate way. Perhaps that is a confusion we could address.

**DR. BROWN**: I think you are right in your example, but it is not the way it often plays out. I will try to make that clear. Howard asks how to use shares as a planning tool. What I think we are doing now is a kind of hybrid method, which I actually endorse, which cannot be defined either as shares or as counts. You do use it as a planning tool. You identify groups that you think are undercounted, and you invest disproportionate effort in counting those groups. That is because you want their shares to be more nearly accurate. So when you talk about undercounted groups, in that sense, you are interested in getting their shares accurate. When I say "disproportionate effort," disproportionate to what? It is disproportionate to what you would invest if you were really interested in getting the national total correct. You spend a lot of effort on small groups. I understand they are undercounted a little more, so you should spend a little more effort in terms of the national total. But we spend a lot.

**DR. PREWITT**: For the record, I want to make it clear that I do not think that, Larry, quite reflects what we do. Anywhere we believe there may be a less-than-total count, we make an effort to go do something about that. A very good example in the current operations is the new construction program. It has nothing to do with racial minorities. That has to do with people who are going to move into areas where we think our address file is inadequate. Therefore, we are going to fix that address file, and we are going to allow them to add their addresses right after April 1. That is a population group that, because of construction of housing and the address list, we feared we would miss. Therefore, we put a procedure in place to do it.

You are quite right; we disproportionately address any population group we can imagine finding—and, indeed, it is not all racial; we worry a lot about Appalachia, white renters, we worry a lot about mobile homes, we worry a lot about any population group that we know to have characteristics that we think are likely to be counted at less than 100 percent. We actually do not sit there and say, are there some groups that are just below that margin—that is, we think we are going to miss 2 percent here and over here we are going to miss 12 percent, and we spend a lot more time on the 12 percent undercount group than the 2 percent. We actually do not do that. At least we do not have the capacity to fine-tune it that well.

This is a disproportionate attention to the groups that we believe to have the largest undercounts, absolutely. You are right. In the total aggregate, if we counted every white and we missed 12 percent of the Indians, we would end up with a larger count. You are absolutely right. Since there is concern about shares on the basis of—but the design itself is to sort of push everywhere we can to the 100 percent.

**DR. BROWN**: Actually not. For example, in terms of the design that was built into the selection of the PES sample—

**DR. PREWITT**: I am talking about the enumeration operation.

**DR. BROWN**: Ail right. But in terms of the PES design, there really was an emphasis on certain groups. The PES design could have been constructed with a goal of total numerical accuracy count, national, and it was not constructed that way. It was constructed with an emphasis on certain subgroups.

**DR. FIENBERG**: I agree totally with Larry and Don—and this goes back to the earlier exchange—that virtually every comparison is a shares issue. But Bill's example brings us back to what was in 1990 a hidden tension because of the structure of the debate, and here has become a more explicit tension, as an evaluation criterion. If you say that what you are interested in is a comparison of groups in the population, then you have one criterion for evaluating distributional accuracy in shares. If you say that we are interested in geographic accuracy, down to very low levels, for example, you have a totally different version. The question is to what extent you can meet multiple criteria of that sort simultaneously.

What we know from the PES in 1990 is that it did a lot better than the census for dealing with the demographic groups. It measured a differential undercount that nobody is willing to stand up and say is not there. Everybody agrees that it is there for blacks, it is there for Hispanics, and it is there, to a lesser (and somewhat more problematic) extent, for Asian Americans and American Indians.

Now the question, I think, we face is whether or not you can eliminate that differential, or at least essentially wipe it out as an aggregate bias, and not make anything else worse. If you could do that, then you have solved one distributional-shares problem, and you have not made the other one worse. If you could make it better, that is even an improvement, in every sense. I think it is very appropriate to ask about those, but let us not confuse the fact that these are different shares problems and that different articulations of the problem are quite different.

So Bill's articulation is absolutely right. If we set as our goal to improve geographic accuracy, we would not count the American Indian group better.

**DR. SPENCER**: The question of counts or shares gets to the issue of a loss function. I am not sure it is a scientific question. It strikes me as one of agency goals. That strikes me as political values, or statistical values in the sense of family values and not numbers. What should the goal of the census be? I think we could be given alternative vectors of errors for geographic areas or for subgroups, and we could say which vector of errors we would prefer over another one. We might be able to summarize our preferences by a mathematical function, which we can refer to as a loss function. Different individuals would have different functions. I do not think anybody's function would wind up being totally shares or totally counts. I think it would be some mix of the two. We have talked about some extreme examples, which are pretty artificial, I think.

In practical terms, when you think about the adjustment decision that the Census Bureau is going to have to make in the near future, as to whether to adjust the census counts or not, we have been talking about what information they have to go on, and what indicators. I think indicators are the right way to go, to try

to get a handle on the error in the census. You can use A.C.E. to get a handle on the error in the census and then use the indicators to get a handle on the error in A.C.E., and try to weigh the two.

How are you going to factor shares into that analysis? You are going to be lucky to be able to know, even roughly, what the indicators tell you in terms of the accuracy for counts. To figure out shares, at what geographic level, even for states, is going to be very, very difficult. I do not know how you would do it.

**DR. MULRY**: I am going to sort of echo a little bit what Bruce said. It seems to me, for an operation for the census, from a practical level, numeric accuracy makes the most sense. It would just be really weird, all the decisions you have to make if you are trying to think of distributional accuracy, all the people in the field that make these decisions. To go for the count makes the most sense. Then you start talking about adjusting. The distributional accuracy has to have more—not necessarily more weight than numeric accuracy, but has to play much more strongly into the decisions that you make about an A.C.E. adjustment.

**MR. PETRIE**: The shares-versus-counts issue is one that is simmering north of the border. It does not have much to do with the political side, but the money side of the equation. Money and politics are both of considerable interest, and money sometimes, in our case, I think wins out more than politics. Transfer payments from the federal government to the provinces are essentially based on population shares. It is a zero-sum game. So the result is that when literally hundreds of millions of dollars, or billions in some cases, end up being allocated based on some statistical formula, people tend to get interested and start looking at it. That has happened in the case of Quebec, where the fact that we did start adjusting our census as the basis for producing our population estimates resulted in some quite large impacts on provincial revenues.

Again, our approach to dealing with it has been simply to say that we are doing the best job possible and enumerating the population, getting the best possible count. We do have a method for adjusting the count. Here is the method. We will get down in the mud with you and wrestle on the details of the methodology. If you can show us a better way, we will adopt it.

So there is flexibility there. There is room to take other information into account. The point that was just raised—often, when we make a decision on a current set of estimates, we will get better information later, either in terms of better methodology or a better source of data. So a flexible method that incorporates that, that is transparent—as transparent as these things can be; it is difficult to explain to a technical crowd, let alone to politicians or journalists—our approach has been that it is very important to convince people that they have a credible census, that it is doing the best possible job of enumerating the population, and that the methods it is using to produce the population estimates are also credible and open to suggestions for improvements.

**DR. HAUSER**: I want to throw something out. It is a little bit oblique to the shares-versus-counts thing. But I wonder if anybody has any thoughts about the relationship between counting and adjustment in the census and, say, the use of census accounts, adjusted or not, and small-area intercensal population estimates

and the kind of information that is used in them. That speaks both to the issue, it seems to me, of adjustment and to the issue of evaluating the census.

I have not heard anything in all of these discussions about any sources of data, except this broad, global demographic analysis, other than the census itself and the A.C.E. I wonder, for example, whether there are useful ways to think, in terms of the goals, about the difference between small-area estimation between censuses and the use of population estimates from the previous census. In one case, you are dealing with change over time. In the other case, you are just dealing with differences. But, conceptually, I am not sure there is that much of a difference.

What conceivable role could indicators, outside this box we have been looking at, play with respect to our evaluation, either of A.C.E. or of the census itself? They are certainly used in intercensal estimates.

I would just like to throw that out and see if anybody has any thoughts.

**DR. NORWOOD**: I am waiting. Bruce?

**DR. SPENCER**: The postcensal estimates do form one other comparison for evaluating the census, at a rough level. What you do is, you take the 1990 census, and you can adjust it for undercount locally, and then take the estimates of postcensal change, which are based on vital statistics, and estimates of migration, which are, maybe, not too bad at the state level and large cities—below that, they are less good—and use that as one of the indicators that you look at for the demographic analysis kinds of comparisons.

That just answers a little part of your question.

**DR. NORWOOD**: Any other comments?

## CONCLUDING COMMENTS FROM INVITED GUESTS

**DR. NORWOOD**: I think we have reached the time now for us to sort of do a round robin and give our invited guests an opportunity to say whatever it is they would like to say as a final statement to us. We will start with Marty Wells.

**DR. WELLS**: Actually, I am from a social statistics department.

There are a couple of important points. The first one is that I think it is very helpful for data to be available. As we saw yesterday, there was some discussion about what the counts meant or did not mean. But we could not have had that discussion if the data were not available. If the Bureau could make data available as quickly as possible, it would be great.

The other thing this panel should do is get a laundry list of the evaluations that are taking place at the Census Bureau, and then have some way of evaluating those evaluations—maybe critiquing them or asking for more. It would be crucial to know what they are doing along the way.

A couple of things have come up—evaluation of movers between post-strata, evaluating the matching, the unduplication. These seem to be fairly important topics—also evaluating the missing data imputation, those methods.

I look forward to a loss function analysis of all this. That is another issue, I guess for another day. But I think that will really tell the story.

I mentioned to Larry that this panel has a tough job. It is sort of like being a theater critic and leaving before the last act. You have to make this recommendation and write a review when you really do not know how the play ends.

**DR. STARK**: I have said a lot of what I had hoped to say. I want to add one more comment about OCR and then talk about some more general things.

My understanding is that the error rate of the best handwriting recognition algorithms, OCR algorithms, is on the order of a couple of a percent, and that is for mixed upper- and lower-case letters and digits written in blue boxes that disappear from the scanner. So it is not surprising that something like 86 percent of the forms might require some rekeying.

What concerns me is that the internal uncertainty measures that the OCR algorithms produce do not correlate with the actual errors that they commit. That is, they do not know when they are in trouble. They do not flag their errors very well. This means that if the Bureau is relying on the OCR algorithm to say, "Hey, I'm in trouble, I need someone to interpret this "M"—many errors will not be flagged that way by the algorithm itself. So I am really hoping that there are careful evaluation studies of the accuracy of the OCR, and that some sample of the forms in which no errors were flagged are rekeyed, or perhaps if there is some error that is flagged on a form, the whole form is rekeyed—something, to somehow get a handle on the size of that.

**DR. EDDY**: Before you go on, could you clarify the source of your claim?

**DR. STARK**: The person I spoke with is [name not clear], who, I believe, is at AT&T. I also spoke with [name not clear] at UC-Berkeley, who knows a bit about real-time handwriting recognition.

**DR. EDDY**: I am sorry, that is not the part I was interested in. I was interested in the claim that the Census Bureau is not properly handling the flagging of errors.

**DR. STARK**: I was not claiming that they were not properly handling anything. I am saying that OCR algorithms, as a general principle, are not good at catching their own mistakes. The error flags that the algorithms raise are not well associated with the errors that the algorithm actually makes. So the algorithms do not know when they are in trouble. That is almost an axiom, because if the algorithms knew when they were having difficulty, you could write an improved algorithm that would take advantage of that to do a better job.

**DR. EDDY**: I will not disagree with that, but I want to report what, I think, John told me yesterday. I had made the statement about the 80-odd percent. It turns out it is 86 percent. Maybe it was Jay who told me. He clarified that by saying that 20 percent of the fields actually go through keying—20 percent. So the 2 percent error, I think, is—I believe that. I believe that 2 percent error. I think it is not a problem if 20 percent of the fields are being marked as errors. I think the basic problem is solved, but I certainly endorse, as you do, the evaluation of it. I just wanted it to be clear that there is not a big problem.

**DR. STARK**: My concern is that the part of the problem that is detected by the algorithm might not be the largest part of the problem because of this lack of association between the diagnostics that the algorithms produce and their actual errors. I am hoping that a human is looking at some fraction of these at random.

Moving on from that issue to more general things, my deepest fear is that there are a number of changes for 2000 that could negatively impact the accuracy of the census. I do not know whether they will or will not. That remains to be seen. Some of this has to do with the multiple modes of propagation of forms, including the Internet and "Be Counted." Some of this has to do with the data acquisition, the OCR and the new data capture methods. Some of this has to do with the new differential coverage that we expect to happen from all of the complete-count committees and so forth that are being formed.

That being said, if the census is compromised, if the accuracy is compromised, that, a priority, improves the argument in favor of adjustment. The issue then becomes—we need a way to decide whether adjustment is an improvement or not. I believe pretty strongly that in 1990 it would not have been an improvement. But a lot of the basis for that belief is the evaluation studies that were done. I am really concerned that the same sorts of evaluations of things like matching error, fabrications, and the like, happen this time again. Otherwise, we might find ourselves in a situation where the census is not as good as we had hoped, and we do not have any way of deciding whether the adjustment is an improvement.

I really urge that all the data be made available, that evaluation studies be done and be made available as well.

**DR. NORWOOD**: Thank you.

**DR. BAILAR**: I just have a few comments to add to what I have already said this morning. The first thing I would like to do is thank Howard and his staff for providing all of us with some very useful papers.

I would like to ask for some early evaluations of the census counts. What has happened with race and ethnicity? I noticed in the dress rehearsal that a lot of people did not choose multiple race, but there was not that much publicity, either, in the dress rehearsal. I think it would help a lot if we knew what the situation really was.

I would say the same thing about the multiple ways to respond. What do we know about how many people selected different methods? What do we know about the Census Bureau's ability to unduplicate those? I think that will tell us a lot about whether we are going to see a problem or not in the census itself.

I would say that I do not think the Census Bureau can do all of the evaluations and the PES that they did before in the same time frame, but it would be useful to do some very select ones that would give people the ability to have some kind of sense of whether the PES ran into major difficulties or not.

**DR. SPENCER**: Let me start with a couple of very minor things. Steve and other people talked about possible changes in the quality of the census returns, as the return date gets later. I wonder if that could be used as a post-stratification variable. In principle, it could. It might be interesting to see what it does. If people who returned their census questionnaire much later, after a month or so, where they have to be interviewed, are representing, as it were, more uncounted people with the same characteristics than people who return earlier, that could be detected. You could group people by early returners or late as a post-stratification variable and see if it made a difference. I have no idea if it will. It is just a cheap observation, since I do not have to do it.

Another one, also minor, has to do with what you mean by error. An example was, what do we mean by the correct residence for a person? This is something Howard said yesterday. It was determined in A.C.E. by the belief of the P-sample respondent, as opposed to getting what you call theological truth. Without getting into the theology, I think you are better off defining things, when you can, in terms of an objective criterion—certainly for where you were on a particular day. You may not be able to observe it with certainty, but I think you are better off defining things that way and then allowing for errors in your measurement. I would much rather see the object of the estimation defined as well as you can, and then, where you fall short of that, that is measurement error. It is not a huge point.

Another example would be, in PES-C, what you would really like to have is the number of outmovers. You have a match rate for outmovers. You would like to have the number of outmovers in each post-stratum. You are using the number of inmovers as an estimate of the number of outmovers. When you look at the error in that, you say there are two components to this error. There is my error in estimating the number of inmovers, and then there is the actual difference between the actual number of inmovers and the actual number of outmovers.

Getting down to what I think are bigger points, it is really the question of the evaluation and the real-time adjustment decision. Barbara had asked for getting some of the key evaluations done as soon as possible. I do not know what is possible and what the constraints are, but I certainly, in principle, support that idea. I think, for example, sampling variances might well be known before the adjustment decision date. That is something good to look at. Other things, I think, will depend on some of the evaluation studies, which depend on evaluation follow-up and other things. There is a much longer lead time needed for that, and they may not be available.

The question about the indicators, when you do not have these particular evaluation studies—and even when you do, because you are only getting parts of the error, and you cannot do a full loss function analysis, which shows you how different levels of errors interact, whether you are getting counts or shares. I would like to see work go ahead, people brainstorming, about what indicators would be interesting to look at, and also what the indicators tell you about possible accuracy or inaccuracy, and what the fallibility is of these indicators. We would be foolish to trust any single one of them by itself, but you want to look at the whole set of them.

Some internal indicators are things like the last resort rate, missing data rate, some of the quality assurance rates that may come out. People at the Bureau know much more about this than I ever will.

External indicators we have talked about—sex ratios from demographic analysis. We can also look at demographic analysis numbers. Bob Hauser's suggestion about looking at postcensal estimates is a very good one as well. That gets into the geographic distribution.

**DR. MULRY**: Trying to synthesize all these errors when you look at the census quality, trying to synthesize all your indicators – one goes one way, one goes the other way—is going to be your challenge, I believe. Bruce mentioned the closeout

rates and the last resort rates—all the missing data. When you do not have a self-response—one thing is someone filling out a mail return or an interview with a household member in nonresponse follow-up—when you do not have at least that, that should be an indicator of some of the quality of the census.

The other thing that I have not heard mentioned is what happens with the primary-selection algorithm. We have talked about the problems of having all these different modes of responding—the Internet and "Be Counted" forms and everything. What happens with the primary selection algorithm? How many forms do they throw out? How many times do they combine households across forms? Nobody has said what is going to be available on that, but there should be something available from the primary selection algorithm that would give you insight into your duplication problem.

As far as the A.C.E. is concerned, as I have said, I think this housing unit matching—the listing that they are doing, and then matching against the census address list—gives you a lot of insight into the quality of the foundation for the A.C.E., and also what has gone on with the census.

What would be really nice to see is the return rate for those listed blocks that the A.C.E. listed. I know you are getting them by tracts, but if you get them for those blocks, you would have what the geocoding error was going to be and some idea about the duplicate housing units and so on. Then, if you had the return rate for those blocks—if there is a lot of duplication, you could have return rates that are higher than the number of housing units listed by both operations.

I am concerned about the mover match rate. I think you ought to be very concerned about what happens there, and also the missing data rate and the non-interview rate with the A.C.E., and just generally look at the quality control of the interviewing, those results. There is supposed to be some quality control on the matching, too: how many qualify as being extreme blocks, where they look at the difference between enumeration rate and the non-match rate? How many have to be reviewed from that?

I think you will have some data that will give you some insight into what is going on. The problem is going to be synthesizing all that. I wish you a lot of luck on that.

**DR. NORWOOD**: Thanks. Don?

**DR. YLVISAKER**: Thank you. I will not have very much to add, I think, after what others have said. One thing is that the more data we are allowed to look at, the better off we are. I think that point has been made.

In terms of doing things like loss functions, from a statistical, scientific point of view, we can set up problems and take loss functions and do data analyses, and so on, and were data available, we could be doing these kinds of things a lot, and maybe gaining something in total. Some specific loss functions I think are presumptuous on our part. There are so many users and so many share problems and so many things going on that our focusing on specific things is—it is a data-analytic device and much fun and maybe a lot can be learned, but we are not solving people's problems.

The real loss functions that I am concerned about are, in fact, the Bureau's loss functions, which are very much subjective. That is, if we do this, then what are the consequences? If we do that, then what are the consequences?

The big things I see in this current census—I am very much taken with the publicity business. I realize the downside of that, but I think one is building on something for the future. I am hoping that there will be a real effect to publicity, the schools program, that sort of thing. I think this is all positive.

The other large item I see is the multi-race issue. It is the first time we have seen it. We are very new at this kind of business. I have concerns, which I have already expressed.

Things that I have heard so far that I will endorse—there is the treatment of movers. Movers across post-strata are of some concern. I am sure this point has been made. I will not add anything further to that.

I think that is about all I have to say.

**DR. BILLARD**: Most of the comments that I might have made have already been made. I will add my concerns also about the potential—next time we will be more concerned about overcounts than undercounts, in connection with all of these "Be Counted" efforts, get-out-the-vote types of efforts. I guess it remains to be seen how that will work out. If we think that the interest out there is so low that the mail return rate is going to drop off anyhow, maybe there will not be a problem with these potential duplicate counts. But I think it is a little naive on my part to hope that.

Coming back to what I would like to see out of the panel, for myself, the more I read and the more I listen and the more I try to learn about what is going on, the more I think I do not know anything at all. I guess what I am saying is that I get more confused. There is just more out there than I could have possibly imagined. So, in some sense, I would like to see a summarizing list, I suppose, of the different methodologies, itemizing, in a shorthand way, the strengths and weaknesses of each, to help me try to pull together all that I am trying to learn.

But I would also like to see a third sort of column, so to speak. We look at what we all agree we do not know, but would like to know. We recognize that there are some things even in the methodologies we have that we really do not know how to deal with. What I would like to see is which of these we know are impossible to do, that we just cannot do. Maybe theoretically we know it is not possible. I do not know. Maybe it is just operationally. But maybe there are some that are doable, but we do not yet know how to do them. If these could be identified, it would give us something to work on or think about for the future.

But having said that, as I said a minute ago, the more I have learned, the more I feel even more ignorant about it. But, simultaneously, the more impressed I am that the Census Bureau is managing as well as they are. I think it is a real credit to them. We have just been focusing on the statistical issues here, but there are a lot of non-statistical issues, as many of us know. It just boggles my mind when I think about how that is being handled. I am glad it is, and I am glad it is by them and not me.

**DR. NORWOOD**: Thank you. Steve Fienberg.

**DR. FIENBERG**: All my points have not been covered.

I did not come to the meeting with prepared remarks, and although I have a quote from Santayana, I will not read it.

I also did not plan to talk about lessons from 1990, but after you were given a brief homily on some of them yesterday, I thought it appropriate to at least fill in some of the details. Then I want to give you a few comments about thinking about what 1990 really says about 2000 and A.C.E., and then give you my issues and concerns.

Let me begin by recalling why it is we are so focused on dual-systems estimation. That is because of the concern of the differential undercount in the 1980s, largely for blacks and only secondarily for other groups. We measured it. It was, as Larry and others have said rightly, a distributional problem, but a different one than a lot of people have been focused on. We actually learned not just about blacks, but the fact that you could measure that differential in a very substantial way for other minority groups.

We have been talking about those evaluations from 1990 as if they were some massive block and tied all to the secretary's marching orders. I would like to remind people of P-16/Part 1.[5] P-16/Part 1 was the evaluation of the evaluation post-strata. That is the report that I think people should really be paying attention to about what we learned and why it worked, or why we think it worked—or why I think it worked—in 1990. The author is sitting down below.

P-16/Part 2 really then carried that and asked questions of numerical and distributive accuracy. It provided whatever evidence we had to glean about how that was also being handled, but that was not the starting point.

Everybody has talked about the errors and model failures. I am not here to tell you that errors do not exist or that the model is correct. The model is wrong. All models are wrong. So the question is, how good is it, and how can we evaluate it?

[Transparency]

One of the issues here is to point to the extent of error in the 1990 PES. I just want to give you the numbers that go behind the comment. David Freedman yesterday stressed the PES net number of 5 million and talked about 3 million errors—which I will not grant, but I will take as a starting point for discussion. If you take the PES gross—and I read this out of David's most recent publication, collaboratively—and take the 10 million EEs and the 15 million omissions, they net out to the 5 million. Take 25 million; 3 over 25 is not a very big proportion.

But I ask you this question: what if the net was zero? That is, what if the number that you got for the nation was exactly correct? Would you care? I hope the answer is yes, because if you got the number for the nation correct, but you got the group shares wrong and the geography shares wrong, that is exactly what we have been talking about.

So you could take the 3 million and compare it to a base of zero, and you would still want to use adjusted counts, if other criteria showed that that was right.

David also yesterday talked about the computer error. Since Ken introduced it in his opening remarks, let me tell you the facts. David talked yesterday about

---

[5]P-16 refers to one of the "P" studies carried out by the Census Bureau to evaluate the 1990 PES.

two sets of numbers. He said if you take the adjusted numbers as proposed in July of 1991 and run them, two seats change. They go from Wisconsin and Pennsylvania to California and Arizona, respectively in fact, if you look at the order in which they go. Then he told you that if you took the computer error correction into account and used the CAPE post-stratification, which, as John told us, was for a totally different purpose, and took those numbers in random, only one seat changed—that is, the seat that went from Wisconsin to California. In fact, the seat in that allocation does not move from Pennsylvania to Arizona.

The fact is that there are three sets of counts. The middle one is the one that David did not tell you about, which is what would happen if you fixed the computer error and ran the numbers again. The answer is, two seats change. One seat goes from Wisconsin to California, and the second seat goes from Pennsylvania to Arizona. The numbers are public. You can find them by going to this Web site. This is the only reason I really came up, to give the Web site. Those who do not recognize what the extension is about, it is this.

That is a Web page for *Who Counts?*[6] If you quickly scroll through the book reviews and a few other stray comments and get to the bottom, there is a pointer to a paper called "History, Mythmaking, and Statistics: A Short Story About the Reapportionment of Congress in the 1990 Census." On it are the three sets of figures and the results of running it in the apportionment formula.

The problem with block-level adjustments—we had that discussion yesterday, but it is important to know why we are doing it. We are not doing block-level adjustments because we care about block-level numbers. It is a device. Block-level adjustments are there as a device so that numbers will add up, whatever way we choose to aggregate things for other purposes. When the Academy panel met in the nineties—I am sorry Ben King had to disappear, because he was there at the time, we had a lot of discussion about whether or not there was another way to do this so that we did not have to put people into blocks, because then we would not have to argue about block-level accuracy. The Census Bureau prevailed in its notion, and I lost in my notion that what we really needed was a black box that produced the right aggregate numbers that would always be consistent.

So that is the past.

I was actually very impressed by the materials in the book that we reviewed. There are details there I did not know, some because they were just created, but others because I did not understand them. I thought the exposition moved us forward a lot, and that exposition is crucial. I think the current activities are working to clarify and refine the results. My best guess is that they should lead to improvements. I do not have a lot to add to the others' suggestions on the A.C.E., but there are lots of lacunae. I did not agree with some of the choices. Marty disagreed with others. I would guess that each of us disagreed with an overlapping but different set.

---

[6]The Web page Dr. Fienberg cites is `http://lib.stat.cmu.edu/~fienberg/WhoCounts.html`, which deals with his co-authored book: Anderson, Margo J. and Stephen E. Fienberg (1999). *Who Counts? The Politics of Census-Taking in Contemporary America*. New York: Russell Sage Foundation.

I think the outcome is far from certain. The question is, is there an alternative? And I do not see one on the table. I want to stress that. Demographic analysis is not the quick fix. Howard Hogan gave us some of the reasons. They are given in much greater length in one of the analyses we did. They are just not there to do this job that you are talking about doing, at any level, except perhaps at the national level for blacks.

So here we are in 2000, and the Bureau is under way. The census is going, although not the biggest parts of the operation. I was struck as I read the materials by the reference list. It was great material, but none of the things that were referenced seemed to have occurred outside the Census Bureau. I thought it appropriate to mention that there is research that goes on outside the Census Bureau, and some of it actually has relevance, I believe.

Let me take the topic of correlation bias and heterogeneity and say that nobody has solved your problem, but if you look at the paper by Kadane, Meyer, and Tukey, in 1998 *JASA* [*Journal of the American Statistical Association*],[7] or Richard Cormack's recent exchange in the *Journal of Epidemiology*, with Ron Regal and [Ernest] Hook,[8] or [Juha] Alho's work on heterogeneity[9] or, I dare to add, my paper in *JRSSA* [*Journal of the Royal Statistical Society, Series A*] last year, with Johnson and Junker,[10] there are clues about what you can do and what you can learn, and the direction of the biases that are associated with heterogeneity and correlation structure.

That leads me to suggest that there is actually a need for a short-term research plan, at the Bureau and outside. I do not think the Bureau can do this research between now and when it needs to get things done. I do not think it could commission it all on the outside, but I think it could commission some, and it could probably rally the community in some ways that it has not, that could be supportive of the in-house effort. Most of the empirical stuff has to go on inside the Bureau. We cannot, outside, do that for you. But I do think that there are things that could be done that would at least inform the analyses and help out, and maybe would help out in the longer term evaluation, too.

I have two major concerns about census process. One we have talked about repeatedly, and that is how we are going to take stock of the effects of all of these local improvement efforts. But something else that was not mentioned that I think really does need some consideration by the panel—and it hardly got mentioned— the census consists of two parts. It consists of the short form and the long form.

---

[7]Kadane, Joseph B., Michael M. Meyer, and John W. Tukey (1999). "Yule's Association Paradox and Stratum Heterogeneity in Capture-Recapture Studies." *Journal of the American Statistical Association*, pp. 855–859.

[8]See Cormack, Richard M. (1999). "Problems With Using Capture-Recapture in Epidemiology: An Example of a Measles Epidemic." *Journal of Clinical Epidemiology*, pp. 909–914; and Hook, Ernest B. and Ronald R. Regal (1999). "Recommendations for Presentation and Evaluation of Capture-Recapture Estimates in Epidemiology." *Journal of Clinical Epidemiology*, pp. 929–933.

[9]See, e.g., Alho, Juha M. (1994). "Analysis of Sample Based Capture-Recapture Experiments." *Journal of Official Statistics*, pp. 245–256.

[10]Fienberg, Stephen E., Matthew S. Johnson, and Brian W. Junker (1999). "Classical Multilevel and Bayesian Approaches to Population Size Estimation Using Multiple Lists." *Journal of the Royal Statistical Society, Series A*, pp. 383–406.

Bob reminded us about the importance of the long-form data. That is where all the content is. We are supposedly doing that on a sample basis. Does anybody have an evaluation planned to see how well we fulfill the sampling plan? It sounds like a trivial issue, but it is not. Just begin to think about all those extra forms that we are adding in and all the other devices, and what it does to the one-in-six sample. I have no idea. Indeed, an evaluation plan for the long form seems to me long overdue, even if this may be the last time we see it.

Which brings me to the collateral issue: do we have any other data? The answer is, we may actually have some data from the American Community Survey [ACS], and that may be of value for evaluation here as a different kind of touchstone. Actually, it is interesting, because we have asked the question in reverse. Can we use the long form as benchmark for the ACS, which is an important public issue if you are going to stress the Academies [reference not clear]? But we can glean information from the ACS about the census process because they are out there now gathering data, and it is going to continue through the census process.

Evaluation criteria: it strikes me that we need the list. It needs to be public. It comes in two parts. There are the things you are going to do before you release the numbers on April 1, 2001, presuming that you make a decision that the quality is of sufficient magnitude to release the adjusted counts, as well as the raw counts. But then you need the other ones. I think that list has to be—it is not going to be definitive, but we need it in public. I think that is what other people have been asking about. But I do want to suggest that it comes in two parts. It is not obvious which things will go in the first part, beyond the simple quality control checks.

Finally, I was struck by something when Bill Eddy and I did our little piece of data analysis and our search through the P.L. 94[-171] file. Something we did not mention was that, as we did that and we looked at the numbers on the Census Bureau's Web site, there are lots of 1's. In particular, the second block cluster we identified was one that had—because we found it in the P.L. 94 file—1 Hispanic and 70 American Indians. In other settings, when I would say that, the people from the Census Bureau would hide under their chairs, because that appears to violate virtually every reporting rule that relates to confidentiality.

There is, of course, the issue of what those numbers are. In fact, it is my belief that those numbers are correct only for the block total, and there has been data swapping—that is, a confidentiality edit has been applied to those numbers.

Many of the public uses of the products of the 2000 census will come from public census products, as opposed to confidential census products. There is a major activity under way called the American FactFinder. The question is, will you look at it and evaluate the quality and the level of error that is going to be associated with what will be released in it?

It is a tough job. As Marty said, I am glad you are doing it.

**DR. NORWOOD**: Ken Prewitt has a few words.

**DR. PREWITT**: It was suggested yesterday that they would be wise, but I am no longer promising to meet that standard.

Another word or two on the issue of counts and shares. As I roam the country and talk to lots of people about the census and its uses, I am actually impressed by

how many places there are in the system where the absolute count is what matters more than the share. Obviously, everything is comparative, but where there are threshold effects shot through this system—Detroit was mentioned, but there are dozens and dozens of programs and activities and investments where they have to do with threshold effects, the total numbers. The Gap [retail chain] does not care how many units it puts out there, but it only wants to put them out there in communities of at least 50,000. So however many they find that are 50,000, that is where they are going to go. Every town in the country that wants a Gap wants to be 50,000. That is a made-up example, but not widely made-up.

One runs into lots and lots of decisions. The Social Security people, of course, want to know absolutely how many old people there are going to be 15 or 20 years from now. The educational planners want to know how many kids there are. Obviously, you can say that is all a share basis. You have Social Security and educational planning, and so if there are more old people and fewer young people, then you—that is actually not how the system works. People are actually making all kinds of very critical public and private decisions based upon counts, not shares.

One of the things that most impresses me about the public relationship to the census is the importance for many, many groups of social recognition. They actually want to be counted because they want to be recognized. I spent a fascinating half-day with Arab-American leaders. They actually want to be counted, not because they are in some sort of share contest—except in the largest, most generic sense—they simply want the rest of us to know how many of them there are here and how well they are doing. The American Indian count problem is really much, much more about social recognition than shares, I promise you. The census measured about 500,000 American Indians in 1900. That number is now close to 3 million. That is not fertility. That is more and more people wanting to claim their heritage.

So do not underestimate the importance of counts for a complicated dynamic called social recognition. We are the only place to provide it to groups. It is way, way beyond just whether they are going to get some share of a federal largesse. That is a reality about our society.

That does not mean this is a simple process. Obviously, the census data have a deep responsibility for reapportionment and redistricting and allocation of federal funds and so forth—a deep, deep responsibility. That is why it is in the Constitution. We do not trivialize the importance of getting the shares as close to accurate as we can. But do not underestimate the power of counts themselves in this society. The boosterism, the Thomas Jefferson story, and so forth—it is really beyond boosterism.

A quick word, if I may, on the overcount issue. Just to be reassuring, we are not unmindful of the overcount complication, given the greater degree of public investment in the census. I should say, on the Internet, we really think we have put enough firewalls in that we will not have an Internet problem. You cannot get into the Internet file unless you have a bar-coded questionnaire in front of you. You cannot fabricate a bar-coded code number that we do not already have in our files. So I do not think the Internet is going to be a source of an overcount problem.

We have spent a lot of time worrying about how to handle the "Be Counted" forms. We have restricted the amount of time they are going to be out there, and so forth and so on. That does not mean it is still not an issue, but we do have internal processes, of course, to make sure the "Be Counted" forms are not going to be counted unless we can get them back to an address that we have identified, either as a bar-coded address there already or—we will go out and field-test it— if there is no address, we cannot use the "Be Counted" form. So we are not unmindful of the fact that new procedures could lead to an overcount. We have spent an awful lot of time, I think, trying to protect from that.

A third sort of small thing I wanted to mention—and I think it feeds into this very important conversation we are having about the number of different indicators we are going to be looking at as we struggle with making the big decision—is something we did not have a chance to talk about, but you may be interested in. We have pretty well finished our work on the non-city-style addresses. That is a 20-million-household address file, based on the non-city-style areas. Most of that has now gone back out to the communities and our LUCA [Local Update of Census Addresses] Program for checking. They have come back in. We have reconciled. They have appealed where they wanted to appeal. The OMB appeal process is now well under way. Indeed, it is so far under way that the numbers I am about to give you we are fairly confident about. It looks as if we will add about 10,000 addresses in that process. That is on a base of 20 million. That is down to the fourth decimal point. That is 1/20th of 1/100th.

I am not promising that other operations in the census will be that accurate. Nevertheless, we take some satisfaction in the fact that what was, after all, a quite difficult field operational process, in which we now have real data—insofar as the communities, the cities, the local planning groups, and so forth, are giving us their real judgment, and we have an OMB appeal process and so forth—we now are finding that, at least for that operation, the accuracy rate was remarkably high. Will the city-style addresses prove to be as accurate? We do not know. We have not gotten there yet. But at least a 20-million address file turns out to be—we did a quite good job on it, operationally.

Those are just odds and ends. I mention that one as a way to, if you will, lead into my only interesting point, I think, or concluding point.

With respect to the DSE, obviously we have been appreciative of this session and conversation, as I knew we would be. Many things we will go back and think hard about. There has been a good conversation about the mover problem, a good conversation about the search regions, a good conversation about some of the matching issues. These things we will go back and, insofar as we can make ongoing improvements, we will certainly try to do it. is the apportionment date and April 1 is the

But, as Janet said in leading off this session, finally, a decision has to be made in real time. We did not invent the fact that December 31 redistricting deadline. The reapportionment date, December 31, was, after all, put in place in 1932, when the country was half the size and a lot easier to count. We have made, as a bureau, obviously, enormous productivity increases since the 1930s in order to

still be within that same time frame, counting twice as many people, and so forth. But we are living within real time. That is what the law says. It is a law that obviously could be changed. You could imagine a longer time period for doing the census. Nevertheless, for now, we will have the redistricting data out by April 1, as we are obligated to do.

Therefore, it seems to me—and now I will speak personally, in terms of the decision that the Census Bureau will make—the director of the Census Bureau, when this decision is made, could well be someone in this room but not me. There is an upcoming election, and there is reason to believe that there would be a different Census director if there was a different administration. Indeed, there is no particular reason to believe that, even if there was a continuation of the same party in power, there would be the same Census director. I think that is too bad, because I would like whoever has to make this decision to have heard this conversation and to have benefited from this interaction with the panel. So I hope the person is in the room, if it is not me, one way or the other.

**DR. NORWOOD**: We fully expect it to be you, Ken.

**DR. PREWITT**: As I look at making the decision, it seems to me that we can use a very simple $2 \times 2$ table. You have a census, good or bad; you have an A.C.E., good or bad. If the current thinking of the Census Bureau is that the census is not going to be very good and it is going to have a lot of measurement error in it, but we have an A.C.E. that is good—an easy decision—we will adjust and report adjusted numbers. If you have a good census, as best we can tell from the indicators available to us—response rate, level of public cooperation, address file, and so forth—and a bad A.C.E., it is an easy decision. We will not adjust. That takes care of those two cells.

What do you do if they are both good, based upon whatever evidence we are able to accumulate? It is actually a very difficult decision. There you are asking yourself, at the margins, is A.C.E. going to improve? It would be my own thinking, just speaking candidly as if I happened to be the Census director, that you would really want to make sure that marginal improvement was sufficient to pay the complicated legal costs, political costs, public-relations costs, and all kinds of things. If the census is basically good, good enough for the purposes for which it is normally designed—that is, if there was not much of an undercount correction/adjustment factor—then my feeling is, you probably do not adjust, if they are both good.

Let me say about the cell where the A.C.E. is not good, we had a dress rehearsal experience with the A.C.E., and we sweated. We deliberately sweated, because we ran into some things in our South Carolina dress rehearsal site, as most of you know, where we had to ask ourselves the question of whether the census was so bad—or whether the A.C.E. was bad or the PES was bad—as Jay Waite put it so well, did we have a problem with the camera or a problem with the thing being photographed? We knew something was not working right. We said to ourselves— and this was not an easy decision—we are going to make this decision in real time. We are going to set exactly the same deadline as we would have in the real census. If we cannot sort it out within that period, if we do not come to some sort

of reasonable agreement and conclusion internally about whether it is a camera problem or a problem with what we are filming, then we would not be able to adjust.

We struggled. Howard and his team struggled for hours and hours and hours and hours, to meet that deadline. At the end of the day, we realized that we had taken a good picture of a very bad census that had a lot of anomalies that we had not anticipated. But we struggled with that question. The issue of whether the A.C.E. itself will be good is a decision we will struggle with when we actually have the experience. We are obviously presuming a high response rate. What if we get a low response rate on the A.C.E.? What does that mean? So do not imagine that that is an empty cell. We hope it is an empty cell, but theoretically it is not an empty cell.

Now you also have the problem of bad/bad. You have every reason to believe that you had a bad operational census. All kinds of things go wrong. You have natural disasters. You have public relations disasters. You were not able to hire the labor pool that you needed. But you do not know if the A.C.E. is good enough to sort of correct for that. That is a very tough cell to be in. It is very hard sitting here today to know how we will invoke the criteria to struggle with a decision in that cell. We think that is a low-probability cell. A lot of things would have to go wrong in the next year that we think will not go wrong. Nevertheless, we have to theoretically hold it open.

I would say that, under the circumstances we feel we have had, using as many data points as we can use, as many indicators, and so forth, for a census that has serious measurement problems and an A.C.E. that worked well, we will adjust. If we have a very good census and an A.C.E. that worked well, we will have a more complicated decision at the margins. If we have a good census and a bad A.C.E., we obviously will not adjust. If we have bad/bad, we will reconvene the panel and ask for help.

That is kind of how I see it. I would just conclude with one other thought. There has been a lot of talk about the response rate. We are putting a lot of time and effort into trying to improve that response rate. Some people in this room—indeed, I will not name the person—he is not confident that we are going to improve much on the response rate. He is showing such lack of confidence in my leadership on the response rate issue that he has wagered me. I would say that I am open to any wagers from anyone in the room. In fact, I made one the day before yesterday with the Government Accounting Office, which has just, as you know, written a report saying that they think that we have been too optimistic about the response rate.

We set a number—it was a friendly bet—and then agreed that it is that number plus 10 percent going in both directions. If we get 4 percent, we will add 40 percent to the base number. If we go down 4 percentage points—that is a wager I am willing to make with anyone in the room, on exactly those terms. We set a number at 61. If it is 62, you will need that number plus 10 percent, 63, so forth and so on, and I owe you that much more going below 61. I am very confident. I just invite anyone who wants to join the wager to do so.

Thank you, Janet. Thank you, panel.

**DR. NORWOOD**: Thank you very much, Ken. I want to thank everyone who has been here and all the people who have worked so hard. I especially want to thank Howard and his staff for all the hard work and for being on the hot seat for so long.

It is not over, Howard. We are still in operation. We will be after you.

I want to thank Raj [Singh] for all the help he has given us in working things out.

The meeting is now adjourned.

# REFERENCES

National Research Council

 2001a *The 2000 Census: Interim Assessment*. Panel to Review the 2000 Census, Constance F. Citro, Daniel L. Cork, and Janet L. Norwood, eds. Committee on National Statistics. Washington, D.C.: National Academy Press.

 2001b *Proceedings, First Workshop of the Panel to Review the 2000 Census* (October 6, 1999). Committee on National Statistics. Washington, D.C.: National Academy Press.

 2001c *Proceedings, Third Workshop of the Panel to Review the 2000 Census* (October 2, 2000). Committee on National Statistics. Washington, D.C.: National Academy Press.

U.S. Census Monitoring Board, Congressional Members

 1999 *Unkept Promise: Statistical Adjustment Fails to Eliminate Local Undercounts, as Revealed by Evaluation of Severely Undercounted Blocks from the 1990 Census Plan*. Report to Congress. Washington, D.C.: U.S. Census Monitoring Board, Congressional Members. (September 30).


**U.S. Census Bureau Materials Provided to Workshop Participants**

Cantwell, Patrick

 2000 Accuracy and Coverage Evaluation Survey: Missing Data Procedures. DSSD Census 2000 Procedures and Operations Memorandum Series Q-19 U.S. Census Bureau, Washington, D.C. (January 13).

Childers, Danny R.

 2000 Accuracy and Coverage Evaluation: The Design Document. DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1. U.S. Census Bureau, Washington, D.C. (January 11).

Childers, Danny R., and Debbie Fenstermaker

 2000 Accuracy and Coverage Evaluation: Overview of the Design. DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-2. U.S. Census Bureau, Washington, D.C. (January 11).

Griffin, Richard

    2000 Accuracy and Coverage Evaluation Survey: Dual System Estimation. DSSD Census 2000 Procedures and Operations Memorandum Series Q-20. U.S. Census Bureau, Washington, D.C. (January 12).

Griffin, Richard, and Dawn Haines

    2000 Accuracy and Coverage Evaluation Survey: Post-Stratification for Dual System Estimation. DSSD Census 2000 Procedures and Operations Memorandum Series Q-21. U.S. Census Bureau, Washington, D.C. (January 12).

Haines, Dawn

    2000 Accuracy and Coverage Evaluation Survey: Synthetic Estimation. DSSD Census 2000 Procedures and Operations Memorandum Series Q-22. U.S. Census Bureau, Washington, D.C. (January 12).

Hogan, Howard

    2000 Accuracy and Coverage Evaluation: Theory and Application. Prepared for the February 2-3, 2000 DSE Workshop of the National Academy of Science Panel to Review the 2000 Census. U.S. Census Bureau, Washington, D.C.

Navarro, Alfredo

    2000 Accuracy and Coverage Evaluation Survey: Targeted Extended Search Plans. DSSD Census 2000 Procedures and Operations Memorandum Series Q-18. U.S. Census Bureau, Washington, D.C. (January 12).

Schindler, Eric

    2000 Accuracy and Coverage Evaluation Survey: Post-Stratification Preliminary Research Results. DSSD Census 2000 Procedures and Operations Memorandum Series Q-23. U.S. Census Bureau, Washington, D.C. (January 12).

# LIST OF INVITED PARTICIPANTS

Barbara Bailar, National Opinion Research Center, Chicago, IL
William Barron, Deputy Director, U.S. Census Bureau
Lynne Billard, Department of Statistics, University of Georgia
Robert Fay, Chief Mathematical Statistician, Director's Office, U.S. Census Buireau
Stephen Fienberg, Department of Statistics, Carnegie Mellon University
David Freedman, Department of Statistics, University of California at Berkeley
Hermann Habermann, Statistics Division, United Nations, New York, NY
Howard Hogan, Chief, Decennial Statistical Studies Division, U.S. Census Bureau
Charles Jones, consultant to U.S. Census Monitoring Board, Congressional Members
Ruth Ann Killion, Chief, Planning, Research, and Evaluation Division, U.S. Census Bureau
Benjamin King, Boca Raton, FL
Donna Kostanich, Assistant Division Chief, Decennial Statistical Studies Division, U.S. Census
     Bureau
Robert Lerner, consultant to U.S. Census Monitoring Board, Congressional Members
Mary Mulry, consultant to U.S. Census Bureau
Kenneth Prewitt, Director, U.S. Census Bureau
J. Gregory Robinson, Chief, Population Analysis and Evaluation Division, U.S. Census Bureau
John Rolph, Marshall School of Business, University of Southern California
Bruce Spencer, Department of Statistics, Northwestern University
Philip Stark, Department of Statistics, University of California at Berkeley
John Thompson, Associate Director for Decennial Census, U.S. Census Bureau
Preston Jay Waite, Assistant Director for Decennial Census, U.S. Census Bureau
Martin Wells, Department of Social Statistics, Cornell University
Kirk Wolter, National Opinion Research Center, IL
Donald Ylvisaker, Department of Statistics, University of California at Los Angeles
Alan Zaslavsky, Department of Health Care Policy, Harvard Medical School