



Proceedings, Third Workshop: Panel to Review the 2000 Census

DETAILS

101 pages | 8.5 x 11 | PAPERBACK
ISBN null | DOI 10.17226/10280

BUY THIS BOOK

AUTHORS

Committee on National Statistics, National Research Council

FIND RELATED TITLES

Visit the National Academies Press at NAP.edu and login or register to get:

- Access to free PDF downloads of thousands of scientific reports
- 10% off the price of print titles
- Email or social media notifications of new titles related to your interests
- Special offers and discounts



Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. (Request Permission) Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences.

PROCEEDINGS, THIRD 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.

The project that is the subject of this report was supported by contract no. 50-YABC-8-66010 between the National Academy of Sciences and the U.S. Census Bureau. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

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, Third Workshop. Panel to Review the 2000 Census*. Committee on National Statistics. Washington, DC: National Academy Press.

Printed in the United States of America

©2001 by the National Academy of Sciences. All rights reserved.

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.

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

**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 KALSBECK, 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*

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 third workshop. (See National Research Council, 2001b and 2001c for proceedings of the first and second 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.

The documents provided by the Census Bureau for the third (October 2, 2000) workshop were drafts, containing background text, proposed topics for inclusion, and table shells, indicating the kinds of analysis the Bureau proposed to conduct to inform its decision process for March 2001. The documents contained no results from the census or the A.C.E. Each of the sixteen documents provided corresponds to a document in the "B-series" of memoranda that the Bureau released March 1, 2001, containing full text and tabular and other analytic results. References to these documents in the workshop proceedings are to the **final, published 2001 version** of each. References by speakers to page numbers in the draft documents have been retained, although they do not necessarily, or likely, correspond to the published version. In addition to the completed versions of the 16 documents for which the Census Bureau provided drafts at the workshop, the

Bureau released another 3 documents in the B-series. All of the documents released on March 1, 2001, are available at <http://www.census.gov/dmd/www/EscapRep.html>.

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.

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 | 1 |
| Opening Remarks of the Chair | 1 |
| Opening Remarks of Director Prewitt | 2 |
| Planned Decision Process | 8 |
| Review of the Quality of the Unadjusted Census | 10 |
| Remarks on Possible Duplication | 22 |
| Review of the A.C.E. Operations | 31 |
| How Accuracy of A.C.E. and Unadjusted Census Counts Compare | 33 |
| Role of Geographic Accuracy in Adjustment Decision | 41 |
| Review of Measurements of A.C.E. Quality | 47 |
| Demographic Analysis | 59 |
| Continued Discussion of A.C.E. Quality Measures | 65 |
| Forming an Overall Assessment | 71 |
| Final Comments | 79 |
| References | 88 |
| List of Invited Participants | 91 |

PROCEEDINGS

OPENING REMARKS OF THE CHAIR

DR. NORWOOD: I would like to welcome you all to this meeting. I think it is an important meeting. As you all know, the role of our panel is to evaluate the census, but one extremely important aspect of that task is to evaluate census coverage, including any possible disproportionate undercount and to evaluate any steps that the Census Bureau might take to adjust those numbers.

The purpose of this workshop is to help us to get ready to do that job. We have asked members of the Census Bureau to come to this workshop to explain to us how they will evaluate the census and any possible undercount and, if they should decide to adjust the numbers for the undercount, to come and tell us the scientific preparations that they are making so that they will be in a position to make a professional decision about the adjustment, if they decide to do so, and about the quality of the numbers that they might release.

In particular, we are very interested in finding out what kinds of evidence they will develop in order to provide a scientific basis for their decision.

So the meeting, I hope, and I am sure, will be very useful to our panel. We want to know how the Census Bureau will make its determination and then, if it adjusts, what data and evaluation it contemplates to determine whether the adjusted numbers are better than the unadjusted ones. The panel will have to decide how it will approach this issue and this meeting is one part of the information gathering that we are doing. We have invited a number of scholars to help us at this meeting and they and the panel will participate in the discussions, asking questions, raising issues, and, I hope, giving us the benefit of their thinking.

Let me, first, tell you who they are. We have, over here, Barbara Bailar, who is, I believe, a senior vice president at NORC, the National Opinion Research Center, at the University of Chicago. Lynne Billard is in the Department of Statistics at the University of Georgia. Jeff Passel [from the Urban Institute] is not here yet; we will introduce him when he comes in. Allen Schirm is from Mathematica—the Academy also has a panel to help to advise the Census Bureau on what it might do in 2010, and Allen is on the 2010 panel. Michael Stoto is a demographer with the George Washington University School of Public Health. Marty Wells is from Cornell. Don Ylvisaker, from UCLA, is also on the 2010 panel, and there is Alan Zaslavsky from Harvard.¹

We also are very fortunate to have the director of the Census Bureau here, Ken Prewitt. We are also privileged this morning to have Barbara Torrey with us, who is sitting over there, who, as you probably all know, is director of the Division of Behavioral and Social Sciences and Education, which is the governing body over the Committee on National Statistics [at the National Research Council].

As I have said, the people around the table, I hope, will be very interactive, because that is the purpose of this meeting. At the end of the day, we will provide a chance for anyone in the audience who has any questions or wishes to make a comment briefly to do so.

¹Zaslavsky was added to the Panel on Research on Future Census Methods (“the 2010 panel”), of which Schirm and Ylvisaker are members, in autumn 2001.

2  PANEL TO REVIEW THE 2000 CENSUS

Before we begin, I want to emphasize, first, the panel remains open to the examination of any scientific methodology that it believes would be useful in carrying out its task.

Secondly, the panel remains open to all points of view. Although we have done a good deal of work to understand the census operations, we have not yet made any decisions about the evaluation we have an obligation to perform. We hope this workshop today will help us in that endeavor.

Third, we are all perhaps too much aware of the fact, which I think will become clear during this discussion, that the data required for effective research on these issues will not be available publicly until the material has been put together by the Bureau. We are also very much aware that the Bureau has been undertaking a tremendous operational effort and it is getting things together as fast as it can, but it always takes a good deal of time. It is a monumental task. On the other hand, I am sure that our census friends fully understand that it is in the public interest for us all to have the information needed for effective evaluation of the decisions that we will have to make.

Finally, I would just like to take a moment to compliment Ken Prewitt and the rest of the Census Bureau staff for your cooperation in putting together this meeting and the previous meetings and for your professionalism and openness, and for your assistance, really, in helping us to understand what is going on in the census. We will start by having Howard Hogan introduce these issues. The group around the table has a whole set of papers that develop the methodology—of course; the numbers are not there, because they are not available yet. That is where we are starting.

OPENING REMARKS OF DIRECTOR PREWITT

DR. PREWITT: Just a word or two, if I may, Janet, and John [Thompson] will spend a few minutes bringing everyone up to date on the operations before we turn to Howard. I would like to put some introductory comments in a very, very broad context.

Some of you who have good memories will know that after the 1990 census the term “failed census” was fairly frequently used. It was used loosely, that is, without defining what constituted a successful and/or a failed census. Indeed, clearly, 1990 was not a failed census; a failed census would be one that was not used, could not be used, and, clearly, 1990 was used to reapportion, to redistrict, and as the base for intercensal estimates and everything else, so clearly it was not a failed census.

Nevertheless, the nomenclature stuck a bit and that is a very bad thing, it seems to me, for the Census Bureau and for the statistical system, to have an operation such as the 1990 census so described. It was not a failed census, not only in terms of the fact that it was used but it was not a failed census even operationally, even though the [net] undercount, as we know, should [not] have moved back up, having declined for the last two censuses.

But it does seem to me that it is very, very important in 2000 to bury that label once and for all, because it is simply, as I say, not healthy for the statistical system,

certainly not for the Census Bureau, to live in an environment in which we could even imagine the census failing.

John will talk more about where we are operationally, but I do think that one of the things that has been accomplished in 2000 is to replace that label with a set of labels that better describe the census as “successful” or “good” or what-have-you.

It does not mean that you still cannot actually have a failure in Census 2000 and that could obviously now happen in the software and in our data cleaning, data matching, and so forth, and we will talk about that a bit today, but we have every reason to believe that is a very, very, very low probability, so we do not any more think of the census along that dimension.

I mention that because 1990 introduced something else: Secretary Mosbacher’s decision not to use the corrected data, the adjusted data, overruling the recommendation of the Census Bureau, sending Barbara Bailer to the University of Chicago, and many other consequences. Secretary Mosbacher put into his decision criteria a sentence, which I paraphrase, but basically goes as follows.

One of the reasons that he decided against adjusting the 1990 census was, he said, that it would open up the possibility of manipulating the data for political reasons, i.e., somehow the Census Bureau would be able to pre-design a census with a known partisan outcome. Mr. Mosbacher was careful to say that he did not believe the Census Bureau did that in 1990, but it certainly set the possibility that that could happen in future censuses and, therefore, there should not be an adjusted or corrected census.

To my knowledge, and I think I have looked enough into the literature to be confident of this, that is the first time a senior official of the United States Government put on the table the idea, the prospect, that the Census Bureau could pre-design a census in order to have a known partisan outcome.

As you know, although Mr. Mosbacher spoke in the conditional sense, that conditional sense rather quickly got erased and from then on, for the last five or six years, the assumption has been that, indeed, the Census Bureau not only could but would and probably was designing the census, knowing it would have a given partisan outcome.

Obviously, people in this room familiar with the census operation know the impracticality of that and also know that it is extremely difficult to predict the partisan consequences, even if the Census Bureau were of a mind to do that. It is extremely difficult to predict the partisan consequences of adjustment or non-adjustment, and I do not need to talk about all that in this room.

The point, however, and this is why this meeting is so very, very important, is we need to bury that phrase just as we needed to bury the phrase that the 1990 census was a “failed census.” It is very bad for the federal statistical system and the Census Bureau to live under the cloud that somehow it has the will and the capacity to design a census knowing beforehand what the likely partisan outcome of that census will be. Having said that, there is no doubt that census numbers do have partisan consequences. We are not talking about whether they have partisan consequences or not; we are talking about whether the Census Bureau would design a census in order to achieve a particular partisan outcome.

4  PANEL TO REVIEW THE 2000 CENSUS

What is so very important about this meeting, and its deliberation in ways that I will just hint at in a second, is that it is the opportunity to say: yes, census numbers have political consequences, that is why they were put in the United States Constitution, but that is very, very different from saying that the Census Bureau itself has a partisan motivation, partisan capacity, partisan interest, or anything of the sort.

Unless we can get rid of that in the vocabulary that surrounds Census 2000, with all of its contention, and so forth, unless we can get rid of that vocabulary and that kind of descriptor, we have done something harmful, it seems to me, to the Census Bureau and federal statistics. That is why we put as much attention to this workshop and, indeed, to the work of the panel as we have.

I should say, by the way, obviously this is quite orthogonal to the argument about whether we should or should not correct. That is a scientific decision and it will be made as best we can make in ways we will talk about. There will be critics of that. We welcome the critics. That is not what the question is. The question is not whether it makes sense or not; the question is if it makes or does not make sense, whether that decision is being made on anything other than professional scientific criteria.

As you know, the Census Bureau has made a preliminary determination that it is feasible to use dual-systems estimation in census 2000. However, as you also know, we have not yet determined whether the A.C.E. [Accuracy and Coverage Evaluation Program] will meet our expectations and, indeed, we will evaluate both the census and the Accuracy and Coverage Evaluation early next year to decide which set of data to denominate as the P.L. [Public Law] 94-171 data [redistricting data]. I want to say that again. We have determined that it is feasible to use the Accuracy and Coverage Evaluation. We have not determined whether we will. That is what the workshop is obviously about. There are many, many people in Washington, D.C., and other places that do not believe that, who believe we have already made up our minds. Well, it is simply not the case.

We will do all the kinds of stuff that Howard [Hogan] will describe in order to make that determination and the panel will, obviously, be examining that and decide in its own judgment whether the kinds of things we are bringing to bear are the right kinds of things to bring to bear.

We have issued a feasibility document, as you know, and circulated it to all of you and that is where we set forth why we think it is feasible [Prewitt, 2000]. As you also know, partly responding to Secretary Mosbacher's decision after the 1990 census, Secretary Daley did issue a federal regulation that delegated the power, the authority, to make the decision about whether to adjust or not to the Census Bureau. That did occasion some conversation, active conversation, in the Congress and in other places in Washington.

I should say, to me, it was odd that it occasioned that conversation. We obviously issue the apportionment numbers, and we do a lot of complicated technical things to get the apportionment numbers out without thinking that we would check those numbers with the Secretary of Commerce, and we do not see anything odd about doing the same thing with the next set of operations, that is, the Accuracy and Coverage Evaluation operations.

Indeed, we think that all that Secretary Daley's decision was, was to reestablish the 1980 pattern, when Vince Barabba and the professionals at the Census Bureau indeed made the decision. It was not until 1990 that for some reason—not “for some reason,” for reasons we know—the decision got moved to the secretary's office, and it is that anomaly, if you will, in statistical practice that was corrected by the federal regulation notice.

From our point of view, of course, it is where it belongs and the Census Bureau will make the decision, obviously. In making that decision, there is a special committee that has been formed, outlined in the feasibility document—we call it the ESCAP committee, the executive committee—for this process.² It will make a recommendation to the director and, I stress, “whoever that happens to be” when that decision is made. Clearly, as you know, the director's term is coterminous with the Administration's term and, therefore, one way or the other, something has to happen after January 20th, irrespective of the party that wins the election.

The whole idea about how the decision is going to be made and the delegating of authority to the Census Bureau was really made behind a veil of ignorance; that is, the decision about how to make the decision was made not knowing who would be making the decision.

The ESCAP committee will deliberate in January-February, hopefully making the decision by early March. Obviously, that decision will be based on only technical considerations and scientific analysis to the best of our capacity.

Now I want to return to the larger theme about Secretary Mosbacher's observation following the 1990 census about the possibility of designing a census to have a known partisan outcome. One of the things that the Census Bureau has tried to do over the last several years is to dissuade people from making that accusation by being as transparent as it can be.

Indeed, we determined that in this highly politically charged atmosphere it was very important for the Census Bureau to try to be transparent, consistent with good statistical practice, though at the edges we have actually done things that would not have been prudent from an operational or statistical point of view in order to display even more openness and transparency.

Just a few factoids: I have testified before the Congress 17 times as census director. I have not looked up comparative data but my guess is, Janet, that that is reasonably unusual for an agency head, in less than two years to testify 17 times. We have responded to over 150 letters from the House Subcommittee [on the Census] in the last two years and provided a massive amount of data in response to it. There have been a number of field visits by the GAO [U.S. General Accounting Office], the [Congressional] Monitoring Board, the I.G. [Commerce Department Inspector General], and others—a total of 522 field visits during our operations.

The GAO itself has testified nine times and issued nine reports over the last two-year period. The Department of Commerce Inspector General's office has issued 25 reports on the census operations. Indeed, by my count (it is a rough count, obviously), there are well over one hundred people in those formal oversight operations who are full-time—full-time monitoring the census operation.

²ESCAP stands for Executive Steering Committee for A.C.E. Policy.

I should say that one of the things that this census monitoring operation failed to do, and I certainly say that publicly in front of our friends from the GAO and the inspector general's office today, they wrote a large number of reports over the last two years talking about the things for which the Census Bureau was ill-prepared.

The one thing they forgot to put into those reports, and this is not just a flip comment, the one thing they forgot to put in those reports is that we were ill-prepared for the amount of oversight we had to deal with. There is not a single person, who ever said, "And you had better staff up for these 25 reports or these 522 field visits or these 17 testimonies." No one ever asked, "Are you staffed to do that?"

It turned out that to be responsive to all of that, we had to deflect some of the important operational management time from Jay Waite and John Thompson and other people, because we simply had to be responsive to that.

I would say to the 2010 panel members here today, do not let that happen again. Make sure that the 2010 census is planned knowing that it is going to be as heavily scrutinized as the 2000 census has been. In addition to those formal oversight processes, of course, we have been meeting with our advisory committees throughout the decade, soliciting input. We have tried to be open with the press. I have held 40 press conferences since I became Census Bureau director and, of course, there is the NAS panel.

We have also issued our executive state of the census report on a weekly basis and given that to all of our stakeholders. All of our decision memos have been widely circulated. We have obviously pre-specified as much as we can, especially the A.C.E. process—the panel has obviously been paying attention to that. We have extensively documented—I probably do not have to remind you about the documentation that Howard and his team have generated.

We are currently providing the agenda and the minutes of our ESCAP meetings to the subcommittee and to other interested stakeholders.

I just want to emphasize that to make the simple point that I think that is the price we had to pay and will continue to pay; that is, in order to claim that we are transparent we had to be as transparent as we possibly could be, and, as I say, there were plenty of times in the field operation period, and there will certainly be times in the next phase of the census, where doing that did subtract some of the attention we should have been giving to the job itself.

Nevertheless, if it helps bury forever, or at least for the foreseeable future, the idea that the Census Bureau has got some sort of team out there figuring out which political party is going to benefit from this kind of post-strata structure, then it was well worth the price.

Now a word or two about the stuff that Howard will be talking about. We, as both Howard and John will emphasize, will be looking at a large number of processes, of data sets, of our own operations, during the tough period between January and February, when the decision will be made about whether to correct the data or not. We will be looking at those data on a flow basis, and we will make all data that we use in that decision process available to the public.

There will be pressures on us, intense pressures, and they will start the minute this meeting is over. It will probably start the minute I finish talking. There will be intense pressures on us to release some of those data earlier than we think is prudent to do so. We will resist the best we can those pressures, and I want to set forth some of the reasons that we will try to resist those pressures.

We believe that premature release of some data sets would simply do more harm than good. We have had some very unfortunate experiences already in the census where we have prematurely shared data with some of the oversight apparatus and found that they were not analyzed correctly. We spent a lot of time trying to get them analyzed correctly, and, in the meantime, of course, there is an awful lot of press attention and other kinds of attention, so it creates public confusion, quite honestly.

We also think that no data should be released until we have verified them, until we have done all of our quality checking on them. That takes time. We do not want half-formulated data floating around where people have this interpretation and that interpretation until we are as certain as we can be that the data are cleaned, and we simply want to minimize the amount of incorrect conclusions that can be drawn.

Also, there is the distinct possibility that any kind of early release could invite the appearance of manipulation; that is, the very thing we are trying to get away from could be aggravated or accelerated by that process. What we need to insist on, going into the January-February period, is that we are not selectively releasing data in order to try to create one or another assumption or predilection about whether we are going to correct or not.

I will give you just one example that John Thompson used in a session the other day. Let us say that we find that demographic analysis suggests that there is a large differential undercount, so we release that. We could then be accused of having released that in order to sort of set the predicate that we should be using the A.C.E. That would not be why we would do it, but we could get accused of that, so it is our judgment at this stage that all of the data will be released but none of them should be released until we have looked at them, examined them, weighed them, made our judgments about them, and then shared them at the same time with everyone, to the public, to the panel, to the subcommittee, the Monitoring Board, and everyone who has an interest in these sets of data—certainly to the litigation process, which, of course, will be interested in these data as well.

We are very concerned that we do not have bits and pieces of data sets out in ways that could create public confusion or could lead someone to interpret what we have done in such a way as to suggest that we were being affected by political considerations.

Finally, I would say, obviously, and I appreciate Janet's opening comments, we have a lot of work yet to do—very, very major work. It is obviously out of the field and into the offices but we now have to do the kind of data analysis and data cleaning and correction, and so forth, that goes into finally producing the apportionment counts and the redistricting data. We must somehow have the opportunity to deliberate and argue about these data amongst ourselves before we

talk about them. Every decision that we make will be documented. Every datum that we use will be released. There will be nothing “secretive” about it, but we do have to still get the work done. We are on a very tight time schedule to meet our next two major deliverables, which are the apportionment count on the 31st [of December, 2000] and the redistricting data by April 1st, 2001.

I just wanted, Janet, to introduce the day that way. To be redundant, I do not mind being redundant on this, I feel so very strongly about it, as I hope everyone in the room does, I want to say that whatever we should have accomplished by the time Census 2000 is over, of all the things we want to accomplish with Census 2000, one of the things we must—must—accomplish is that the Census Bureau itself is not partisan.

The data may have partisan consequences but the Census Bureau is not partisan and I really urge those in the room, especially those on the panel, who may be critics of dual-systems estimation—and that is fine, as you all know, we have no trouble with that—we really hope that the criticism of dual-systems estimation is articulated on scientific and technical grounds and not because somehow the Census Bureau can manipulate data, which it does not know how to do.

As I have said many times publicly, we do not have experts in voting behavior, we do not have experts in redistricting, we would not know how to go about trying to design a census that would have a known partisan outcome, and especially we do not even know who will be running the Census Bureau or which particular party will have appointed him, when this decision is made. The shallowness of that argument is, I think, transparent itself, but, nevertheless, it sits there in the atmosphere, and I really hope that it is laid to rest. That is why we take this day to be so very important as a kind of collective effort by the scientific community to have a legitimate debate about the methodologies. We hope to use this National Academy of Sciences’ process to set aside what has been, I think, a decade-long and extremely unfortunate charge leveled against the Census Bureau. Thank you.

DR. NORWOOD: Thank you very much, Ken. Now I would like to turn to John Thompson. We could not possibly start any discussion without knowing where we are, John, and we count on you to tell us about all the things you have done and still have to do.

PLANNED DECISION PROCESS

MR. THOMPSON: I will be pretty quick so that we can give Howard a lot of time. I will just mention a few things. We have finished all of our major field operations. As Ken said, we are out of the field, we now have the data back in the offices. There will be one more field operation, which Howard will probably talk about, and that is a follow-up operation as part of the A.C.E.

We are right now in the process of closing down our local census offices. All but 43 have been closed, and our schedule calls for 36 being closed by the middle of October and the remaining 7 by the end of the month, and we are moving right along on that.

On our data capture, we have been capturing the data in two phases. We call it pass 1 and pass 2. The first pass we collect and capture all of the information, short-form information, including short-form information from long forms. That is finished.

We are now doing the second pass, where we go back and capture the long-form information. What this entails is reprocessing the long-form data. We have already scanned them into our computers and we have stored them. We reprocess them and run our optical mark recognition software and optical character recognition software on them, and then we send what we cannot recognize to clerks for data entry. That is also moving along on schedule; in fact, it is probably moving along just a little ahead of schedule. We are very pleased with that.

Data capture centers: we are scheduled to have all the Title 13 materials out of our data capture centers by December 1st. This will allow us to go in and start the de-installation process in order to return the centers to the states that the leases require.

We still, though, have quite a bit of work to do. We are now in the process of doing a lot of computer editing of our data files. There are a couple of major deliverables that we are shooting to achieve. One is—and you will hear some of the Census Bureau people talk about this—what we call our “census unedited file.”

This is basically a file where we pull together all the information we have collected and we have a file organized, one record for every housing unit in the United States, and for many housing units we have had—I should not say “many,” but a certain proportion of housing units—more than one response because of our census process allowing multiple responses.

We have a process in place that puts those responses together into one response for each household. We are running that right now and producing our census edited file. The census edited file is a big deliverable, because that goes into the computer matching and then the clerical matching for the Accuracy and Coverage Evaluation—Howard will probably talk a little bit about that.

Jay Waite tells me that we have started the computer matching for the A.C.E. and that is well under way. The next big step for the A.C.E. is clerical matching, which will start October 11th, and we are on schedule to hit that deliverable.

The census unedited file also goes through a process that should not be that surprising, an editing process, where we edit the file for inconsistencies, we do statistical imputation to correct for missing data and nonresponse to the characteristics. That produces the next major deliverable, called the “census edited file,” which then is used to tabulate the census tabulations, including the apportionment and the redistricting data.

I will just finish the update by noting that we have announced that we anticipate a surplus of approximately \$300 million. The surplus resulted from a number of factors, primarily because we did get a higher-than-expected response rate and in our field operations our management of our field offices was more efficient than we anticipated and they did not have to hire as many people. This allowed us to realize some savings. We will probably be talking about that later in the year as we analyze these data more.

With that, I am going to basically conclude. We are looking forward to your discussions and hearing from Howard. I would make one final footnote for Howard. What you see today is our current thinking of the data that we are going to evaluate.

As we start deliberating and analyzing the data, we may find that some of the data will not be needed or may not be available. We may also find, as the case may be, that we need some additional data, which we will also be documenting and discussing. With that, I am done, Janet.

DR. NORWOOD: Thanks very much. I think, from what I know, the census is probably one of the largest operational efforts ever. Certainly it is the largest in the federal statistical system and I know how hard you have worked on it, so you must be relieved to be able to tell us that a good part of it, at least, is over.

MR. THOMPSON: We still have quite a bit left to do. My friend, Jay Waite, down there, and I are not claiming victory yet.

DR. NORWOOD: I know enough about the preparation of data to know that there is a lot more to be done. I would like to turn now to Howard. Let me say, before Howard begins, that I would like this to be, for the rest of the day, really, as informal as possible. I encourage those of you sitting around the table—certainly, the panel, I know, does not need any encouragement to raise any questions and to intervene—but I would like to underscore that they should and that the others of you, all of our invited guests, please do the same, because what we want to accomplish today is a full understanding of the approach that the Census Bureau is taking and if we think there are things that should be done that Howard has not talked about, it would be useful for Howard to know about them.

If we do not understand something, we ought to speak up. If you have any trouble in getting recognized, please somehow put your name tag up on end so that I can see it and I will be sure to try to call on you. Howard?

REVIEW OF THE QUALITY OF THE UNADJUSTED CENSUS

DR. HOGAN: Thank you. I would like to begin by thanking the NAS and the panel for inviting us here today to publicly release our plans. I would like to say, walking in today and looking at the name tags and faces, I felt I was among friends—perhaps not everybody who agrees with everything we did, but friends, nonetheless.

Amongst the thanks, especially along this back row here, when people introduced themselves and they said, “Census Bureau,” “Census Bureau,” “Census Bureau,” you should have been able to match those names to the names of the documents. They put in a tremendous amount of work to get ready for these documents and to document what we are doing. I would also like to thank the chair and the staff for giving us this opportunity.

As John said, we finished [A.C.E.] interviewing. We have started the computer matching process. We have our clerks trained and they are now doing some practice clusters and everything else, so that we will be able to begin clerical matching as soon as things get through the computer matching pipeline, so we are on a roll here and very happy with where we are.

The documents you have are designed as input to the Executive Steering Committee for A.C.E. Policy, or ESCAP. I am going to go through them very briefly and then we will go through them in great detail.

[NOTE: The documents provided by the Census Bureau for the October 2, 2000, workshop were drafts, containing background text, proposed topics for inclusion, and table shells, indicating the kinds of analysis the Bureau proposed to conduct to inform ESCAP. The documents contained no results from the census or the A.C.E. Each of the sixteen documents provided in the "B-series" of memoranda corresponds to a document that the Bureau released March 1, 2001, containing full text and tabular and other analytic results. References to these documents in the workshop proceedings are to the final, published 2001 version of each. References by speakers to page numbers in the draft documents have been retained, although they do not necessarily, or likely, correspond to the published version. Note also that, in addition to the completed versions of the 16 documents for which the Bureau provided drafts at the workshop, the Census Bureau released another 3 documents in the B-series (Griffin, 2001b; Mulry and Spencer, 2001; Navarro and Olson, 2001). All of the documents released on March 1, 2001, are available at <http://www.census.gov/dmd/www/EscapRep.html>.]

DR. HOGAN [cont'd]: The first one, which is pretty much how I will be talking today, is the one by me, and it gives an overview [Hogan, 2001]. This is followed by another overview, written by Jim Farber [2001a]. This synthesizes the data in one place, sort of the *Reader's Digest* version of the documents.

There is a document on the quality of the census processes [Baumgardner et al., 2001]. There is a document on demographic analysis [Robinson, 2001]. Then there is another document on demographic full count review [Batutis, 2001].

There are a number of documents that are on the various A.C.E. processes, and there is, finally, a sort of orphan document on the multiplicity estimator that we use to measure the service-based enumeration of the population that includes what many people call the homeless [Griffin, 2001a].

I will be referencing these documents as we go along today, but I will principally be following the document that I wrote [Hogan, 2001]. These documents say where we are today, what we plan to do. We worked very hard to get here. In planning for the A.C.E., we tried very, very hard to pre-specify everything to say exactly what we were going to do, lay it out there, and, barring very unusual circumstances, that is what we plan on doing. Pre-specification we felt was quite important for how we came up with the numbers that we are preparing for possible use for adjustment.

The documents here we do not really feel are appropriate for complete pre-specification. The idea of making this decision, of deciding which data sets are more likely to be more accurate, the corrected or the uncorrected, we do not feel calls for pre-specification. We are taking what we think is our best shot today. If more data flow in that we had not even thought about but would make one believe one set was or was not more accurate, then those are data we plan to collect and we plan to put, first, in front of the ESCAP and then, as the director said, later, publicly.

Similarly, this is a fairly ambitious pile of reports here; not everything may get done. At some point we may decide that something is impossible, that data we thought were easily gotten were impossible to get, and we, as the staff to the ESCAP and, I believe, the ESCAP, are going to make the decision based on the information we get in front of them.

I really did want to make a distinction between my last few presentations, where pre-specification was very important, and our philosophy here, which is we want to put before the decision-makers the best information that we possibly can put before them. I assume that you have all read all the documents that we provided.

DR. NORWOOD: Our panel usually does its homework.

DR. HOGAN: Yes. Now I am going to pretty much walk through document one, which is the one signed by me. It is the data analysis to inform the ESCAP recommendation. If you turn to page 2 of that document [Hogan, 2001], you can see the structure of the document and, not coincidentally, that follows, or vice versa, the structure of today's agenda.

I will first talk somewhat about the quality of the uncorrected census; do a review of the A.C.E. operations, did we do what we pre-specified, did we do it well; review A.C.E. quality, that is, the various kinds of errors you get in dual-systems estimation—matching bias, correlation bias, et cetera; and then part of that review of A.C.E. quality is a discussion of how one might synthesize those errors into an overall measure.

It is not on the agenda but I will probably at least briefly talk about it later, the assessment of the service-based enumeration results, and then a discussion about performing the overall assessment. That is the structure of the paper and the structure of today's discussion.

Literally turning now to page 3, at our last meeting the director pretty much laid out one way of looking at this problem, which is this 2×2 table and I am a big fan of 2×2 tables myself. This 2×2 table, the census was very good or the census was not so good, the A.C.E. was very good, the A.C.E. was not so good.

The decision process said, well, if you had a very good census, would you need to even use a good A.C.E.? If you had a poorer census but a good A.C.E., what would you do? A good census and a poor A.C.E., it is probably clear what you would do. A poor census and a poor A.C.E. is a nightmare for many of us, and we hope we are not in that quadrant.

The first step in reviewing whether to proceed with this whole adjustment/correction decision is to look at the census results to see what we know, to see if the level of the undercount, the net undercount, is significantly changed from what we saw in previous censuses, to see if the differential undercount was significantly changed from what we have seen in previous censuses.

In other words, this controversy, both statistically and policy-wise for our nation, was driven to a great extent by a problem, a problem of differential undercount, a historically documented problem. The A.C.E. was designed with a view to how this census might likely come out in terms of the undercount.

We have made a number of improvements in the census: paid advertising, local updates of census information, a simplified questionnaire. One of the things that I think made a huge difference was the pay scale. All of these things might indeed result in a sort of net coverage of the census that is quite different from what we got in previous censuses. That is the beginning point of reviewing the decisions.

Also, we have to look at not just the net national coverage, what can be measured by the demographic analysis, for example, but also the gross errors, the omissions, the fictitious and other erroneous enumerations, the level of census imputations, and we have to review census heterogeneity. It is one thing to say in 1990 that the undercount was 1.8 percent net national; how did that distribute by geographical area, by group, by city, by whatever else, how did that heterogeneity result from overcounts or undercounts, whatever? The review of the decision process, a review of the data in anticipation of the decision process needs to look at what the census is that we might adjust. This decision process we see (or at least it can be thought of) as in two parts, one of which is comparing the census—I am now on page 4—to demographic analysis and external reviews and then an internal review of the census data.

On page 4, the documents that we relate to a comparison with demographic analysis [DA] and demographic estimates are the demographic analysis results documents [Robinson, 2001]. You need not pull that out, I am just referencing it for you. There is also the demographic full count review document [Batutis, 2001]. The review of Census 2000 will begin with basically the demographic analysis results, which are available, as almost all of you know, black/nonblack.

We have documented in previous censuses, together with published articles, the levels of uncertainties, the kinds of limitations that DA might have and, clearly, any comparison of the census and the DA must start with the limitations of the DA as well as the limitations of the census, and many of those that we documented in 1990 remain in 2000.

I do want to mention two things that maybe were not mentioned for 2000, one of which is race reporting. The demographic analysis estimates are based largely on vital registration records. Vital registration reports race in its own way. For example, in birth registration statistics, race of mother, race of father are reported, not the race of the child.

In Census 2000 we now have the option of people reporting more than one race. As those data flow in, the people doing demographic analysis need to understand what that means in terms of comparisons of the demographic historic data series to the census as collected in 2000, so that is a new wrinkle that we will have to take into account in any comparisons we do.

The other wrinkle, which may not be important at this point but it will be later in my discussion, is that demographic analysis measures the undercount of the total population. The Accuracy and Coverage Evaluation measures the undercount of the household population, the difference being, obviously, the group quarters population. The group quarters population includes mental homes, prisons, college dormitories, and other group settings. This may make the comparisons of the census coverage as measured by demographic analysis and the census coverage as

measured by the A.C.E., I will not say difficult, but at least you have to be cognizant of the fact that they are measuring slightly different universes. For some age groups, principally 18 to 29, those differences might be important and might have to be taken into account.

Certainly that is where we will start at this. Demographic analysis, A.C.E., and census at some point will become sort of the triangle of what we are trying to reconcile and understand, the levels of the undercounts as measured, levels of the population, the patterns of the undercounts, how they work out.

One of the most important things from demographic analysis is, of course, the sex ratios. How do the sex ratios of the census compare to what you would expect from demographic analysis? How do the sex ratios of the A.C.E. compare to what you would expect from the demographic analysis? Our role here is to make sense of all of these things.

Let me mention two things and then I will pause here long enough for questions. Historic demographic analysis, starting with birth registration and death registration, is basically available for black and nonblack and that is the data series that we have all been looking at for many, many years.

In addition, the Population Division in the U.S. Census Bureau has various other population estimates that start with census 1990 and carry forward at all sorts of levels, which includes, I believe, Hispanic and other groups besides black and nonblack. Traditionally the population estimates have not been corrected for undercount, but taking these estimates together with what we know about the 1990 undercount we can come up with some other benchmarks for some of these other groups and say, well, is the Hispanic population as measured by the census or the A.C.E. consistent with what we would expect?

In addition, the Population Division and the Housing [and Household Economic Statistics] Division have different programs in terms of housing unit stock, all sorts of other things that are not as well known in the undercount community—they are extremely well known in other communities—that will let us understand the Census 2000 and understand the A.C.E.

Finally, in a program that I think is quite wonderful and not designed for the A.C.E., but we are going to ride its coattails, we will be bringing in state demographers, local demographers, to look over the census results and really understand at a local level how the census worked, bringing in local information, providing a wealth of data on local problems, state problems, regional problems.

The purpose of this—what is called “full count review”—is to alert us to problems in time to correct them, to correct them before the A.C.E. ever comes along, but it is going to provide a lot of information that those people, first, the ESCAP and the Census Bureau and, later, outside people will want to look at to understand how the census really worked at the local level and understand the heterogeneity of the census and how that might have been picked up or not picked up by the A.C.E.

DR. NORWOOD: When will that local review take place?

DR. HOGAN: I will refer to John there. Is there anything before I go on to the next document?

DR. OLKIN: Just a comment. I was struck by Ken's comment about the use of the terms "failed census" and "successful census," but you use the terms "good" and "bad" census, so maybe you had better think of some new words—because if you keep using that, everybody will use it.

DR. PREWITT: Thank you.

DR. YLVISAKER: I would also raise what is new terminology to me, and that is "corrected" and "uncorrected" census. I do not quite understand the change from "adjusted" and "census." That is, I do not view "uncorrected census" as a neutral term.

DR. HOGAN: My viewpoint and I am happy throughout the day to call it the "adjusted census" if people are more comfortable. The words float around to me. We would go forward with it only if we believed the adjustment represented a correction, so I have been using them interchangeably. I understand the criticism. I will today try to use the word "adjusted." I have no problem with that at all.

DR. STOTO: Howard, I am glad to see your focus on the implications of a change in definition of race and ethnicity in this. I think it is important that you have noticed that. I wonder if, either now or at sometime later today, you will be able to say something about exactly how you are going to address that in demographic analysis? The other variant of that is whether anybody has looked at the results from the dress rehearsal census or other data to suggest how big a problem that might be?

DR. HOGAN: We have started and we have people here who can address it; it just depends on how we want to use the time—Greg Robinson is here as well as John Long.

We have looked at the dress rehearsal (I know Greg has). Unfortunately, it was two, what we would call not very representative, sites in the nation, Sacramento and South Carolina. Also, some of the publicity that surrounded Census 2000 itself in terms of groups saying "do mark ours," "do not mark theirs," was not present in the dress rehearsals, so until we get the data in I am not sure you can generalize from the dress rehearsals.

At the very beginning one can simply look at people who mark more than one race and put them into two categories. Remember the actual demographic analysis, the traditional data series, is black/nonblack, so someone who chooses white and Asian does not affect that, it is only people who choose black and something else. If that is, indeed, a large group, then that can really fuddy the historic series.

If the reporting [differences] between black and nonblack are small, then the historic series will be largely preserved, although some of these other comparisons from 1990 on, say, the Asian population, might change.

I believe, going into the census, for example, that for the American Indian population the people who might mark white and American Indian might be quite large, and that is represented in some of the ways we post-stratified for the A.C.E. Does that answer your question or do you want Greg to address it in more detail?

DR. STOTO: It goes in the right direction. The issue is how many people mark black and something else versus just black, and what is the difference between those two? It may be only a couple of percentage points but, of course, that is the nature of the undercount as well.

DR. NORWOOD: Why don't you continue?

DR. HOGAN: All right, at this point of the analysis we will have at an aggregate level how the census compared to the demographic analysis, the demographic estimates, and the full-count review.

The next step is just to look over the census in somewhat more detail. It is internal measures. That is in the [B-3] document [Baumgardner et al., 2001].

Basically it says we have run the census and in running the census we have kept track of a lot of things. Breaking it down in very crude terms, running a census means you compile the address list, you deliver questionnaires, you follow up for nonresponse and for coverage errors, and then you process the data. If it were that simple, we all would not have jobs.

For each of those activities we have measured directly what was processed through that, what came in through the various local updates of census information, what was unduplicated, who mailed back questionnaires and to where they mailed them back, where nonresponse follow-up went very well and where it went very, very well. We have our coverage edit follow-up where we telephoned people, for example, who had households bigger than seven or more (there are only six slots on the questionnaire). We have information about how that went, who got called, when we got wrong numbers or could not get through. We have the data from the "Be Counted" program. We have a lot of information coming in about how well the census went nationally, statewide, local, whatever, and many of these activities had quality assurance programs that actually controlled the quality and, as a spin off, provided some measure of the quality.

This is a mass of data, but we want to really look through these data, again, to understand gaps or special localized problems, gaps in the address lists, duplicates in the address lists, levels of missed geocoding—this relates to the quality of the census but will also relate to the quality of the dual-systems estimate and the synthetic adjustment later—level of housing/person duplicates, as I mentioned, large household follow-up, how often do we get through to these households, what were the results?

It says a lot about the census but will also later explain if we, for example, on the A.C.E. have a lot of non-matched children. People say, well, gee, you had a lot of unmatched children, this is new. If we also have data from these processes that say (and I have not looked at the data, so I am just, as with everything today, speculating) we got through to a lot of these people or we did not get through to a lot of these people, this ties together with the consistent story of how the census worked.

DR. OLKIN: You keep using the term "quality assurance." "Quality" is such an amorphous term. Can you describe a little bit more carefully what....

DR. HOGAN: In the census we really had two sorts of quality assurance activities. The one that I am really not talking about, but let me mention it, is we had an extensive software quality assurance of testing, test decks, or whatever, to make sure the programs were written according to specs and ran according to specs.

In addition, we had more traditional quality assurance programs. For example and I will probably get this slightly wrong but I will have the spirit right, in non-response follow-up we hired hundreds of thousands of off-the-street interviewers.

We had a rule that when they started their assignments, for one out of every 10 of their first 70 interviews we called the respondent back to make sure that that interviewer had gone to that house, had done his or her job, had not fabricated, and then we had some other targeted activities. That is an example of where we had designed a quality assurance program to make sure that this activity was in control.

On a number of these other activities we have similar processes. They are not extensive re-interviews; I do not want to exaggerate. What they do in the non-response follow-up quality control is just to make sure the persons [interviewers] went to the households, that they did not completely get lost or make up data. We do not re-interview everybody and double-check that they understood and asked every question; it is really fairly basic quality control, but it gives us data back on how the census was run, not just nationally but locally, and that is the kind of stuff that I am referencing here.

DR. BROWN: I have two questions. Is this the document where there was a housing match done between the census maps and the A.C.E. maps?

DR. HOGAN: No, this is different. Had we not done an A.C.E. at all, we would still have monitored the census mailout, the census enumeration process, the census address list compilation processes; we would have had various quality assurance management measures that said how well that census was done before we even dreamt of an A.C.E.

DR. BROWN: That is what I thought. You just had said something that I kind of misinterpreted or thought could be misinterpreted.

I want to pick up on Ingram's question just a little more. Quality is a very amorphous kind of concept. I applaud all of these kinds of data, I think they are very useful and it is very good that we have them, but have you thought about how to use them? Some of it is easy to use. If you find that things are going very well, they are going very well, so it is pretty clear, and if you find big flaws, whole areas that were unreported or the data got lost, well, you know, something major needs to be followed up but, in-between, there is a big gray area.

Suppose you find—I think we expect from the dress rehearsal to find a noticeable proportion of proxy data. What do we make of that? Have you any thought? Is that good, bad?

DR. HOGAN: I have a number of thoughts, and I will go back to dress rehearsal, because this is really quite a bit of my model. In the dress rehearsal, as the panel remembers, and probably a lot of the audience, when we got the PES [post-enumeration survey] results in, especially for South Carolina, they were counterintuitive, they did not fit historic patterns at all.

We looked at them and asked what is going on here? We did not understand what they were saying. Had we just had the A.C.E. results, I think we would have just pulled out our hair, or something. However, we had information about how the census had been collected, how the address list had been built, possible gaps in that address list, so by reviewing not just the dual-systems estimates from the PES [post-enumeration survey] but also information from the census, both this kind of information and also demographic projections for the area, we could put the information together into a consistent pattern of what had happened.

In this case it was a pattern that was consistent between the census, the A.C.E., and the demographic analysis that said what the A.C.E. is measuring is probably a real phenomenon that actually happened in dress rehearsal and not a fluke, an artifact, of how you did the dual-systems estimate.

I see really very much, based on the dress rehearsal experience, that at the end of this is not, yes, it is a great census or it is a super great census, but here is what we know about how the census was conducted, here is a story. From demographic analysis we have a story about how good the census was. From A.C.E we are going to have a story.

Putting them together, and I really think of this as a triangle, putting these together at the end of the day, one very important criterion is was the A.C.E. finding consistent with measures of census quality, measures of census coverage external to the A.C.E.? If it is not, I personally, and I would think the ESCAP as a group would be very interested, but if this triangle fits together into an overall pattern, then that passes one level of reassurance.

DR. BROWN: Let me make a stab at what you are saying in terms of how I understand what is going on here. The confusion in my mind has resulted from the fact that I put emphasis on the word “assurance” in the statement “quality assurance.”

The fact is, these data, or pieces of them, at any rate, do not assure quality; they are statements about quality that need to be interpreted in terms of assurance or non-assurance in conjunction with the other pieces of the triangle that you have talked about. These are important data but by themselves do not assure quality.

DR. HOGAN: You are absolutely right. When we called it the quality assurance program, because it was designed to make sure the census did what it was designed to do, that does not mean at the end of the day (nor can you directly infer it from the data), that there was no fabrication in nonresponse follow-up, that nothing could have gotten by our airtight system of calling back 10 of the persons’ work loads or, indeed, that it was even evenly applied.

I am using “quality assurance” in sort of the way it is used in the Q.A. literature but, no, each one of these is a little piece to the puzzle. At the end of the day, do you have a picture is very much my approach here.

DR. BAILAR: What happened when you did that follow-up and you found out that the interviewers had not been there? Did you re-do the assignment?

DR. HOGAN: I think that was the plan. Jennifer Rickert is the Q.A. person on my staff and she can tell us what the plan was.

DR. PREWITT: This is extremely important. During the operations, and I think John and Jay can explain a lot of these things in great detail, during the operations we found parts of the census that we thought were substandard or below par, or whatever.

We actually went out and worked on those, so we actually improved the quality. We got a higher-than-expected population unknown count back and we went back out into the field to figure out why that was, and we actually improved the census.

But Larry is absolutely right. All of those operations where you still had time and field apparatus and resources to go out and improve the census in August, we

did it. We found, for example, unexpectedly, some small areas where we think we did miss hunks, and we went out and, I guess, added 83,000 cases to our address file in August, so there was a kind of overall improvement of the quality of the field operation.

That is somewhat separate, obviously, from what Howard is trying to get at here, so there was a bit of an assurance element in this, but at the end of the day you still have the data, so you talk about proxy results or something to which a lot of attention was drawn, vacancy rates.

Some people argued that we had too-high vacancy rates in certain areas. It turns out that we went back and looked at those areas and those were all areas where you would expect high vacancy rates, high seasonal homes. There is no automatic cutoff, vacancy rates are vacancy rates, and buildings get destroyed during the time before you do the census and after you assemble the address file. We went back into the field in a large number of cases, with the assistance of the I.G.'s office, the GAO offices, and lots of other people, and looked at things and, in that process, documented that the vacancy rates were exactly what we would expect, given local conditions.

All of that did go on, Larry, all through July and August right up into the early part of September. That is different from now looking at all of these data and making a decision about the overall "quality" of the census.

MR. THOMPSON: I think that is good. We used to call these operations quality control and now we are calling them quality assurance, because we are doing more than just sampling and spec repair. We try to look for patterns in the data that would help us. As you know, with quality control or quality assurance you do not get perfect quality, but we did not want to send that [message] to people doing the program, saying "quality maybe." We designed it with the purpose of trying to build in as much quality as possible.

DR. YLVISAKER: This thing about South Carolina, the level in South Carolina was such as to bring it to your attention. Could you speak a little more about South Carolina? You said you discovered things there, and I am asking about the level at which they were discovered.

DR. HOGAN: In South Carolina, basically, when we got our dual-systems estimates back, the undercount of owners was higher than the undercount of renters in a very confusing pattern. We had never seen that before. Literally, my first thought was that someone had coded zero for owners when they should have coded zero for renters—I mean, it was that striking. It was quite curious.

We then went to what we knew about how the address list had been built in South Carolina, the gaps in that methodology, and it turned out that if you followed the story through, how that address list had been built, it could very well (and probably, indeed, did) result in the kinds of patterns that the PES measured. That method of building address lists, by the way, we changed, and I want to point that out. In this argument I mentioned in my oral remarks that quality assurance is one thing we are going to look at. This document and this whole data analysis does not focus in on Q.A. results; that is one small piece. It is more what do we know about the processes.

For example, one thing is the census mailback rates, the census mail response rates, those patterns, having nothing to do with quality assurance that I know of, how did those compare to 1990 and previous censuses and what might that tell us about the level, quality, and patterns of coverage in Census 2000 and any comparisons you might make between 2000 and 1990?

Going back to dress rehearsal, dress rehearsal for large households, that is, households in the dress rehearsal that had more than five people, we mailed them a questionnaire [to obtain data for the additional people], and virtually none of them mailed them back, which is why for Census 2000 we switched to the telephone.

We knew that, and when the A.C.E. or the PES found many, many non-matched children, we had a perspective for that. We said, "I understand why there are a lot of non-matched children in the dress rehearsal PES, because anybody who had more than five people in the household, we did not get the information from them." I am not saying Q.A. is central—that is one little piece of this puzzle—it is what do we know about the pattern in the census, the level, the quality overall and, again, does this fit into a consistent story? I hope that came near to answering your question.

DR. YLVISAKER: At the aggregate level, yes. I am a little concerned about when the local people come in, they are interested, also, in regions and various other items that are not so familiar to them, say, as owners and renters, so they think more in terms, I would suspect, of regions. I am wondering about the level to which one goes there looking for anomalies?

DR. HOGAN: In terms of this analysis I am not sure there is a particular level that we are looking at or for, nothing too very fine, certainly not block or tract, but trying to understand state and regional differences that might affect the census quality and the A.C.E. quality.

DR. PREWITT: May I add to that, Janet? To go back to the South Carolina experience, because it is very important in terms of the general thrust of today's meeting and where I tried to start these comments, we conducted ourselves during the South Carolina (and all the dress rehearsals but, in particular, South Carolina) as if we were on exactly the same deadline we will be on.

Ironically, South Carolina was the one site that was not to have been adjusted. We were supposed to run the experiment with one unadjusted site and two adjusted sites. South Carolina was the one unadjusted site. Nevertheless, we treated it as if it had to be adjusted within the same time frame, the nine months from Census Day.

We actually said to ourselves in this process that if we could not explain this anomaly, which, initially, as Howard just said, must be a coding error, something does not look right, if we could not explain it in the time available to us, that will be available to us in 2000, we will not adjust this census, and we explained that to the secretary and we said this is live ammunition time, because we hit an anomaly in the data, we could not explain it, and we would not have trusted—because we would not know if it was a problem in the address file from the census or whether it was a problem in the administration of the PES, so we actually made that decision.

It turns out that using all of the stuff that Howard just talked about, including field people knowing something about the local conditions, the number of mobile homes that actually appear on the roles as owner units and not rental units, we could explain these patterns.

This is why having the state demographers sitting there during this period is so critical to us, because if we suddenly hit an anomaly that does not make sense, given our historical experience or the demographic analysis, then we will turn to them and say, "Can you help us make sense of that?" And we have our own data as well.

I just want to repeat the principle that is at stake here today, that the Census Bureau's decision about whether to adjust the data will be based upon its assessment of to what extent are we convinced that by applying the dual-systems estimation we are making a better census evident? If we think we are not, we will not. If we think we are, that is really what we mean by the "good/bad."

As Janet has taught us all, a census is an estimate of the truth. We have now done a series of things that we think get that estimate closer to the truth, coverage improvement, coverage edit, computer edits, field edits, quality assurance, all those things we think get the estimate closer. DSE is one more step in that process. If we do not think it gets the estimate closer, we will not use it. If we think it does get it closer, we will use it.

DR. NORWOOD: I just wanted to say that the way that I look at quality assurance is that it is one piece of data that we have used, the Census Bureau has used, for years and years in just about every survey that it does. Those data are used as just one piece of information for those of us, at least when I was in the government, to use to interpret the data, because if there were things that did not feel right or did not look right, you went back to try to find out what had happened, and one of the ways—there are lots of others—was to try to see what happened with quality assurance. That is why I used to feel very strongly that it should never be cut back, it should be only expanded, so I am pleased that you are getting at that.

One of the things that I would like you to talk about a little bit more, and maybe that should be for more detailed discussion later, is the whole unduplication process. We could defer that until we get into more detail, because you are giving the overview and I do not want to interrupt that, but I do not want to have that omitted. I think it is terribly important.

We do not really know, but some of us suspect, that there may be more duplication this time around than in the past because of some of the steps that have been taken to try to assure increased reporting and, therefore, the way in which that is handled is going to be a very important part of this whole issue of evaluating the quality of the census, but we can delay that until we get into more of the specifics.

DR. NORWOOD: We have a lot to cover. Howard has enticed us with some of the beginning things, but there is a lot more. Okay, Howard, continue.

DR. HOGAN: All right—unless there are further questions on the census before we go on.

DR. NORWOOD: Are we going to get back to this or do you want to talk about duplication now? Why don't you tell us a little bit about that, since we have time?

REMARKS ON POSSIBLE DUPLICATION

MR. THOMPSON: Let me talk a little bit about that, since I have been worrying about it. This census, in a direct departure from previous censuses, allowed respondents more than one opportunity to be included in the census. For example, in 1990 the only way you were included in the census if you lived in a housing unit was to fill out your census questionnaire or wait for a nonresponse enumerator to come by and enumerate you.

In Census 2000 we provided opportunities for more than one opportunity to respond. You could have filled out your census form. If you thought you had not been counted, you could go and get a "Be Counted" form, fill that out, send it in. You could also call the telephone question assistance number and give your interview over the phone. Also, if you were in a household that, say, spoke Spanish as the primary language, you could have requested the Spanish questionnaire and filled out that. You could have also filled out your English questionnaire.

Finally, we had a fairly—not fairly—we had a really what I considered an outstanding effort on the part of local governments to review the address lists and give us updates. We got quite a few address updates through our Local Update of Census Addresses [LUCA Program], and we were very pleased with that.

Of course, the problem with this—not the problem—the potential, when you have multiple sources for response is that there is the opportunity for people to have responded more than once and to have been included more than once.

What we are doing now is, we have done a couple of things. One is, basically, we have a program that we run, we call it our primary selection algorithm [PSA]—I am not going to say too much about the algorithm, because it is what we call a census confidential algorithm. What this algorithm does is it takes the responses that we have gotten, if we have multiple responses for a housing unit, and it either puts the responses together, if we believe there is evidence that it is the same household that is responding, or if there are different households that are responding, it will select what we believe is the household that has the best chance of being the most accurate household.

This is the algorithm that we are running right now in preparing the census unedited file that is illuminating potential duplicates. That is basically where we are right now.

DR. NORWOOD: Have you tested it? Have you had experience with it before?

MR. THOMPSON: We ran this program in the dress rehearsal. We have tested this program quite a bit. It is probably the most well tested program that we have. We have gone through formal exercises and specifications and requirements review. We have run test decks through the program. We take this program so seriously, along with our other programs, that we have people in Howard's division who have programmed it independently of our production programmers to validate the output, and Howard also has people who are reviewing the output of the program

to make sure it ran correctly, just as we run the program state by state. All of our programs generate certain outputs that allow you to validate what input came in and what output is coming out, so we can check that as well, so this program is something we take very seriously and we have tested it and we are checking it as it runs.

DR. NORWOOD: Howard, am I right that if you do not catch a lot—if there should be, and I do not know whether there will be—if there were a lot of duplication, a significant amount of duplication, and some of it did not get caught, wouldn't that affect dual-systems estimation?

DR. HOGAN: Yes, it will affect it in a couple of ways. First, let's assume, using whatever processes they have done in the censuses on duplicated people, housing units, and whatever else, and we have a file, if after that point there are still duplicates in it, or fictitious things or anything else, but let's focus on duplicates, then we have to, as part of the PES or the dual-systems estimates, as part of the E-sample, specifically, measure the number of duplicates on a statistical basis, based on the sample, and subtract that out from the census count.

What we are trying to do, as explained in my last talk here, is to, first, try to measure the number of people correctly counted in the census and then we are going to try to measure the proportion of all people who were correctly counted. That is the essence of the dual-systems model.

The first step is to figure out, of the census count, how many are correctly there and, as part of the A.C.E., we have our E-sample, and we look for duplicates. There is a clerical operation—after we have matched everybody in the A.C.E. sample [independent P-sample] to the census sample [E-sample of census enumerations in the A.C.E. sample block clusters], we have some leftover people, and one of the first things we do is check to see if they are duplicates.

The effect of duplication on the A.C.E. is in a couple of ways. One is we are measuring it on a sample basis and so it affects the variance. If the duplicates that get through the process to the DSE [dual-systems estimate] are clustered or highly clustered, that can have an important effect on the A.C.E.

Secondly, and this is true whether or not we adjust, the level of duplication, if it is uneven, contributes to heterogeneity in the census and to the extent that that unevenness is not related to any of our post-stratification variables, then that unevenness will remain after we ratio-adjust the census, the actual final step.

To the extent that there is duplication that is clustered, it can raise the heterogeneity in the unadjusted census and it can raise the heterogeneity in the adjusted census and, of course, if, in a post-stratum, the level of duplication or other census errors is such as to create a gross overcount, that is, if there are more people duplicated than missed, that means we will have—we would expect to have a coverage correction factor of less than one; that is, when we go in there we would say in this post-stratum there should have been 10 million people, based on our base assessment from the dual-systems estimate, the census counted 10,500,000 and we now have to reduce the census measures by half a million and then we will spread that out across the post-strata, so there are essentially three effects of this.

DR. NORWOOD: Of course, you hope that you will catch them all, as John has said with your wonderful algorithm. And I do, too.

DR. HOGAN: We also hope we get all the missed people, too. No, some will slip through and, one thing, and this is what I am really talking about in this section, is in this census we actually have captured the names for everyone. In 1990 we had names for only the PES sample, so we had no independent check on how well we were measuring the duplicates.

This time we have the opportunity to draw samples or do some research outside the A.C.E. sample, perhaps not as good as the A.C.E. sample, because it will not be backed up, probably, by clerical matching, but it will give us an idea of what the A.C.E. is measuring and how that relates to something outside our sample; again, an independent verification of what we have measured. That is what I was really referring to here.

DR. BROWN: That would be an evaluation study that would happen some time in the future?

DR. HOGAN: We are seeing if we can get some information in time for the ESCAP. It is a matter of looking at when files are available and when programs can be written. It probably would not be backed up by any clerical matching, so it would be very gross.

DR. BELL: John described testing and that seems like testing of the algorithm itself in the programming, but the question of evaluating to what extent it is finding the right people, in other words, not selecting out people who should be in and not selecting in people who should be some place else or who should not be at this household, what data are going to be available for that?

MS. KILLION: The A.C.E. is not the only thing where we run the program and evaluate it, finishing all the evaluation we would like later. There is an evaluation of the primary selection algorithm, I believe, on the books. The results of that will not be available until 2003, late 2002, at the earliest, so the kind of careful after-the-fact evaluation to really see if it did what it was intended to do is well after this decision has to be made.

DR. BELL: I am not sure that that is critical to this particular decision. Is that based on field follow-up work? I see a nodding "yes" from Ruth Ann [Killion].

DR. BROWN: I have two questions here. I want to try to separate them. One really does not have much to do with the statistical issues but it has been raised, I mean it is a natural question, and Ken is here, so it is really a question for Ken or John.

This census had, as John said, a number of new kinds of opportunities for people to respond with the risk that it might encourage duplicate response. How much duplicate response and how it is filtered out has been the topic here, but sort of overall, is there a way to evaluate whether these programs were good programs and "good" could be interpreted in various ways.

In particular, the "Be Counted" forms were used a lot less than one might have worried that they would be used. That is good in terms of duplication but does that mean that it was an unnecessary program, to begin with?

DR. PREWITT: Let me start. To speak specifically to the “Be Counted” forms, Norman is not here today, and Norman Bradburn, of course, has been a voice of great concern about the “Be Counted” process, including the opportunity for fraudulent or aggressive use of the census to be counted.

We had in place an algorithm, if you will, for some threshold issues in terms of both the overall use of and clustering of the “Be Counted” forms and we had in place a corrective procedure. When the “Be Counted” forms came in, we did not have to use the corrective procedure. We were very pleased at that, as a matter of fact.

Therefore, we thought that our standard de-duplication process would be able to clean it—you know, there is still stuff, but it will clean it up. It is not as if you suddenly had a city some place that said, “My gosh, let’s really go after it,” so forth and so on. We were fairly confident that we did not hit real danger thresholds on the “Be Counted.” I think the same thing is true with the telephone assistance stuff. It patterned the way you would expect it to pattern. Its language use was roughly what you would expect, and so forth.

In terms of kind of compromising the quality of the census, we are sort of pleased, but now you ask the other half of the question. Well, yes, but given the amount of effort, attention, concern, and so forth, is it worth the investment? We will evaluate that, and I think the jury is still out on that.

Obviously, every real case you can get in that you would otherwise not have gotten in is an improvement of the census, even if it is a reasonably small number. Whether, as a mechanism of coverage, the “Be Counted” or the telephone assistance turns out to be the most important thing, it is hard to say at this stage, we just do not know. We are fairly confident that the language program improved the quality of the census, and certainly we learned some things that we would do differently in 2010 about the language program, without going through that now.

We certainly learned about the Internet; we have no reason to believe that there could not be a huge multiple of those 70,000 cases in 2010, and we are pleased with what we learned about that in the census environment, so we are quite confident of being able to use Internet filing into the future. Do you want to talk about how we are going to assess the “Be Counted?”

MR. THOMPSON: We are going to assess all these programs that we ran. We are also going to be looking at the kind of census we can put together for 2010, taking into account the environment in 2010, which will also include what kind of census can we build in conjunction with the American Community Survey, which presents a whole new range of opportunities for innovative reengineering of the census and how we can also build in some improvements to our geographic data base.

We have some initiatives where we want to bring our TIGER [Topologically Integrated Geographic Encoding and Referencing System] file, for example, into line with true geographic locations, true GPS [Global Positioning System] coordinates, which will then allow us to do a lot of very innovative things and build in the address list and send enumerators out, but this is the kind of program we have to get going in the next year to put it together.

There are lots of things we are going to look at, including the effectiveness of the programs and how they would build into this potentially new census-taking environment.

DR. NORWOOD: Howard?

DR. HOGAN: If that completes the questions on comparing the DA to the census and the census to the DA, we will then proceed to....

DR. NORWOOD: Excuse me. Yes, Marty?

DR. WELLS: When you are in the room trying to make a decision for your meeting, how are you going to weigh the local issues of quality to the more national issues? You can imagine a local demographer coming in and saying there is a problem in this area, sets of local demographers having issues, where is the cutoff? I guess that is the question I have. How do you weigh one against the other with a single decision?

DR. HOGAN: First, I think, let me mention, again, we have this program with the Population Division and the state demographers to come in well before the A.C.E. to do just that, to bring localized knowledge to the table and see if we can get localized problems corrected. I think that that is a very good program and I hope that clears these up. When we get to the decision on the A.C.E., the adjusted versus the unadjusted files, then we have to ask whether the adjustment made an improvement, reduced the problem, was basically neutral, or made the problem worse. I believe—but we will wait for the data before we claim victory—many of these local issues are going to be situations where the A.C.E. might make a small improvement, because the post-stratification might capture some of them, but the truly local problems, if they are not caught by the full-count review, are going to there before we correct the census, adjust the census, and they are going to, probably, be there after we adjust the census.

The dilemma will really be if we have an area where we make some things worse and, demonstrably, can actually show that we are making something worse. That is certainly imaginable. Then we have one decision to make and we have to weigh whether the overall improvement outweighs any degradation that might happen locally.

In a sense, that is a price we pay for pre-specification, in my mind, where we have said this is what we are going to do and we will either do it or we will not. In a relaxed, more academic, world one might say, well, I will adjust here but I will not adjust there, because my gut feeling is it would be better, but with pre-specification, which I think is very important (I am not running that down as a paradigm), one of the prices of that is we are making one decision to apply one methodology and balancing what it can improve and what it might make worse.

DR. PREWITT: For everyone's benefit, sort of the philosophy that we have, at least thus far, discussed in the ESCAP, sort of goes as follows. We have a dichotomous decision to make: Yes/no. However, there may be a dozen or more things that are going to go into making that decision. We have decided not to treat each of those individual components, the variance, the match rate, or the correlation bias as dichotomous decisions; if above this threshold, then, yes, if below that threshold, then, no. We really are thinking that the only way to do this intelligently is to

look at the pattern of data. We know that at the final moment we have a yes/no decision to be made, but we are not treating each of the data sets that are going to feed into that as themselves threshold issues, and that is just our judgment about how to best go about this process.

Every member of the [ESCAP] committee is probably here today—13 people? And every one of them will have a different set of expertise. Some will be talking to the state and local demographers, making adjustments. Others are looking at the statistical patterns. They are all looking at things in order to bring that judgment to bear; that is how it will actually happen.

I just want you to know that we are not sitting here with some sort of magic—a dozen different thresholds and if they are above or below that magic point, then X. It is really going to be an issue of judgment, statistical and professional judgment.

DR. BELL: I want to follow up on something Howard said. I agree with what you said about if there is a place that is local, where the census was very bad, A.C.E. will probably be relatively neutral or perhaps improve it a little bit. There is a flip side to that, which is a place where there is a big problem, can also cause influential cases, or outliers, to the A.C.E. Are you going to discuss at some point today your plans for influential block clusters and, also, how that is going to come into the decision that is finally made? Now may not be the right time for that.

DR. HOGAN: I forget where we are going to get to that. Let me say a few words now and, if I forget later, at least I will have said this.

We are handling influential clusters in the A.C.E. in three ways, one of which is we are trying to identify them as they happen, identify them coming out of, say, the housing unit matching, the person matching, the follow-up, so that as part of our matching and follow-up process we can gain information to see if we are measuring something that is real in the census or an error in the A.C.E.

At the end of this, to the extent that what we are measuring is on the A.C.E., we will have identified that and, I hope, corrected that. At the very least we will have—the staff who do that in Jeffersonville, Indiana, who do this work are going to document that, so we will have that.

Related to that, we also have (and I will talk about this later) the targeted extended search, where, if the local census problem is a missed geocoding problem, we can expand the search area and take care of that in a statistical way. Third, assuming there was a problem in the census and not in the A.C.E. and it is not handled by the targeted extended search, then we do have a pre-specified weight reduction mechanism that says if the influence of this block cluster is more than—I forget what the cutoff is—5,000? (which is the difference between omissions and erroneous inclusions), then we will downweight that, and that is pre-specified, a much lower cutoff than we had in 1990, again, to handle it statistically.³

³The weight of a cluster on an American Indian reservation was trimmed if its weighted net error (weighted omissions minus weighted erroneous enumerations) exceeded 6,250; elsewhere, the weighted cluster net error had to exceed 75,000 for the cluster weight to be trimmed (see Farber, 2001a:29).

The final [point] is all this information—which clusters we downweighted, which ones we had to do additional field work on, and the journals from the matching clerks that say here are the problems we encountered, here are the unique situations out there—that will be documented and will be made available to the ESCAP to make its decision. There are three or four levels of handling them that are out there.

DR. HOGAN: We have, coming out of the preliminary housing unit matching—our clerical staff have been identifying and documenting problems. They have been finding questions they want to ask, the results of quality control and re-listing, yes, so this has started and will continue.

DR. BROWN: I am really just looking ahead, Howard. You can tell me you are going to talk about this later and I will remember to ask you. I mentioned it before already, but I do not see it in any of these reports. Do any of these reports give the data that describe the address list match between the two mapping processes, the census and the A.C.E.?

DR. HOGAN: No, not really, and the reason is that the way we process the A.C.E.—let me review that for the other people—is we have the A.C.E. listing of housing units and, based on the census files [Master Address File, MAF] as they existed in January, we matched the housing units from the census to the A.C.E.

The reason we did this was to help us draw our targeted extended search sample to facilitate the matching clerks and also to allow us to identify which housing units had already returned a census questionnaire to allow us to start the telephoning, because we did not want to telephone until they returned the census questionnaire.

This was an operational match; it was not, really, in my mind, a statistical match. The reason is it was to the January census [MAF] files. Why is this important? Let's take the parts of the country where we do update/leave, where part of the job is to find out addresses that the census did not have and add them to the files and leave a questionnaire.

We do not have that on this file nor as part of the various processes where the interviewer goes out to do two addresses and finds out that they are the same and deletes one. We did not have that. Any statistical inference from that match is really an inference of the quality of the census address list before the census began, and I do not know what ESCAP or anybody else could infer from that.

DR. BROWN: It is information of the same kind as some of the other Q.A. information. You might discover that in some areas the address lists were very poor, demonstrably poor, as of January, and in some other areas you might discover that they seemed to be okay as of January. It depends on what else happened in the process.

DR. HOGAN: We can think that through, again. I mean the extent of the person matching, which is against the final census file in terms of both person and households as opposed to housing units, should give us much the same flavor and, really, in my mind, that is where the focus should be. To the extent there are housing unit issues, those data can be made available to the ESCAP, but it was not part of this package for the reasons I gave.

DR. BROWN: I guess I want to ask another question at some point, but I will ask it now. That has to do with the LUCA process, which is also an address list issue. There are a lot of data there, I guess, as to what was added and what was not. The question is, how does the addition of those addresses—differentially, some areas were very active and some areas were inactive—how does that affect the census process? Areas that were active, does that increase their census count? Does it decrease their undercount?

MR. THOMPSON: That is sort of along the same lines of our Partnership Program and some of the other programs we had. We hope and we believe that through the partnership and the efforts in reviewing our files and giving us input it made the ultimate census better. That is something that we are sure happened. What the A.C.E. is doing is measuring the overall quality of the final census, which includes the efforts of our many partners to improve it on the local level. Is that what you are getting at?

DR. BROWN: Well, yes, but what I really want to do at some point, and this may really be evaluation data rather than something that can be looked at now, but to get some sense as to whether these processes really did help or not, or did you end up adding a lot of addresses that were not real addresses and caused more problems than....

MR. THOMPSON: We really want to evaluate the LUCA Program, we want to evaluate, to the extent we can, our partners who made a lot of effort, to try to ascertain and understand what went on in the census, so we do want to evaluate the LUCA Program. That is one of our major goals.

DR. YLVISAKER: I am a little bit confounded with expectations, consistency, and so on, that is, we find many people sitting over in this town and we did not know we had those people. They also have a program going on. Was it the program or did we not understand.

DR. PREWITT: The number of programs that happened around Census 2000, it is going to be extremely difficult to parse them all out. You have LUCA, which is an address file initiative, you have local promotional efforts, you have some areas where a lot of the promotional effort went into staff recruitment; that is, we did better in some areas with our enumerator staff, we were oversubscribed. There were other areas where we were scrambling. We had some areas where we moved enumerators from one place to another. I mean, it is so many things that were happening in this two- to three-month period.

Undoubtedly, the effort external to the Census Bureau, whether promotion, public service announcements, that kind of stuff, whether it was address file work in LUCA, whether it was recruitment work for the enumerator staff, all of those things affected the census, no doubt about it.

We think, in the aggregate, they improved the census. That does not mean in particular places we did not have a city planning unit that put a lot of addresses in that actually confused our census workers when they got out on the streets. We have a census file where apartments are 1, 2, 3, 4, 5, and a utility happens to count those as A, B, C, D, E, so we have to sort out whether it is 1, 2, 3, 4, 5, or A, B, C, D, E, and they simply got it from their utility firm. There is no doubt

but that some of those processes will have put into the census—this goes back to Larry’s first question about the “Be Counted” forms, and so forth—some of those processes will put into the census things we wish were not in there. Our primary selection algorithm, all of our work right now, is to try to get out of it, get rid of it.

In the aggregate, I think it certainly improved the census. Now we do not have a measure of that yet. We know we got a better-than-expected mailback response rate, a seriously better-than-expected response rate, and we know we were able to finish our field work on schedule, which means we had the enumerator staff, we had levels of cooperation during the enumeration process, the follow-up process, so all of those things actually happened, but sorting out the impact of any particular one of those things, we are not going to know that between now and February.

MR. WAITE: We may never actually know the exact implication of something like a LUCA, for example, because when we are out there doing nonresponse follow-up [NRFU] and when we are doing update/leave, our enumerators have a list of all of the known addresses that we have at that time. If there is another address they notice that is not on that list, their instructions are to add it.

If a community had added a bunch of addresses before that by LUCA, that would be on their list. Then you are going to be reduced to speculating, well, if LUCA had not put it there, would your NRFU interviewer have found it? The truth is we will probably never know for sure for a specific address, but a lot of our field operations started with what we knew at the time we started that field operation, so for communities that supplied addresses, that made sure that it was on the list. It is not a certitude that it would not have been on the list if they had not submitted it, so to try to parse those things out gets to be very difficult.

We had operations at virtually every step that were designed to identify any addresses that might be existent on the ground that were not in our files.

DR. YLVISAKER: So maybe this is more about the “Be Counted” forms, which are not necessarily tied to addresses so much and whose effects are not really understood very well, I suspect. We are saying we would like to evaluate and some day find out whether they mattered, and so on, but at some point you are also going to be looking at results and deciding whether this is correct or not.

DR. PREWITT: With respect to the “Be Counted” forms, of course, we did not put any “Be Counted” forms in the file where we could not go out and validate the address. We went back out to the field to make sure. There had to have been a housing unit there that was not already on our address file from which we did not already have a response.

DR. YLVISAKER: But I am talking about a city whose population suddenly seems to be much larger than you might have thought based on 1990. Was that due to “Be Counted” forms, was it due to the fact that the city grew? We do not really know. This can be an inconsistency or it might not be an inconsistency, but we do not really know, because it is confounded.

MR. WAITE: Theoretically, that is certainly true, but, as Ken mentioned before, we had a procedure in place, an algorithm in place with a triggering mechanism, where we knew how many “Be Counted” forms came in from each jurisdiction in total, and we knew something about what the 1990 population of that jurisdiction had been.

We had a mechanism in place, if there had been a huge jump in the jurisdiction, that would have triggered an additional process—not now, not during A.C.E., but during the census itself, way back in June—to have gone out there and tried to figure out what was going on. We did not feel there were any communities where that was the case. It does not mean that, theoretically, there could not have been, but we did not observe any that actually existed, so we did not implement that process.

DR. HOGAN: Let me add one thing. While I think the initial match of the January housing unit file is very hard to interpret, we do have a final housing unit match that will be done as part of the evaluation program. We will start as soon as we are done with people and that will evaluate the quality of the final census address list and, I think, will allow linking back to which process initially put this address on the list, so as part of that evaluation some of the questions that you are asking might be answered.

DR. BROWN: That is an evaluation study?

DR. HOGAN: Yes, that starts right after we are finished with the people.

REVIEW OF THE A.C.E. OPERATIONS

DR. HOGAN: I am now going to turn to page 8 of my paper [Hogan, 2001], where we are going to begin our discussion of the review of the A.C.E. operations. We need to ask several questions. We have looked at the census, how well it went, how well its operations went, how well it compared to the DA, and now we are going to turn to the A.C.E. and we are going to ask, in my mind, several questions.

First, did we do what we said we were going to do? Then, did we do it well? Then, how well was well? On page 9, section 3.1, “Proper Execution of the Steps Between Processing and Estimation” [Hogan, 2001],⁴ this, as I said, is did we do what we said we were going to do?

One of the concerns lurking in the background (Ken mentioned it earlier), is somehow this whole process was subject to outside manipulation, that we were going to pretend to be doing this and, at the very end, numbers were going to leap out, you know, “guesstimation” work, or whatever.

I think one thing we should be able to lay to bed very clearly is that what we did here was the result of the work that we pre-specified and, if there were any deviations, how they occurred and why they occurred.

There are three documents here, “Missing Data Results” [Cantwell et al., 2001]; “Decomposition of DSE Components” [Mule, 2001]; and “Dual System Estimation Results” [Davis, 2001]. Turn to “Decomposition of Dual System Estimate Components.”

When you find it, I have ripped off the last page, because it is easier to compare it to what is on page 12, but essentially, whether you rip it off or not, for each of the post-strata and post-strata groups, and a number of other aggregations and breakdowns, we are going to start essentially with what we get coming out of the clerical processing: the interviewing, whatever; the number of resolved cases; the

⁴The final version of Hogan (2001) does not contain section numbers.

sample size, weighted, unweighted; correct enumerations; the number in the E-sample; what is left over—you see the top of the prototype example; the P-sample; the movers, the nonmovers, the outmovers; the matched ones, the number of nonmovers, the number of in-movers, the number of outmovers; what is left over; out of scope; and then total matches and N.P. is the total number in the P-sample; the correct enumeration rate; the matched rate; and the correction ratio for what is resolved.

We have every bit of information for these cases that we would hope to get. Then what it does is walk you through in, I think, considerable detail, every step of the estimation process that we specified, the weighting for the household non-interview, what that added.

The numbers here, I think, are completely phony, but you will see that, for example, under the E-sample that is blank. Why is that blank? Because we have no household noninterviews on the E-sample; the sample was a sample of the cases the census got.

We then go to the resolved cases, added Census Day noninterviews [NI], then characteristic imputations, how many cases were added to this cell because we imputed the characteristics and only the characteristics? Line 5, characteristic imputation and residence; that is, we could not tell whether they actually lived there on Census Day or not, we had an ambiguous response, and how we added those, whether we added them as being a household on Census Day or not. Then we have characteristic imputation residence—ICE is imputation cell estimation.

We walked down one after another. We had to impute the match rates if we know you are there; there are some cases where we do not know; if you link to whether a census record was there; what we had from what the clerks did and what we imputed; then characteristic imputation and correct enumeration—it is on the census, the E-sample side. You should not have very much characteristic imputation on the E-sample side, because we are taking results of the census imputation as given.

We go on down there to where we had to impute whether they were correctly in the census or not, because we could not resolve it in the field. There is the number of targeted extended search cases, how they were coded, and the final numbers that will feed into the dual-systems estimate.

Each row here is defined in detail in the memo itself and it would bore me to tears, if not you, to go through this in the detail that it needs to be specified and has been specified. If you do have questions, I would be more than happy to defer them to my colleagues.

To me, the story here is, at the end of the micro-records, we will know here is what the microrecord said. That is line 1. At the end of line 12, these are the numbers we put into the dual-systems model to compute our correction factors. If any hanky-panky arises, it should be clear in one of those lines; each one of those has an estimation specification, a procedure, a program, that says this goes from that.

After you read this and verify this, I think—am very confident in this statement—you will agree that we did what we said we were going to do. There was no outside whatever.

DR. OLKIN: Howard, are there measures of uncertainty when you go through this whole process?

DR. HOGAN: This is practically an accounting process. This is just take a sample and...

DR. OLKIN: But, still, a lot of these are estimates, which have a certain—when you keep adding....

DR. HOGAN: Right, and we do have measures of variance as sort of the end result.

DR. OLKIN: Of the whole system?

DR. HOGAN: Right, yes.

DR. OLKIN: Where is that kind of—you know, how much leeway is there in these numbers? I always get nervous when I see four-place numbers.

DR. HOGAN: As I said, I conceive of this as really an accounting mechanism that says these programs were designed to do this. They did this, they added one person here and two people here, and there was no fudging or hand waving at the end.

At the end, then, we have our estimates, and we then, in the next document, we get variances, C.V.s [coefficients of variation], confidence intervals, for those estimates. This shows two things: first, that we did exactly what we showed and, also, if there was a really big bump somewhere, if a characteristic imputation really did a lot more than, say, you might expect, looking at this, you will say, “I didn’t expect characteristic imputation to do much and, by gosh, it really did a lot, that might be something I’ll want to spend time investigating.”

Whether it is a variance problem or a bias problem or a computer programming problem, if a whole lot got added based on characteristic imputation, that might be a flag to some people. Let me repeat one more time, this is, to me, an accounting problem that says here are our pre-specifications, here is what we did. They agree, they match up. Let’s put that one issue to bed and then turn to the statistical issues.

DR. EDDY: How many of these tables are there going to be?

DR. HOGAN: Lots. It started out with 448 [post-strata]—we will provide them on a spread sheet—and then various rollups, which are specified in the document to things that might be interesting to some people, but at the very least there will be 448, which is the level at which we will compute the DSE’s and then summations of that.

HOW ACCURACY OF A.C.E. AND UNADJUSTED CENSUS COUNTS COMPARE

DR. HOGAN: Let me go on to the next document (we can take these two documents together), that feeds into document 9, “Dual System Estimation Results” [Davis, 2001] If you go to the back of any of those tables there, say, Table G-1, what you have here—and, again, we will provide this for each of the 448 dual-systems estimates—is all the numbers that you will need to reconstruct the dual-systems estimates.

You will have the number of census data-defined people, the number of census insufficient information [III], and the census count [C], that will allow you to construct the first little component— $(C - II)$. Then you will have the number of P-sample nonmovers, in-movers, out-movers, weighted nonmovers, in-movers, out-movers, matched movers, nonmovers, in-movers, and out-movers.

You can then compute the second component, and then you will have the E-sample total and correct enumeration, weighted and unweighted, so you can construct the third component. Put the three components together. The paper copies will be truncated to some reasonable printed thing—I think the spreadsheet thing is going to be as many decimal points as your hearts desire.

You should be able to say here is the statistical model that you said you were using, I can reproduce it, based on the data that come out of the decomposition. We did this in 1990 and I used this a whole lot in my work in understanding what went on.

So we will then have the dual-systems estimate; its standard error; coefficient of variation; then the coverage correction factors; and the percent net undercount. So here you have it. You have the formula, you have the data, you should be able to reproduce 448 correction factors.

DR. BROWN: Will you be able to collapse some of these 448?

DR. HOGAN: Yes, absolutely. First, just to verify the computation and everything else, you need the 448, but then we will certainly do things like roll them up to the 64 post-strata groups; that is, collapsing over age and sex.

DR. BROWN: No, the question was really will it be 448 or will it be somewhat less?

DR. HOGAN: It might be somewhat less. We do have collapsing rules that say if the sample size gets too small, 100 P-sample people and I think 10 out-movers—so 100 P-sample cases, in which case we have fewer than 100 or so, we will collapse.

There are only a few groups where the sample size is even close, in my mind, with 448 strata and about 300,000 housing units, so well over half-a-million people. Most of them will be so far above 100 that it is not even going to be a problem.

Some of the smaller ones—it is no longer Asian and Pacific Islanders, it is Hawaiians and Pacific Islanders—by some of the age and sex groups we might come out too few. American Indians not living on reservations—we are able, for the American Indians living on reservations, because we knew where they were, we could oversample and ensure a sufficient sample, but for those off, it is sampling a rare population, so there are a few where we might have to collapse, and it is pre-specified.

Again, this does two things. First, it is the accounting procedure, the thing does it. Second, it gives you, actually, the base results, the coverage correction factors that will be used, the undercount rates, that one can begin, then, by summarizing them to various levels, to compare to demographic analysis, and it gives some measures of uncertainty, the coefficients of variation that one can start analyzing relative to the size of the populations.

DR. YLVISAKER: Any red flags at all as you read 448 of these? Did you discover that imputation seemed to be awfully large in this post-stratum, or things of that variety? It seems to me that you might find entries where you say, ah, look at all the imputations sitting here or some other statistic on one of these pages that looks strange to you.

DR. HOGAN: Yes, the first thing I am going to do, being one of the first people to look at it, is to check things like the log of the outliers. The first program should be well running and well tested, but I want to make sure the computation was done right. You can go back to the decomposition and work that back to see where it first jumped out, check the outlier blocks to see if we had a particular problem that might have resulted in this, or whether as many.

DR. YLVISAKER: Outlier blocks did you say?

DR. HOGAN: Outlier block clusters.

Again, we are going to have identified block clusters that looked unusual and document what the census did, what the A.C.E. did, in those. Although there is not a neat mapping between cluster and post-strata, there is an approximate mapping. One can say that these came back, probably, to this set of problems, but we have looked at these problems, we have our best numbers there, it then flows logically, or there was a problem here where something was, say, miscoded, or whatever, and did not come through.

For us it will be our first look at the process. We have the microdata, the decomposition, the DSE, and we should be able to link the two in a thread.

One other document here, which will come up again later, is, "Missing Data Results" [Cantwell et al., 2001]. Again, it lays out in even more pages the missing data models, the missing data cells. Remember, we have essentially three kinds of missing data in the A.C.E.: we have whole household non-interviews where we got a refusal and could not get an interview; we have cases where we got an interview but we did not get sex or we did not get age or race, or Hispanic origin; and, finally, we have cases in the P-sample where we could not determine whether the person lived there on Census Day or, having determined that he/[she] lived there on Census Day, we could not determine definitively whether he/[she] linked to the census or not.

On the E-sample side, we have a census enumeration where we could not determine whether it was, say, fictitious or not, it did not match to the P-sample. We went to the field for follow-up, we tried to find people who said this guy never existed or the guy did exist, it comes back unresolved on the E-sample, we do not know whether it is correct or not, we have to impute, and we do that using our imputation cell estimation—I will call it ICE from now on.

This lays out in, I think, a fair amount of detail how that is done. I know the last time you had some questions about our imputation methodology. We knew at that point that we were going to use the ICE approach, but we had not laid out all of our cells and had not laid out all of our methods. We have now, in a previous document [Cantwell et al., 2001], and this lets you verify it: Here are the specs, here is what we plan to do, here is what we did, they should agree perfectly.

If there are particular questions about this, I have a number of people here who can help me answer them.

[No response]

If not, we can move on.

As I said, what we are doing in 3.1 [Hogan, 2001] is did we do what we said we were going to do, and I think that is one thing I think we can really, I hope, as a group, lay to rest with a great consensus.

The second question is did we do it well? Did we control the person interviewing; did we control the matching; did we control the follow-up results; were the levels of missing data not inordinately large?

To address this, did we do our work well as professionals, we have documents 5 [Byrne et al., 2001], 6 [Childers et al., 2001], and 7 [Cantwell et al., 2001]. Five is person interviewing results. Six is person-matching and follow-up results. Document 7, again, is the missing data results. Documents 5 and 6, but also 7, are analogous, at least in terms of this discussion, to document 3 [Baumgardner et al., 2001], "Quality of Census 2000 Processes."

This is where we can simply look at the interviewing, how well was the interviewing done? Did we have a large number of field noninterviews? Did we have a lot of problems reconstructing Census Day address? How well did the CAPI [computer-assisted personal interviewing] instrument work? What was the timing on the interviewing, was it close to Census Day, was it close to Halloween? What was the quality control of the person interviewing? How well did we conduct this simply as a survey?

This document [B-5, Byrne et al., 2001] lays out the kinds of standard survey results that you would like to know. For example, flipping to Table 2, when did we do the interviewing? Here it is by week: when we did the interviewing. Overall, when were the interviews conducted? When was the telephone interviewing conducted? When was the personal visit interviewing conducted? Then there are some basic summaries, the median and the mean, number of weeks from Census Day. How close to Census Day did we get out there and do our work?

This is going to be at what level of detail? I mean, we could reproduce it at the end of the 448, if we need it—it gets kind of boring. Again, were there patterns, geographic patterns, in how we did it in the 13 census offices. If you go to Table 7 [Byrne et al., 2001], you can see it clearly lays out census dates, census regional offices, or A.C.E. regional offices. The statistics are the same even though the office number was different: completed interviews, sufficient partial interviews, refusals, vacant, non-existent on interview day, interviews.

Were the procedures uniformly applied? We do not expect perfect uniformity; Atlanta is indeed different from Los Angeles, and the regions around them are different. Were there localized problems? Was there a breakdown in what we did?

The first part of this document, then, is simply, as I said, basic survey stuff: when did you do your interviews; did you get complete interviews; did you get partial interviews; did you get refusals; how do the data look to a survey statistician?

Then, as we did on the census side, we also have quality assurance results from the A.C.E. person interviewing. The quality assurance program, again, was designed primarily to make sure the people were visiting the right housing units and getting an interview. They could fail to visit the right housing unit, either

because they were cheating or falsifying, whatever, or they were simply lost, so we have a quality assurance program to make sure that they were doing their jobs, and we will have the results of that Q.A. program, the Q.A. results, which show the extent to which we identified these problems and the extent to which there were patterns, regional or otherwise differential, in the fabrication or falsification or simply going to the wrong address.

Again, our Q.A. was targeted on did they visit the right address, did they ask the questions; it was not a re-interview. As a Q.A. program we did not sample them and ask all the questions again to make sure that they did everything else. We really depended on the CAPI instrument for that. Several of you got to see that instrument; it really forces the interviewers to go through, if they even show up at the door, and ask all the questions; no matter how often the respondent says, "I already told you that," they have to ask it one more time and they can blame the computer, which is what all good bureaucrats do.

We also have the estimated quality rate coming out of the Q.A. interviewing and the limitations of that kind of data—it is just documented here. Then there are lots and lots of tables.

At the end of reading this, you should be able to see not just did we do the interviewing as we said we were doing the interviewing, but we maintained (I hope it shows) a high level of quality.

DR. ZASLAVSKY: There is a concern that Larry has brought to our attention about the potential for the telephone interviewing to introduce correlation bias because of the fact that the only way to be interviewed by telephone is if you returned the census form, so that to the extent that there is some sort of mode effect associated with either the telephone interviews being better or worse or with the telephone interviews happening earlier, so that there is less potential for moving and, therefore, a greater opportunity for somebody to be interviewed during A.C.E., there is the potential for that introducing some sort of correlation bias.

The question is, is there anything, either in these tables—these tables could identify for us the extent to which there were telephone interviews and where they occurred, which post-strata they are occurring in, and things like that, but beyond telling us to what extent there was the potential for that bias, do these tables or any others provide any information that might help us evaluate whether that occurred?

DR. HOGAN: I cannot think of any directly, no. I understand the concern, and at one level I share it; it was a question I asked myself when we went to telephone. I think, if you or Larry have ideas, if we could agree on a mechanism where this might have occurred. Theoretically, whenever you have interviewing at the same time you have interviewing going on, A.C.E. interviewing at the same time as census interviewing, there is the potential for something going wrong.

I have been struggling to figure out how, exactly, that might relate in correlation bias, for example, what the statistical or other mechanisms would be. We do the telephoning only, as you pointed out, for some of the people in the audience, if we have verified we have got a questionnaire back for that address ID. Occasionally there might be a misdelivery problem and we got a questionnaire back from a neighbor but we interview only when we got a questionnaire back from that address ID.

The immediate problem of correlation bias is, I think, alleviated. There might be a more subtle mechanism, which you have alluded to, in terms of how movers get weighted or not weighted or nonresponse gets weighted or not weighted; that, I think, if it is there, is very subtle, and I might like to hear from maybe you or Larry how that thread would wind its way.

DR. ZASLAVSKY: I guess the notion is that you identify missed enumerations by way of the match rate and so the idea is that you have two housing units right next to each other and one returned the census form and the other did not return the census form. The dual-systems estimation assumption is that the A.C.E. would have the same probability of finding the two, finding a person in those two housing units.

Obviously, there is a certain type of correlation bias that is associated with the type of person who is in the household but, at least in operational terms, typically in 1990 an interviewer would have gone out to those two housing units not knowing whether anybody had responded.

In this situation, now, the unit that responded, there was a phone call, there were attempts made by a phone call and, if that was not successful, then there was an enumerator knocking on the door, so two chances to find those people and, also, the telephone offered an opportunity to find them a little bit closer to the time that they were actually there, whereas the household that had not responded got only the one opportunity, which is not as good as two, which is not a given, but certainly seems possible, to add to correlation bias.

DR. HOGAN: I think there are, in my mind, two things going on. One is it could lead to differential nonresponse in the A.C.E., depending on whether you mailed back your questionnaire. Then there is our missing-data model that missed that aspect of the non-differential nonresponse. I think that is a very real question and I think if we have not already looked at it, we need to see how we could go about it.

The other one, more subtle—and I am not sure how it plays out, which is what you alluded to earlier—is you will get a different number of movers for the telephone universe and the non-telephone universe just because of the lapse of time, and how that might perk its way through the DSE and perhaps, at the end of the day, cause a problem.

If, indeed, it does, it is a fairly subtle problem and one that, again, if you or Larry have ideas about how that might work, it might be....

DR. BROWN: Yes, the whole problem is very subtle, because if you went into this process, the A.C.E. process, thinking there is no such thing as correlation bias, then you could find this, because I think you are going to find correlation bias. None of us believes that there is no correlation bias; we just hope that it is not too big so as to disturb the final results.

The simplest thing you could do would be to produce tables as in the preceding complete set of tables, thinking of one set of tables for the telephone call-back universe and another set of tables for the personal follow-up, and if there were really no correlation bias, those tables should look the same, statistically.

DR. HOGAN: I am not sure that I agree.

DR. BROWN: By post-strata.

DR. HOGAN: Again, I am not sure.

DR. BROWN: I am not sure, either.

DR. HOGAN: I mean, we are telephoning a very specific universe, people who mailed back their questionnaires and who are willing to give a response over the telephone.

DR. BROWN: Right, so....

DR. HOGAN: And I would expect that the census coverage of that group would be very different from people who did not mail back the questionnaire or people who might be reluctant to give....

DR. BROWN: If that coverage is different within post-strata, isn't that correlation bias?

DR. HOGAN: Not necessarily. It could cause correlation bias, but correlation bias occurs only if that response is correlated with response to the A.C.E. If we gave up after telephoning, it certainly would be. If we got a complete interview for everybody, otherwise, and were able to get people into the A.C.E. equally well on the interview side as the telephone side, it would not be, so the question is to what extent in cleaning up the universe we also get the people who were reluctant.

DR. BROWN: I think we agree that mailing back—in fact, you are using it as a post-strata variable at some level as a split—mailing back your census form is a covariate that predicts being caught. Now, you used it in post-strata, but you used it only as a qualitative split into two pieces, so now when you do the telephone interview you are looking at the universe of people who mailed back the forms, so they have different response rates and different probabilities of being caught in the A.C.E., because it is conditioned on that covariate.

So tables for them should look different. I mean, if we agree at this level, then tables for those people should look different from tables for the other kind of people. Now, maybe we are defining correlation bias differently. That is a kind of correlation bias, since that is not a post-strata variable.

DR. HOGAN: It is a form of heterogeneity within—but it is an important distinction—it is a form of heterogeneity within the post-strata that can result in correlation bias if the A.C.E. interview has the same kind of heterogeneity. And it is likely, to some extent, that it does, that these later people might have moved out by then and, therefore, their chances of being in the A.C.E. are less because of the way we structured it, so it can result in correlation bias to the extent that the personal follow-up does not do a really good job of capturing them.

There is a mechanism there but it is not solely because the census has heterogeneity; it is that the census heterogeneity is correlated with the A.C.E. heterogeneity.

DR. BROWN: We can sort of pinpoint exactly how we want to describe what is going on, but the trouble is, I do not know how you measure that, because the simple attempt to measure it is the one that I said, you could construct separate tables for the telephone response group and that later group and try to compare those tables. The trouble is, I do not know what that comparison should look like, and I am stuck.

DR. HOGAN: That is where we are stuck as well. Constructing the tables is certainly possible. What one infers from them is much more difficult. To the extent that this feeds into correlation bias we pick it up when we start looking at correlation bias specifically in some of the later things, but how to isolate this aspect of what caused the correlation bias from other aspects, the heterogeneity, the telephone—we also had a number of letters we turned over to David Whitford of people saying “I already participated in the census, I’ll never participate in your A.C.E.,” which is another form of correlation bias in the other direction.

We will have evidence of correlation bias at the end of the day. It would be surprising if that problem vanished. How one traces a thread back to the telephones, to the heterogeneity, to the publicity of the refusals, I do not know.

DR. ZASLAVSKY: When I posed the question, I had no idea, either. I was hoping that somehow you guys had some magic bullet.

DR. HOGAN: I do not think we have anything specifically on this. We are measuring overall correlation bias, but that is different from this.

DR. ZASLAVSKY: Not all of your people who respond to the census give you phone numbers, right?

DR. HOGAN: No.

DR. ZASLAVSKY: So you do have a comparison of people who gave you phone numbers and people who did not who were both in the census.

DR. HOGAN: Yes. That is a possibility I had not thought of.

DR. ZASLAVSKY: This is actually one of the few cases where correlation bias could go in either direction, because you do not know with the later group that is being done by personal interview whether maybe there is a mode effect and it could go either way.

DR. HOGAN: I had not thought about that, but it may be that the people who mailed back with or without phone numbers would give us a handle.

MR. WAITE: In thinking about the time that that interview took place, you should not be under the illusion that all personal interviews take place on exactly the same day, either. There is a lot of variability between the first personal interview that was done out there and the last personal interview that was done out there.

DR. BROWN: But the average date for personal interviews is a lot later than the average date for the telephone interview.

DR. BELL: And it is done blind to whether or not they returned the census form, whether or not there was a good count from the census.

MR. WAITE: Somehow, because I know they returned the census form, I am going to elicit a different response, is that it?

DR. BELL: No, I am just saying the telephone interview is different because you know they returned the census form [when] you make that telephone call. Whereas if it is in an apartment building where you would not make the telephone call, whether you go to that apartment building early or late or whether or not you go to a particular unit early or late is independent of whether or not you got good response from the unit. There is a firewall in the operations that gives you independence.

DR. HOGAN: There are two things when you say you know when they returned the questionnaire, there are two different “you’s” there. We at the headquarters certainly do know. I am not sure that that was ever conveyed to our interviewers, and they are using exactly the same questions, walking through the exact same skip pattern, so it is just differentiating those two things.

DR. BELL: Absolutely.

DR. NORWOOD: I think that on this note it might be useful for us to break for lunch.

ROLE OF GEOGRAPHIC ACCURACY IN ADJUSTMENT DECISION

DR. NORWOOD: We seem to be a pretty good group. Many of us, at least (not all), have returned [from lunch], and perhaps we ought to see who else is out there. I want to be sure the members of the panel are here.

Okay, Howard, where do we start?

DR. HOGAN: After a brief correction, unless there are further questions, we are about ready to go on to section 4, “Review of A.C.E. Quality” [Hogan, 2001]. I have a brief correction. This morning, when we talked about the outlier cutoff, I said it was 25,000. It was actually 75,000, so let the record show the right number.

DR. NORWOOD: I have a question that is related to that part of your discussion on page 12 of your paper. There is a sentence that says: “Our analysis and comparisons will focus on the distributions of the C.V.’s at various geographic levels. We will not be comparing the C.V. for any particular city, county, or other substate entity to that entity’s 1990 level.” That is fine.

It is this next sentence I want to talk about: “The Census Bureau is required to decide whether the A.C.E. numbers as a whole are superior to the unadjusted counts and considers the C.V. of any given substate entity to be irrelevant to that determination.”

The question that I have is, does that mean that the Census Bureau has decided that it has to make a decision that only one set of numbers is better than the other, or, since the Census Bureau has to issue two sets of data, does it see any possibility of saying for this level, say, the national level, or for some level that has hundreds of thousands of people in it, we can say that the data are better or not better, and that if you are going to be allocating funds to some very small area, we cannot tell you whether there is a difference?

I guess what I am asking is what is the Census Bureau’s position on that, or is that open?

DR. PREWITT: That is a deputy director question.

DR. HOGAN: I think there are two answers to that, and one is a deputy director question, and that is, what is the policy of the government about what is official and what is not and what that means.

The other—and I will let one of these two gentlemen answer that—is to what extent we will try to inform the users, whichever set we release, about the accuracy at whatever levels.

DR. NORWOOD: But you are releasing two sets.

DR. HOGAN: Yes.

MR. THOMPSON: What we are releasing, we are releasing if we decide that the adjustment will make an improvement.

DR. NORWOOD: Let's make the assumption you have decided to adjust.

MR. THOMPSON: Then we will release the adjusted numbers and we will denominate those numbers as the official P.L. data [for redistricting]. We will also release those numbers in what we call our summary file 1 and other products. We are required by 1998 appropriations law to release the unadjusted numbers for redistricting and for Summary File 1, which we plan to do.

Summary File 1, for those of you who do not understand what I am talking about, is the first detailed summary file we put out, which has very basic short-form data on it. Then we have a Summary File 2, which has more short-form data, cross-classified, and there are Summary Files 3 and 4 for long-form data.

The only thing we would print two of would be, basically, the P.L. file, redistricting, and Summary File 1; everything else would be adjusted and available only adjusted. We would basically be saying the data we support or we denominate as redistricting data are the adjusted data, so that is what we are endorsing.

We would also describe the accuracy of the adjusted data. I have not really answered your question, but basically we are saying this is the data set to us.

DR. NORWOOD: It does not really answer my question. I do not know what is going to happen, I do not know. First of all, I do not really know what the data are going to show. Secondly, I do not know whether you are going to adjust or not.

But let's make the assumption that you will adjust. I could envision—I do not know whether it will happen—I could envision that you might say, well, if you get down to, oh, say, a school district, the data at that level, you have a much smaller area and we can tell you that we think this or we think that about the two sets of data.

If you got up to the state or some substate level that had 400,000 or 500,000 people in it, I would think that would be a different level from a couple of blocks. My question is whether you would consider that or not. The Census Bureau has always had the position that there is one set of data, right?

MR. THOMPSON: In this world that is what we have pre-specified and what we are supporting, one set of data is our official data. Now, if you take the data for redistricting, basically we are putting the data out at very small levels to be added up. There is no set formula that prescribes how you add up these block-level data.

Our position is, when you add those data up to an area the size of a congressional district or for other uses, the data will be more accurate. At the block level we do not anticipate there would be a detectable difference.

DR. PREWITT: There are two things, I think, Janet. The precision with which we would know, of course, is not available to us until January-February, after all the evaluations are in and all the scholars have looked at the data, and so forth—you could readdress your question.

In that sense, based on what we will have available to us in February or early March, this is current thinking. It could evolve with evidence and so forth. Certainly, current thinking is we will make only one decision. Obviously, we have

reflected on what would happen if it looked as if the A.C.E. correction were superior for some parts of the system but maybe not for other parts.

Well, what do you do about that? Our current thinking is that we can make only one decision. We cannot decide that, well, you are going to adjust some post-strata and not other post-strata, for example; it would not make sense.

When I say that we have to approach this as just a statistical operational decision, we also have to be able to stand behind these data and be able to—there will be endless court cases, no matter what we do, and we have to be able to stand behind them.

It is at least our current judgment that we can better stand behind a single decision than a number of different decisions, certainly, based on what we will know in February-March. That is the current thinking.

DR. NORWOOD: I wanted to clarify that, because I could see several different approaches.

MR. THOMPSON: Yes, right. As Howard said before, that is the limit of pre-specification. Where we are right now is we had to make a decision and we pre-specified that we were going to correct everything. That is basically what we laid out in our plans and our procedures.

DR. NORWOOD: If you adjust, you will adjust everything.

MR. THOMPSON: Right.

DR. NORWOOD: I knew that, but you will also issue only the unadjusted numbers for particular parts, is that what you are saying? I had thought that if you adjusted, that you would also have to issue the unadjusted numbers.

MR. THOMPSON: We do to meet the requirements of the law. We are required to release the unadjusted data for the redistricting files and for Summary File 1. That is basically all we plan to release unadjusted data for, should we decide to release adjusted data. We plan to release only one set of adjusted data—one set of data, adjusted—for long-form data.

DR. NORWOOD: But set aside the long form, that is not coming up for a while. Summary File 1 is at what level of aggregation?

MR. THOMPSON: Summary File 1 goes down to block for some tabulations. It is the first (John Long could speak to this much better than I can), we put out two detailed summary files for short-form data. The first that comes out is Summary File 1, which has basic tables, not detailed cross-classifications—John, why don't you say a few words about that?

DR. LONG: The Summary File 1 is, in fact, the first one. It has basically what is on the P.L. data plus some extra detail for the basic race groups. What we are going to do for Summary File 2 is then to go into more detail for a large number of other groups that we did not have that information for, particularly by race, and go into greater detail for those categories, because there is a lot of detailed race information on the short form itself.

DR. NORWOOD: The Summary File 1 would be issued and for, say, political districting, there would be all the pieces that they need in order to determine a district, for blacks, for whites?

DR. LONG: Only those pieces will be there in the public law file itself, the P.L. data. Those actually come out before Summary File 1. Summary File 1 is basically for other uses that include an extra amount of detail beyond what we actually have to have. For redistricting, we actually need to know only if a person is above or below 18. In Summary File 1 we will have the full set of age data for them, but it is basically still the same 100-percent characteristics and that level of detail.

DR. NORWOOD: The geographic areas?

DR. LONG: The geographic areas are no more detailed than geographic areas that are in the P.L., because the P.L. is, in fact, down to the block level. For some of the tables, in fact, it will not be as detailed as the P.L.

DR. NORWOOD: But for all practical purposes, you will have—if you do adjust—you will have an adjusted and an unadjusted number down to the block level?

DR. LONG: That is correct.

DR. NORWOOD: That is what I did not understand.

DR. LONG: We can get the panel the full set of information.

DR. NORWOOD: The other parts of it are somewhat less important, but that is what I was trying to get at.

DR. PREWITT: Just for full disclosure purposes, there is a very complicated twist that has to do with this new federal regulation that has now been issued by the Department of Commerce. If the ESCAP committee recommends to adjust and the Census Bureau director does not take that recommendation, then that regulation reads that the adjusted data would, nevertheless, be released.

Therefore, there is a fourth cell, and the fourth cell is that the ESCAP committee does not recommend to adjust and the Census Bureau does not adjust, and then, presumably, the adjusted data would not be released under those conditions. They may, some day, be released by the Census Bureau for research purposes, but they would not be released, certainly, as part of the P.L. process.

DR. COHEN: I think the statement probably is not completely accurate that Janet was referring to. If you have the C.V. of a substate entry and you are checking the numbers for anomalous patterns, certainly the variance of the adjusted counts is something to take into consideration to try to judge how anomalous a certain data point might be, right? So it must be part of the decision process.

DR. HOGAN: Well, to a limited extent. Remember, first, in terms of computing the variances, we will first compute the dual-systems estimates for the 448 post-strata, which are not directly comparable to the 357 [post-strata in] 1990, but there is sort of a link there, so we can ask if doubling the sample size, more or less, reduces the variances, how was that compensated for by the target extended search? Did we get the gain that we thought?

Although the race domains are different in the 357 estimates, we had urbanicity and now we have metro—whatever, there are some differences—you can make some comparisons, and that is going to be a very important analysis and the kinds of questions you are talking about, did the variance really go up for American Indians on reservations, or for whatever, that is a number you can really sink your teeth into and work your way, again, all the way up the line to the outlier strata or whatever else.

However, once you get to geography, it is a synthetic estimate and the sampling variance of the synthetic estimate, the variance that we will be able to compute, is a function of only two things: the variances of the 448 and the composition of that area.

The variance for the state of California and the variance for the county of Ventura are probably not going to differ by a whole lot, because they probably have roughly similar characteristics, so looking at did Prince George's County have high variance versus according to Anne Arundel County is not going to get you hardly anything.

That is why these statements are here: we really need to look at where the variances are in the states compared to 1990—were they about right?—were the variances among counties, in general, about right?

If you had a tract that had the same proportion of Asian or black or whatever as the state of California, you would have the same sampling variance as California, so you do not want to carry that down too far.

It is explicit, I think, in the background document (maybe implicit in what I wrote), this sampling variance coming out of the A.C.E. is not the synthetic variance or bias, or however you want to talk about it—we will talk about that later. You get the individual variation across areas and we are not going to measure that here and, therefore, the kind of thing you are talking about is probably not appropriate.

DR. ZASLAVSKY: I think you started to address what I was going to ask, which is, since this is variance of synthetic estimators, how are they really comparable if the post-strata are defined differently? I could make the variance of a synthetic estimator very small by having only one post-stratum and use the entire national sample to estimate its rates.

If you really want to tell where you have gotten variances down, don't you need to have comparable post-stratifications?

DR. HOGAN: To a certain extent, yes, and, of course, that was the dilemma we faced in our research to build these post-strata, to choose these post-strata. If all you looked at were variances, you clearly would just choose one.

However, I think, at least at one level, you can step back and say, yes, the precise definition of the African American domain for Census 2000 does not completely agree with 1990, but it is close enough that you can compare the variance on the black populations and see if there was a reduction or about the same, or increased, just as I think almost everybody is going to look at the level estimates for the black population in 1990 and 2000 and ask if they roughly measure the same thing.

As I mentioned, urbanicity is not quite the same as the metro area; they differ (as my friends in geography will tell me). The same kinds of cities, you know, New York is going to show up in the top of both, Los Angeles in the top of both, so you can make comparisons to say we doubled the sample size hoping to get the variance down. How well did that work? In terms of looking directly at variances, those are the kinds of questions we will be looking at here.

Once you feed the 448 into, say, California, then you have an entity that is, in some sense, the same California we had in 1990, you have sort of standardized on California, and then you can compare the variances at that level in, I think, a fairly meaningful way.

DR. YLVISAKER: In terms of variances, Howard, could you say a few words about the quality of the A.C.E. data, and that has to do with, say, is there a study in these chapters that has to do with the quality of the data in a differential sense? That is, we believe the census has a differential quality to it and do the A.C.E. data have—is there a study directed in that same way?

DR. HOGAN: Not an overall one. We may come back to this a little bit, but a number of things that I talked about earlier say that the interviewing—the matching—is a good one, where we have how the interviewing went by census regional offices.

DR. YLVISAKER: But the census regional offices are not very detailed.

DR. HOGAN: No, and we had some other by states, or whatever else. There is only so far down you can go in detail with a sample survey. We plan to go down awfully close to that line.

DR. YLVISAKER: I am thinking about local census offices, for example, that are in, say, heavily minority districts, which are known to be hard to enumerate. Now, if we look at the quality of A.C.E. data for such things, is there a study in that direction?

DR. HOGAN: I am not sure how close we get to that in, say, noninterview rates for states [and] census regional offices. So far we have not specified to go much lower than that.

That is an interesting idea and we can look at it; we will have to figure out basically how many clusters we have in each LCO [Local Census Office] and whether what you get out of that you can interpret, is probably a good idea.

DR. YLVISAKER: I am not sure how to interpret it, but the question would be are we trying to fix those problems that are hard to fix with very poor data, basically? That is, we are going into areas that are hard to count and we are getting A.C.E. data that are not very good and then we are fixing those problems?

DR. HOGAN: I think what may help on that—it goes back to what I talked about earlier, the decomposition, which is by post-stratum. That does not directly get at the local geography but it at least says the following post-strata were traditionally hard-to-count post-strata and, in Census 2000, perhaps had high undercounts measured in these post-strata, what was the level of missing data in those post-strata, what was the level of item imputation? It does not directly address the geography issue that you raised but, at least, it addresses the extent to which the weaknesses in the A.C.E. are correlated with the undercount that we are trying to measure.

DR. NORWOOD: Why don't you continue?

REVIEW OF MEASUREMENTS OF A.C.E. QUALITY

DR. HOGAN: Well, let me review the bidding here a little bit. What we have done so far is we have gone through the census compared to DA [demographic analysis], we have compared the census operations, and we looked at the A.C.E. operations, verifying that we did what we said we were doing to do and we controlled the process.

We are now going to proceed with essentially three or four steps. The next step is to look at the individual components of A.C.E. errors, what we can learn about them based on the data we will have available, how that might compare to 1990, synthesize those into, perhaps, an overall scale or an overall measure of A.C.E. quality.

Then we will try to compare those to the overall quality of the census and figure out what we know about the relative quality of the two. That will be input into the final step, which is to gather all the information together for the ESCAP and make a decision.

What I am going to do now is talk a little about the various components of A.C.E. error, one at a time. These components have been out in the literature at least since Bob Fay wrote the 1980 report—I think he is the first one, to my knowledge, to really write it.

DR. FAY: As a co-author.

DR. HOGAN: As a co-author, okay—Bob Fay co-authored the 1980 report.

The individual components of A.C.E. quality are variance and misreporting on census data address, A.C.E. fabrications—these have been discussed—and various slight variations for about two decades now [reference not clear].

The question is, on each one of these, what do we know and when will we know it? The first one (we have already talked about this) is the variance. We will have, I think, measured the A.C.E. variances for the post-strata level and also for the political geography.

There are two caveats I want to mention. The most important one is, once we get to geography, it is the sampling variance of a synthetic estimator and, therefore, traditionally, in the direct estimator, the smaller the geography you get, the higher the variance gets. With a synthetic estimator, the sampling variances remain pretty much the same, although other synthetic variances, synthetic biases, tend to creep in, but they are not measured here, that is my point (not that they do not exist).

Secondly, to some extent, the way we are going to measure the sampling variance may pick up some of the other variances going on—not totally, by no means totally—but to the extent that interviewers do whole clusters, matching clerks do whole clusters, some of the variation that they introduce through that clustering process will be picked up by the way we are measuring sampling variance, which is a cluster-based method (not totally, but to a certain extent).

I think we have pretty much covered sampling variance, haven't we?

[No response]

Okay, the next component is what I call consistent reporting of Census Day residence. I am on 4.1.2 of my paper [Hogan, 2001]. The relevant background document is "Person Matching and Follow-up Results" [Childers et al., 2001]

Here is where we would like—let me try to explain this. If someone were selected into the E-sample and asked where he lived on Census Day, he would give the same response as if he had been selected into the P-sample and asked where he lived on Census Day. Primarily, it does not matter if he gives us the right answer as long as he gives us a consistent answer. There are some second-order effects that we could get into, but the primary thing is consistent reporting of where people live.

If, on the other hand, because of the use of proxies or different survey instruments or the difference of respondents, if we ask people on the P-sample, “Where did you live,” and they said “Here,” but they were counted somewhere else, and we went to the somewhere else and asked, on the E-sample, the neighbor, “Did the people live here,” and they say, “Yes, he lived here,” we would have an inconsistent reporting of Census Day address. It is a major concern in the dual-systems estimate, as designed and implemented in the A.C.E.

In 1980 we had an evaluation study that actually went out and measured it—I forget which P [-study] that was—

PARTICIPANT: It was in 1990.

DR. HOGAN: In 1990, yes. There it is, thank you. I thought you said P-1990 and, boy, I was confused—anyway, we had a lot.

That was a study where we went back and re-interviewed people and really tried to nail down where they lived. We will not have such a study in time for this decision. We will have an evaluation study that is coming up to be ready in 2002 or 2003, but we will not have a comparable study for 2000, when we have to make the decision.

We are going to have to assess the level of this misreporting based on fairly tangential kinds of evidence, the reading of how well the instrument worked, the extent to which we think the CAPI instrument really forced people to answer a set of questions, readings on how well the follow-up worked, levels of proxies in the follow-up.

Remember, a lot of the E-sample, non-matched cases are first interviewed to determine where they lived, in terms of A.C.E. processing, in the follow-up—that will be in the late fall. We will also have some information coming out of the matching, but fairly tangential information, that says there was evidence that there was inconsistency in this cluster or that cluster.

By and large, we will not have a comparable study as we had in 1990 that says here is the level where we went out, we reinterviewed, and this is how it differs by minority or non-minority, by owner or renter, or whatever else; we are going to have to judge the quality of the DSE in terms of this error based on our reading of the overall effectiveness of the survey, survey instrument, and the matching. I want to get that on the table.

DR. BROWN: It is not an issue for more data or anything, but just to clarify, this is one aspect of the picture that should be very different in 2000 from in 1990, at least the implications, because of the different treatment of movers and nonmovers. That comes into this issue, right?

DR. HOGAN: To an extent, yes, but, also, because it is the inconsistency, and it is partially addressed later. Even among the nonmovers, that is, people who say, "I live here now and lived here on Census Day," they might have a second address, where, if someone showed up there and asked them the same question, they would say, "Yes, I have owned this summer home," or the neighbor would say they owned a summer home here, but, yes, the treatment of movers, the fact that we are trying to reconstruct the Census Day address can affect this but, also, the instrument itself, the way the questions are asked in the initial PES instrument, and the follow-up form, all would affect this.

On none of those will we have direct data. Direct data will be available only when we go back now, months and months later, and ask them, "Where did you really, really live?"

This is one of the ones where we are going to have to make up our minds based on the evidence we have available.

DR. EDDY: This is a little off topic, but what about people who die between Census Day and A.C.E.?

MR. THOMPSON: It is a lot.

DR. HOGAN: Although we usually call them outmovers—it would be someone who lived there on Census Day who no longer lives there and would have the opportunity of being reported, and I think we have had instances where that did happen, where people said—

DR. EDDY: Now that I go back and think about it, there is also the issue of people who were born. Those are not in-movers?

DR. HOGAN: That is correct, yes. They are not in-movers.

DR. BROWN: People who are born in-between are not a problem, I think, but people who die in-between are a bit of a problem.

DR. HOGAN: The way we are approaching this, because we had a big discussion of this last time and it is worth going over, we get the rate at which movers were enumerated by getting the people who live there on Census Day who have left, the outmovers. They are the ones we match, and we get a match rate for them in the PES-C.

We know that the number of such movers reported tends to be an underestimate, traditionally, and, therefore, we try to estimate the total number of movers by the people who are there now who were not there on Census Day. If the housing unit population was a closed population to immigration or emigration or births and deaths, then in-movers would equal outmovers, and we would have a good estimate of the number of outmovers, an estimate of the mover match rate, and we would put them together. In-movers and outmovers we tend to say are the same people, by and large. There are some exceptions.

People who die are not balanced in the same post-strata as people who are born, so there is no point in bringing in the births to "balance" the deaths.

An issue was raised at the last meeting, I think it was a very good issue, and we went back and looked at it, which is that we have an imbalance with college students that, because of the timing of the census and the A.C.E. interviewing, we will have far more people who moved into the housing units from dormitories and

group quarters than people who moved out, and this group suggested that we were probably overestimating that and, indeed, we were.

We went back and we looked at the instrument and we came up with a way to reduce that problem by not counting people who moved from group quarters into housing units in the age bracket of 18 to 29, where this is a big problem and, I think, brought our estimator back into a better line based on the comments we got at the last meeting. Those were good comments, and thank you.

Now I am not sure where we were in terms of questions.

DR. PREWITT: Just a footnote. As with all of these things, you balance what you can balance, but to get back to the conversation about the telephone interviews in the A.C.E., when we realized that we were going to be able to get many more cases by phone than we thought we would and we raised the kinds of questions you raised earlier, what are the down sides of that, the huge upside is you are getting data much closer to Census Day, so we thought some of these kinds of errors would be reduced by maintaining the flow of telephone interviews.

An example of the larger point, that is, where we are going to have to balance all of this off and we have sort of got it all in front of us.

DR. HOGAN: The next component of A.C.E. dual-systems error is matching error, and I am now on 4.1.3 [Hogan, 2001]. This is an area where I think maybe, I am jumping the gun here. I think we made great improvements and I think we will have pretty good data to show that.

This is where, conditional on having the interview, conditional on having the data coming in, did you make the linkages between the A.C.E. and the census that you should have made? Did you have false matches, false nonmatches, did you call somebody a nonresponder when you should not have conditional on the responses you have gotten.

In 1990 we had a rematch study where we rematched these cases, we took a group and had clerks redo it, and we are, again, going to do that as part of the evaluation after the decision that we are making here has to be made, so in terms of making the decision, we will have essentially two sets of information. We will know the kinds of improvements we made in the matching system, and we will have the quality control, quality assurance, results out of the matching system.

Let me say a few words about each one of those. In the matching system in 1990 we had seven, I think—how many processing offices did we have in 1990? We did the matching in seven offices with seven different staff's and a traveling circus, and it was partially computerized, that is, a lot of the matching was computerized, but, for example, all mover matching was done completely clerically, including geocoding.

In 2000 we have done a couple of things. Because the census has data captured the name and all the information, which is a tremendous asset to us, the matching now will be in one location, Jeffersonville, Indiana. It will be virtually 100-percent computerized. They will have the computer screen with the census results, the A.C.E. results, a very nice point-and-click kind of methodology. They will have computer searches for duplicates and names that will help.

They will have very little paper there and they will even have computerized matching, so we have every reason to believe the matching system will be under far greater control than it has ever been; clerks simply cannot enter an invalid code, because the system will not let them.

On page 14 of your document [Hogan, 2001], it lists the numbers of improvements to the matching system, as well as in document 6 [Childers et al., 2001], if you want to read that.

In addition, we will have the A.C.E. quality assurance, matching quality assurance program; that is, each matching clerk, there is a sample of his or her work that goes—we have three levels of matching, or four levels. We have the computer, but that just takes almost-certain cases. We have clerical matching, which is about 250 clerks. We have about 50 technicians we have been training for a year, and we have about 10 analysts who have been with us for 10, 20, and at least one of them for more than 30 years. Each level of clerk is quality assured by the next level. The clerks are quality-assured by the techs; the techs are quality-assured by the analysts—sort of a triangle.

PARTICIPANT: And the top level you send to Ken?

DR. HOGAN: No, the top level we send to Danny Childers right back there and if he cannot figure out, it definitely is unresolved.

Because of the way we pyramided this, I think we can show an important fact, the fact that we are doing 300,000 housing units, the fact that we have trained 250 clerks, I think we can show the extent that that influenced the matching over and above the kinds of judgmental calls you would have if you had 10 highly trained people doing it, because each level is quality-assured by the next.

I am not saying this is 100-percent true but, really, that is the evidence that it will be pointing in that direction, so we will know the extent that we have controlled the matching system, the extent that large bunches of clerks introduced problems.

We will not know, if we took another group of 10 analysts and trained them for 30 years and then actually let Ken supervise them, we might get a different result than from our 10 analysts—we will not know that, we will not have an absolute version of truth, but we will have information that this was under control and [more] under control, probably, relative to 1990.

We have been talking a lot about PES-C versus B and the way we handle movers here versus how we handled them in 1990. One of the big advantages in the DSE literature on procedure C is that the matching is just a lot easier to control, because you have to deal only with people who lived in that block on Census Day. I think we will have good evidence here on the level of matching error.

DR. BILLARD: I have a question, and it may really be a duplication question rather than a matching question. These algorithms, how do they decide if there is a match? Are they looking at an address, in which case I can see some things, but what about a person? You might have Uncle Louie, who is visiting me in Georgia, and I put him on my Georgia form, but he is also put on his home form back in Saint Louis? But that is an individual, that is not an address. How do you catch those duplicates or those matches?

DR. HOGAN: There are two parts to the answer. First, in terms of the matching system itself, it is a person-matching system. We have everybody that the census counted in this block and we have everybody whom the A.C.E. says should have been living in that block. We have their addresses, we have their names, we have the relationship, and the computer algorithm standardizes all of that, and I can give you someone who can tell you all the details of the standardization, the weighting it is using, the Fellegi-Sunter algorithm. It comes out with a score.

The ones that the computer says are matched, virtually everybody would agree that it really is the right person. It matches about 70 to 80 percent, around there, but I think hardly anybody would quibble about what the computer says is matched.

The other 30-or-so percent it spits out to the clerks, the techs, and the analysts, and they are using judgment in terms of whether it is the same housing unit. They actually have and can bring up on the computer screen the census maps with their map spots, the A.C.E. maps and the map spots to see if that is probably the same household. They can go in there and they will have the relationship to the head of household.

On the census side they will even have an image of the actual handwritten census questionnaire—the computer screen is, of course, all computer ASCII-kind of code based on the optical character recognition, but if they have a garbled name they can actually bring it up and see what the person wrote, so they will have all of that to make a determination within the block whether this was the same person.

Getting to duplicates, the A.C.E. handles duplicates in two ways, and I am using “duplicate” now in the very largest sense of the word, which is where you or someone is counted in Connecticut and Florida as well as people who are counted in apartment A and apartment B. The A and B is within the block or the block cluster, so after the computer and the clerks match all the people they can match, they have some leftover P-sample people, they have some leftover census people, the leftover P-sample people—that is, presumably, where the non-matches are going to be.

The leftover census people, there are two or three things that could be happening there: A.C.E. could have missed them (we never said the A.C.E. was perfect); they could have been really living there and the A.C.E. did not get them; they could be duplicated within the cluster; or they could have been duplicated here and in Florida.

If they were duplicated within the cluster, the clerks do that and then they could do all the Bills in the cluster, they can sort by everybody who is of a given age in the cluster, looking for duplicates.

The ones who are both Connecticut and Florida, and this goes back to what I was saying a minute ago about consistent reporting, we want—let’s say they lived nine-and-a-half months of the year in Connecticut and that is where their residence was—[we want] them to consistently say, if we ask them in Connecticut, “Yes, I lived in Connecticut, that’s my correct residence, that’s where I should have been counted or missed,” and if we ask them in Florida, or ask their neighbors in Florida, that the neighbor would say, “Oh, he lives here only a couple of months a year, I think he has another residence somewhere else.”

We will never, in terms of the A.C.E., be able to say it was duplicated. What we would like to say is in Florida he should not have been counted, he is erroneous, and in Connecticut he was or was not correctly counted. That is how we handle duplicates and erroneous enumerations.

DR. YLVISAKER: How about cases where the primary selection algorithm [PSA] has been applied? Would the clerks know that?

DR. HOGAN: Only tangentially. On the computer screen they have what the official final PSA count was. Those are the people who were in the census for better or for worse.

As I said, they have access and they can bring up the actual questionnaire that the person mailed back in. I think in some cases they might be able to infer that there was not quite a line-up between the questionnaire that had been mailed back and what they are looking at, perhaps because of something PSA did, but there is no flag in there that says this was deleted because of PSA or anything else—I think that is correct. Danny is nodding, so that is true.

DR. YLVISAKER: It could be that there is another submission somewhere that actually matches perfectly, say, with the A.C.E.

DR. HOGAN: Yes, there could be, but if that other—let's say there were two questionnaires for this ID., this address, and the PSA selected the wrong one. The right one is not in the census any more; those people are now, in every meaningful sense, missed and, therefore, when the A.C.E. looks for them, they will say—and correctly say—those people were missed. They will not know why but they will know that they were missed, and the one that was selected in error, presumably, was an error in the A.C.E. on the E-sample side, [they] will say it was erroneously enumerated, you got the one from across the street, or whatever, and put that in.

The clerks will not really know that, that is not part of their job, and it would have to be a very alert clerk. The analysts could probably figure it out; they have been with us long enough.

MR. THOMPSON: And even if they did know it, there is not really any mechanism for them to reach back into the bag and undo the PSA and put the right person in. They are not “fixing” the census; they are really just assessing what it is.

DR. HOGAN: The next level, component, of A.C.E. error—I am on, I think, your page 15, 4.1.4, “A.C.E. Fabrications” [Hogan, 2001]—let me be very clear about what I am talking about. This error is the A.C.E. interviewers making up fictitious people or, I guess, what could occasionally happen is a respondent could tell the interviewer about a fictitious person. It probably does not happen too much, but it probably does happen.

Obviously, if the A.C.E. P-sample includes a person record that does not correspond to a real person, the chances of finding him in the census are remarkably small. To the extent that there are these A.C.E. fabrications, that inflates the measured undercount; it creates a non-match with obviously no chance of creating a comparable number of matches.

It was a big problem in 1980. I think in 1990 we were able to control it—some of our evaluation studies showed that we were controlling it. In 2000 we are going

to have a couple of things on this, one of which is the nature of the instrument, the CAPI instrument—I think some of you went out with it [observing interviews].

I think it would be easier to do the interview than to curbstone, because every key stroke is recorded and the time you started the interview and the time you stopped the interview and the time you started the next interview was all recorded. For an A.C.E. interviewer to curbstone, to make up cases, would be a challenge—which is not to say they have not tried.

We also, as I said, have the quality assurance process of the A.C.E., which was targeted on finding fabricators among the A.C.E. interviewers. That was one of its primary purposes, and we had the results of that, what it found, its probability of finding more being out there that were not found. I think there we will have some data that I hope will show that this was not a problem in 2000. Those are the data we will have; it will be a reading of the CAPI instrument and a reading of the A.C.E. quality assurance rates.

Document 6 [Childers et al., 2001], which talks about A.C.E. quality assurance, talks about that.

DR. COHEN: Just one last question on matching. The expert matchers, they will know the decisions of the routine matchers when they make their assessments, so you do not have any assessment of matching error where you have people actually working independently and just kind of starting from scratch and seeing what—and there is no field work at all?

DR. HOGAN: No, not for this. There are evaluation studies that will include independent rematch studies and evaluation follow-up. Those will be done and available in a couple of years but certainly not by March.

Missing data now—I think we have touched a little bit on this, so I think we can go through this a little bit more quickly. This relates to document 7 [Cantwell et al., 2001], which we talked about earlier. We will have the levels of missing data, the whole household non-interviews, the unresolved match residence and enumeration status. The enumeration status is on the E-sample side. Unresolved match and residence are on the P-sample side.

Missing post-stratification variables, that is, characteristics, age, sex, race, whatever. We already went through the various documents, the document that talked about the various kinds of things you can look at, where you will be able to see the effect of the non-interview on the weights, the distribution and characteristics of the imputed people, people we had to impute status for, and results of follow-up matching and whatever else. All that will be there for you to review, together with, again, one of our benchmarks here will be the results from 1990, where we evaluated the missing data and, also, we ran some alternatives in 1990.

I think, again, we will have the basic stuff. You can see the effects of missing data and whatever else. We are not planning on coming up with “reasonable alternatives” and trying to run two, three, or four alternative dual-systems models under two, three, or four missing data plans by the time we have to make the decision. I want to again make that very clear; that is not part of the plan.

I am willing to take more questions on this. We covered some of this earlier, so I am glossing over this a little bit quickly.

DR. NORWOOD: I think we did talk about missing data at some length before. Why don't we move on?

DR. HOGAN: Okay. The next one, page 18, 4.1.6, "Balancing Error" [Hogan, 2001], let me say a little bit about this, because it is one of those that is not intuitive. Essentially, we want the same rules applied in the P-sample when we say a person was correctly enumerated [in the E-sample].

We select a person at random, we try to go to the census to see if that person is correctly enumerated. The same rules that we apply on the E-sample side, when we have an enumeration and ask if that enumeration was correct, it is easiest to think of in terms of expected value. We would not want someone, if we selected him on the P-sample and we went to find his record in the census, and we found that record and said, yes, that was a match, that person was correctly counted, we would not want, if that same record had fallen into the E-sample, to say, no, that is an erroneous enumeration. We cannot, at the same time in the model say this person matches to this record and, therefore, this record is correct and, on the census side of this, say, no, that record is not correct, it is erroneous, by applying two different definitions. We would not want to say we cannot find this person in the census because he was missed and, at the same time, have a record that he could have and should have matched.

It was a real problem, and I think if I go back in history, you can understand the problem. In 1980 we had different search areas for the E-sample and the P-sample; that is, in the E-sample we selected a set of census records and we searched everywhere in the enumeration district to see whether they should have been counted in that enumeration district, were they duplicated in that enumeration district, but to be correctly counted, you had to have been counted in that enumeration district and nowhere else. If you were counted somewhere else, you were erroneous.

On the P-sample side, when we selected a person, when all the dust was settled, we actually wound up looking at one, two, three, four, five, six, seven, eight, nine, ten enumeration districts trying to find you, so after we got past the second or third enumeration district, it was possible that we would have matched the case to a census enumeration that, had it fallen into the E-sample, we would have said that was an erroneous enumeration—correctly enumerated and erroneously enumerated do not go together. That is a geographic balancing error and it was a big problem in 1980.

In 1990 we virtually solved this problem, because the P-sample and the E-sample overlapped almost perfectly, the same records were pretty much on both. We had a surrounding block search that was pretty much identical on both. It was virtually eliminated as a problem, with the possible exception of when we had a mover (then this overlap was not perfect). Everything we saw in 1990 said that we had solved this problem.

In 2000 we changed our design a little bit, and this problem might begin to creep back in. It is something we really have to look at. We cannot assume that since it was solved in 1990 it is solved forever. In 1990, if we had a person, we searched the surrounding ring of blocks always and even it did not really matter. They might have been counted with their grandmother erroneously one block

away; we would look for it. Occasionally we found it but often we did not. It was pretty boring for the clerks.

In 2000 we changed the nature of the surrounding block search. It is not focused on any kind of random person who got counted in the wrong place but only the results of mis-geocoding; in other words, rather than just seeing if a person might happen to have been counted across the street, we are going to look for him across the street only if we have reason to think the census mis-geocoded him.

Why? Because the payoff for the surrounding block search really is in variance reduction. If you do not do the surrounding block search and the census has moved an entire apartment building across the street, you get, depending on which block fell in the sample, a huge number of misses or a huge number of erroneous enumerations if you restrict the definition to only one block.

By doing an extended block search you can reduce the variance quite a bit, because many of these census missed geocodings are literally right across the street, but our rules say we are going to look across the street only if the whole household was missed in the P-sample [or] if the whole household was mis-geocoded in the E-sample. We are not looking for individuals any more, we are looking for households. It is a new level of complexity that we did not have. It is called a targeted extended search, because we are targeting which households we are looking for rather than looking for them all.

In addition, we are going to do this on only a sample basis. That adds another level of complexity. A block where we have very few missed geocoding errors might be selected in with a fairly low probability. A block where we clearly have a lot of missed geocoding errors will be selected in with certainty.

We have reintroduced a concern with balancing error that we have to look at, and we are going to look at this basically by looking at the results of the targeted extended search, what we found in the surrounding blocks in terms of misses, geocoding errors, what we had to impute in terms of the missing data for a targeted surrounding block search to see that our design, our new design, our TES design, is working. That is, as I said, a new design; it was not done in dress rehearsal, even, it was not done in 1990, and we have to see if it is working.

DR. BROWN: What data are you going to use to look at this?

DR. HOGAN: Basically, the results of the TES to see what we geocoded across the street, weighted, unweighted, what we matched across the street, weighted, unweighted.

DR. BROWN: So it should, presumably, sort of balance out?

DR. HOGAN: Sort of, yes. Never perfectly, because we have misses on the one but, yes, there should be some rough balance there. At the small post-strata level it will not be, because we have the variance problem, but as you aggregate up, it should more and more balance.

DR. OLKIN: Howard, are you going to go on to a new section?

DR. HOGAN: Not if you have a question.

DR. OLKIN: I want to go back to an older section.

I have a lag effect in my brain. I am still on fabrications. You keep reading about examples of Post Office delivery people dumping mail under the pressure

of time, and fabrications might become a bigger item than we think of, and the estimates are pretty broad. Is there no way beside what you are doing to get a better handle on the degree of fabrications?

DR. HOGAN: There are two fabrications going on here. The one I talked about a minute ago is the A.C.E. interviewer with the CAPI instrument fabricating, which has virtually nothing to do with the Post Office. Then we have measuring fabrications in the census and the census either presumably created some fabrications and, then, as part of the E-sample, the extent to which we can find them and correctly identify them. Actually, we are getting ahead of the story, because we are going to get to that soon, so you are going forward, but that is fine.

DR. EDDY: Can we go back to the CAPI fabrications, since we are there? You mentioned these time stamps or something in the interviews and it would be pretty painful to fake them.

DR. HOGAN: In the initial interviewing. The follow-up interviewing is paper and pencil, I do not want to exaggerate there.

DR. EDDY: I guess I am wondering what sort of check is actually done on these records. Fine, they are time-stamped, so if somebody looks, they discover it, but who is looking?

DR. HOGAN: Not systematically. Under what conditions do we look for those? I know I had one where I went back and looked but....

DR. EDDY: So it was 13 per day and after 8:00 p.m.? So nobody does actually look at the times.

DR. HOGAN: Not for all 300,000, but it is available if we have suspicions.

DR. EDDY: By the way, you should use smaller laptops for those CAPI things.

DR. HOGAN: Well, next time everybody will have this tiny little whatever.

Any more questions on that, because I think now we are turning to the next set of questions. I am now on page 19, 4.1.7, "Errors in Measuring Erroneous Census Enumerations" [Hogan, 2001]. This is where normally [we] have the leftover people in the census, people the census got and the A.C.E. did not get. We send those to follow-up and then we have to determine, for example, are they fictitious or are they duplicated, did they even live in that block?

Again, here we will have, at best, tangential information. We will have the ability (I do not know if it is in any of these) to see the level of census fabrications we were able to find and sort of the pattern of census fabrications we were able to find.

We will not have, because we will not have the evaluation finished by then, a third or maybe it is a fourth read by now of whether that person really lived there or not. We will not know whether we over- or underestimated or misestimated the level of census fabrications. Some people were concerned in 1990 that the census had far more fabrications than the PES measured.

Here, as in the inconsistent reporting the census data addressed, we will have some information in terms of how the follow-up went, the kinds of census errors the follow-up had, does that look reasonable, is that pattern of census fabrication, census duplication, consistent with what we would expect?

It is possible, with the census now having data-captured all the names, as I mentioned earlier, even outside the A.C.E. blocks, we actually could, at some level of aggregation, measure duplicates outside the A.C.E. and, therefore, have two measures of the same thing to verify that the A.C.E. was measuring duplicates well.

DR. OLKIN: I do not know whether or not any of the kinds of studies that Educational Testing Service does on deciding when one student has cheated on an exam are similar to the kind of model of fictitious [cases]—because they look for patterns. I do not know what kinds of patterns they look for, but some of those studies might be informative in terms of deciding whether someone is following a comparable set of questions.

DR. HOGAN: But remember, what we are doing here is measuring the errors that the A.C.E. made in measuring the census errors. Although we do not have an artificial intelligence program, we do have these analysts who are looking at the whole pattern within the cluster. Very often census interviewers who fabricate tend to do a whole mess at once, so they will be using human intelligence to try to discern patterns or likelihoods of census fabrication.

The question in terms of how good the A.C.E. was is did they do a good job on that and for that we do not have any specific measure. We will have their reports, we will have the patterns that they found, and we can look it over to see if there is any indication that this got a lot better or a lot worse than 1990.

DR. PREWITT: We do not have any way of dealing with fabrications, I do not think, do we, where the respondent fabricated as against the enumerator?

DR. HOGAN: In a sense we do. To some extent people make it easy for us. I mean, we do get census questionnaires that clearly say Donald Duck, that clearly say Fido. There are smart alecks out there that we do find and identify in the mailback questionnaires. That tends to be a small proportion of the problem.

More importantly, though, and it is not really fabrication, there is a tendency that I think virtually every parent can sympathize with, for parents to put down their college students. It is not exactly fabrication. They know that the rules kind of say that their daughter should not be counted at home but it is really hard. I had a freshman this year and it was really hard, but I answered it honestly.

So we do have that, it is not exactly fabrication, but we do have people misapplying the rules, and there we really depend on either the A.C.E. interviewing or the follow-up, all those very detailed problems to sort that out.

DR. BROWN: A person who insists on misinterpreting the rules, if they consistently misinterpret twice, there is nothing to catch it?

DR. HOGAN: Exactly. This is the exact problem. If at home the person consistently says his daughter lived there, there is nothing that balances that at the university. If we went to the university, which we do not, and we said should that person have been counted there, they would have correctly said yes, she was in this dormitory.

That is an exact case where you have an imbalance on where the person lived in the P- and the E-samples and that does introduce error, absolutely.

DR. NORWOOD: It is perhaps one of the problems that we have [as] a democracy, because what we are doing in this country is saying to people, “You identify yourselves, you tell us the answers to things,” and often that gives us difficulties, but we continue to do it, because we are a democracy and we feel that is the way you do things. It is another interesting [thing] to think about, I think, particularly in the long form.

DR. PREWITT: Look, the political science literature is full of analyses of strategic voting. One of the issues that occurred to us, it certainly occurred to me, during all the promotional effort about the census is could you have strategic census completion, and there was a bit of that, obviously, in the discussion about the multiracial item.

More generally, you could use the census form to make a strategic contribution to this, that, or the other. I did what I could do anecdotally on that and, as best I can tell, there was very little strategic use of the census form as a way to create some kind of outcome that you otherwise would not create.

Theoretically, it is an interesting issue, just as there is strategic voting.

DR. NORWOOD: It certainly gets us back to the demographic issue, which I think several people—and now Jeff is here—did want to revisit a little bit. How are you really coping with the change in demographics? It is not a Census Bureau issue, but, nevertheless, you have to cope with it.

DEMOGRAPHIC ANALYSIS

DR. HOGAN: We wanted Jeff [Passel] here for this discussion, so we glossed over it earlier.

DR. LONG: In the demographic analysis paper that you have [Robinson, 2001], there is a table in which we tried to show one of the ways we hoped to deal with the issue. For demographic analysis, we said before that the race detail that we actually have to work with is black/nonblack.

The tables in the back of that section have a column—and there is a Table 6 that talks about model 1 and model 2. There are several ways, of course, for tabulating when you have multiple responses. The main one we are using in the P.L. data is to actually tabulate all those possible combinations, so there are 63 combinations that we can see with the OMB race detail plus the extra categories that the Census Bureau uses. So using the complete information would be one way.

Another way would be to tabulate persons who marked only the specific category that you were interested in and marked that alone, so a person who marked black or African American and did not mark any other category we would put in that response. If a person marked several categories, we might just lump him into a two-or-more race response. That would be one way of doing the analysis and that is, in fact, what we call here model 1.

Another way of doing the analysis would be to say if they marked the section called black or African American, we would use that, no matter what else they did. If they marked that plus some other races, we could still count them as that. If

you are looking at only that response and you are trying to worry about something that is mutually exclusive here, you could actually handle it that way.

Model 2 in this sense does that. What, in effect, we have is a maximum and a minimum of what you would get by several different ways of trying to classify that particular response.

Since it [is] still too early to know and we have only a little bit of fragmentary information from a few previous studies such as the dress rehearsal and now a little bit from the American Community Survey, it is hard to know what the total amounts are going to be, but there is speculation that particularly for the black and African American population the differences may not be particularly large for those two groups. We will have to wait to see what happens.

In any case, we have a maximum and a minimum to look at. If those numbers are not too far apart, there should not be much difference in the determinations that come out of it. If they turn out to be quite far apart, then we will have another problem to deal with.

DR. ROBINSON: That is right but it is almost important, I think, as Howard mentioned earlier, that is another reason looking at sex ratios will be especially important this time. As long as the shifting is relatively the same for males and females, that will more or less cancel out if you are looking at sex ratios. We do not think there will be as much play in the data for sex ratios as for total; we will not know until we see the results.

There will be this problem of interpreting trends and coverage by race groups from 1990 because of this added dimension of multiple race.

DR. YLVIKAKER: How many multiple race forms did one pick up?

DR. LONG: Since we have not tabulated the data yet for the census, we do not know that answer. The guess is it is on the order of a couple percent, but no one really knows yet.

DR. NORWOOD: There are efforts by some groups to have only one race, only one category listed, so whether that had any effect or not, nobody knows. [What about] Hispanic, which is becoming increasingly important in this country?

DR. LONG: In some ways, with this Hispanic category, this is something we have been doing for a long time, actually having an independent race and Hispanic-origin question, so that you can, in fact, be Hispanic or not Hispanic within any particular race group. That gives you a large matrix to look at already from that perspective.

DR. NORWOOD: You have a longer history of that.

DR. LONG: That is right. It is sometimes difficult to deal with, but we have managed to do that pretty well and I think the public has become rather comfortable with handling it.

DR. BROWN: But in general, you do not have demographic estimates for any Hispanic category that are of the same nature as you do....

DR. LONG: Not in the long term, but let Greg talk about some of the other things we try to do for Hispanics, at least with the data that we do have available.

DR. ROBINSON: Right, this clearly is an important population and can we develop some kind of benchmarks for that population? What we are going to plan

to do this time is, first, to utilize our regular population estimates we do by race and Hispanic origin. These would not be totally independent of the census but we produce estimates of the Hispanic population that are benchmarked on the previous census, 1990, carried forward with estimates of components of change, so we have an estimate of the Hispanic population on April 1, 2000, which is a benchmark.

One thing we can do is we can adjust for undercount; at least 1990 for that population using the PES results, so that will give us a rough benchmark that is not truly independent like national demographic analysis for blacks and nonblacks, but we can use that to roughly assess how close are the census results and the A.C.E. to this benchmark.

The other thing we can do—again, one of the main reasons we do not do demographic analysis for Hispanics is that we do not have the historical time series of births and deaths and immigration going back in time, but for the population under 20 we do have births. We can develop some fairly rough estimates of coverage of the Hispanic population using the demographic technique.

In fact, there is a paper in our Population Division working paper series, where we did some illustrative Hispanic undercount estimates for the population under 10 in 1990. We will do that. These are not the same caliber as our national estimates but there are some benchmarks we will be developing to get some rough assessment of the consistency of the three numbers, the A.C.E., the census, and the demographic benchmarks.

PARTICIPANT: How can you identify them?

DR. ROBINSON: It is on the birth certificates. Again, it is the origin of the mother or father. The National Center for Health Statistics—John [Long], check me on this—gives the origin of the mother for Hispanics right now. It comes from the administrative data.

DR. NORWOOD: We break that out for employment at times when we run the CPS [Current Population Survey]. We have been doing that for some years.

DR. BROWN: I have a separate question. You have essentially acknowledged it but I sort of want to point out, I find the concept of using these kinds of demographic benchmarks to decide the validity of the adjustment in 2000 kind of questionable, because the benchmarks themselves depend on which number you took in 1990; that is, if you believe the adjustment in 1990 was an improvement or not.

In some sense your decision about 1990 almost drives the decision in the comparison.

DR. ROBINSON: No, for the total demographic estimates, for the total population, they are independent of—

DR. BROWN: They are independent. It is the Hispanic, it is the portions of the population where you use the 1990 numbers and carry forward.

DR. NORWOOD: They are unadjusted.

DR. BROWN: But he is talking about he might use the adjusted numbers.

DR. ROBINSON: Yes, again, this is just one of the many pieces we would use. This would not have the same standing, really, as the other groups. This is

something that we have not done before. We showed this [because it] may give us some information to the overall assessment of results, and we can do it, but it is certainly not part of the main tables in this prototype report.

We really do not show anything specific for Hispanics, because that does not have the same independent basis of the national demographic analysis, which goes back to 1935 for births and deaths and then for Medicare we use for the population over 65, essentially independent of the census we are evaluating but, true, for the Hispanic benchmarks it is based on 1990 with the adjustment carried forward, so it has problems in measurement of the population of 1990.

DR. STOTO: The table that John called our attention to with the model 1 and model 2 is probably about the best that can be done at this stage and we are just going to have to wait to see what the numbers look like. I just think that in terms of looking forward to 2010, this is going to disintegrate very fast, because vital statistics are not changing until around 2003.

For the next three years we are going to have different definitions of race and ethnicity in our births and deaths in the population, so all of this stuff about Hispanic is going to be multiplied by all these different groups. I suspect that this approach will be much less informative 10 years from now than it will be this time.

DR. LONG: Yes, I think it is quite true that we are entering a new world of complexity, and we are not too sure where we are going, but it is not simply the fact that the demographic measures are complex, the society is becoming more complex and we need to find a way to deal with it.

DR. PASSEL: That is a good place to pick up. I think it is important to keep in mind what demographic analysis is and what it uses and what it does and what it can do. It was an extremely valuable tool in the world of 1970 and before in the United States, where we basically had a black-and-white society with very low levels of immigration and very strong divisions, very strong color lines, if you will, people were on one side or the other. As we have come into the 21st century the country has changed. It is getting more and more problematic to compare data from one census to the next with regard to race and so forth.

Having said that, there is still a lot of power in demography. People get one year older each year—mostly (some people do not). Approximately for every boy there is a girl. That works, too, and when it does not, we know what the deviations are.

In that sense, demographic analysis provides some important benchmarks to compare the census against and the A.C.E. There are certain things, and this may be as good as we can do. I think that we have to wait and see what the data look like. My own suspicion is that a lot of the multiracial people, a lot of the people who we are not sure which side of the black/nonblack line to put them on are going to be children, young children, whose parents filled out the form. That suggests for the adults this may be better and this approach will enable us to get at that. I would be surprised if there were a big effect on the sex ratios of the multiracial population, but that is built in here, too.

I agree with Mike [Stoto], but for a different reason. More and more the black population is being increased by immigration. There are more and more people

coming into the country whose conception of race is based on what it was in their home countries, and the way they answer these census questions may not be what we expect or what we would have them answer, but they give what for them are perfectly valid and meaningful answers. That is just a piece of society changing.

Over time, I think that is correct, the ability of demographic analysis to pick up very small deviations, very small differences, will deteriorate, but I think it is still extremely valuable for picking up large deviations. If there are places where demographic analysis is out of line with either the A.C.E. or the census, it is not to say which one of them is wrong but it is a flag to say that there is something wrong here and it may be demographic analysis but it may be the A.C.E.

In that sense it is one more arrow in the quiver and even as problematic as the other race pieces are carried forward from 1990, they still provide what may be potentially valuable indicators of problems in one data system or the other.

DR. HAUSER: I have a couple of comments. This may carry more interest for the future than for the past, but there are some very interesting data available from the Adolescent Health Study on racial identification—[Microphone intermittently on]. My recollection is that you are much more likely to get one race [reported] at home than at school.

The other question I have is about this business of the Hispanic population, because you have immigration, and estimating [net] immigration becomes a terribly serious issue in the Hispanic [population]. It becomes a terribly difficult and serious issue there.

The other thing I wonder about is whether you would feel comfortable doing anything with the sex ratios in the Hispanic population, given the amount and nature of immigration.

DR. ROBINSON: We have not looked at the sex ratios as closely as we do for the black population, because we are often, there, looking at the low sex ratios implying differential undercoverage. The sex ratios for the Hispanic population were actually above one for many ages because of the immigration.

Certainly we could look at sex ratios as another tool to assess how the sex ratios look compared to what we might expect, if that is what you are asking, but you are right, these are the various different tools we have to assess, here are the census results, how consistent are they with what we expect, and what our estimates are.

Another thing I will also say is doing this projection for the Hispanic population, you are right, it is one of our weaker components, immigration, so if in doing this projection we find that for some age groups that while the census results are somewhat different from our projection, it could be a problem that our components have changed and give us some information. We need to check that this may have understated immigration of some group.

We actually did this in 1990, and we found that there may have been some problems in our components, but that is certainly a rough benchmark to use for this group.

DR. STOTO: The graphs that are much more obvious in this paper than in most of the other ones, I think, are really very helpful, and I think they come naturally out of this approach because they are tracking things over time. I think they

will help us to understand which of the two models is more reasonable, understand trends in that.

I think probably to the extent that this time series approach can be used in other areas, that would be helpful as well, recognizing that the design....

DR. NORWOOD: What you are saying, I think, is that for purposes of evaluation of both the census and the A.C.E. and their differences, that demographic analysis can be useful to point out differences but that one needs to be rather careful in interpreting them and, certainly, they would not be a method for adjustment by themselves. Is that right?

MR. ZITTER: Actually, demographic analysis, without the concern of measures of undocumented aliens, the illegal component of Hispanics, particularly, could be [useful]. In 1970, when there first was word of this, we started making adjustments. We do not know, but by 2000 we have something like 5 million undocumented aliens. The point is, it is a component with a lot of unknowns.

DR. ROBINSON: That brings up the uncertainty. In the first-time development for our 1990 demographic program we had some measures of error in the estimates and our error band around the undocumented was one of the biggest by far, so that is definitely true and that affects [some] age and sex groups more than others. Any of these interpretations needs to be couched in terms of what is the uncertainty around the numbers.

DR. BROWN: Are you going to have any measures of uncertainty in these data?

DR. ROBINSON: Yes, we will. We do not have them here, but, similar to what we did in 1990, we will have error bands and certainly bands around the demographic estimates for 2000.

DR. BILLARD: I have a comment and then a question. I remember the Hispanic question was an ethnic question as distinct from a race question. For many people, Hispanic is their race. Even as we have been talking demographically here, we have not really said it that way but we are sort of implying that Hispanic is one group, black is another, and nonblack another.

Somehow we are going to have to sort that out. I certainly know that sitting here in the U.S., when I think of Hispanic and I think of white I am thinking of one thing, but if you sit in Spain and you use the word "Hispanic" and white, let me tell you that is a very different thought that they have as to what we have here, and yet we will have immigrants from many of these different countries.

Either somehow or other the questions have to be sorted out differently and/or it would seem to me, if I were a Spanish immigrant and I came to that, I would leave one of those questions unanswered, and maybe there are a lot of unanswered questions there.

What do we do with those missing data? They have sent in a form but they have not answered all the questions. Do we impute the missing answers or do we go back as a nonresponse follow-up on those who left out certain questions?

DR. LONG: Some other time we can go into the details of the imputation process, but that is one of the reasons we do leave independent the ethnicity versus the race question, because we know that people respond very differently to those

questions. For some, they do believe that they have, in fact, answered the whole set of questions once they have said they are Hispanic. When they get to the race question they may leave it blank or they mark “other.”

Either of those are possibilities. The “other” possibility is acceptable, [but] blanks are not acceptable for the census so we usually do imputation.

DR. BILLARD: If the projections are 12 percent Hispanic immigrants, or Mexican, or whatever, some figure of that order over the next however many years—I do not know what it is, I do not even know if I have my 12 percent correct—that is a big chunk of the population.

DR. NORWOOD: But that is not our problem here.

DR. BILLARD: Yes, that is true.

DR. LONG: And the A.C.E., of course, does measure that. That is where we would, I think, go to the most differences. I think demographic analysis, as has been pointed out, has much less to say about the Hispanic population than it does about the total population and the black population, but it can say a little bit.

DR. BROWN: A different aspect of this, the tables that you have laid out group the age categories the same way as the A.C.E. does, and that is convenient and maybe is the only convenient way to do it if the only goal is to compare the demographic adjustment to the A.C.E. adjustment, but I think there is another role that demographic analysis can play, sort of more like the quality assurance kind of role, feeding in basic data.

For that it might be useful to have the complete array of ages or some subdivisions and simply compare census to DA total numbers, let’s say. You might find, for example, that census undercounted considerably in the 18 to 24 age group but not so much in the 24 to 29 age group, and that would say something about how the census worked with respect to college population, just as an example. Maybe for that purpose, for looking at just the census versus DA, it could be useful to have more detailed breakdown of ages.

DR. ROBINSON: Thank you, yes, and that is one of the strengths of the DA. We have estimates by single years of age, so we will be comparing to the census to get an initial evaluation of that. In fact, for the demographic estimates for 1990 for ages 18, 19, 20 and, I think, 21, there were net overcounts in the census, so this age group where you can be at college....

DR. BROWN: That presumably may be reflecting this double reporting.

DR. ROBINSON: This report was written for the comparison to the A.C.E. but we will have the age detail for all groups, right.

CONTINUED DISCUSSION OF A.C.E. QUALITY MEASURES

DR. NORWOOD: Howard, we are making great progress. I think we are on correlation and synthetic biases; we are on error measurement and then pulling it all together.

DR. HOGAN: I want to go through correlation bias, then synthetic bias, then a few other little biases very quickly, and then get to pulling it all together, in two senses—one of which is pulling it all together in a statistical model and [the

other is] pulling it together as a decision process. In my mind, those are really two different things.

Correlation bias, we all know it is there. It would be surprising if somehow we ran a PES where it disappeared. We measured this in 1990 in comparison with the demographic analysis, specifically the demographic analysis sex ratios, although we just had the discussion on the DA that there are some limitations and uncertainty there.

If you accept the DA sex-ratio estimates, you can measure the correlation bias, at least among males, by taking the A.C.E. estimates for females, figuring out how many males should be there and finding out how many males A.C.E. found. We intend to use essentially the same models that we used in 1990 again in 2000. Bill Bell is working with us on that, who did the 1990 work.

There were some concerns about the 1990 approach, because correlation bias is one of the errors and if you have compensating errors, concern was expressed that the method that we had might mismeasure correlation bias, because there might be some compensating errors there.

We have looked at that—Bill, specifically, has looked at that—and we think, since we are using the sex ratios rather than the levels, that to the extent that there is matching error or whatever else in the A.C.E., they probably affect males and females very similarly and what we are able to measure using the 1990 methodology is probably a valid measurement of correlation bias. Those are the methods we plan to use.

Also, then, we can talk about how that might distribute to a lower level, when we get there. There is not much to add over 1990, but I am here and Bill is here, so I am sure we can answer questions.

DR. NORWOOD: I do not sense any burning enthusiasm to ask questions.

DR. HOGAN: All right. That gets us, then, very quickly—boy, we are catching up on the agenda—to synthetic—

DR. BROWN: This is really something just stating the obvious, that the DA estimates are only one way of trying to get at an evaluation of correlation bias. The problem is that the other way is all about involved evaluation studies, or maybe there is something else that I am missing.

DR. HOGAN: There are only sort of two ways of getting at it that we know, one of which is with respect to some outside measures, such as demographic analysis, or triple system or something where you go beyond the dual-systems model, and demographic analysis is what we have available. I do not see an evaluation study going well beyond that.

The other way we have looked at is essentially to go into the Alho approaches, where you use logistic regression, or whatever, and you effectively make your post-strata as small as possible; rather than 448 you can effectively have far more and approach correlation bias that way, but there is a real limit to how far you can go down that road, because the essential underlying correlation-capture probabilities are still present, even if you had not 448 but 4,448. You would probably still tend to underestimate certain types of people.

You can move part of the way in terms of those logistics but not, we think, extremely far. There is nothing against it except for the time. The external measure, we think, gets us a lot closer. Except for those or triple system, I am not sure what is out there to measure correlation bias.

We have a few little glimmers of things. I mentioned letters from people refusing because they had already been counted in the census, refusing A.C.E., but I would not take that one very far down the road.

PARTICIPANT: So we could say we do not understand correlation bias very well in 2000. Is that a fair statement for you?

DR. HOGAN: No, I think we understand it. What we do not know is how to measure it, except by comparison to an outside standard, such as demographic analysis. I think we have a good understanding of correlation bias, but that is different from being able to measure it precisely.

PARTICIPANT: Quantitatively by April?

DR. HOGAN: Quantitatively? Using comparison demographic analysis, yes.

DR. COHEN: Are there ethnographic studies, they are not quite comparable to what was done 10 years ago, but are those possibly able to provide any information about correlation bias?

DR. HOGAN: Yes, we did have the 1990 ethnographic studies that shed some light on it, but they were very spotty, since they were not random samples of anything, but those are still available. We are not going to have new ethnographic studies available in early winter.

DR. HAUSER: The fictional post-strata are about the same size as those in 1990. Wouldn't that be an appropriate kind of benchmark to use in getting a handle on the extent of correlation bias with the larger post-strata?

DR. HOGAN: I am not sure that I follow.

DR. HAUSER: You have a much larger sample now, so suppose you went to post-strata that were of the same size. You talked about a much larger number, but that would be a number that one might argue would be kind of reasonable.

DR. HOGAN: It is something that could be done. I have a feeling and, until one does it, you would not know, but I have a feeling that that would estimate just a little bit of the correlation bias. Going from, say, 500 to 750 is going to measure some additional correlation bias, but not very much.

If the A.C.E. comes in here, demographic analysis here, some of these other measures, my guess is, are going to get here. It is not [to] say that they are not worth doing, in theory, but I do not think at the end of the day people would be very satisfied that that really picked up their main concerns about correlation bias in the dual system.

The next one, and the last of the major components of A.C.E. error, is synthetic error; that is, when we distribute the undercount locally, we are making some assumptions that there is a geographic uniformity. We have national estimates of black owner large city undercounts; are there differences subnationally?

In this discussion I think there are a couple of things. One distinction that I think is important for only the very smallest level is a distinction between what, in my nomenclature, and I know it is not universal, I think of as synthetic bias,

which means there are different capture probabilities across areas and synthetic variation or synthetic variance that says sometimes some people will be counted and sometimes some people will not be counted.

At the very small levels, that synthetic variance effect, it is a stochastic process being counted in the census, can be very important, obviously. When you summarize much above tract [synthetic bias comes in], because there was actually a variation in people's chances of being counted.

The other thing that I think is really worth pointing out with synthetic error is it is there before we do the A.C.E., before we do the adjustment. Matching error was not there before we did the A.C.E., there is no matching error in the census; if there is error in the A.C.E. we have introduced it into the adjusted numbers.

If there is error because of A.C.E. fabrication, it was not there in the unadjusted numbers, it was introduced to the adjustment. The synthetic error arises because of heterogeneity in the census, it essentially is there. There might be differences in the net census undercount for Dallas and for Atlanta, I do not know. If there are, it is there before I adjust and to a great extent it is going to be there after I adjust.

It is, in a sense, a very different kind of error than some of these other ones we have talked about. If you do not want any matching error, you have a choice: Do not put the A.C.E. in, if the matching error is too bad, do not do it.

Synthetic error, if you see there is really great variability across the nation in how the census achieved coverage, choosing not to go forward with the adjustment does not mean that problem went away, it is still there. It is a more tricky problem in how errors might net out whether the application of the A.C.E. synthetic estimates might make it better or perhaps might make it worse, but it did not create the problem, and I think that is quite a distinction we need to explore in our thinking about this.

There is not a lot of—there are some data, but there are not a lot of directly relevant data on synthetic basis, synthetic variability. We have results of the A.C.E., but, of course, it is a sample survey. In any city block, in one city or area, there are only a handful of sample cases, certainly, within a post-stratum and, therefore, one cannot get much directly from the PES or the A.C.E., so then you have to go to other things that are available nationally, available for a broad base.

These are mentioned in your documents on page 22, that is, B-1 [Hogan, 2001], the one I am basically lecturing from, things like census allocation rates, census mail return rates, census substitution rates. Each of these will exhibit patterns, geographic patterns, and we can look through some simulations, artificial population simulations, to the extent that these patterns are picked up or similar to what we get through our A.C.E. post-strata or whether they are completely tangential and there are things going on out there that we are not picking up. We intend to do that.

Limitations of this, of course, are things like census allocation rates may not be distributed, for better or for worse, like net undercount.

PARTICIPANT: Would you say something about allocation rates?

DR. HOGAN: Good question. Let me see if I remember. When the census has a record, believes that there is probably a person there, some indication that

there is a person there but we did [not] get data for that person, someone mails back a questionnaire and indicates there are two people but does not really give the information, then we have to do a whole person substitution into that household, and we have records of this for the whole country.

It is an indication of problems in census-taking that might be differential geographically, so it is something one can use to simulate to see to what extent the post-strata and the methodology we are using, the A.C.E., might mimic these patterns, that they follow our post-strata or they do not.

DR. BELL: But census substitution rate is something similar, it sounds like?

DR. HOGAN: Yes. I will have to get a census person to get them straight.

MR. THOMPSON: I think the allocation rate is simply the missing data rate for a particular characteristic and the substitution rate is what you just described.

DR. HOGAN: Okay, so the substitution rate is whole person substitution, allocation rate is where we had to impute characteristics.

There might be other things one could look at beside these three, but the basic idea is you try to find a variable available outside the A.C.E. sample that is related somehow, in your mind, to the net coverage rates and try to see how well the A.C.E. sample post-strata design can work in capturing that local variation.

DR. BELL: I have not looked at the papers you cite that have done something like this. It sounds as if they are trying to estimate these rates from coverage correction factors. Is that right?

DR. HOGAN: What they do, most of them, is they scale these rates so they have the same scale as the net undercount rate and then they simulate an adjustment. They would simulate a synthetic adjustment for California, for Georgia, whatever, based on this—they construct a national undercount scale from these very scales to agree with the PES, or whatever.

Then they do post-stratum level estimates and they carry that back down locally, and then you can actually compare the undercount that you got for Ventura County or for Georgia together with the synthetic estimates you get from the carrying-down process.

DR. BELL: But they are not using the undercount coverage correction factors, they are using, in essence, a substitution rate correction factor, right?

DR. HOGAN: Yes.

DR. BELL: I misread that from your paper.

DR. HOGAN: Sorry.

DR. BELL: That answers half my question. I guess the other half is how well do people believe this sort of really tells you about the extent to which undercount is going to be heterogeneous as opposed to these other things?

DR. BROWN: It does not tell you whether—I mean, there is an assumption in the analysis that the census allocation rate is heterogeneous in the same way that coverage is heterogeneous. This is just a pure assumption. Then what you find out is, if that assumption is valid, how well does a PES methodology work?

It essentially constructs some artificial populations in the United States with artificial coverages and then you go back and you look at those populations and you see how well did the PES methodology work at the state level, at the county level, and so on.

DR. BELL: So I guess my question is, do the people who have thought about this more than I feel as if this is valuable?

DR. BROWN: David Freedman thinks it is valuable. I think his paper is pretty convincing, but there are other people who do not think it is convincing at all. That is the Statistical Science paper, it is a discussion paper. It is something you can do.

DR. ZASLAVSKY: Or you can say that about almost anything else.

DR. BROWN: It is something you can really do, that is for sure, and you can really say something about it and then you can discuss whether you think—one of the things you can discuss is whether you think these things look like coverage, and then there are a lot of other things you can discuss as well.

DR. ZASLAVSKY: You do not really have to believe that these are distributed just like coverage. You can say here are some other variables that are generated by real processes that we control as little as we can control undercoverage.

DR. BROWN: Right. They certainly do not have to be distributed like coverage with respect to the geography. They should just be variable in the same general fashion that coverage is, that is all.

DR. STOTO: I am trying to understand the relevance of this. I think you were, first of all, ambiguous about whether you called this synthetic bias or synthetic variance. It sounds to me as if it really is a variance, because if it is going to be high in one place, it has to be lower in another place to balance out.

DR. BROWN: It does not, necessarily. It can result in bias. There is variance, obviously there is variance in here, too, but there can be bias as well.

DR. HOGAN: At the very local level, at the block level, then the pure variance, I think, is important, but once you aggregate to most important levels, then the bias, where I mean people actually had different probabilities of being counted a priori, before we decided which interviewer to send down that street, is what dominates the discussion.

DR. BROWN: It is mixed up with correlation bias in that sense, so it is hard to separate out the two terms in that analysis.

DR. HOGAN: They are obviously related, because they both relate to heterogeneity in the census. I think they are distinct phenomena that, at least in my mind, I can keep separate.

In a perfect world, and this goes back to our earlier discussion, if the A.C.E. had uniform coverage, then you could have local variation in census coverage, so no correlation bias. I am not saying that exists, but that is certainly part of the model.

On the other hand, you could have a correlation between the census and the A.C.E. capture probabilities, that caused correlation bias, but if that heterogeneity in the census was not geographically localized but had a little bit of this group and a little bit of that group, then that would not necessarily cause synthetic bias.

Really, they are two separate things.

DR. STOTO: But is something being proposed to be done about this, some adjustment or correction?

DR. HOGAN: What is being proposed is we are trying to look at it, trying to figure out its effect on the quality of the unadjusted census and its effect on the adjusted census, so that when we make our decision we can say this is a problem or we do not have a really good handle on it, but we think it is about bigger than a bread box, and it is one that we can either ignore or one that we really have to come down one way or the other on.

DR. BROWN: Are you proposing to do some analyses like this?

DR. HOGAN: These synthetic ones? Yes.

DR. BROWN: Okay, this is not just a speculative section, but one that is going to be tied to a paper.

DR. HOGAN: There is a paper, I am sorry. The paper is B-14 [Griffin and Malec, 2001], so we are working to get this going.

DR. BELL: One thing I have considered—I have not tried doing it—and thought about is whether one could do something directly with the A.C.E. estimates of coverage; for instance, taking match rates in each of the A.C.E. blocks and forming sort of clusters of blocks that are as large, say, as congressional districts and looking to see whether those clusters have similar match rates, say, adjusted for what the overall A.C.E. would say.

I do not know if anybody has ever tried anything like that or not.

DR. HOGAN: You would be doing something like a direct congressional district DSE, something like that?

DR. BELL: Something like that. It would not tell you anything specific to post-strata, because I do not think you could disaggregate down to post-strata at that level of geography, but perhaps by saying, okay, the direct estimates of population are systematically higher than what the synthetic estimate is saying for this cluster of 20 or 30 blocks that would make up a congressional district and, for this cluster of 20 or 30 blocks it is consistently lower.

DR. YLVISAKER: I have done some of that, but not with a geographic component.... [Microphone off]

DR. BELL: I did a little calculation and it seemed to me as if that could lead to a reasonable estimate of what the synthetic bias might be at a level like that. You do not have a lot of information about any particular, say, congressional district if you have only 20 or 30 blocks going into it, but you would have 435 of those, so you might be able to estimate a variance pretty well.

DR. HOGAN: We could think about seeing if it could be done. That is worth thinking about, certainly.

Anything else on synthetic error, variance, bias, however you would like to characterize it?

[No response]

FORMING AN OVERALL ASSESSMENT

DR. HOGAN: The next section is on other measurement and technical errors, I just want to mention them. Since the A.C.E. is based on a ratio estimator, it has an initial bias. We think our sample sizes are such that that would not be a

problem but we will look at it—it might be a problem for some of our very smallest strata.

DR. BROWN: Go back and look at this. I see there is a number 30 that you cite from Cochran in 1963. It is easy enough to just see whether that number really is a good number. I have something in mind. There are other places where 30 has been used in the statistical literature from the 1950s and 1960s. That may have been okay in the 1950s and 1960s in terms of the accuracy that could be gotten then, but in terms of the accuracy that we are looking for now, 30 is not good.

Degrees of freedom, people think of 30 in terms of degrees of freedom, and there is an npq equals 5 or 3 rule, and that rule, I know, is not a very good rule.⁵

DR. HOGAN: I completely agree. I just threw that in to show that I could find Cochran.

DR. BROWN: Yes, I know, I understand. These things tend to become institutionalized in ways that we accept and probably should reexamine.

DR. ZASLAVSKY: People have ways of estimating bias using the bootstrap, and so on.

DR. BROWN: Yes, there is no problem, it is not hard to double-check or triple-check these kinds of things now.

DR. HOGAN: I completely agree. Contamination error: this is if, somehow, having conducted the A.C.E. listing and whatever else in the block, it means that within the sample blocks the census gets conducted differently, that has been a concern. In 1980 we actually [were]—or Bob [Fay] was—able to demonstrate that that was a problem in 1980. It was not a problem in 1990.

There was concern when we had the ICM [Integrated Coverage Measurement] that it might have been a problem. I am not going to flat out say that it is not a problem, but it has never been, except in 1980, a big enough problem that we have ever been able to measure it. It tends to be a very subtle problem if it is there at all. We acknowledge that it is out there and might be related to the telephone, and we will look at some of the timing on the telephone, but probably you have nothing explicit on that.

Finally, among the sort of technical problems is the inconsistency in the post-stratification. This goes back to some of our earlier discussion, that someone can, in the A.C.E., mark down that he is Hispanic, while in the census his neighbor said he was not Hispanic, or in the census he is Hispanic and in the A.C.E. he does not answer the question and we impute him as non-Hispanic.

We have our post-stratification variables for the cases where there is a match. We have what the census said, imputed and non-imputed, and we have what the A.C.E. said, imputed and non-imputed. We can see how similar they are.

Some of the concerns about reporting, I think, we have guarded by our choice of mainly race domain. For example, by and large, Hispanic is going to be a race domain, or race-ethnicity domain, regardless of how many multiple races the person chooses or does not choose, so if someone reports Hispanic and either does

⁵ npq is the variance of a binomial random variable, sample size n , with success probability p , with $q = 1 - p$.

not report his race or reports it inconsistently between the census and the A.C.E., that will not affect our post-stratification.

Similarly, for example, for American Indians, people who mark American Indian and some other race who are not associated with reservations or other Indian country will be put into domain seven with whites and some other people who mark five races, simply because we think there is so much inconsistency in people suddenly discovering a Cherokee grandmother, especially given the option of multiple reporting.

I think we have guarded against a lot of this in our choices of post-stratification, but it is something that is easy to look at, just tab it and look at it. Later on, not for evaluating the A.C.E., for evaluating the census we can actually do the full 63×63 and learn something about how consistent this new race reporting is.

Then the next two steps have to do with how do you put these components together into some sort of overall quality measure or quality measures for the A.C.E., and whatever you come out with, how do you compare that with what was the overall quality of the census?

We have a lot of data from both sides and how do you put it together? The next two sections talk about our direction for doing this. I think I am on your page 5, "Synthesizing A.C.E. Quality." This relates to document B-13 [Navarro and Asiala, 2001].

What we are trying to do, at least let me say where we are and where we hope to go and how we hope to get there, is bring in each of the components, what we know about them, the level of error, the uncertainty that that component contributed in 1990, how what we know compares in 2000. Do we have evidence we have improved it, do we have evidence that it got worse? We will try to quantify that into essentially a total error model.

Now we have Bruce Spencer working with us and, I am very happy to say, we now have Mary Mulry working with us on this—she worked all weekend on this (we just brought her on board and it is very good).

We face a couple of challenges, which is one of the reasons we brought Mary in, one of which is that it should be pretty evident by now that each of these components that I have talked about we do not necessarily have a nice quantitative measure for. We have some evidence on each one, what happened in 1990, what happened in 2000, did it get better or worse, whatever? Setting up the model to use those data in a very reasonable way is a challenge.

Secondly, and I think it is also very important, we do not have a lot of information about how the pattern of some of these errors might have changed, and let me give you an example, and this is my personal opinion. I think we made great improvement in matching technology. We have not started the matching yet, but I believe we probably will reduce matching error. I might even be able to convince you that we reduced it a lot.

Now, I would have a hard time convincing you that we [reduced] it a lot more for Hispanics and somewhat less for whites and a little bit more for Asians, how did that improvement impact the various groups that we are going to have to assess accuracy over.

On each one of these we are going to have information about did we have, generally, an improvement or, generally, a deterioration? Some of them, say, correlation bias, we might have black/nonblack, but putting together this information with 1990 we will probably have to make some fairly simple underlying assumptions in terms of the pattern remained the same but we scaled it differently, things like that.

We will probably have to put together a fairly straightforward model, where you can just say I think this might be half as large and this might be twice as large as 1990, so that we can get results out. The 1990 approach, which had a number of evaluation studies that dragged on for months but also required an elaborate resampling process out of covariance matrices and distributions, we certainly will not repeat, I want to make that clear. We might shortcircuit it in various parts and make some assumptions that we can do part of that, but not the whole thing from data collection through our total error model.

We are going to have the kinds of data that I talked about here for this decision and those are the kinds of data we will have to decide on.

Related to the total error model, or what I call synthesizing the quality of the A.C.E., is how do you compare that to the census? In 1990 we had the loss function analysis, where we had models at all sorts of levels assuming with and without correlation bias for states, for large cities, for cities of 100,000—I forget how else we split it—we had model after model, squared error versus absolute error, went through all of them and had a number of results there.

That is probably not going to happen in time for us to make this decision. What we are going to have to choose is a model or two that makes a lot of sense, that we can work with, and set it up so that when the data come in we can plug them in and get some statistical assessment about what, based on everything we know from the statistical data, what these loss functions say in terms of the relative loss, the relative risk between the census and the A.C.E.

Not too far down this road we hope to make great progress soon and that is why we brought Mary in and we have Bruce working with us. If anybody has any suggestions, especially about simplifying assumptions, I would be thrilled.

DR. NORWOOD: So much of this is a problem of—[microphone off]—and how much of it is that you really will not have the data in time. Do you anticipate that you will have to rely on 1990 data?

DR. HOGAN: To a certain extent we will have to rely on 1990 data. I envision a process—and more than details have to be worked out—where we have, say, the 1990 matching error and that broken across whatever groupings and going in and somehow saying, no, this is much better and what would be the effect of that being much better, and working along those lines.

The 1990 [error analysis] had, I think, depending on which set of studies, 13 evaluation post-strata, where we actually tried to measure the errors in each of these 13 post-strata. We are not going to have that, so to a certain extent the patterns of how these errors display across subgroups, we are going to have to rely on things like the 1990 results.

Part of it is the input data, part of it is the 1990 approach was incredibly elaborate with every possible way we could do it, this way, this way, this way, crossed by that way, that way, that way, and this way, for this area, this area, and this area. It was just an incredibly time-consuming process that just went on and on and we do not have the luxury of repeating that.

DR. BELL: The detailed thorough analyses are going to come later, but do you have a sense of what kind of a handle you will have on how much of the total error in the adjusted estimates is systematic in some form or another? The reason I ask is that you could imagine at some relevant level of geography the total error in the adjusted and unadjusted estimates is in kind of the same ballpark, but then how you feel about the adjusted estimates might depend on the extent to which the error is systematic. You might feel one way if it seems to be mostly random and another way if the systematic component is nontrivial.

DR. HOGAN: That is a good way of looking at it. For some things I think we will have a feeling for it and for others we will not. Matching error, since it is now centralized in one office with one set of clerks and one set of analysts, probably what you are going to get will be random error. You will no longer have the Albany office probably doing something different from the San Diego office, so you have eliminated a source of systematic error.

The interviewing, our only information on whether it is systematic or not will come from some of the things I talked about before, how did the proxy rates, the noninterview rates, vary across geography, how did they vary across post-strata, that might tell us was this a general across-the-board problem or one that was fairly localized and random.

That does not completely answer your question, because I think we have to think more about your suggestion.

DR. BILLARD: Howard, you said you do not have the luxury of time to look at all these loss functions over all of these different axes and information cells, if you had the time—[microphone off]—necessary to do them over all of those things? Maybe one works better than another, I do not know.

DR. HOGAN: Two questions, would I do it and is it necessary? Being a researcher, of course, I would do it every possible way.

Is it necessary? Some of them I think we went overboard with. Some of it I think may not be relevant for the decision at hand in the spring, which is a decision that focuses fairly directly on the uses of these data for congressional redistricting and maybe other state and local redistricting.

Some of the loss functions we did in 1990 were maybe relevant for other uses beside the one that this group has been asked to make a recommendation on. I might do a few more than we have planned but I personally, except in a research mode, would not go back and do them all.

DR. YLVISAKER: I guess the problem is not so much running the different loss functions; the problem is targets, so that you can talk about some sort of truth. These targets, I think, are awfully problematic at this juncture. We can discuss what the targets ought to be. Loss functions after that might be of some use.

DR. HOGAN: Yes, I think there are two issues and I am glad you differentiate. There is one that is simply the technical question, to use squared error or absolute error. I agree that that is probably not where the action is.

The other is how you come up with the targets and that gets back to how you use all the information you have thus far. In coming up with the targets you need to do a couple of things. You have the census results, you have the A.C.E. results, you have the variance on the A.C.E. results, you have your estimate of A.C.E. bias, and the variance on that. I guess you could go on down the line there, but let's stop there.

In doing these targets, the key things are the bias of the A.C.E. estimates and the variance, or at least the uncertainty, around those. Depending on how you read what we have done so far today, you might come up with different biases and certainly different variances around those.

Plugging those into the loss functions is where we might come down differently, not, probably in squared or weighted squared, or whatever.

DR. YLVISAKER: Looking in various places for consistency and what looks right to us, and so on, there is one way to avoid truth and that is to simply look at the difference between the census and the synthetic estimates. I do not know if anyone did that in 1990, but you have 7 million blocks, some were adjusted up, some were adjusted down, and you have a pattern over the entire country that could be looked at.

It has nothing to do with truth so far—I can induce a variety of truth to this—but if you were going to have a pattern of where you have actually adjusted up and where you have actually adjusted down, is this a meaningful pattern? I do not know that applying some target of the variety that has been mentioned, which seems to me awfully problematic, is going to give me a number that I am going to believe in a lot more than what I just mentioned.

DR. HOGAN: That is an extremely good point. There are really two things going on here and let me emphasize where I think you are going. One is that many people find it useful to take the individual components via a total error model, via a loss function, and come up with a scalar with ranges of uncertainty. That was very important to the 1990 decision, for a number of reasons.

In approaching the [2000] decision, the ESCAP would like to see that, because it was so very important in 1990. However, in a later section, which we might as well get to now, we actually have to make the decision. This is one input, this is one thing that we are going to weigh. Some people might weight it heavily, some might weight it less heavily.

The real decision, in my mind, is you take what this says and the uncertainty around this kind of analysis and do you believe the targets or do you not believe the targets? You compare that with demographic analysis and, again, do you believe demographic analysis, do you not? With what you know about the census, how the census was applied, the results of the full-count review, any local information. You put in your mind all this information together and, as a decision maker, the question for me is does the story, does the pattern, does the level that the A.C.E. is telling me about where the population is, make more sense than the unadjusted census?

So this is an input, but I do not want to make this the goal of everything we have done so far, where we come up with λ and, if it is positive, we adjust and, if it is negative, we do not adjust. It is one input that some people find very useful and some people might be totally confused by it.

DR. YLVISAKER: I would argue against loss functions as a decision mechanism in basically any form other than as some sort of contributor that somebody wants to muse over the weekend about, or something like that, because there are too many users and there are too many uses, and so on, for it to be a determinant in any form.

DR. HOGAN: I do not think we think of it as a determinant. We think of it as one more piece of information, one more way of looking at the data we have got in here.

DR. YLVISAKER: We are not talking about a single number but hundreds of them, say?

DR. HOGAN: I do not know if we will have hundreds of them. We will have somewhere between one and 100. We [will] clearly have several. We will clearly have loss functions looking at the count. We will have loss functions looking at proportional shares. Clearly we will do that. We will have them at the congressional district level, and we are going to work to have it at the level of the average state legislative district.

DR. YLVISAKER: With targets chosen as you mentioned before?

DR. HOGAN: Yes, with our reading of these data. The target, at least with my understanding of the loss function, is fairly complicated in the risk [meaning not clear], but you bring in the undercount the A.C.E. had, that difference, together with your biases and variances around that.

DR. YLVISAKER: The biases, which we do not really have.

DR. HOGAN: We have our measures of the biases with the uncertainty around them.

DR. YLVISAKER: The uncertainty from 2000? I mean, maybe with large uncertainties, yes.

DR. STOTO: One way to avoid the difficulty of the loss function is not to use a loss function but to think about patterns. Presumably, what we are talking about here is a change that would reduce bias at the cost of some variance. Of course, that is not the same across the board. Variance is more important in small areas and bias is more important for larger groups, and there are different kinds of biases that, presumably, would be more important than others—if the undercount is concentrated in certain demographic groups or in certain parts of the country or in certain socioeconomic groups, and so on.

One thing, maybe, to think about in doing this would be think about different patterns of bias that might be there and different patterns of variance that would be introduced, think about the tradeoff between them. What are the things that we are comfortable with? At what point do we think that we get a better deal by adding some variance to reduce bias? Maybe even discuss them with politicians and decision makers to get a sense of what this is all about.

DR. HOGAN: Leave aside discussing with politicians, I think we do need to return—and this may be a good time—to where we came in. We, the Census Bureau, did not go down the dual-systems route without a history and it is the history of the pattern of differential undercount.

While we certainly hope that the A.C.E. or whatever can make the census better overall, that is certainly part of its goal.

The principal reason this whole discussion started 25 years ago was the differential undercount. I think in looking at the patterns, the patterns of biases, the patterns of variances, it is very important in our decision process. First, does this census have a differential undercount? If the answer is no, that is going to be very important evidence. If the answer is yes, and the A.C.E. can make this a lot better, all other things being equal, then that is very strong evidence to go forward (a lot of other things would have to be equal).

We should not look at this in terms of mean squared error. It arose from a history.

DR. STOTO: I guess another way of saying what I said might be to think about what kinds of tables and charts you might be able to generate when the data are in hand that show something about the differential bias and how much that can be reduced and to show something about the uncertainty and the estimates at different levels and how much that might be increased, to sort of think about being clear about the pros and cons of the different options that are on the table.

DR. HOGAN: I am getting really close to the end. Let me say just one thing—I will not even dwell on it very much. Our uses of time have been proportional, I think correctly proportional, to the importance. The A.C.E. is not the only statistical adjustment that we are proposing for applying after the apportionment counts. We also have the multiplicity estimator for the service-based enumeration, an estimator where we show up at, say, a homeless shelter—it is a little bit more complicated than this.

We show up at a homeless shelter and we ask who is here now and how often do you use homeless shelters. The person says once a week, the person gets a weight of 7. If he says he is there all the time, he gets a weight of 1. This is a statistical model that we will be conducting and looking at, making sure it was under control and, assuming it was, bringing that into the census files as well as the A.C.E.

There is a paper here that documents that, if anybody is particularly interested. We can discuss it, either online or off line, but I did want to make sure that it was not completely forgotten in the discussion.

I think that brings me to the end—in more ways than one.

DR. BELL: On that last topic, I did not see anything that really was an attempt to evaluate the multiplicity assumption or how good that estimator was going to be. Is there anything planned?

DR. HOGAN: No, apparently not. It was tested in the dress rehearsal, we have the results in the dress rehearsal.

DR. NORWOOD: I want to, first, thank you, Howard. You have done yeoman's service.

DR. HOGAN: You are welcome.

FINAL COMMENTS

DR. NORWOOD: I would like to go around the room and ask our invited guests for comments, if they have any, that they would like to make. May I start with Marty [Wells]?

DR. WELLS: It seems from the discussions today in the morning that there are a lot of methods and tables to find the errors in the census pretty easily—it is easier to find those. In the afternoon we talked about finding errors in the A.C.E., and a lot of those were a bit harder to find. It seems it is human error in decision making—when you can find errors in one place and the other errors might be a bit more ambiguous, it is easy to lean to where you can find the errors.

Could you just comment on the way that you are going to think about this, because it just seems that with the A.C.E. a lot of the discussion was not as detailed as we had this morning and how are you going to weigh the two, the ambiguity in evaluating the A.C.E. and the more systematically evaluated census, the way you are going to balance that in your committee.

It just seems that the A.C.E., a lot of the evaluation and quality issues you talked about, they do not seem to be as solid and on as firm a foundation as the evaluations of the census. It is easier to find errors in the census, maybe, than it is to find errors in the A.C.E.

DR. HOGAN: Here is the way I think of it. First, for both of them we have comparison with the demographic analysis estimates and that has been the traditional standard there, and it can be equally applied or misapplied to the census and to the A.C.E. at the aggregate level.

In terms of other errors, my reading of these documents is we are spending a lot more time on finding errors in the A.C.E. than in the census. The kinds of errors we have when we have the results of quality control on the one and quality control on the other, some operational measures on the one and operational levels on the other, when we actually focus in on the components, we are really, I think, putting the A.C.E., at least in my viewpoint, under a far more intense spotlight than on the census processes, so I am not sure that I agree with your premise.

DR. WELLS: But it seems that the A.C.E., in some of the evaluations it is just not clear, the correlation bias, synthetic bias, various things. You cannot get as detailed an assessment there.

DR. HOGAN: Outside of the A.C.E. and demographic analysis comparison, the same kinds of things apply to the census when we do not know much about the coverage from the census errors.

The one thing I think makes it a little more difficult is, to the extent that the A.C.E. is successful and has reduced the gap here between the count and the true population, you are measuring a smaller residual. We start out with, at least using 1990, a gap of about 2 percent. At the end of the PES we had it down much smaller and, therefore, seeing how really close you were at the end it was a little bit harder with the PES.

Again, my reading of this is we are putting the A.C.E. in certainly as intense if not a far more intense spotlight than the census.

MR. THOMPSON: The A.C.E. we put under a much more intense spotlight than the census. We do not even try to measure the response error in the census and the synthetic error in the census or the correlation bias in the census before we apply the results for apportionment. We are putting the A.C.E. under a much more intense spotlight than the census. Those errors are in the census, too. We just do not try to measure them, because we do not have the facility to do it.

The A.C.E. being a sample, we have the ability to try to understand it a little bit more than the census. I think I am with Howard, I think we are putting the A.C.E. under a much more intensive spotlight.

DR. BILLARD: My reading of what you were saying, Martin, was that we have to separate out operational errors from statistical errors, and we did spend a long time this morning talking about operational errors, and it was the statistical errors that we were talking about in A.C.E. We did not happen to talk about the operational errors in A.C.E.

DR. NORWOOD: The papers talk about the procedures and all the same operational errors. In fact, we know more about A.C.E. because it is a statistically designed sample survey, but that is a useful comment, because I think it is important, if that is the impression that has come across, that you be evenhanded.

DR. YLVISAKER: I would just go a little further with that, because A.C.E., I think, has to be held to much higher standards and they are arguing that it is. I missed in that section on the evaluation of A.C.E. information about how the quality of A.C.E. data might track the quality of census data, the perceived quality of census data; that is, in local places, where census counts are bad, we are going in and, presumably, correcting them by looking at those people.

If we are using data that are of about the same quality, then I do not expect to see a lot of corrections. I think that means it has to be inspected and maybe even more than appears in that particular section.

I had some questions about the attention paid to adjustment results. We are looking at 448 times 51-or-so pieces of paper as local people come in, and we look to see if we can see inconsistencies, and so on. My first question is what is an inconsistency, and my second question—of course, the first question has no answer but the second question is what is the remedy, and I think you satisfied me somewhat in that regard; that is, a remedy is that once we have the data we put out our A.C.E. numbers and that is it.

I think I got assurance from you on that score, so I will not raise that particular issue, but I am a little disturbed by the stress that has been put on numeric accuracy as opposed to distributive accuracy. I guess I would point in that direction to simply how one is going to look at these two things. We are looking for consistency. How does one look for consistency? One compares.

As I say, distributive accuracy is really what happens in the world. We do not say, well, let's see, did somebody get up to 50,000 here, that means the census is better or the A.C.E. is better, or anything of that nature. It is not done in an absolute sense, it is done by making comparisons with 1990, with everything else, so I think distributive accuracy should carry more weight than numerical accuracy.

I am a little disturbed that it has been made as a preliminary decision that A.C.E. is more accurate. Then I guess it has been solved that this is now the corrected and non-corrected form of the census.

DR. PREWITT: May I ask a question? This sounds more flip than I intend it, so forgive me for that, but I do not know how else to frame it. Let us say that we have now finished the basic enumeration work, we are doing our data editing, and so forth. We think we did a first-rate job. We do not think we are going to field A.C.E.

Would you feel good about that decision?

DR. BROWN: This is a hypothetical question.

DR. YLVISAKER: Say it one more time, please.

DR. PREWITT: The hypothetical is we finished the census, the basic enumeration census, and we think it went pretty well, we counted almost everybody, and we are tired. We are coming in here today to announce we are not going to do the A.C.E. Would you feel good about the census?

DR. YLVISAKER: I have a problem with the census. I think Howard mentioned the problem, it has come up in a couple of ways. When we run the primary selection algorithm we decide that this is the census. If we go back, we say, well, that is the way it came out. We do not go and look to see what the other form said. This is how the census came out. Howard mentioned this with respect to the variability across post-strata; that is in the population, this is the census, after all. We might find it in A.C.E., too, but that is what we are finding, it is the census.

DR. PREWITT: Let me then ask you, did the census include imputation? What if we had stopped the census before we did the imputation?

DR. YLVISAKER: I did not raise any questions about imputation.

DR. PREWITT: I am just curious about at what point in the process do you decide that we now have the best count we can have? We could have stopped prior to imputation.

DR. YLVISAKER: I guess my basic problem with adjustment is and will continue to be that we do not know precisely where to put the people. We can run these things, there is no doubt that we can run it, but we do not really know where to put people and there are some pretty convincing arguments that we have got a whole pile of people and we are not quite sure where to put them, and if we have a mechanism for putting them, I say fine, but I do not know that that is a better picture than the census.

DR. BILLARD: I have not read all these through as carefully as I would like to. I must say I am very impressed by the material we have been given and the presentations and listening to the discussion today.

I am left here sitting and thinking—well, one reaction is that I am amazed at some of the sorts of things that can go wrong or the types of data that you should get that you do not get, I am amazed at how many of these have been thought about by the Bureau and by different people and have answers to them. I am not sure I could have thought of all of those places where it can go wrong myself.

I am reasonably comfortable from what I have heard that systems are in place to check things like the matching, movers, and all of these things. I am not going

to sit here, and I cannot imagine the Bureau would, either (maybe they would), and say that the processes in place are the best they can be. I mean, you can always improve with time and, anyway, I have not sat and looked at those algorithms, I have not used them, how can I evaluate them in that sense?

If we are going to see results afterwards about how they work and how they do not work, that is the best we can ask for and I am pleased to hear that we will have them.

In terms of the A.C.E., historically, we do know historically—and demographic analysis has shown us—that there is a differential undercount, so it seems to me important that we do, again, the best we can in terms of trying to measure that.

My understanding is the methods being proposed here will help measure that. I think the same process errors that you were talking about a minute ago relative to the census, my sense is that they are also in place for the A.C.E., so I do not have a problem with that.

I do not know what the biases will be, what the variances will be. I do see that the sample sizes are roughly double—is that right?—what they were in 1990, and that has got to be a huge improvement right there, even if that were the only change. That would seem to me a huge improvement. We just have to get the best count, the most accurate. It may not be the absolute correct count but it has to be the most accurate that we are able to do and it sounds to me as if we are on track.

DR. BAILAR: I think in previous census years, where we thought about the correction or adjustment of the census, we have used only one measure of the accuracy of the census and that came from demographic analysis, so we were looking at primarily was there an undercount, was there a differential, so I was glad to see some emphasis on looking at the errors in the census itself. I think that is really the first time that has been done.

Even so, some of this may not be available at the time when you are making your decision, but will come out as part of the evaluation studies, but there will be some things that are available, and I think they are useful to look at to give you indications of the quality of each of these different pieces.

I really am just in admiration of the kind of preparation that has gone into this, with all the tables and prespecification. It is obviously a tremendous amount of work thinking about how those tables are going to look, how they are going to be prepared, at what level, and so forth. I think it is very good, and I am sure it is going to give you a lot of insights into where you have problems, where you want to look further, but it is also going to be, I think, a very good indication to those critics of the process that you have prespecified things as far as you can.

I do think your procedures are simplified over 1990, and I think if this is the year when adjustment actually is done, that that is a good thing, though you may want to go back to something else. If you do it once, you are probably going to do it from then on. I think then you can go back and look at other ways of doing things, but I think this year it has to be, probably, as easy as you can to explain it to the public.

I think the only thing that dismayed me today was listening to some of the conversation where, even if you are able to make the adjustment this year, that

your problems really are not over yet, because everything is getting more complex and I am referring to some of the things that were mentioned about the race and ethnic groups, where this is getting so blurred that you have to think about new ways of looking at that.

DR. SCHIRM: A couple of points. One is more of a procedural thing. It came up several times today that some of the evaluations will not be completed in time to really have a bearing on the decision, and that is sensible. That will require some reliance on results from the 1990 census or the dress rehearsal or other research, and that also seems quite sensible.

I think, just procedurally, it would be helpful to make it formally clear and transparent where and how results have been used from other evaluations, be it from 1990 or the dress rehearsal. As I said, I think those other results should be used and used systematically, but just making it clear where they have been brought to bear on the decision.

The one other point I want to make is just to emphasize a topic that came up here at the end, which is that I would encourage spending time trying to identify patterns of errors in the adjusted estimates by as many characteristics as you can look at.

Here the notion is that we may be willing to tolerate equal or even more total error if we can remove certain systematic errors that we find particularly intolerable, or it could very well be that the introduction of certain errors may be intolerable even if total error is reduced. I think it is helpful to do everything that can be done to look at the adjusted estimates and the systematic patterns in them.

DR. ZASLAVSKY: I will try not to repeat things other people have said. This is obviously very challenging stuff, and it is impressive to have such a list of things that you can do before the evaluations. It is like what do you do before the doctor comes when you are in a remote rural area.

I guess it seems as if a lot of the things that have been laid out, we have not said explicitly how you would tie all these things together, but, in some sense, what we are saying is that what you are looking for is process evidence of a reasonable degree of homogeneity of error in the A.C.E. and a relevant degree of heterogeneity in census quality, and relevant heterogeneity implying heterogeneity at the levels where you are actually trying to measure, so heterogeneity between New York and Boston is not relevant, because those do not correspond to post-strata, but heterogeneity between rural areas and urban areas is, because that is part of what you are trying to measure. Maybe some of that could be made a little more explicit. Part of what is lacking is that you have said we are going to look at all this information and look for some patterns, but what are we looking for in those patterns? I think you have some ideas about that and maybe that could not be formally prespecified, but the rationale for looking for certain kinds of things could be articulated a little bit more explicitly before the process gets under way.

The other comment I would make is that I do not want—I will be the advocate for not downplaying the quantitative comparisons and loss functions and those types of analyses. I find them very useful and important, especially if we can identify levels at which they are likely to be useful, not excessively small levels

of geography but levels that correspond to not necessarily geographical but large population blocks for which we know there are likely to be variations and for which we can estimate whether we have been able to reduce those by adjustment or whether we would not be able to.

I know that some of what we would like to have for that is missing. We do not have direct bias estimates for the A.C.E., so some of that will have to be based on 1990, updated to some extent with the process information we have now compared to what we had from the last census, and there could also be some sensitivity analysis to see how much those comparisons actually depend critically on the bias estimates, or whether, under some range of reasonable assumptions about the biases, the evidence points pretty much in one direction.

Also, if we are dealing with the synthetic estimation issues, some comparisons of direct estimates to synthetic estimates, again for fairly large areas for which those comparisons could be meaningfully made may be helpful. I would like to see those be part of this, also I realize that stuff is not to be applied mechanically as a decision mechanism.

DR. STOTO: I have two things. First of all, I would like to second what a couple of others have already said about how impressed I am with the degree of preparation and thoughtfulness that have gone into these papers and the presentations and how important I think these are.

I was last involved with this issue just after the 1980 census. It was just so hard to try to sort these things out after the fact. Having all this done in advance, I think, is going to be really very positive. I really want to commend you all on this.

The second thing is, I think that one of the questions that we really have not addressed today, or maybe when I was not here, was how we are thinking about the decision process. Sometimes we were talking about being able to prove that the A.C.E. and the things that go with it are better than the census, and I think that another way of thinking about it is what is the best possible estimate we can make about the size and characteristics of the American population at a variety of levels, and not to give prior weight to one side or the other, but just to think about how, from the scientific point of view, can we do a good job of estimating the population?

There may be some people in town who do not want to think about it that way, but, as scientists, I think that is what we have to keep on the table.

DR. NORWOOD: Any final comments, Ken?

DR. PREWITT: I guess it is not a question that has to be answered right now, but if those who think we should not adjust, the question we put to them, in effect, is, if we are going to adjust, are we doing the right thing; that is, is this a reasonable strategy for trying to make the adjustment decision?

Obviously, if you think we should not adjust, then that is fine, I appreciate that argument, but the question is, if, is this the right way to approach that decision? I think that is extremely important for us to hear back on.

The other thing I would say picks up on what Mike [Stoto] said. We actually have approached this as a process by which we are trying to get an estimate that is closer to the truth and, indeed, we, even after nonresponse follow-up, did about five

or six major field operations, each one of which we thought got it closer, coverage improvement strategies, doing something about the population count, a number of things, even before we got to the data processing stage, or we would not have done them. If one of them had collapsed, we would not have put it into the final count. If some things do not work out, we do not use them.

Then we also do imputation. We put whole persons into the census before we report the apportionment count. We, again, do that because we think it is a better estimate. We see the A.C.E as simply one more step in that process. It is obviously a demanding step, and that is why it is getting the kind of attention that it is getting, though it surprises me how much more this process gets than some of the other things we have done to try to get the estimate closer—it is just surprising to me that somehow this one takes on a life, a political life and a public life, way out of proportion to almost anything we did in getting the first 98 percent counted, but, nevertheless, that is the environment in which we find ourselves.

I think the most important thing I would like to say—well, before I say that, let me come back to the panel and, obviously, express the Bureau's appreciation. I am aware this is all pro bono, you are doing this because you have professional statistical commitment to a quality census. It is extremely valuable to have had the panel as a part of this process over the last year and a half, extremely important to have the Academy have the 2010 committee to begin to move from where we already are in terms of how we will approach 2010.

It is appreciated and, obviously, we will all read your report carefully. The entire world will read your report carefully, and I say that because I appreciate the magnitude of the burden on you. You are also going to be scrutinized just as we have been scrutinized. A lot of people are going to read the committee's report very carefully, looking for nuances, looking for hints, directions, suggestions of internal disagreements, and so forth. This will be a highly visible report for not just the statistical community but for the political community as well.

We do appreciate that there is a very heavy burden on you, and we also realize that you would like to make a timely report, and we will do whatever we can to facilitate your making a timely report.

Again, just to repeat what I said this morning, I would hope that your report, or somebody's, if not yours, the Monitoring Board or the subcommittee, would sooner or later systematically address the question of political manipulation.

Just to go back to that for a moment, if data are collected by an inefficient organization, then why believe in them? That is why it was very important to us to try to prove that we were an efficient organization in doing the census. But it is even worse if data are collected by a corrupt institution, a politically corrupt institution. Then why believe in it?

We really have an obligation, it seems to me, to say to the society when Census 2000 is over, whether we correct or not, adjust or not, there was a reasonably intelligent operation to try to collect information about 275 million people, plus or minus, and it was not a corrupt operation, because if the American public is allowed to believe it was either done by an inefficient, ineffective organization or done by a politically corrupt organization, we somehow are not servicing or making

the kind of statements it seems to me we have to make and should be able to make publicly. I just want to emphasize that once again, the real importance I put on that.

I guess my only other comment, then, Janet, is I hope you believe us, and I think maybe most of you do—I am sure the panelists do—this is not a predetermined decision. We have not made the decision. We actually are going to look at all of this. Obviously, if we did not think that the A.C.E. and dual-systems estimation would improve the census, we would not have spent the money, the trouble, and the effort to go do it.

Obviously, we go into it on the assumption that it will make an improvement, but if it turns out that we think it did not, then we will simply—either it did not because it did not pan out the way we hope it will pan out or because it is not necessary, because, lo and beyond, we did count 100 percent of the American population (doubtful, really doubtful), because I got too many letters from people saying, “No matter what you do, you are never going to count me,” and some of whom actually sent in one hundred dollars (that is the fine).

They sent in checks, saying, “I’d rather pay than be counted.” We got some of them through proxy (maybe even all of them). We did send the hundred dollar checks back.

I only say we are doing it because we believe in it, obviously, we are a statistical agency, but that is very different from saying we are going to do it, that is, make the decision about whether to adjust or not, based on our analysis in a very, very tight time frame. We, after all, have statutory deadlines and we intend to meet those statutory deadlines.

I do hope that the panel and everyone else who is interested in this process appreciates that [it] is a really honest self-scrutiny about the quality of the census and whether the A.C.E.-adjusted numbers will or will not improve it. I think, Michael [Stoto], your formulation of it is correct. It is not is this one better than that one, but do the two of them together produce a better estimate than either of them independently of the other one. That is exactly how we are approaching it.

DR. NORWOOD: Thank you.

Let me just say that I think we, the panel, have benefited from hearing all of this. I would like to particularly thank Howard and his whole staff. I know they have done a great deal of work and I know what that takes. I have had enough experience in statistical agencies to know how each of those papers represents massive amounts of work, so I want to thank you for being so cooperative and for providing us with information on this whole process.

I spent this summer, a large part of this summer, revisiting the early experience of this country. I suddenly, I do not know why, got interested in the revolutionary period and the period thereafter. I read a number of things, including a very good set of biographies of George Washington, of Thomas Jefferson, of Alexander Hamilton, and I am trying my best—and if I did not have all these things to read I would have completed one on John Adams that I am about three-quarters of the way through.

That is where this whole approach to the census started. It started, really, because of all the difficulties of trying to figure out how to balance power. They succeeded, I think, quite well, but it does mean that it left the Census Bureau with a difficult job.

It is also interesting to me that in spite of their belief that a census was a fairly easy thing to do, you just count people, nevertheless, when the census was done, George Washington and Thomas Jefferson were among the first to say that it was not enough people.

It is interesting that instead of the argument being among the various states or even within the various states, the argument was between the United States and the other countries of the world in trying to show that we really were a growing country with growing power and, in order to have growing power, we had to have more people.

The 4 million or so who were counted they felt was clearly an undercount, because, as they said, some people just do not want to answer and we cannot find them, and there are all kinds of reasons. It is a very interesting thing to go back to.

The point is that they succeeded, I think, in moving the country ahead, and my hope is that whatever you decide to do, and, I might say, whatever our panel decides to say about what you do, which is also very much up in the air, I think you are trying very hard to move things ahead in a very professional manner.

I also want to say that the census is important not just for all its uses. The way in which the census is done is critical to the federal statistical system, and I care a lot about the federal statistical system. If the census is considered to be politically dominated, one way or the other, it seems to me that the entire statistical system will be very seriously affected, because I think that people in this country will lose confidence in everything that the government puts out.

This is really critical, and it is not critical in terms of whether you adjust or you do not adjust or whether we think if you do adjust that you did the right thing or not—I do not know what the answer is to all those things, clearly we have to look at an awful lot of data to know that. What is important, I think, as you started out with this morning, Ken, is to understand that this has to be a completely professional, unbiased approach.

We will try to do whatever we can as a panel to give you our best scientific evaluation of all the processes, and I just hope that we get through all of this, as well as the long form, in the not-too-distant future.

I want to thank all of the people here, especially our invited guests who came and participated, all of the others who are here. I think this has been a very useful meeting.

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, Second Workshop of the Panel to Review the 2000 Census* (February 2-3, 2000). Committee on National Statistics. Washington, D.C.: National Academy Press.

U.S. Census Bureau Materials

NOTE: The documents provided by the Census Bureau for the October 2, 2000, workshop were drafts, containing background text, proposed topics for inclusion, and table shells, indicating the kinds of analysis the Bureau proposed to conduct to inform ESCAP. The documents contained no results from the census or the A.C.E. Each of the sixteen documents provided corresponds to a document that the Bureau released March 1, 2001, containing full text and tabular and other analytic results. References to these documents below are to the final, published 2001 version of each. In addition to the completed versions of the 16 documents for which the Bureau provided drafts at the workshop, the Census Bureau released another 3 documents in the "B-series" of memoranda (Griffin, 2001b; Mulry and Spencer, 2001; Navarro and Olson, 2001). All of the documents released on March 1, 2001, are available at <http://www.census.gov/dmd/www/EscaRep.html>.

Batutis, Michael J.

- 2001 Demographic Full Count Review: 100% Data Files and Products. DSSD Census 2000 Procedures and Operations Memorandum Series B-16. U.S. Census Bureau, Washington, D.C. [author of draft provided to workshop participants was listed as Gretchen Stiers]

Baumgardner, Stephanie K., Darlene A. Moul, Robin A. Pennington, Rebecca I. Piegari, Herbert F. Stackhouse, Kevin J. Zajac, Nick S. Alberti, Jennifer W. Reichert, and James B. Treat

- 2001 Quality of Census 2000 Processes. DSSD Census 2000 Procedures and Operations Memorandum Series B-3. U.S. Census Bureau, Washington, D.C. [authors of draft

provided to workshop participants were listed as Frank Vitrano, Tracey McNally, Rebecca Piegari, and Carrie Johanson]

Bell, William R.

2001 Accuracy and Coverage Evaluation: Correlation Bias. DSSD Census 2000 Procedures and Operations Memorandum Series B-12. U.S. Census Bureau, Washington, D.C.

Byrne, Rosemary, Lynn Imel, Magdalena Ramos, and Phawn Stallone

2001 Accuracy and Coverage Evaluation: Person Interviewing Results. DSSD Census 2000 Procedures and Operations Memorandum Series B-5. U.S. Census Bureau, Washington, D.C. [authors of draft provided to workshop participants were listed as Patricia Feindt and Rosemary Byrne]

Cantwell, Patrick J., David McGrath, Nghanha Nguyen, and Mary Frances Zelenak

2001 Accuracy and Coverage Evaluation: Missing Data Results. DSSD Census 2000 Procedures and Operations Memorandum Series B-7. U.S. Census Bureau, Washington, D.C.

Childers, Danny R., Rosemary L. Byrne, Tamara S. Adams, and Roxanne Feldpausch

2001 Accuracy and Coverage Evaluation: Person Matching and Follow-up Results. DSSD Census 2000 Procedures and Operations Memorandum Series B-6. U.S. Census Bureau, Washington, D.C.

Davis, Peter P.

2001 Accuracy and Coverage Evaluation: Dual System Estimation Results. DSSD Census 2000 Procedures and Operations Memorandum Series B-9. U.S. Census Bureau, Washington, D.C.

Farber, James

2001a Quality Indicators of Census 2000 and the Accuracy and Coverage Evaluation. DSSD Census 2000 Procedures and Operations Memorandum Series B-2. U.S. Census Bureau, Washington, D.C.

2001b Accuracy and Coverage Evaluation: Consistency of Post-Stratification Variables. DSSD Census 2000 Procedures and Operations Memorandum Series B-10. U.S. Census Bureau, Washington, D.C.

Griffin, Richard A.

2001a Accuracy and Coverage Evaluation: Service Based Enumeration Multiplicity Estimation. DSSD Census 2000 Procedures and Operations Memorandum Series B-15. U.S. Census Bureau, Washington, D.C.

2001b Census 2000: Missing Housing Unit Status and Population Data. DSSD Census 2000 Procedures and Operations Memorandum Series B-17. U.S. Census Bureau, Washington, D.C.

90 ■ REFERENCES

Griffin, Richard A., and Donald J. Malec

- 2001 Accuracy and Coverage Evaluation: Assessment of Synthetic Assumptions. DSSD Census 2000 Procedures and Operations Memorandum Series B-14. U.S. Census Bureau, Washington, D.C.

Hogan, Howard

- 2001 Accuracy and Coverage Evaluation: Data and Analysis to Inform the ESCAP Report. DSSD Census 2000 Procedures and Operations Memorandum Series B-1. U.S. Census Bureau, Washington, D.C.

Mule, Thomas

- 2001 Accuracy and Coverage Evaluation: Decomposition of Dual System Estimate Components. DSSD Census 2000 Procedures and Operations Memorandum Series B-8. U.S. Census Bureau, Washington, D.C.

Mulry, Mary H., and Bruce D. Spencer

- 2001 Overview of Total Error Modeling and Loss Function Analysis. Prepared by Abt Associates. DSSD Census 2000 Procedures and Operations Memorandum Series B-19. U.S. Census Bureau, Washington, D.C.

Navarro, Alfredo, and Mark Asiala

- 2001 Accuracy and Coverage Evaluation: Comparing Accuracy. DSSD Census 2000 Procedures and Operations Memorandum Series B-13. U.S. Census Bureau, Washington, D.C.

Navarro, Alfredo, and Douglas Olson

- 2001 Accuracy and Coverage Evaluation: Effect of Targeted Extended Search. DSSD Census 2000 Procedures and Operations Memorandum Series B-18. U.S. Census Bureau, Washington, D.C.

Prewitt, Kenneth

- 2000 *Accuracy and Coverage Evaluation: Statement on the Feasibility of Using Statistical Methods to Improve the Accuracy of Census 2000*. U.S. Census Bureau. Washington, D.C.: U.S. Department of Commerce. (June).

Robinson, J. Gregory

- 2001 Accuracy and Coverage Evaluation: Demographic Analysis Results. DSSD Census 2000 Procedures and Operations Memorandum Series B-4. U.S. Census Bureau, Washington, D.C.

Starsinic, Michael D., Charles D. Sissel, and Mark E. Asiala

- 2001 Accuracy and Coverage Evaluation: Variance Estimates by Size of Geographic Area. DSSD Census 2000 Procedures and Operations Memorandum Series B-11. U.S. Census Bureau, Washington, D.C.

LIST OF INVITED PARTICIPANTS

Barbara Bailar, National Opinion Research Center, Chicago, IL
Lynne Billard, Department of Statistics, University of Georgia
Robert E. Fay, Senior Mathematical Statistician, Director's Office, U.S. Census Bureau
Howard Hogan, Chief, Decennial Statistical Studies Division, U.S. Census Bureau
Ruth Ann Killion, Chief, Planning, Research, and Evaluation Division, U.S. Census Bureau
John Long, Chief, Population Division, U.S. Census Bureau
Jeffrey Passel, Urban Institute, Washington, DC
Kenneth Prewitt, Director, U.S. Census Bureau
J. Gregory Robinson, Chief, Population Analysis & Evaluation Staff, Population Division, U.S. Census Bureau
Allen Schirm, Mathematica Policy Research, Inc., Washington, DC
Michael Stoto, School of Public Health, George Washington University
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
Donald Ylvisaker, Department of Statistics, University of California at Los Angeles
Alan Zaslavsky, Department of Health Care Policy, Harvard Medical School