Kathleen P Zveare
12/02/1999 10:00 AM

To: Kenneth Prewitt, William G Barron Jr, Nancy A Potok, Paula Jane Schneider, Cynthia Z F Clark, Nancy M Gordon, John H Thompson, Preston J Waite, Robert E Fay III, Howard R Hogan, Ruth Ann Killion, John F Long, Susan Miskura

cc: Maria E Urrutia, Fay F Nash, Phyllis A Bonnette, Patricia E Curran, Ellen Lee, Betty Ann Saucier, Jeannette D Greene, Margaret A Applekamp, Jane F Green, Sue A Kent, Mary A Cochran, Linda A Hiner, Carnelle E Sligh, Lois M Kline, Angela Frazier, Linda K Bonney

Subject: Meetings for ESCAP

You should have received a memo (attached) letting you know that we would be contacting you about the Executive Steering Committee for A.C.E. Policy meetings.

The meetings will take place the 2nd and 4th Wednesdays starting December 8 from 10-11:30 in Rm. 2412/3.

Attendees:

K. Prewitt
B. Barron
N. Potok
P. Schneider
C. Clark
N. Gordon
J. Thompson
J. Waite
B. Fay
C. Bush
H. Hogan
R. Killion
J. Long
S. Miskura

- acegburn.wpd - acepchar.wpd

Kathleen P Zveare
12/02/1999 10:00 AM

To: Kenneth Prewitt, William G Barron Jr, Nancy A Potok, Paula Jane Schneider, Cynthia Z F Clark, Nancy M Gordon, John H Thompson, Preston J Waite, Robert E Fay III, Howard R Hogan, Ruth Ann Killion, John F Long, Susan Miskura
cc: Maria E Urrutia, Fay F Nash, Phyllis A Bonnette, Patricia E Curran, Ellen Lee, Betty Ann Saucier, Jeannette D Greene, Margaret A Applekamp, Jane F Green, Sue A Kent, Mary A Cochran, Linda A Hiner, Carnelle E Sligh, Lois M Kline, Angela Frazier, Linda K Bonney
Subject: Meetings for ESCAP

You should have received a memo (attached) letting you know that we would be contacting you about the Executive Steering Committee for A.C.E. Policy meetings.

The meetings will take place the 2nd and 4th Wednesdays starting December 8 from 10-11:30 in Rm. 2412/3.

Attendees:

K. Prewitt
B. Barron
N. Potok
P. Schneider
C. Clark
N. Gordon
J. Thompson
J. Waite
B. Fay
C. Bush
H. Hogan
R. Killion
J. Long
S. Miskura

- acegburn.wpd - acepchar.wpd

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) # 1

December 8, 1999

Prepared by: Maria Urrutia and Genny Burns

The first meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) was held on December 8, 1999 at 10:00 a.m.

Persons in attendance:

Kenneth Prewitt
William Barron
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura
Raj Singh
Maria Urrutia
Genny Burns

I. Purpose of the group

The purpose of the group was briefly discussed. The charter is attached.

1. Provide advice to the Director of the Census Bureau on issues related to A.C.E. and adjustment.
2. Provide the Director and staff with background and related information on A.C.E.
3. Document rationale and reasons for decisions
4. Provide policy guidance on adjustment issues
5. Issue recommendation in February 2001 on whether adjusted counts should be released for redistricting.

II. Proposed agendas for first four meetings

The first three meetings will provide participants with a basic understanding of coverage measurement. After that, perhaps starting in the third meeting, the 1990 process will be reviewed. Topics included in this review will be the issues associated with the adjustment of the 1990 Census and its effect on the adjustment of the postcensal estimates.

Subsequently, relevant issues will be presented to the group as well as points of progress. John Thompson will work with Jay, Howard, and Susan to identify issues and will bring these to the committee.

III. Presentation - A.C.E. Design and Dual System Estimator (DSE)

Howard Hogan gave an overview of and responded to questions on the A.C.E. design and DSE. Handouts describing post enumeration surveys and the DSE in detail with formulae and text were distributed and are on file with these minutes.

The DSE is used for correcting the coverage error in the census. Aspects of the estimator are: DSE model in theory, application to census in general, and application to A.C.E. The basic model is as follows:

Basic DSE Model

List A (Census)	List B (A.C.E./PES)		Total
	In	Out	
In	N_{11}	N_{12}	$N_{1.}$
Out	N_{21}	N_{22}	$N_{2.}$
Total	$N_{.1}$	$N_{.2}$	$N_{..}$

Thus,

$$\frac{N_{11}}{N_{.1}} = \frac{N_{1.}}{N_{..}}$$

and

$$N_{..} = \frac{N_{.1}\, N_{1.}}{N_{11}}$$

where

$N_{1.}$ is the number of unique people correctly and completely enumerated in the census, $N_{.1}$ in the A.C.E., and N_{11} in both the census and A.C.E. An estimate of $N_{1.}$ is obtained from the census. Components of the DSE may include sampling and/or nonsampling errors.

The A.C.E. actually consists of two samples. The first is a sample of the population, known as the P sample, which measures omissions in the Census. The second is a sample of Census enumerations, known as the E sample. The E sample measures erroneous enumerations in the Census.

In 1980, the P and E samples did not overlap. In 1990, an overlapping sample design, based on the same blocks for both samples, was implemented.

To estimate the net undercount, it is critical to measure (i) the rate of erroneous enumerations in the initial phase of the census, and (ii) the rate of P sample matches to census enumerations in the A.C.E. block clusters. Followup operations will be used to determine erroneous enumerations by identifying duplicates, geocoding errors, fictitious persons, and illegible names. These operations will also be used to determine if a nonmatched person was correctly enumerated. All E sample matched cases will be assumed to be correct. E sample nonmatches will be followed-up to determine whether they were correctly enumerated. P sample nonmatches will be followed-up selectively.

For A.C.E., a sample of block clusters is selected averaging about 30 housing units each. Some blocks will have fewer than 30 housing units while others may be larger and require subsampling within the block. In September, 1999, maps were given to interviewers which contained only the physical boundaries but no housing units. Interviewers were required to map spot the location of each housing unit on these maps and also to complete independent listing books (ILBs) with housing unit information in each cluster. These ILBs were keyed and resulted in the sampling frame for A.C.E. interviewing. In July, 2000, the A.C.E. interviewers will visit each housing unit selected in A.C.E. sample to find out who lived there on April 1, 2000. This will give us the people in A.C.E. blocks which will then be linked with the person records from the census unedited file (CUF) that are in the E sample. This will not result in a one to one match since there will be some unmatched records in the P sample and some from the E sample. These unmatched census records will be sent to Field Division (FLD) for verification. This verification could be the third visit to a household since it could have previously received a census nonresponse followup and an A.C.E. interviewing visit.

For some people in either sample, the information collected will be insufficient, resulting in unresolved cases. For these cases, the probability of a match or correct enumeration

will be assigned through estimation based on the corresponding rates from similar people with resolved status. A similar procedure will be used to handle mover cases. The hot-deck methodology to estimate missing characteristics, such as race, sex, and age, will be used. For whole household nonresponse, a weighting approach will be implemented.

In 1990, the population was divided into 357 poststrata or estimation cells to classify persons into groups that were as much alike as possible with respect to coverage error. Each person can belong to only one poststrata since they are mutually exclusive partitioned. For cases where the poststrata gets too small for estimation or publication of results, collapsing is implemented. Poststratification variables are being determined for the 2000 Census.

Synthetic estimation will be used in conjunction with the poststratification. Coverage factors will be computed using the following formula:

$$CF_j = \frac{N_{Tj}}{N_j^c}$$

= estimate of total population/complete census count (including erroneous enumerations and imputations)

where

j = poststrata numbers 1...n (For 1990, n=357)

These results for the j^{th} poststratum are applied to the census figures in the j^{th} stratum to form a synthetic estimate down to the block level. After adjustment, the numbers will not be integers. A controlled rounding will be used to obtain integer numbers such that each rounded number is within $\pm$ 1 of unrounded numbers.

IV. Future Discussions

The following topics were identified for further discussion in future meetings.

1. How the guidelines were developed for sending A.C.E. cases to followup.

2. The term, erroneous enumerations, includes some cases with insufficient information for matching and are not necessarily in error. Thus, this term may need to be changed for clarification in meaning.

V. Discussion Points for Next Meeting

1. How to operationalize the A.C.E. and DSE
2. How to measure each component of the DSE
3. More details on A.C.E. and DSE
4. What variables are used to impute unresolved cases.

ESCAP Committee

Kenneth Prewitt
William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson, Chair
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura

cc:

Teresa Angueira
Ed Gore
Ed Pike
Catherine Miller
Fay Nash
Miguel Perez
Maria Urrutia
Genny Burns
Carolee Bush
Donna Kostanich
Raj Singh
David Whitford

Kathleen P Zveare
12/16/1999 02:16 PM

To: Kenneth Prewitt, William G Barron Jr, Nancy A Potok, Paula Jane Schneider, Cynthia Z F Clark, Nancy M Gordon, John H Thompson, Preston J Waite, Robert E Fay III, Howard R Hogan, Ruth Ann Killion, John F Long, Susan Miskura, Kathleen P Zveare
cc: Maria E Urrutia, Fay F Nash, Phyllis A Bonnette, Patricia E Curran, Ellen Lee, Betty Ann Saucier, Jeannette D Greene, Margaret A Applekamp, Jane F Green, Sue A Kent, Mary A Cochran, Linda A Hiner, Carnelle E Sligh, Lois M Kline, Angela Frazier, Linda K Bonney, Carolee Bush, Rosalyn R Harrington, Geneva A Burns
Subject: Re[2]: Meetings for ESCAP

This message is to confirm the rescheduled ESCAP meeting.

Date: December 20, 1999

Time: 4-5 p.m.

Room: 2412/3

Agenda: Sample Design and Dual System Estimation

Attendees:

K. Prewitt
B. Barron
N. Potok
P. Schneider
C. Clark
N. Gordon
J. Thompson
J. Waite
B. Fay
C. Bush
H. Hogan
R. Killion
J. Long

Please cancel the 12/22 meeting. Thanks.

_____ Reply Separator
_____
Subject: Re: Meetings for ESCAP
Author: Kathleen P Zveare at DMD
Date: 12/15/1999 11:04 AM

We are going to reschedule the next ESCAP meeting scheduled for Thursday December 22. Please let me know your availability for Monday December 20 from 4-5. Thanks.

_____ Reply Separator
_____
Subject: Meetings for ESCAP
Author: Kathleen P Zveare at DMD
Date: 12/2/1999 10:00 AM

You should have received a memo (attached) letting you know that we would be contacting you about the Executive Steering Committee for A.C.E. Policy meetings.

The meetings will take place the 2nd and 4th Wednesdays starting December 8 from 10-11:30 in Rm. 2412/3.

Attendees:

K. Prewitt
B. Barron
N. Potok
P. Schneider
C. Clark
N. Gordon
J. Thompson
J. Waite
B. Fay
C. Bush
H. Hogan
R. Killion
J. Long
S. Miskura

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) # 2

December 20, 1999

Prepared by: Genny Burns and Kathy Stoner

The second meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on December 20, 1999 at 4:00 p.m.

Persons in attendance:

Kenneth Prewitt
William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
John Long
Ruth Ann Killion
Donna Kostanich
Raj Singh
Carolee Bush
Genny Burns
Kathy Stoner

I. Presentation - A.C.E. Poststratification

Howard Hogan recommended that the agenda topic for this meeting be 2000 Census poststratification since this is an issue of immediate concern. He briefly discussed the history of poststratification and then described the proposed plan for 2000 along with the objectives for choosing poststratification variables. Input and guidance on policy issues in developing poststratification is needed from the committee. Handouts were distributed describing the objectives, recommendation, and background information on poststratification. These handouts are on file with the minutes.

The 1990 Post Enumeration Survey (PES) had a 357 poststrata design which started with a cross-classification of the age/sex, race/Hispanic origin, tenure, urbanicity, and region variables. The 2000 A.C.E. will have a larger sample size than 1990 which should allow

for more poststrata and should result in smaller variances.

Poststratification serves dual purposes of grouping people to form estimation cells that lead to reduced correlation bias in the dual system estimation (DSE) and of grouping people with similar net undercount or coverage probabilities for synthetic estimation purposes, down to the block level. Thus, the poststrata should be operationally feasible for both DSE estimation and for synthetic estimation. Poststrata should differentiate geographic areas and are required to have a minimum population size to control variance and reduce ratio bias. Also, there needs to be a minimum sample size. If these minimum requirements are not met, groups will be collapsed according to expected sample size within a poststratum. Groups that cannot be assigned to a category based on data or logic will be assigned to the largest of the logical groups since the larger groups will usually be disaggregated on other characteristics so the bias and variance should be smaller.

It is important for poststrata to be defined on variables that are reported consistently in the Census and the A.C.E. The poststrata variables should be well documented and thoroughly discussed in advance of receiving the data. The recommended poststrata variables, based on research from 1990, are race/Hispanic origin, age/sex, tenure, urbanicity/type of enumeration area (TEA), and mail response rate. Since the urbanicity/TEA variable is the only explicit geographic variable included, there is concern about creating the potential for bias due to geographic variation in undercount. However, the decision to exclude other geographic variables is based on research from 1990 results where region was included but appeared to add about as much variance as it reduced bias.

In 1990, only one race could be selected by the respondents but in 2000, for the first time in census history, multiple responses to the race question will be permitted. The Census 2000 questionnaire has 15 possible race responses. For estimation purposes, the 15 responses will be collapsed into 6 major race groups for which persons with a single race essentially place themselves. Allowing persons to self-identify with multiple races complicates the details for assigning persons to a race/Hispanic origin group. Thus, a hierarchy is proposed to assign persons to one of 7 race/Hispanic origin poststrata. Although data from 1990 and Dress Rehearsal (DR) have been researched, many of the decisions on how to classify persons into one of the 7 poststrata must be based on previously observed demographic factors and professional judgment.

The DR revealed inconsistencies in reporting more than one race in A.C.E. and Census which led to the need for broad racial poststratification categories. Also, the A.C.E. sample size will only support a limited number of race/Hispanic categories.

A decision memorandum will be prepared announcing the poststratification design in advance of the 2000 Census implementation. The recommended design was discussed with selected members of the National Academy of Sciences (NAS) Expert Review Panel on Census 2000. It will be formally presented to the entire Panel at a future meeting.

II. Issues Regarding Poststratification and Multiple Race Groupings

John Thompson volunteered to summarize the issues and concerns brought up at the meeting and distribute these for comments and further discussions. These issues are described as follow:

- The mail response rate variable is different from the other variables recommended for poststratification since it is based on a Census operation attribute rather than a respondent attribute. This will be the first time a poststratification variable has the quality of being operationally dependent rather than respondent dependent. Since the Census Bureau has some control over this variable, it will be important to document that no purposeful influences are introduced into the coverage estimates.

- It is important that consistency underlie the definition of racial poststrata. For example, the rationale for collapsing options for Asians should be consistent with other groups.

- A proposal was made to treat Hawaiians in a manner similar to that of the American Indians. People reporting Hawaiian and one or more other races and who are Hawaiians living in Hawaii should be classified as Hawaiian while those not living in Hawaii should be treated as Pacific Islanders.

- Perhaps Dress Rehearsal data should be run to test mode effects on reporting multiple race for A.C.E. and census data collection. Look at consistency of responses across race groups to determine how much they vary.

- Collapsing guidelines are an important component of the DSE methodology and it is essential that these are well documented.

- There was concern expressed regarding the lack of geographic poststratification variables. Subsequent discussion resulted in a proposal to consider some regional poststratification variables in the next round of discussions.

III. Next Meeting

The next meeting will be held on Wednesday, January 5, 2000. Agenda topics will be treatment of movers and other differences between 1990 and 2000.

ESCAP Committee

Kenneth Prewitt
William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson, Chair
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura

cc:

Teresa Angueira
Ed Gore
Ed Pike
Catherine Miller
Fay Nash
Miguel Perez
Maria Urrutia
Genny Burns
Carolee Bush
Donna Kostanich
Raj Singh
David Whitford

Kathleen P Zveare
01/04/2000 03:01 PM

To: Margaret A Applekamp, William G Barron Jr, Phyllis A Bonnette, Geneva A Burns, Carolee Bush, Elizabeth Centrella, Cynthia Z F Clark, Mary A Cochran, Patricia E Curran, Robert E Fay III, Angela Frazier, Nancy M Gordon, Jeannette D Greene, Linda A Hiner, Howard R Hogan, Sue A Kent, Ruth Ann Killion, Lois M Kline, John F Long, Susan Miskura, Nancy A Potok, Kenneth Prewitt, Betty Ann Saucier, Paula J Schneider, Rajendra P Singh, Carnelle E Sligh, John H Thompson, Maria E Urrutia, Preston J Waite
cc:
Subject: Agenda for tomorrow's ESCAP meeting

The agenda for tomorrow's ESCAP meeting scheduled from 11-12:30 in Rm. G-316/3 is as follows:

1. Treatment of Movers

2. Other Differences Between 1990 and 2000

January 5, 2000

A.C.E. - PES
2000 - 1990 CHANGES

(Talking Points)

SAMPLING

* **SAMPLE SIZE**–approximately 300,000 housing units vs. 170,000 in the 1990 PES.

* **WITHIN STRATA SAMPLE**--Designed to have at least the same or better c.v. than in 1990.

* **SAMPLING PROBABILITY**--Units in sample will have much closer to equal sampling probabilities than in 90. Minority groups will be only slightly differential and we're increasing the sampling rate in potential problem clusters.

* **2 STAGE SAMPLE FOR SMALL BLOCKS**--Our small block sample in 1990 was a "clunky" operation. For 2000 it has been greatly refined.

* **COVERAGE (GROUP QUARTERS)-** We are not including group quarters (college dorms, institutions, military reservations, etc.) in the A.C.E. universe. We did not include some of them in the 1990 PES. We feel we cannot do an adequate job measuring their coverage. We will use a "rigorous" enumeration methodology in the initial enumeration.

LISTING/HU MATCHING

* **INITIAL HOUSING UNIT MATCHING AND SUBSAMPLING METHOD**--In 2000 we will match our A.C.E. address listing to the January, 2000 version of the Decennial Master Address File. This will make subsequent subsampling of large blocks much easier and less time consuming.

PERSON INTERVIEW

* **AUTOMATED A.C.E. INSTRUMENTS--**A large reason we are able to deliver the coverage measurement products earlier than we did in 1990 is that we are doing the A.C.E. interview using laptop computers: Keying is no longer needed and quality control checks are quicker.

* **A.C.E./NRFU OVERLAP**-We are doing A.C.E. "telephone CAPI" interviews while nonresponse followup is underway. In these we call mail respondents to the census who have given us their phone number and conduct an A.C.E. interview with them.

* **TIMING and STAFFING**--In 2000 we are allowing 6 weeks for FLD to complete the A.C.E. Person Interview and 2 week for the Nonresponse Conversion operation. In 1990 we

interviewed for 6 weeks and nonresponse followup was an unplanned operation which did not immediately follow the interviewing.

Since the A.C.E. follows immediately after NRFU, we will use our best initial count nonresponse followup interviewers in A.C.E. interviewing.

The A.C.E. Person Followup in 2000 will be done over a months time as it was in 1990.

* **PES-C**--Movers were a problem in the 1990 PES when we tried to find the census questionnaire for in-movers. In 2000 we are planning a PES-C approach which uses in-mover counts and demographics but outmover match rates.

MATCHING

* **COMPUTER MATCHING**--Improvements have been made in address standardization and parameter estimation. In the latter, we've incorporated theoretical advances in record linkage models presented in recent statistical literature.

* **SEARCH AREA**--We are planning to have the cluster be the search area except in exceptional areas where we will do an additional targeted surrounding block search. In the 1990 PES the search area included the surrounding blocks.

* **CENTRALIZATION OF MATCHING AND PROCESSING OPERATIONS**–We are doing clerical matching at one site. It is essentially a paperless operation–maps, housing unit, and person information is accessed by computer. In 1990 we needed to access huge numbers of paper maps, address listing books, microfilm of census forms, and actual census and PES forms. This required a much larger staff which we dispersed throughout our 7 processing offices.

PERSON FOLLOWUP

* **A.C.E. FOLLOWUP CASES–** We have cut the percentage of A.C.E. cases that need to be followed up in the field. In our tests we gained quantitative evidence that we can trust many of our initial interviews--especially those with census household members.

ESTIMATION

* **POST STRATIFICATION METHOD**--We will use the same variables as in the 1990 PES and add a mail return variable and account for update/leave areas.

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 3

January 5, 2000

Prepared by: Maria Urrutia and Genny Burns

The third meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on January 5, 2000 at 11:00.

Persons in attendance:

Kenneth Prewitt
William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Raj Singh
Gregg Robinson
Signe Wetrogen
Carolee Bush
Maria Urrutia
Genny Burns

I. Part 1 of the Presentation - Movers

Howard Hogan discussed the plans for handling movers in the census and A.C.E. The census will count people who resided in housing units on April 1. The three to four month time span between the April 1 Census date and the June A.C.E. interviewing allows for moving activity. The people who are in the same unit on Census Day and at the time of the A.C.E. interview are more straightforward to match but the ones who are somewhere else are more difficult to process. Particularly difficult are those cases that include deaths or births, persons who emigrate or immigrate, and individuals who moved to or from group quarters (GQs). (A.C.E. includes only housing units, not GQs.)

The three procedures described below for capturing movers were discussed.

Procedure A is based on defining the residents of the A.C.E. sample housing units as of Census Day. Therefore, respondents are asked to identify all persons who were living or staying in the sample housing unit on Census Day. These persons are then matched

against names on the census questionnaire for the sample address. From this information, estimates of the number and percent matched for non-movers and out-movers are made. The advantage is that matching is simpler since it is performed at the sample addresses. The disadvantage is that movers may be understated. Estimates for movers will not be based on the complete mover universe and will be biased to some degree.

Procedure B is based on where the residents of the A.C.E. sample housing units actually lived on Census Day. Thus, current residents of the A.C.E. sample housing units are asked where they lived on Census Day. Those residents determined to be movers are matched at their Census Day addresses. This often involves determining (geocoding) the location of Census Day addresses in the Census records. The advantages are (1) since the actual movers are being interviewed, a lower nonresponse rate and potentially more accurate data are obtained; and (2) a more complete mover universe is incorporated into the A.C.E. process. The disadvantages are more complex matching, Census Day address recall biases for inmovers, and geocoding problems.

Procedure C is a two-step process as follows: (1) Determining the number of movers from the current residents of A.C.E. housing units, and (2) Estimating the match rate for movers based on the Procedure A interview. Therefore, the match rate is estimated by determining or reconstructing Census Day residents of the A.C.E. household and matching them to Census records. The advantage is that it produces good estimates of the number of movers. The disadvantage is the match rate may not be representative of the entire mover universe.

The person match rate for Procedure C is calculated as:

$$\frac{M_{NV} + (\frac{M_{OV}}{N_{OV}})N_{IV}}{N_{NV} + N_{IV}}$$

where M_{NV} = the weighted number of matched nonmovers in the census

M_{OV} = the weighted number of matched outmovers in the census

N_{IV} = the weighted number of inmovers in A.C.E.

N_{OV} = the weighted number of outmovers in A.C.E.

N_{NV} = the weighted number of nonmovers in A.C.E.

Although Procedure B was used in 1990, it will not be applied in 2000 partly based on research conducted in Dress Rehearsal (DR) and because Procedure C fits better with the timing of the operational flow. There is no ideal method for handling movers. This is a very difficult part of DSE methodology. However, Procedure C is judged to be the best blend of operational feasibility and accuracy. As with all our procedures, this will be included in the A.C.E. evaluations.

II. Part 2 of the Presentation - Differences Between the 1990 PES and 2000 A.C.E.

Howard distributed the attached handout describing the 1990-2000 changes. The effects of these changes on design were briefly discussed and it was noted that these changes have resulted in modest improvements to the design. It was also noted that these changes have been discussed with the Bureau's statistical advisors.

The major points raised during the discussion were as follow:

(1) The sample size of approximately 300,000 housing units is larger than that in 1990. Since small blocks will undergo a two-stage sampling process, their weights should be better controlled than in 1990 and there should be fewer outlier clusters. The goals of this two-stage sampling process are to attempt to reduce the contribution of small clusters to the variance of the DSE and to ensure that the workload can be efficiently managed.

(2) Housing units in A.C.E. initial sample blocks have been independently listed and will be linked to the Decennial Master Address File (DMAF). A disadvantage of listing earlier than in 1990 is there are many changes in addresses between the listing phase and Census Day. An advantage is that it alerts us to problems sooner so geocoding problems, e.g., early evidence of A.C.E. geocoding error, can be addressed and potentially corrected. The initial housing unit match will be conducted in February but information from this process will not be released in order that the independence to the A.C.E. not be compromised.

(3) Unlike 1990, where matching processes were conducted in several processing offices, the matching operations will be centralized in one processing office in 2000. This should result in better control and more consistency over the matching, especially in handling difficult cases.

(4) The rules regarding which persons go to A.C.E. followup (FU) have been modified. Detailed discussions of these rules were deferred to a future meeting where a complete presentation of the rationale for sending people to FU will be conducted.

(5) In 1990, permanent staff conducted the Nonresponse Conversion (NRCO) but in 2000 this process will be conducted by temporary interviewers.

III. Other

Howard mentioned that there is a change from the discussion two weeks ago in the grouping of American Indians. There were three stratum groups: (1) American Indians on reservations, (2) American Indians on tribal/trust lands, or (3) all other. It was decided to combine American Indians on tribal/trust lands with the all other grouping, i.e., groups 2 and 3 will be combined. Thus, the two strata will now be American Indians on reservations and those off reservations.

IV. Next Meeting

The next meeting will be held on Wednesday, January 12, 2000. Agenda topics will be 1990 evaluations of the PES and the associated decision processes. This discussion will include major issues and concerns with PES methodology and the steps that have been taken to address them.

A copy of the Federal Register Notice which documents the decision for not adjusting the 1990 Census will be distributed. Also, interest was expressed in viewing the training video on matching at a future meeting.

ESCAP Committee

Kenneth Prewitt
William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson, Chair
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura

cc:

Teresa Angueira
Ed Gore
Ed Pike
Catherine Miller
Fay Nash
Miguel Perez
Maria Urrutia
Genny Burns
Carolee Bush
Donna Kostanich
Raj Singh
David Whitford
Gregg Robinson
Signe Wetrogen
Magdalena Ramos

There was no agenda developed or used for the January 12, 2000 meeting.

January 12, 2000

Chronology of Events for the 1990 Census Adjustment Decision

1980

- 1980 Census taken
- Formation of Undercount Steering Committee
- Decision was made not to adjust the Census. Undercount Research Staff formed to conduct coverage measurement research

1985

- The Bicentennial Census, a National Academy of Sciences Committee on National Statistics report, recommends method for adjustment of the count and improving accuracy in 1990
- Census Bureau tests computer matching technique in Florida test

1986

- Census Bureau conducts Test of Adjustment Related Operations (TARO) in Los Angeles to determine feasibility of adjustment of 1990 census. Bureau concludes that it is technically feasible to adjust the Census, but there are operational concerns.

1987

Spring
- Census Bureau announces it has developed a feasible method for undercount adjustment using a 300,000 household Post Enumeration Survey (PES)

October
- Commerce Department cancels the plans for the PES for adjustment

1988

Fall
- New York City sues the Commerce Dept for reinstatement of the PES and adjustment

1989

July
- Government signs stipulation agreement with New York plaintiffs reinstating PES and adjustment methodology; decision on which results to use set for July 1991. The settlement had 3 outcomes:

    1. A PES would be conducted and evaluated to see if it could be used to correct the Census count by 7/15/91. The Secretary of Commerce would make decision.

    2. The Secretary of Commerce would publish guidelines he would follow to make decision.

    3. An eight member Special Advisory Panel was appointed to advise the Secretary. Four members on each side of the argument.

July	r	Reestablish Census Undercount Steering Committee with Paula Schneider as chair.
1990	r	1990 Census taken
March	r	Guidelines are issued by the Secretary, challenged and upheld
Winter	r	Unadjusted 1990 Census results released
1991		
Jan. 1-April 4	*	Redistricting
June	r	Completed extensive evaluation of the PES. Twenty-one evaluations on the sampling and non-sampling error and eleven for demographic analysis.
June 21	r	Census Undercount Steering Committee issued recommendation that Census be adjusted.
June 28	r	Census Director Barbara Bryant recommended to Secretary Mosbacher that the 1990 Census be adjusted. Special Advisory Panel rendered a split decision (4 members for adjustment and 4 against adjustment).
	r	Under Secretary Darby recommended not to adjust the 1990 Census.
July 22	r	Secretary Mosbacher decided not to adjust the 1990 Census. He directed Bureau to review results of PES and see if PES could be used to adjust the post-censal estimates.
Summer	r	NYC requests a trial to determine if Secretary's decision was "arbitrary and capricious"
August	*	CAPE committee formed to direct research on potential adjustment of post-censal estimates
Fall	*	CAPE issues report indicating that states are improved while results are not conclusive for small areas
	r	Decennial Census Improvement Act creates National Academy of Sciences panel to study improved methods for 2000 Census
October	r	Computer error discovered in 1991 estimates
1992	r	Two CNSTAT panels were formed to study improvements for 2000 Census
Spring	r	1990 Census lawsuit goes to 13 day trial
Dec. 29	r	Census Director Barbara Bryant decides not to adjust post-censal estimates on the basis of the PES. Adjustment would make distribution for states better but cannot make any conclusions for entities with less than 100,000 population. So Director decided not to adjust since could not improve coverage in all areas. She did announce that the Federal Statistical System could use adjusted numbers for survey controls.
1993		
Spring	r	Judge McLaughlin rules, holding that the decision of the Commerce Secretary was not "arbitrary and capricious"

Sept. 29 r BLS decided to use adjusted population counts for CPS controls. Rest of Federal Statistical System followed.

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 4

January 12, 2000

Prepared by: Maria Urrutia and Genny Burns

The fourth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on January 12, 2000 at 10:30. The purpose of this meeting was to discuss the 1990 Census experiences regarding use of statistical methods to adjust the census. Two aspects were discussed: (1) a chronology of events, and (2) the decision process.

Persons in attendance:

William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Raj Singh
Gregg Robinson
Signe Wetrogen
Carolee Bush
Sally Obenski
Maria Urrutia
Genny Burns

I. Chronology of Events for the 1990 Census Adjustment Decision

John Thompson presented the attached chronology of events for the 1990 Census adjustment decision. The following handouts for the discussion were distributed and will be on file with these minutes.

(1) Bureau of the Census, "Assessment of Accuracy of Adjusted Versus Unadjusted 1990 Census Base for Use in Intercensal Estimates," Report of the Committee on Adjustment of Postcensal Estimates, August 7, 1992.
(2) Department of Commerce, Office of the Secretary, "Adjustment of the 1990

Census for Overcounts and Undercounts of Population and Housing; Notice of Final Decision," Federal Register, Part III, July 22, 1991.

(3) Obenski, Sally, "Summary of C.A.P.E. Technical Findings", January 11, 2000.

(4) Obenski, Sally and Fay, Robert, "An Analysis of the Consistency of the 1991 Mosbacher Guidelines to Census Bureau Standards", DRAFT, January 11, 2000.

(5) Thompson, John H., Memorandum for CAPE Committee, Addendum to August 7, 1992 CAPE Report, November 25, 1992.

(6) Thompson, John H., Chronology of Events for the 1990 Census Adjustment Decision, January 12, 2000.

In 1980, the Bureau of the Census and the Department of Commerce were sued over the decision to not use statistical methods to adjust the census. Because of problems with the Coverage Measurement Survey, the Bureau was opposed to adjustment and this was upheld in court. After the 1980 Census, the Bureau formed the Undercount Research Staff.

The Census Bureau planned a dual track approach for the 1990 Census based on conducting the best possible census while having processes in place for adjustment. A decision on which track to pursue was to be made before the census. A sample of 300,000 housing units was allotted for the Post-Enumeration (PES), 150,000 of which were targeted for research and evaluation and 150,000 for adjustment purposes. In October 1987, the Commerce Department announced that the PES would not be used for adjusting the 1990 Census. This led to multiple suits being filed which were aimed at directing the Bureau to use the PES for adjustment purposes.

In 1989, a settlement was reached in litigation with the following results:

(1) The Bureau would conduct the PES and the Secretary of Commerce would decide whether to adjust the census by July 15, 1991.

(2) The Secretary of Commerce would publish guidelines that would be followed in reaching a decision.

(3) A panel of experts, four members on each side of the litigation, would be formed to advise the Secretary of Commerce.

To support the analysis of the guidelines, the Bureau conducted various evaluations. The Census Bureau senior technical staff, the Undercount Steering Committee, reviewed the evaluation results and recommended that the 1990 Census be adjusted. Director Bryant reviewed the Census Bureau's technical decision and based on the research and on her analysis

recommended adjustment. Senior Department of Commerce management recommended against adjustment. The Special Advisory Panel rendered a split decision.

The Secretary of Commerce reviewed the recommendations and decided not to adjust and published this decision in the Federal Register Notice on July 22, 1991.

The Secretary of Commerce directed the Census Bureau to review the results of the PES to determine if these could be used to adjust the post-censal estimates. The Census Bureau formed the Committee on Adjustment of Postcensal Estimates (C.A.P.E.) to direct these efforts. The post-censal estimation adjustment decision was delegated to the Director of the Census Bureau. As part of the C.A.P.E. review, a computer error was found and corrected. The C.A.P.E. issued a report indicating that adjustment would make distribution for states better but could not find any differences between the adjusted and unadjusted data for entities with less than 100,000 population. Since adjustment was not demonstrated to improve coverage for all areas, Director Bryant decided not to adjust post-censal estimates on the basis of the PES. She did decide that the Federal Statistical System could use adjusted numbers for survey controls.

II. Mosbacher Guidelines

The handout, "An Analysis of the Consistency of the 1991 Mosbacher Guidelines to Census Bureau Standards," was discussed. The details of the decision made by Mosbacher are discussed in this handout. Following are the key points.

(1) The Secretary's decision was based on criteria that required the adjustment be shown to be better at all levels used. The effect of this principle was that the unadjusted census estimates were assumed to be better a priori.

(2) It was noted that the Secretary's decision rested solely on the concept of distributive accuracy. Also, it was noted that the wide variety of census data uses necessitates that both numeric and distributive accuracy are important to consider.

III. Next Meeting

The next meeting will be held on Wednesday, February 23, 2000. The agenda topics are a summary of the February 2-3 National Academy of Science (NAS) discussion on poststratification and actions the Bureau will take as a result of this discussion.

ESCAP Committee

Kenneth Prewitt
William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson, Chair
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura

cc:

Teresa Angueira
Ed Gore
Ed Pike
Catherine Miller
Fay Nash
Miguel Perez
Maria Urrutia
Genny Burns
Carolee Bush
Donna Kostanich
Raj Singh
David Whitford
Gregg Robinson
Signe Wetrogen
Magdalena Ramos
Sally Obenski

Kathleen P Zveare
02/23/2000 07:03 AM

To: Margaret A Applekamp/DIR/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Elizabeth Centrella/DSSD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Angela Frazier/DMD/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Jane F Green/DSD/HQ/BOC@BOC

cc:

Subject: Agenda for Today's ESCAP Meeting

* * * R E M I N D E R * * * *

The agenda for today's ESCAP meeting which is scheduled from 10:30-12 in Rm. 2412/3:

The summary of February 2-3 NAS discussion on post-stratification and actions the Census Bureau will take as a result of this discussion.

Issues Raised at the NAS Panel Meeting, 2-3 Feb. 2000 and Census Bureau Actions

March 30, 2000

Issue: Census Adjustment Objective/Numeric vs. Distributive Accuracy

1. Need for clarification on the overall objective of adjustment. Is the objective to improve overall numeric and distributive accuracy or to improve demographic distributive accuracy while not adversely affecting geographic distributive accuracy?
2. Need for a clear explanation on how numeric vs distributive accuracy affect decennial planning and evaluation.
3. Need for documentation on the effects of adjustment on small areas, especially blocks.

Actions: The first two issues are being addressed in large part in the bureau's A.C.E. feasibility document being prepared by the Associate Director of Decennial Census.[1] Additionally, two analyses supporting the feasibility paper examine distributive accuracy. [2] An assessment of the Mosbacher 1991 decision criteria examines the relative merits of numeric v distributive accuracy, and an assessment of small area accuracy discusses key aspects of assessing distributive accuracy in both the census and the PES. The last issue regarding small area accuracy should be resolved upon completion of a research project on block-level accuracy conducted by Dr. Bruce Spencer and Ms. Joan Hill. [3]

Issue: The Adjustment Decision Process/Performance Indicators

1. Unanimous recommendation for a public discussion of the evaluation process that the Census Bureau will follow in determining whether to release adjusted redistricting data.
2. The need to identify what performance indicators/data will be available to inform decision-makers prior to April 1, 2001.

Action: The bureau will document the decision process, including identifying those data that will inform decision-makers as to whether to release adjusted data. The plan will be available for the panel's review by September 2000.[4]

Issue: Heterogeneity

1. An assertion was made that nothing has changed since 1990 that would affect heterogeneity. Additionally, a request was made for a document listing the changes made since 1990 that address heterogeneity issues, including a brief discussion of the expected effect of each change. This is important because with

LUCA and Be Counted and other changes, the census could be more heterogeneous than in 1990, something the A.C.E. cannot control.[5]

2. Document the strengths and weaknesses of models and assumptions used in the A.C.E.[6]

Action: Dr. Howard Hogan's staff is developing documentation to address these concerns.

Issue: Evaluations

1. The need for a summary listing of the planned A.C.E. evaluations and how they will be used to estimate total error.

Action: Such a listing can be derived from draft study plans involving the total error model drafted by PRED. The study plans should be close to or finalized by the fall panel meeting. The plan for the total error model is in progress.[7]

Issue: Poststratification

1. The need for an outline of the decision process. This would address the concern expressed that the race combination decisions seem ad hoc.
2. Choose labels carefully for poststrata.

Actions: Dr. Bob Fay will be assisting DSSD staff in documenting the decision process. Dr. Hogan will be including a rationale for the bureau plan for combining race groups.[8]

Issue: Movers

1. The need to have a brief document outlining why the bureau chose PES-C--including the weaknesses of this approach and why they are acceptable.
2. The need to look at the assumption that outmovers = inmovers numbers, that is, the need to ensure the consistency of the P- and E-samples. Therefore, the bureau needs additional information on:
   - ! The movement of college students from dormitories to housing units.
   - ! Internal migration, e.g., persons moving from FL to NY from Census Day to A.C.E. interview day. This is related mainly to the regional variable but also involves the mail, MSA/TEA, and even tenure variables.
   - ! Number of outmovers for which the bureau can get matchable materials.
   - ! Net migration among poststrata for the April to June or July timeframe.
   - ! Match rates for movers, before and after imputation.

Actions: Dr. Hogan will prepare a document discussing the strengths and weaknesses of PES B versus PES C and why we selected PES C. Additionally, his staff is assessing the P- and E-sample consistency issue and will include their findings in the analysis.[9]

Issue: Missing data

1. The need to approximate what would have happened in 1990 if the bureau had used the ratio estimator versus the logistic regression model.
2. Provide more details about imputation cell estimation.
3. Provide more details about the characteristic imputation
4. Noninterview (NI) adjustment to whole household (HH) noninterviews need some scrutiny (probably because not enough details in background materials)

Actions: The bureau will not approximate what would have happened in 1990 had the ratio estimator been used because there was so little missing data that the difference would have been minimial. As for the other issues, the detailed specifications in progress should address these and any other lingering concerns of panel members and invited guests. [10]

Issue: CAPI by Telephone Interviews

1. The need to address the concern that early interviews (especially by the phone) may have different expected values for missed persons and different accuracy for mover reports.
2. Is either operational or model independence being violated?
3. Is there anything to the concern that telephone CAPI will suppress reporting of children?

Actions: Dr. Hogan and his staff will assess the need to examine these issues further.[11]

Issue: Definitions/Clarification

1. Dr. Norwood requested an explanation of the differences among demographic analysis, population estimates and population projections.

Action: Dr. John Long will prepare this explanation.[12]

Issue: **The PES "Error Chart"**

1. The need to understand the numbers presented and to refute the arguments.

Action: Bob Fay and Sally Obenski are preparing a document that explains the numbers and responds to the underlying issues.[13]

End Notes

1. Thompson, John, "The Accuracy and Coverage Measurement Evaluation Survey: A Statement on the Feasibility of Increasing Accuracy Through Statistical Methods," Draft March 31, 2000.

2. Obenski, Sally, " An Analysis of the Consistency of the 1991 Mosbacher Guidelines to Census Bureau Standards," Draft February 23, 2000. Obenski, Sally, "Analysis of C.A.P.E. Findings for Small Geographic Areas," Draft February 24, 2000.

3. Spencer, Bruce and Hill, Joan, "Accuracy of Block-Level Estimates of Population," Draft XXXX.

4. TBD

5. TBD

6. TBD

7. Spencer, Bruce, "Components of Error Needed for the Total Error Model," Draft February 13, 2000.

8. TBD

9. TBD

10. Cantwell, Pat and Ikeda, Michael, ...Specifications for Missing Data Model, XXX.

11. TBD

12. TBD

13. Fay, Bob and Obenski, Sally, "An Assessment of Wachater and Freedman's PES Statistics and Issues," Draft March XX, 2000.

February 2-3 NAS Panel on Dual System Estimation: Bureau Summary of Topics, Discussion, and Closing Statements

April 3, 2000

Summary Objective and Scope

The purpose of this summary is to document the key issues that were raised and discussed during the panel's February meeting on the Accuracy and Coverage Evaluation Survey's (A.C.E.) underlying methodology, Dual System Estimation (DSE). This summary is an internally-generated analysis that identifies issues raised by the panel and invited guests to assist decennial managers in focusing discussion and prioritizing action plans. Neither the panel members nor the invited guests has seen or reviewed this document. The information in the document is not an official representation of participants' positions. It is the Census Bureau's best recollection of the discussions ensuing over the two days and will only be used for internal planning. Further, this document is not intended to be a detailed transcript with every comment attributed to a panel member or invited guest. An official document will be prepared and distributed by the panel. Consequently, other than a few of the principals, such as the panel Chair, names have not been used.

The summary reflects an analysis by Census Bureau officials of the discussion over the two-day period that has been reorganized into topics. It begins with opening statements, provides an overview of the A.C.E. design, describes key topics, associated issues, and perception on consensus, and concludes with invited guests' closing remarks.

Background

The Panel to Review the 2000 Census was convened by the Committee on National Statistics, National Research Council, in the fall of 1998, at the request of the U.S. Census Bureau. The panel, which is chaired by Dr. Janet Norwood, former commissioner of the U.S. Bureau of Labor Statistics, is charged to review the methods, procedures, and results of the 2000 census.

The panel will review features of the census that affect the completeness and quality of the data, such as the Master Address File (MAF), follow-up for nonresponse, proxy responses, race and ethnicity classifications, and other areas. The panel will also review the statistical methods, operations, and results of the planned A.C.E. and DSE methods that the Census Bureau intends to use to evaluate the coverage of the census and to produce adjusted counts in the spring of 2001. The primary focus of the panel to date has been on the A.C.E.'s design and methodologies.

To assist the panel in its evaluation, several meetings were planned to address critical aspects of the A.C.E. The panel met in October 1999 to discuss the sampling and estimation methodology of the A.C.E. In February 2000 the panel met to discuss statistical and operational issues regarding DSE. In the fall of 2000 the panel plans to meet to review the process and performance indicators that Census Bureau officials will use to decide whether to release adjusted redistricting data in the spring of 2001.

Opening Statements

Opening statements were made by the panel chair, Dr. Janet Norwood, the Director of the Census Bureau, Dr. Ken Prewitt, and the Associate Director for Decennial Census, Mr. John Thompson. (Mr. Thompson's opening statements are summarized in the numeric versus distributive accuracy section.) Dr. Norwood stated that the purpose of the meeting was to obtain as much information as possible about DSE and the A.C.E. design, *not* to revisit the 1990 adjustment issue. She further stated that the meeting was not to be about politics, that differences of opinion were to be expected, and that the panel had no official position on adjustment at this time.

Dr. Prewitt opened by telling the panel and invited guests that the Census 2000 was currently on schedule. He then made several statements about adjustment and accuracy. He stated that criticisms of adjustment can be grouped into (1) accusations that the Census Bureau is pursuing partisan politics; (2) concerns over operational feasibility; (3) concerns over public acceptance; and (4) differences in opinion about improvements realized from adjustment. He stated that the first concern is misplaced but is interested in examining the others. Consequently, Dr. Prewitt asked for a precise framing of the issues by the panel and other contributors, particularly if during that process verifiable facts can be identified. He then stated that the Census Bureau's position on distributive versus numeric accuracy was to favor numeric, as it was difficult to maximize both.

Dr. Prewitt concluded by asking two questions: (1) How would you design a census to achieve distributive accuracy? (2) What are the facts about congressional seats shifting as a result of the computer programming error discovered in 1992?

A.C.E. Design Overview

Dr. Howard Hogan, Chief of the Decennial Statistical Studies Division, was the Census Bureau presenter of the A.C.E. design. He first provided an introduction to the underlying methodology, DSE, used in the A.C.E. He included the criteria that defined an application being "in" the census: (1) appropriateness of enumeration; (2) uniqueness; (3) completeness; and (4) geographic correctness. Dr. Hogan then outlined changes and improvements in the A.C.E. over the 1990 Post Enumeration Survey (PES).

Improvements included:

- To minimize 1990 concerns about block size differences, the Census Bureau used the most recent address listing during sample reduction for the A.C.E.;
- To obviate 1990 variance concerns about small blocks, in 2000, the Census Bureau will reduce the small block universe whenever possible by using a clustering algorithm that will group smaller blocks with larger block clusters; and,
- To reduce transcription and keying errors, the Census Bureau will use Computer-Assisted Person Interviewing (CAPI) for the A.C.E. person interview. A relatively small percentage of the overall sample size will be interviewed via CAPI by telephone after their census questionnaires have been received. CAPI by telephone will facilitate training for the person interviewing that starts in July right after Nonresponse Follow-up with the Nonresponse Conversion Operation going into early fall.

Other changes included:

- Increasing the sample size and sampling probabilities should reduce variance levels from 1990 and improve small block sampling;
- Group Quarters (e.g., college dormitories) will not be included in the A.C.E.;
- Two variables will be added to the 1992 poststratification design (i.e., mail return and Metropolitan Statistical Area/Type of Enumeration Area (MSA/TEA));
- The approach for handling people who have either moved out of or into the A.C.E. sample (i.e., movers) has changed from PES B to PES C;
- Instead of searching the areas surrounding the entire A.C.E. sample for matches, the search area will be limited to a targeted extended search;
- There will be overlap between the A.C.E. and Nonresponse Follow-up during the CAPI interviewing by telephone; and,
- In 2000, because of the increase in sample size, the Census Bureau will not use the statistical method of smoothing to offset the effects of high variability.

Numeric versus Distributive Accuracy

The relative merits of numeric versus distributive accuracy was a pivotal topic in 1991 discussions about adjustment. Improving numeric accuracy refers to getting the total population as close to “truth” as possible. Improving distributive accuracy refers to getting the allocation of the population to states or other geographic units as close to “truth” as possible. Both are important to uses of census data.

Issue

While improving both aspects of accuracy is important, it may not be always possible, even at larger geographic levels. Complicating the issue is that improving distributive accuracy can apply to both demographic groups as well as to geographic units. Consequently, it is important to clarify the overall objective of adjustment, including how numeric versus distributive accuracy affects decennial planning and evaluation.

Bureau Statement

Mr. Thompson elaborated on numeric versus distributive accuracy. He explained that in planning for a census or any other survey, the Census Bureau looks for design procedures to improve numeric accuracy. If executed perfectly, such procedures will improve distributive accuracy. However, the bureau does not *plan* a design based on improvements to distributive accuracy; again, it focuses on increasing numeric accuracy. In 1990, PES evaluations almost exclusively focused on distributive accuracy while evaluations of other coverage improvement programs focused on numeric. In 2000, evaluations will focus on both aspects of accuracy for all coverage improvement programs, including the A.C.E.

Discussion

Several participants stated the importance of distributive accuracy as a goal--that is, getting the proportion of the population or share closer to "truth." In response, Dr. Hogan asked how one would go about planning for improved distributive accuracy. The Deputy Director, Mr. William Barron, asked whether the Census Bureau should walk away from numeric improvements in the American Indian count, for example, if distributive accuracy were not improved. Further, one participant pointed out that the formulae for the DSE are for counts and were not designed to improve distributive accuracy. In response, another participant stated that formulae are not determinative and that to evaluate the methodology, one must work through the counts to get to the shares and the uses of census data. Likewise, one participant concluded that everything is about shares, and another stated that operational decisions should focus on numeric accuracy, but whether to adjust should be a distributive accuracy question.

Another area of discussion was whether the appropriate share metric should be geographical or demographic. Several participants pointed out that the emphasis on counts for the undercounted population groups is really a sense of shares and is somewhat of a hybrid. From that perspective, the Census Bureau can use distributive accuracy as a planning tool. Because of increasing tension between multiple criteria, e.g., accuracy of groups v geographic areas, one participant suggested one objective could be to reduce the differential undercount without adversely affecting the rest of the census. This prompted two additional comments. One participant stated that if shares are the ultimate goal, then the Census Bureau should measure a given demographic group across all geographic areas. Another asked which vector of errors the Census Bureau preferred--demographic groups or geographical areas? He added that one vexing problem would be how to factor shares into performance indicators used to make the adjustment decision.

Consensus
Participants focused mostly on the importance of distributive accuracy but seemed to recognize the complexity of whether the objective is to improve the distributive accuracy of demographic groups or geographic units or both. Other than suggesting that the Census Bureau could plan for improved distributive accuracy by focusing on groups, no one addressed Census Bureau concerns about how one can plan for improved distributive accuracy.

Decision Criteria/Performance Indicators

In 1990, the Census Bureau conducted and assessed 21 evaluations of the PES and 11 evaluations involving demographic analysis prior to the July 1991 adjustment decision by the then Secretary of Commerce Robert Mosbacher. Synthesizing the volumes of data involved using a total error model and loss functions to determine whether the error in the census was greater than the error contained in the PES due to sampling and nonsampling (e.g., matching, recall) error. Such a comprehensive assessment cannot be conducted prior to the April 1, 2001 adjustment decision for redistricting data.

Issue
The Census Bureau will not have completed a full evaluation of the A.C.E. until well after the adjustment decision. Consequently, only limited data will be available prior to the adjustment. The Census Bureau needs to determine the degree to which it needs a formal decision process, supported by performance data from the census and the A.C.E.

The Panel Chair Statement
Dr. Norwood reiterated that the purpose of the panel is not to decide whether numbers should or should not be adjusted but to evaluate what the Census Bureau has done. She then posed the following question: What should be looked at to evaluate the accuracy of the census with and without adjustment?

Discussion
Most participants seemed to acknowledge that repeating the 1991 Mosbacher-type assessment before April 1 was not feasible. However, many did assert the need for some performance data to inform an adjustment decision, while recognizing that some subjectivity is always necessary. Dr. Norwood suggested conducting some evaluations earlier than planned. One participant stated that the Census Bureau should provide (1) a public list of things to do before the numbers are released on April 1, (2) a catalogue of assumptions and error-sources in the A.C.E. (but focus on gross errors), and (3) a list of the planned A.C.E. evaluations and how they will be used to estimate total error. Many participants provided possible performance indicators and evaluation methodologies. The suggested performance indicators included:

- Demographic sex ratios from the census;
- Address list matching from the A.C.E. housing study as an early indicator for the MAF

quality;

- Census indicators, including mail response rates, geocoding errors, erroneous enumerations, duplicates, and last resort cases;
- Intercensal estimates; and
- A.C.E. indicators, including interview rates, noninterview rates, quality control measures, estimates/variance on estimates, numbers of movers, movers across poststrata lines, match rates, missing data rates, and blocks in which matches exceeded enumerations (influential blocks).

As for evaluation methodologies, one participant pointed out that the key was to understand the relation between indicators and accuracy. However, another responded by stating how difficult it will be to synthesize the errors for decision-making. Several suggested that although there was no "right" scientific approach, it was important to conduct sensitivity analyses to better understand indicators and error. Using loss functions was suggested as a means to evaluate the consequences of using adjusted versus unadjusted data at differing geographic and error levels. One participant provided the following guidance: Obtain the best possible numbers, be perceived to have done a reasonable job, and bullet-proof the Census Bureau against the unreasonable because their minds cannot be changed.

A few participants expressed a global concern that high-levels of all types of measurement error in the 1991 PES would be repeated in the A.C.E. However, others pointed out specific concerns: (1) A potential increase in erroneous enumerations due to multiple response options (especially Be Counted) and Complete Count Committees focused on making numbers bigger; and (2) A decline in data quality due to the use of Optical Character Reading (OCR) in data capture.

Consensus

Consensus was reached that the Census Bureau cannot repeat the 1991 Mosbacher-type assessment prior to the decision to adjust, but should have a documented decision process with performance indicators made publicly available. Opinions differed as to the degree and effect of nonsampling error on the 1991 PES. Near consensus was reached on the need for careful controls for Be Counted Forms (i.e., forms that do not have census identification numbers), but no consensus was reached on whether the planned use of OCR would lead to a decline in data quality.

Heterogeneity

Two concerns about DSE are heterogeneity and correlation bias. Heterogeneity occurs when there is a failure of the so-called synthetic or homogeneity assumption in producing the adjusted census counts. As the first step, the population is divided into categories or poststrata defined by a number of variables, such as age and sex, in a set of geographic areas. An adjustment factor is estimated for each poststratum. The synthetic estimate is formed by applying the adjustment factors to the corresponding counts by poststrata at the block level and aggregating the results to

higher levels of geography. The synthetic assumption assumes that the probability of being missed in the census is constant for each person within a poststratum.

Residual heterogeneity, that is, heterogeneity in census inclusion probabilities not explained by the poststratification used to estimate census coverage, has two undesirable consequences. At the poststratum level and above, the consequence is correlation bias; that is, missing the same people that the census missed. For "small areas" below the poststratum level, the concern is that geographic variation in inclusion probabilities violates the synthetic assumption, leading to biased estimates for small areas. Critics of sampling often point to local heterogeneity as being an inherent problem with DSE that demonstrates why the census counts should not be adjusted. For example, the Census Monitoring Board recently published a report on how statistical adjustment fails to eliminate local undercounts.[1]

Issue

An assertion was made that nothing has changed since 1990 although a number of changes that affect heterogeneity were discussed. There needs to be determination of the degree that general documentation is needed on planned models and assumptions and specific documentation is needed on how 1991 heterogeneity concerns are being addressed in the A.C.E. design.

Bureau Statement

Dr. Hogan explained that the initial design did not share undercount data across state lines. Undercounts for states were estimated directly from state-based samples, mitigating heterogeneity concerns. After the Supreme Court decision disallowing sampling to be used for apportionment purposes, the Census Bureau reduced the sample size and changed the design by allowing the sharing of undercount data among states. Consequently, Dr. Hogan launched an intensive research effort to design a poststratification model that would reduce heterogeneity.

Discussion

A presentation was given by Mr. Charles Jones, a Census Monitoring Board staff member, on the results of the Board's Report to Congress. [2] The presentation focused on the effects of adjustment on 1990 block-level accuracy. The statement was made that DSE methodology does not correct small area problems because of heterogeneity. The study used directly calculated block-level DSEs to test the assertion. However, it was pointed out by other participants who had assessed the study that the block-level data used by the Board did not have the associated weights and were not the numbers used in 1990 to form the synthetic estimates. Further, one participant dispelled the notion that the direct DSE was "truth," by giving Ft. Polk as an example. Although Ft. Polk has a large population of multiple minority groups, instead of indicating a high

---

[1]U.S. Census Monitoring Board, "Unkept Promise: Statistical Adjustment Fails to Eliminate Local Undercounts, as Revealed by Evaluation of Severely Undercounted Blocks From the 1990 Census Plan," Report to Congress, September 30, 1999.

[2] Ibid.

undercount as expected, the direct DSE indicated low undercount. Additionally, the participant stated that the Board's study focused on the "tails" of the DSE distribution (i.e., on the extremes) and, if one focused on the bulk of the distribution, it made the case that the DSEs were, in fact, behaving as expected. That is, although the effect was small, adjustment did move the block-level counts generally in the right direction. Moreover, the block-level data are aggregated to higher levels where definitive improvements can be demonstrated.

Several participants generally seemed to like the mail response variable in the poststratification scheme and believed that it would reduce heterogeneity. An assertion was made that little had changed since 1990 that would affect heterogeneity. One participant asked whether the Census Bureau was planning to document how heterogeneity and correlation bias are or are not dealt with in the A.C.E. He suggested that perhaps the research community should spearhead a major effort of this kind. Generally, however, statements indicated that little difference exists between unadjusted and adjusted data at the block-level--both have errors. Focus should not be on blocks but on how blocks are aggregated into tabulations for census uses.

Consensus
No consensus was reached on how serious the effects of heterogeneity are on DSE accuracy.

Poststratification

The 1991 PES design included 1,392 different poststrata. Poststratification is the dividing up of the population into groups with similar capture probabilities. They were formed according to pre-identified variables, such as age, sex, race, tenure, and other variables thought to be associated with differing capture probabilities. A person could be only one poststratum grouping. One of the complexities for 2000 is that the ability for people to select more than one race increases the possible outcomes by almost tenfold.

Because the sample size was relatively small (about 160,000) and the number of poststrata large (1,392), some of the poststrata had high variances that led to the use of a statistical model called smoothing. In part, because of the complexity of smoothing, the Census Bureau eliminated its need by reducing the number of poststrata to 357 during its 1992 analysis of whether intercensal estimates should be adjusted.

Issue
Two issues emerged over poststratification. The first was the adequacy of the Census Bureau's design. However, the Census Bureau made a late change in a variable and was not able to provide detailed performance information on the selected model. A second issue was the treatment of individuals who respond to more than one category. A potential problem is that, given different data collection modes between the census and the A.C.E. (i.e., paper versus CAPI) and as the number of choices increase (e.g., selecting single race, two races, or three), misclassification errors could increase.

Bureau Statement

Dr. Hogan began the poststratification discussion by stating that the 1992 357-poststrata design was used as the baseline design for the A.C.E. The major changes included (1) using region only for non-Hispanic White owners; (2) adding a mail response variable (the real-time mail response rates); and (3) attempting to approximate urban/rural through the MSA/TEA variables. He explained that originally his staff was hoping to get urban/rural status but was told that it would not be operationally possible. Therefore, his staff had to make a late change in the plan, which will be finalized and documented in a couple of weeks. He then identified the following criteria used for selecting poststrata:

- Similar capture probabilities;
- Similar net undercount;
- Permit detection of differences among geographic areas;
- Poststratum cells should > 100 A.C.E. sample cases;
- Operationally feasible to implement in time;
- Minimize classification error;
- Account for changes in the census since 1990; and
- Explainable.

Discussion

A number of questions arose on poststratification, including concerns about the late change in the urbanicity variable, the poststrata sample size, the selection criteria, and the handling of muti-racial responses. Regarding the late change in the urbanicity variable, Dr. Hogan again explained that it was due to determining that obtaining information on urban/rural was not feasible. Therefore, his staff was attempting to approximate the urban/rural variable indirectly through MSA/TEAs. He also explained the variability in sample size (i.e., enough sample to support a robust sample yet include the >100 criteria). Most strata will have large samples. The lower limit of greater than 100 was due to concerns about small, geographically disperse groups, such as American Indians not on reservations and Hawaiians and Pacific Islanders. When asked whether he had looked at combining strata to address this issue, Dr. Hogan stated that he and his staff had looked at several combining options but discarded them to keep a balance between variance and bias. As for determining which poststratification scheme is superior, Dr. Hogan explained that his staff had conducted simulations, looked at variances, and used targets to measure bias.

When asked about the multi-race issue, Dr. Hogan briefly provided some insight into how the muti-race responses will be handled. For example, he stated that American Indians on Indian Country will be coded as an American Indian regardless of having selected another race and/or Hispanic Origin. Any person marking Black and another single race group is coded Black, but

any person marking Asian and White are coded White.[3]

Dr. Hogan readily admitted that there is not much empirical data to support the selected plan. However, Dr. Hogan pointed out that the importance of minimizing classification error cannot be overstated and that, as such, we were using our professional judgement in attempt to control misclassification. It was generally concluded that little knowledge exists on combining racial groups, but a general unease was expressed by several participants that the Census Bureau was taking on such a *de facto* role in deciding how to group racial responses.

Consensus

Because of the late change in the poststratification model, detailed information on how the final model was selected and performance data were not available for participants' review. Consequently, no consensus was reached on the adequacy of the model other than a general agreement that the mail response variable should reduce heterogeneity. As for the multi-race issue, although a general wariness about the Census Bureau plan to handle multi-race responses was expressed by participants, no one seemed to know what to do about it except to suggest caution when choosing poststratification labels.

Treatment of Movers--PES C

People who move present a special challenge for designing a DSE for census application for two reasons. First, people who move are more likely to be missed by the census and by the survey. Second, if a person has a different "usual residence" at the time of the survey than he did at the time of the census, one must decide where to sample him. In the 1990 PES, movers were sampled where they lived at the time of the survey interview. The Census Bureau then searched the census records at, and only at, their April 1 usual residence. This is known as "procedure B" or "PES B." For census 2000, a different procedure will be used, known as "procedure C" or "PES C." The A.C.E. will estimate the number of movers by the number of people who moved into the sample blocks between April 1 and the time of the A.C.E. interview (in-movers). In PES C, the Census Bureau will attempt to determine who lived at the interview address on April 1. If the residents have moved, then interviewers will have to obtain information on them from proxies; that is, either from the new residents or neighbors.

Issue

The move from PES B to PES C represents a significant change in sampling methodology from 1990 that results in design trade-offs rather than improvements. The trade-offs include (a) increased simplicity but may result in inconsistencies between the P- and E-samples, and (b) easier matching but may result in more response error.

---

[3]For a detailed description of the current plan see, DSSD Census 2000 Procedures and Operations Memorandum Series #Q-21, "Accuracy and Coverage Evaluation Survey: Poststratification for Dual System Estimation," January 12, 2000.

Bureau Statement

Dr. Hogan explained the chief differences and the trade-offs between PES B, used in 1990, and PES C, planned for 2000. He pointed out that for PES B, the advantage is talking directly to the person while the disadvantage is that matching is difficult. As for PES C, matching is easy with a reduction in geographic matching error. However, there are more proxy interviews and higher noninterview rates, which could affect the quality of DSE. He stated that there aren't more inmovers than outmovers because the assumption is that the overall number of inmovers equals the overall number of outmovers. PES C tries to compromise between approaches to measure the number of movers and enumeration rate for movers.

Discussion

Most of the discussion focused on the assumption that inmovers and outmovers balance and on the quality of the proxy interviews. The first issue centered on regional differences, seasonal movements, and college students. One participant pointed out that the Northeast Region, for example, doesn't have nearly the inflow of people that the South and the West do. Likewise, another asked about seasonal movements. Dr. Hogan responded to the former by stating that he cannot believe that net migration over 2-3 months is so large as to make assumption about movers problematic. He responded to the latter by stating that seasonal movement was probably not true poststratum to poststratum, and that it is the large net flows between April and July crossing poststrata that inflate or deflate measures; hence, the rates should balance. The last concern was about college students coming out of dormitories not included in the A.C.E. sample because they are Group Quarters and moving into housing units for the summer that could be in the P-sample. One participant questioned whether the P-sample was consistently defined, due to this issue with college students and whether the Census Bureau could compare the match statuses between PES A (similar to PES C) and PES B using 1990 data.

Several participants expressed concern about the proxy interviews. Dr. Hogan explained that tracing is a complex and difficult task that was attempted in the dress rehearsal but was not practical for 2000. Further, tracing proved to be ineffective in the dress rehearsal. One participant pointed out that proxy data could increase response error and hence underestimate the match rate, leading to an inflated undercount.

Consensus

General consensus was reached that these two issues--consistency of the P- and E-samples and effect of proxy data on response error--need to be closely examined and documented.

Missing Data

As in all surveys, there will be nonresponse and incomplete response at various steps. The goal of the missing data process is to improve the DSE estimates. In choosing missing data procedures, the Census Bureau chooses methods that support the underlying DSE assumptions. In 1991, the Census Bureau used a fairly sophisticated hierarchical logistic regression model.

Although scrutinized by critics, two different evaluations validated the model. Moreover, missing data rates in 1991 were negligible.

Issue

The Census Bureau has moved from the 1990 logistic regression model to a simpler ratio estimator model that may not be as accurate. Further, in 2000, missing data may be more of a problem than in 1990.

Bureau Statement

Dr. Hogan stated that imputations will be based on all available information and described the three types of missing data found in the A.C.E.: (1) whole household noninterviews; (2) missing characteristics (needed for poststrata) using the hot deck methodology; and (3) missing enumerator sample. He explained that in 1990 the Census Bureau used a hierarchical logistic regression model, but in 2000 it would use a ratio estimator although the specifications are not yet finalized.[4] Howard further explained that the cell model is easier to verify and to program, which are important considerations.

Discussion

Several participants expressed concern about the move to a ratio estimator. One stated that he was not happy with the move to a unit nonresponse adjusted weight due to the limitations of the ratio estimator. He acknowledged that it was selected for simplicity but stated that there is a danger in making things too simple. He stated that a detailed explanation of the imputation model may help. One participant who had been critical of the regression model in 1990 stated that he would support the Census Bureau on using the ratio estimator for missing data. He stated that he doesn't like the logistic regression model because of the way it treats the bias/variance tradeoff-- at the expense of variance. Another participant pointed out that there is bias in the cell model too, but no sense of variance. He questioned whether we could develop a hybrid, that is use all interactions for some variables and only main effects for others. A final participant commented that the 1990 evaluations showed that the logistic regression model was surprisingly effective. The participant asked if the Census Bureau knew the effect on 1990 data if the cell model were used in lieu of the regression model and wondered what is really being gained or lost.

Consensus

No consensus was reached on the effect of changing models. Generally, participants stated that they would hold judgement until detailed specifications on the missing data model were available.

---

[4]According to the Master Activity Schedule, the detailed specifications will be completed by April 17, 2000.

Telephone CAPI and Interview Duration

In 1991, the Census Bureau did not begin the PES interviewing until the end of Nonresponse Follow-up. In 2000, primarily to ease schedule concerns, the Census Bureau will begin CAPI interviewing by telephone in block clusters that have been enumerated. However, it will overlap with Nonresponse Follow-up in other areas. Additionally, the full data-gathering phase for A.C.E. will extend from May of 2000 well into the fall.

Issue

The length of time between the A.C.E. interviews and Census Day and possible differences between the early telephone CAPI respondents and later respondents could cause an increase in response error and heterogeneity. Additionally, because of its overlap with Nonresponse Follow-up, the use of telephone CAPI could violate independence assumptions. Finally, telephone CAPI could exacerbate the undercount of children, who were a large percentage of the 1990 undercount. Some empirical studies suggest particular difficulty in identifying children as household members through telephone interviewing.

Bureau Statement

Dr. Hogan described the sequence of operations in the A.C.E. First, CAPI telephone interviewing would occur during Nonresponse Follow-up but would include only a small universe, i.e., 10-15 percent, of the cases. Second, the CAPI person interviewing occurs using the same interviewers. Third, the Nonresponse Conversion Operation (designed to resolve nonresponses) occurs. Finally, person follow-up to resolve nonmatches occurs well into the fall. [5] He further explained that the A.C.E. uses the respondents' belief of where they should have been enumerated on Census Day, rather than trying to impose a precise definition on ambiguous cases.

Discussion

Two areas of concern were raised--(1) possible errors due to the lag between Census Day and the A.C.E. interviewing and due to possible differences between telephone CAPI responses and later responses, and (2) the effect of the use of telephone CAPI on independence and responses about children. Addressing the first area of concern, Dr. Hogan explained that the Census Bureau has no data and would have to conduct a study within a study to see if there were differences between early and later respondents, which was not feasible for 2000. Dr. Norwood asked about data from other surveys and Dr. Hogan explained that the Census Bureau had used data from studies to design address and follow-up probes to compensate for recall error. Another participant pointed out that there was a correlation between the timing of interview and capture/response. Regardless of CAPI by telephone, when a survey gets to someone late, people are more likely to

---

[5] CAPI by telephone begins on May 8 and CAPI person interviewing begins on a flow basis as each block cluster is enumerated and ends on August 19. Nonresponse Conversion begins on July 27 and ends on September 1. Person follow-up begins on October 23 and will end on November 21.

move or have a response problem. Therefore, there is an underlying correlation between characteristics of the household and quality. He asked if the Census Bureau had planned experiments on this. To this, a participant commented that people see heterogeneity everywhere and asked why this is even an issue. He pointed out that it sounds like an interesting question but has little effect on estimation.

As for the second major concern--the use of telephone CAPI--one participant asked if telephone CAPI violated the independence assumption. He elaborated by asking if telephone CAPI increases the chance of some people to be included in the P-sample. If they are treated differently, he asserted their data capture probability would be increased and that data quality would differ. Another participant responded that he believed that operational independence is not the issue. He thought that telephone CAPI may compromise model independence, may affect movers, and consequently, evaluation is needed.

As for telephone CAPI and children, one participant pointed out that in 1990, 50 percent of the undercounted were children and that many were left off of the census form and asked if the Census Bureau is concerned about the use of telephone CAPI. Another participant pointed out that the National Immunization Survey results demonstrated how hard it is to find 2-year-olds. A third participant asked what the Census Bureau was doing to increase children's response rates. Mr. Thompson responded that they had initiated a very extensive Census in Schools project. Further, there are quality checks built into the census to ensure that individuals are not left off the census form. For example, the "number of people in the housing unit" number is checked against the number of people listed on the roster.

Consensus

General consensus was reached that these two issues, i.e., possible response differences and the effects of telephone CAPI on the inclusion of children, were legitimate concerns that should be examined.

Closing Comments

The following information represents the Census Bureau's interpretation of key closing comments made by participants. The purpose is informational and the statements and facts presented have not been reviewed by the participants nor have they been verified by the Census Bureau. Consequently, other than the final remarks made by Dr. Prewitt, the names of the participants have been omitted.

Participant 1

! Census 2000 is similar to 1990--and the adjustment is similar. Nothing has really changed.

! To get a handle of magnitudes of correlation bias (CB) and measurement bias (MB) look at 1990 as illustrated by the following chart.

1991 Mosbacher

The adjustment	=	5.3M	
MB	=	3.6M	1.8M (computer error)
Corrected Adj	=	1.7M	
CB	=	3.0M	
DA	=	4.7M	

MB numbers CAPE	=	3.0M
Breiman	=	4.2M
Intermediate	=	3.6M

! Given these levels of error, I question whether any improvement could be made.

! If the Census Bureau had adjusted in 1991, every single state would have gotten an upward adjustment. If two-thirds of the "undercounted" people are really the result of measurement bias, adding people seems insecure. Even with the 1.7 million undercounted, the Census Bureau doesn't know what states they should be allocated to--only what poststrata they are in.

! Even an increase in numeric accuracy is not obvious.

! As for distributive accuracy, in terms of shares, California, Texas, and Florida got upward adjustments while Pennsylvania, Ohio, and some other states got downward adjustments. Given the level of error in the adjusted numbers, it would have been just as plausible if the signs had been reversed.

! How would the Census Bureau defend the production adjustment versus the alternative adjustment cited above? It is difficult when the PES was dominated by measurement bias and heterogeneity. The same problems remain for census 2000.

! In answer to Dr. Prewitt's questions made in his opening statement:

(1) The outcome in 1990 with regard to improvements in adjustment to distributive accuracy was that distributive accuracy had a way to go. As to how to design a census to achieve distributive accuracy? I will talk about that for 2010.
(2) As for moving congressional seats, the 1991 adjustment of 5.3M would have shifted 2 seats; in 1992, when the Census Bureau changed the poststratification design and corrected errors, only 1 seat shifted.

Participant 2

- I recommend explaining DSE using Rick Griffin's paper (page 3)[6] rather than the 2x2 table used today.

- Those people with close to zero probability of being counted either in the census or in the A.C.E., the unreachables, are always going to be a problem. We need to focus on problems that we can control.

- In 1990 the Census Bureau attempted to address all sources of error in the total error model. In 2000, addressing total error doesn't seem quite so integrated into the census process. However, total error analysis is needed for understanding error components and the total error model even more for 2000.

Participant 3

- The Census Bureau should make evaluation data available at the lowest level.

- The Census Bureau should conduct loss function analysis using a total error model.

- The Panel should assess the A.C.E. evaluations looking especially at movers and missing data.

Participant 4

- Based on the performance of other similar systems, I am still concerned about the effect of OCR on quality. No human will be randomly looking at data quality.

- My deepest fear is that changes in the 2000 design will negatively impact census accuracy: multiple modes, OCR, new differential coverage programs such as the Complete Count Committees.

- If census accuracy is compromised, this will favor adjustment. However, we will need quite a bit of information to determine if an improvement is made with adjustment. Therefore, the Census Bureau must have a set of evaluations completed prior to adjustment. Otherwise, there is no way to determine which is better.

---

[6]DSSD Census 2000 Procedures and Operations Memorandum Series #Q-20, "Accuracy and Coverage Evaluation Survey: Dual System Estimation," page 3, January 12, 2000.

Participant 5

- The Census Bureau needs early evaluations of census counts. I am still concerned about multiple modes increasing the probability of duplications, i.e, erroneous enumerations. It is important to look at the census and make a determination on what the Census Bureau expects the adjustment to do.

- I recognize that the Census Bureau cannot do all the evaluations before the adjustment decision; it would be useful to do select ones to get some sense of the quality of the A.C.E.

Participant 6

- The Census Bureau needs to define things objectively. For example, get away from the definition that by correct residence it is "belief of residency." The Census Bureau needs errors in the measurement of the estimate objectively defined.

- As for PES C, two errors are possible--error in the estimate of inmovers means error in estimate of outmovers.

- The Census Bureau needs evaluations in real-time to support the adjustment decision; look at a full set of indicators, including demographic analysis sex ratios, numbers, and postcensal estimates.

Participant 7

- The total error model is of the utmost importance in evaluating the A.C.E. and the DSE. As for indicators to help inform decision-making, synthesizing the indicators is the challenge.

- In the census, the Census Bureau should look at last resort statistics and the results of the Primary Selection Algorithm to get an insight into duplicates.

- In the A.C.E., look at match rates, the return rate for listed blocks, geocoding error, duplicates, and extreme blocks (influential observations).

- By looking at a set of indicators, it could give insight into the synthesizing problem.

Participant 8

- The Census Bureau should provide as much data as possible for review by the statistical community. The more data provided, the better.

- I like loss function analysis although the Census Bureau's choice of loss functions in 1991 was subjective.

- Increased publicity and the expanded Census in Schools in 2000 are positive changes. The move to multi-race and treatment of movers are new and may possibly be problematic.

Participant 9

- I am concerned with overcounts.

- I suggest that the Census Bureau (1) summarize census and adjustment methodologies; (2) identify their strengths and weaknesses; (3) determine unknowns; and (4) assess which of the unknowns are possible to resolve.

Participant 10

- Here is a situation where one should be concerned about all the errors, not just net errors. If there were 25M gross errors as described by Participant 1, even if the net is 0, adjustment would still be worthwhile.

- Also, per Participant 1's statement about changes in congressional seats going from 2 to 1 from 1991 to 1992, participant 1 left out an important piece of data. He stated that when the 1992 357-design was run without the computer error, only 1 seat shifted. However, if the 1392 design is run again without the computer error, 2 seats still move. I discuss this and other topics on my website (http://lib.stat.cmu.edu/~fienberg/WhoCounts.html).

- Block level accuracy is largely irrelevant; it is primarily a device to add up to larger geographical units.

- I was impressed by the thoroughness of the material provided by the Census Bureau, but note that there were no citations outside the Census Bureau.

- I have two major concerns about the census process: (1) Local efforts that could increase erroneous enumerations; and (2) a possible bias (i.e., less long-form data) in the long form due to people selecting unlabeled short forms through multiple modes, such as Be Counted.

- I have identified the following issues: (1) taking stock of local improvement efforts (overcount issue); (2) conducting long-form evaluations possibly using data from the American Community Survey; (3) identifying evaluation criteria--that is, a public list of things to do before the Census Bureau releases numbers on April 1, 2001, including a list of formal evaluations; and (4) evaluating the American Factfinder's confidentiality. On

the latter, I am concerned about the disclosure issue for race.

Dr. Prewitt's Closing Remarks

- Social recognition of groups is a numeric count issue and an important concept.

- I am aware of the overcount issue. The Internet is probably going to be fine because Internet forms require census identification numbers. Blank forms, like Be Counted, will require internal checks.

- The decision on whether to use the DSE to adjust has to be made in real-time with limited input.

- Four potential outcomes could be considered in this decision process.
  - If the A.C.E. is good, and the census is bad--adjust.
  - If the A.C.E. is bad, and the census is good--don't adjust.
  - If both the A.C.E. and the census are good--may or may not adjust.
  - If neither the A.C.E. or the census is good--very tough decision.

Conclusion

Although the panel Chair made it clear that the purpose of the meeting was not to "relive" 1991, that proved to be difficult for two reasons. First, the 1991 PES, the DSE methodology, and the adjustment controversy were the impetus for the emergence of a plethora of complex policy and technical issues generating numerous technical papers over the decade. Second, the A.C.E.'s expected performance is premised on the 2000 census and the A.C.E. performing comparably to the 1990 census and PES. Therefore, many of the 1991 issues about adjustment remain relevant. As such, the preponderance of issues raised by the panel were comparable to ones raised during the 1991 adjustment controversy: How will the Census Bureau know whether it is prudent to adjust? Which is more important, distributive or numeric accuracy? Will heterogeneity be a problem? Are the underlying statistical models valid and robust?

In 2000, unlike 1991 where the Census Bureau was breaking new ground, the Census Bureau has the advantage of a decade of intense research and test results to draw upon. Consequently, anticipating many of the concerns raised during the meeting, the Census Bureau has been in the midst of, or will be documenting, many of the issues raised by the panel, as well as their resolution.

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 5

February 23, 2000

Prepared by: Maria Urrutia and Genny Burns

The fifth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on February 23, 2000 at 10:30. The agenda for the meeting was a summary of the February 2-3 National Academy of Sciences (NAS) discussions on Census 2000 and actions the Bureau will take as a result of these discussions.

Persons in attendance:

William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
Susan Miskura
Tommy Wright
Raj Singh
Gregg Robinson
Signe Wetrogen
Carolee Bush
Sally Obenski
Maria Urrutia
Genny Burns

I. Overview of the Meeting with the National Academy of Sciences (NAS) Panel

John Thompson distributed a summary prepared by Sally Obenski on the meeting with the National Academy of Sciences (NAS) Expert Review Panel on Census 2000 on February 2-3, 2000. He also distributed a list prepared by Bureau staff of policy and technical issues raised at the meeting. These handouts will be included with these minutes and kept on file.

The high points of the summary and list of issues were briefly discussed. John asked that the Committee read the handouts and give comments to Sally especially on topics that may have been omitted. These comments will be incorporated into a document clarifying the issues and responses, decisions, and actions on the major concerns. John also asked Bob Fay and Sally Obenski to gather more information on distributive accuracy in undercounts and expand the summary to include this work.

A list of the topics discussed at the NAS meeting is as follows:

Numeric versus Distributive Accuracy
Performance Indicators/Adjustment Criteria
Heterogeneity
A.C.E. Design Overview
Poststratification
Treatment of Movers–PES C
CATI and Interview Duration
Missing Data

There seems to be broad consensus of all participants that the Census Bureau needs open discussions on whether to release adjusted data and the processes/criteria that will be followed during this process.

A number of issues were also discussed as described in the handout. These issues were discussed during this ESCAP meeting and the attachment describes the actions that will be taken.

II. Next Meeting

The next meeting scheduled for Wednesday, March 8, 2000 has been canceled and is currently being rescheduled.

ESCAP Committee

Kenneth Prewitt
William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson, Chair
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura

cc:

Teresa Angueira
Bill Bell
Debbie Bolton
Genny Burns
Carolee Bush
Gerald Gates
Ed Gore
Dave Hubble
Donna Kostanich
Ellen Lee
Charlene Leggieri
Don Malec
Betsy Martin
Catherine Miller
Fay Nash
Sally Obenski
Miguel Perez
Ed Pike
Magdalena Ramos
Gregg Robinson
Raj Singh
Maria Urrutia
Signe Wetrogen
David Whitford
Henry Woltman
Tommy Wright

Kathleen P Zveare
03/21/2000 10:40 AM

To: Margaret A Applekamp/DIR/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Elizabeth Centrella/DSSD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Angela Frazier/DMD/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Jane F Green/DSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC
cc:
Subject: Agenda for March 22 ESCAP Meeting

The next ESCAP meeting will be March 22 from 10:30-12 in Rm. 2412/3. The agenda is as follows:

Overview of A.C.E. 2000 Evaluations - Ruth Ann Killion/Howard Hogan

The Census 2000 Evaluation Program

Prepared by the
Planning, Research, and Evaluation Division

February 4, 2000

CENSUS 2000 EVALUATION PROGRAM
Table of Contents

February 4, 2000

Introduction[1]

For over half a century, the Census Bureau has conducted a formal evaluation program in conjunction with each decennial census. For Census 2000, the Evaluation Program will assess the effectiveness of key operations, systems, and activities in order to *evaluate the current census* and to *facilitate planning for Census 2010* and the *American Community Survey*.

The Census 2000 Dress Rehearsal, conducted in 1998, included evaluations of questionnaire design, field operations, data processing, and estimation. Over 40 evaluation studies were used to inform the final Census 2000 design. The Census 2000 Evaluation Program more than triples this effort; about 140 evaluations are planned.

These evaluations fall into 18 broad categories covering response rates, data quality, promotion and partnerships, address list development, field operations, coverage improvement, data capture and processing systems, the Accuracy and Coverage Evaluation survey, and others. The evaluations speak to issues of quality, plausibility, feasibility, accuracy, effectiveness, and value, and will provide a comprehensive assessment of the operations and outcomes of the census. For each of these 18 categories, this document provides an "Overview" and a "What Will We Learn" section, followed by a brief description of each planned evaluation.

In addition to the evaluation studies, the Census Bureau will prepare a variety of profiles or assessments based on real-time operations. For example, a quality profile based on the address listing quality assurance results will be prepared.

The design of Census 2000 is by far the most ambitious decennial census in history, particularly in its uses of an open planning process, promotion, partnerships, new technologies, statistical methodology, and alternative methods for hard-to-count populations and areas. Yet, as our nation continues to grow in size, complexity, and the need for rapid and accurate data, all these things and more will need to be further refined and developed to meet the challenges of providing data in the 21st Century - more data needs at lower levels of geography on a more timely basis.

The Census 2000 Evaluation Program is an ambitious program to assist the Census Bureau in evaluating Census 2000 and in exploring new survey procedures in a census environment. In conjunction with the Census 2000 Testing and Experimentation Program, it will build the foundation for making early and informed decisions about the role and scope of Census 2010 in the federal statistical system and its interaction with the American Community Survey and the Community Address Update System. This work will provide critical analysis and information for Census Bureau planning and implementation decisions for Census 2010 and the American Community Survey.

---

[1] The scope of the Census 2000 Evaluation Program is not final; therefore, the information presented in this document may change.

This page intentionally left blank.

A: Response Rates and Behavior Analysis

Overview

These evaluations examine various modes for providing responses to the census. We will study the use of the telephone and Internet as response options along with their use in providing assistance to respondents. The effectiveness of mailing practices and the targeted dissemination of forms will also be assessed. These evaluations focus on respondent behavior and how that behavior impacts response rates (i.e. mailback, telephone, and Internet). Findings from these evaluations will identify methods that can be used in future censuses to improve the overall response rates.

What Will We Learn?

The findings from these evaluations will answer a number of critical questions about how quickly and completely the U.S. population responded to Census 2000. From a technical standpoint, the use of an Internet Questionnaire Assistance (IQA) module will demonstrate the utility of employing the "most current" technologies and provide insight into respondent perception of using this mode for requesting information or completing a questionnaire. Likewise, an enhanced telephone questionnaire assistance (TQA) program that is user-friendly and comprehensive will provide further insight into respondent needs and preferences. Evaluation of the TQA program also will provide insight into questions concerning contractor support.

Analyzing mail response/return rates (by form type, demographics, and geography) and mailing practices, such as tracking undeliverable questionnaires, will provide insight into improving overall response rates. Assessment of the *Be Counted Campaign* will help determine the benefits of targeting geographic areas and/or demographic groups in an effort to improve population and housing coverage. We also will examine the frequency of use of language assistance guides and questionnaires in languages other than English, along with the number of returned non-English questionnaires.

Response Rates and Behavior Analysis Evaluations

(A.1.a) Telephone Questionnaire Assistance Operational Analysis
The Census 2000 Telephone Questionnaire Assistance system was developed with contractor support to provide the following services to respondents: 1) helping them complete questionnaires, 2) providing questionnaires (English forms only) and foreign language guides upon request, and 3) conducting short form questionnaire telephone interviews when necessary. This operational evaluation assesses respondent behavior (i.e., calling patterns), the accuracy of geocoding results, and the quality of data received through the various modes.

(A.1.b) Telephone Questionnaire Assistance Customer Satisfaction Survey
This evaluation focuses on customer reaction to the Census 2000 Telephone Questionnaire Assistance program. It includes analyses in the following areas: accessibility, ease of use, overall satisfaction with the assistance, and appropriateness of the information provided.

(A.2.a) Internet Questionnaire Assistance Operational Analysis
When accessing the Census Bureau website, the Census 2000 Internet Questionnaire Assistance program provides an informational service to respondents. This is the first time such a service has been available to the public. This evaluation assesses the type of service requested, the total number of visits to the site, the time (date, day of week) distributions of these visits, and the topics/pages most frequently searched. Note that Internet Web hits are a poor measure of traffic volume, but in most cases they are the only measure available. They can be used as a relative measure of one page's hits relative to another page's hits, or one server's hits relative to another server's.

(A.2.b) Internet Data Collection Operational Analysis
For Census 2000, respondents have the opportunity to complete the short form questionnaire on the Internet. This is the first time a decennial census has used this data collection mode. Since there is no background data on what might be expected in terms of frequency of use and completeness of the data, a general evaluation of the Internet data collection mode is planned.

(A.2.c) Internet Website and Questionnaire Customer Satisfaction Survey
Customer satisfaction surveys will be used to examine the effectiveness of both the Internet Questionnaire Assistance and the Internet Data Collection programs.

(A.3) Be Counted Campaign
The Be Counted Campaign makes blank questionnaires available at convenient locations for persons who believe they may have been left out of Census 2000. This evaluation will examine person and housing unit coverage gains from the campaign along with the characteristics of those enumerated on Be Counted forms. This evaluation also will assess the impact on the Master Address File through documentation of housing unit adds resulting from this program, and it will evaluate our ability to geocode and process Be Counted forms.

(A.4) Language Program - Use of Non-English Questionnaires and Guides
This study will document how many housing units were mailed the advance letter about requesting a non-English questionnaire, by state and type of enumeration area (e.g., mailout/mailback, update/leave, etc.); how many non-English forms were requested, completed, and checked in; and the frequency of requests for non-English short and long forms. This study also will document the number of language assistance guides requested through Telephone Questionnaire Assistance, Questionnaire Assistance Centers, and the Internet, along with an analysis of which languages were most often requested, whether the requests were clustered geographically, and how many requests for a language assistance guide resulted in a mail returned form.

(A.5) Response Process for Selected Language Groups
This evaluation will provide insight into how Spanish, Vietnamese, and Russian speaking households coped with the census questionnaire in Census 2000. Specifically, we will look at how these non-English speaking long form households were enumerated. We will assess their use of language guides, Questionnaire Assistance Centers, Telephone Questionnaire Assistance, and their experience with the English form.

(A.6.a) U.S. Postal Service Undeliverable Rates for Census 2000 Mailout Questionnaires
For Census 2000, the questionnaire mailout/mailback system provides the primary means of enumeration. This type of enumeration is conducted mainly in urban and suburban areas, but also in some rural areas that contain city-style address (house number/street name) systems. This evaluation examines the rates at which housing units were classified by the U.S. Postal Service as "undeliverable as addressed" (UAA) for varying levels of geography; the occupancy status of those housing units; demographic characteristics for housing units that were deemed undeliverable but had a final status of occupied; the effect that undeliverable questionnaires had on nonresponse rates; and the check-in pattern of UAA questionnaires according to date of receipt.

(A.6.b) Detailed Reasons for U.S. Postal Service Undeliverability of Census 2000 Mailout Questionnaires
This evaluation further examines the issue of the undeliverability of census mailout questionnaires. After the U.S. Postal Service determines that mail pieces are "undeliverable as addressed" (UAA), the Census Bureau will attempt to deliver these cases at the Local Census Office level. This evaluation assesses the quantity of questionnaires designated as UAA and the distribution of the UAA questionnaires according to reason for undeliverability.

(A.7.a) Census 2000 Mailback Response Rates
Housing units in mailout/mailback and update/leave enumeration areas are asked to return questionnaires in postage paid envelopes. Those questionnaires are received and checked in at Data Capture Centers. This evaluation examines mail response rates at varying levels of geography and quantifies information about incoming questionnaires according to form type and timing with respect to critical operational dates.

(A.7.b) Census 2000 Mail Return Rates

Housing units in mailout/mailback and update/leave enumeration areas are asked to return questionnaires in postage paid envelopes, and once all followup operations are complete, those housing units are assigned a final status. Only the housing units that were assigned to receive an update/leave or mailout/mailback questionnaire and had a final status of *occupied* are factored into the mail *return* rates. Data on mail return rates provides more accurate measures of cooperation than mail *response* rates, for which the denominator also includes units that turned out to be vacant or non-existent. This evaluation examines mail return rates at varying levels of geography, quantifies information about incoming questionnaires from occupied housing units according to form type and timing with respect to critical operational dates, and provides return rate data according to certain housing unit demographic characteristics.

B: Content and Data Quality

Overview

For Census 2000, the public will have five ways of providing census data. These modes include mailing back a questionnaire, filling out a census short form on the Internet, picking up and returning a Be Counted form, completing a census interview via telephone questionnaire assistance, or completing a personal visit interview with an enumerator. With this in mind, and the likelihood that the 2010 Census may offer additional options for response, studies in this category will document the characteristics of respondents and the mode by which they responded. Additionally, the data quality of each mode will be assessed. This category includes a Content Reinterview Survey study that will measure response variance, and a Master Trace Sample study. The latter will create a database containing a sample of census records with information pertaining to them from the entire census process. Other research will analyze the imputation process and evaluate multiple responses to the new race question.

What Will We Learn?

The findings from these evaluations will answer a number of critical questions on our process to define content (i.e., what questions to ask) and the resulting quality of data for Census 2000. These findings, in turn, can help us do a better job for Census 2010 and the American Community Survey.

We will learn about the completeness of the data by calculating item nonresponse rates and proxy response rates for all data items on the short and long forms. We also will look at demographic characteristics (such as age, sex, Hispanic origin, and race) of respondents by data collection mode and item nonresponse rates, and document the effects of data edit and imputation processes. We will assess responses to the new race question. In particular, we will recontact a sample of households with responses of two or more races, and ask each person to choose a single race category. This study is needed to meet the data requirements of other agencies that use only single race categories, and for comparison to 1990 Census race data.

We will also gain knowledge about data quality in comparison to external benchmarks by matching and comparing census data to data collected by the following Census Bureau surveys: Current Population Survey, Survey of Income and Program Participation, American Housing Survey, Residential Finance Survey, and the American Community Survey. The results of these matching and comparison studies will also help us to improve the design of future surveys and censuses.

Content and Data Quality Evaluations

(B.1) Analysis of the Imputation Process
To deal with missing data, three components will comprise the imputation process for Census 2000: substitution, edit, and allocation. Rates for each of these components will be produced for the 100 percent data items, for the tenure item, and for select sample housing unit and person items. This analysis will document the imputation process and will serve as a supplement to other evaluations.

(B.2) Documentation of Characteristics and Data Quality by Response Type
For Census 2000, there are five data collection modes available to respondents. Responses to the census may be collected from a mail or enumerator delivered census questionnaire; Simplified Enumerator Questionnaire used during nonresponse followup, list/enumerate, and update/enumerate; Be Counted form; Internet questionnaire; or reverse computer assisted telephone interview via Telephone Questionnaire Assistance. This evaluation will compare demographic differences and item nonresponse rates for these five response modes.

(B.3) Responses to Race Question
The purpose of this study is to create a data file for analytical purposes to allow comparisons between race data collected asking for only one race category and race data collected asking for two or more race responses. The study will include an oversample of households with responses of two or more race categories and ask each person to choose only one race category. This will allow us to measure the effects of this new question compared to the 1990 Census and will provide data needed by some government agencies that still require single race category data for historic comparability studies.

(B.4) Match Study of Accuracy and Coverage Evaluation Survey to Census 2000 to Compare Consistency of Race and Hispanic Origin Responses
The purpose of this evaluation is to determine the reliability of the race and Hispanic origin responses derived from the new race and Hispanic origin question. Reliability relates to the consistency with which responses for individuals are consistent across independent replications of the measurement process. This evaluation will be based on a comparison of census responses to those collected in the Accuracy and Coverage Evaluation survey.

(B.5) Content Reinterview Survey to Measure Accuracy of Data for Selected Population and Housing Characteristics
The Content Reinterview Survey utilizes a test-retest methodology, whereby a sample of households designated to receive the census long form are reinterviewed shortly after they have been enumerated by the census. These households are essentially asked the same question posed on the long form. Then the responses to the census and reinterview survey are compared. This survey assesses response variance and error that result from data collection and capture operations.

(B.6) Master Trace Sample
While most evaluation studies will provide detailed information on specific Census 2000 operations, the Master Trace Sample database will provide information that can be used to study the entire spectrum of operations, along with correlates of error across various systems, for a randomly selected group of census records. This database will contain, but is not limited to: address list information (e.g., source of address), final values for questionnaire items along with their values at each stage of processing, and enumerator information (e.g., number of enumerator attempts before completing an interview and enumerator production rates). This database also will contain information about the data capture system from rekeying and reconciling a subset of Master Trace Sample questionnaire images, the Accuracy and Coverage Evaluation, Content Reinterview Survey, and, possibly, administrative records.

(B.7) Match Study of Current Population Survey to Census 2000
Using the results of a person-level match of responses to the Current Population Survey (CPS) and Census 2000, this study provides a data set about differences between the Census and Survey estimates of social, demographic, or economic characteristics. Its strength is its ability to represent differences arising from non-sampling variation. The study focuses on the difference between CPS and Census estimates of poverty and labor force status (which are measured officially by the CPS) and on differences in reported race/ethnic status (which are measured quite differently on the two questionnaires).

(B.8) Comparisons of Income, Poverty, and Unemployment Estimates Between Census 2000 and the Current Population Survey
This study focuses on changes made to the Census 2000 questionnaire and forms processing systems that were designed to improve unemployment estimates. This evaluation examines whether these changes brought the Census 2000 unemployment estimates (for states, and for various demographic and socio-economic groups) closer to the official Current Population Survey estimates than they were in 1990. This analysis may be extended to compare data, definitions, and collection procedures with the Survey of Income and Program Participation.

(B.9) Housing Measures Compared to the American Housing Survey
In the past, the census and the American Housing Survey (AHS) have had a tendency to produce significant differences for many housing unit items. The purpose of this evaluation is to compare census housing data with data from the AHS and document the data differences between the two.

(B.10) Housing Measures Compared to the Residential Finance Survey
The census and the Residential Finance Survey (RFS) both collect similar housing information on mortgages, taxes, and insurance. However, the RFS collects this information in greater detail and directly from the files of mortgage lenders. This evaluation will compare housing data collected in the census with data from the RFS. Data to be compared include mortgage status, mortgage payments, presence of a second mortgage, second mortgage payments, real estate taxes, and property insurance payments.

(B.11) American Community Survey Evaluation of Follow-up, Edits, and Imputations
This evaluation will examine whether the content edit and followup procedures used in the American Community Survey have a measurable impact on any of the final estimates. This will be done by recalculating these estimates without the effects of the content edit and followup. These results will be used to infer the effect on decennial long form estimates of not conducting a similar content edit and followup procedure in Census 2000.

(B.12) Puerto Rico Race and Ethnicity
The methodology for this evaluation is being developed.

C: Data Products

Overview

The focus of this research is to determine the usability of selected data products and the effects of disclosure prevention measures on them. This will include studies of our data products strategy and of the Census Bureau's new electronic data dissemination system - the American FactFinder. We also will examine the limitations and effects of data swapping and our confidentiality edit – a combination of strategies used to prevent the disclosure of data that can be linked to an individual – on our data products.

What Will We Learn?

We will gain knowledge from these evaluations about the success of our data products strategy in meeting the needs of users and how we can improve it. We also will learn whether the American FactFinder is a usable and acceptable means to obtain census data. In studying our data swapping techniques, we will examine rates for different geographic levels and race groups and document new issues and problems that resulted from multiple responses to the race question.

(C.1) How Variations in Geography and Changes in Race Coding Affect Disclosure Prevention

For Census 2000, the data swapping methods first used in 1990 were refined through better targeting and expanded to include sample data. This evaluation examines variations in the effects of swapping due to: 1) a region's geographic structure, 2) a region's racial diversity, and 3) the number of dimensions used in the swapping.

(C.2) Usability Evaluation of User Interface With American FactFinder

The methodology for this evaluation is being developed.

(C.3) Data Products Strategy

The methodology for this evaluation is being developed.

D: Promotion and Partnership

Overview

During Census 2000, we will use new methods to promote census awareness and increase public cooperation. The primary goal of our comprehensive (and first ever) paid advertising campaign, coupled with an expanded partnership program, is to increase the mailback response rate, especially among historically undercounted populations. The advertising marketing strategy includes messages delivered through print media, radio, television, and out-of-home media (billboards, bus shelters, mobile billboards). The partnership program builds partnerships with state, local, and tribal governments, community-based organizations, and the private sector. Partners are asked to assist in three major areas: data collection support, recruitment, and promotion. In addition, a major school-based public information campaign will be launched to inform parents and guardians about the census through their school-age children. The planned evaluations for this research category will assess the effectiveness of these activities.

What Will We Learn?

These studies will help us understand how people's attitudes, knowledge, and behavior were affected by the paid advertising campaign. We also will compare these data to the 1990 census, which had no paid advertising campaign. We will examine which elements of the paid advertising media were reported/recalled most often by hard-to-enumerate groups, and provide data for Hispanics and for the five major race categories: African-American, Asian, American Indian and Alaska Native, Hawaiian/Pacific Islanders, and White. Specifically, we will look at what impact the marketing program had on the likelihood of returning a census form and (tentatively) whether it increased cooperation during nonresponse followup. The primary goals in studying the Partnership Program are to measure how well national and regional components accomplished their objectives in communicating a consistent census message of program initiatives and to determine which populations were best served by the program. Our assessment of the *Census in the Schools* program will tell us about the effectiveness of census educational materials and whether teachers receiving census materials incorporated them in their curricula.

Promotion and Partnership Evaluations

(D.1) Promotion and Paid Advertising Campaign
The Census Bureau hired the National Opinion Research Center to conduct an assessment of the marketing and advertising campaign by fielding a survey before the campaign began and after the campaign has been launched. From this evaluation, we will assess intended and self-reported response behavior and will establish a baseline and pre- and post-census measures of awareness. We will obtain the actual response behavior for respondents to our survey. We will statistically model what effect self-reported advertising exposure has on the likelihood of responding to the census or cooperating with enumerators. This evaluation also will explore the link between raised awareness, knowledge, attitudes, and response to the census.

(D.2) Census in Schools Program
A post-census survey of school teachers will be conducted to assess the dissemination system for the Census in Schools materials and the effectiveness of the materials in motivating Census participation. Scholastic, Inc. will not conduct this evaluation as previously planned. We will hire an independent contractor to conduct the data collection activities for this program.

(D.3.) Mailout/Mailback Survey of Partners
This evaluation focuses on surveying participants in the Partnership Program by using a customer satisfaction questionnaire. We will assess the effectiveness of disseminating Census 2000 materials to partners, the types and value of in-kind services rendered, the specific partnership activities conducted, and the effectiveness of the program in reaching the hard-to-enumerate population. We also will obtain from non-Federal governments the financial demands placed on them as a result of Census 2000. The sample of partners will be selected using the Contact Profile and Usage Management System database. An independent contractor will be hired to conduct the data collection activities for this program.

E: Special Populations

Overview

The vast majority of U.S. residents live as families or individually in houses, apartments, mobile homes, or other places collectively known as "housing units." However, there are millions of people in the United States who live in group situations such as college dormitories, nursing homes, convents, group homes, migrant worker dormitories, homeless shelters, or even in no place at all. Our evaluations will analyze the effectiveness of procedures to enumerate persons living in different types of group quarters and institutions. Some studies will focus on such things as enumeration at "service based locations" (shelters and food facilities for the homeless; outdoor locations where homeless people sleep). Major evaluations are planned for two operations designed to enhance the address list of special places: the Special Place Facility Questionnaire and the Special Place Local Update of Census Addresses.

What Will We Learn?

The findings from these evaluations will answer important questions on how effective enumeration procedures were in obtaining the count for group quarters. We will be able to analyze the completeness of the Facility Questionnaire and compare the telephone and personal visit operations. We also will assess interview completion rates for these groups and document the proportion of special population facilities that indicated if their administrative records could be made available. The evaluations will include distributions of these populations by type of group quarters, counts of persons at group quarters on Census Day who indicated a usual home elsewhere, and comparison of the predicted group quarters universe from the Facility Questionnaire operation with the group quarters universe as enumerated.

Evaluations for Special Populations

(E.1.a) Special Place/Group Quarters Facility Questionnaire - Operational Analysis
The Census Bureau's initial list of special places was supplemented by field staff and local partners during various operations (such as Block Canvassing and the Local Update of Census Addresses). This evaluation will document the number of special places added by each phase of this master list building operation. It also will document operational results and issues for the Computer Assisted Telephone Interview process used for most special places.

(E.1.b) Facility Questionnaire - Computer Assisted Telephone Interviewing and Person Visit
This evaluation will consist of personal visit reinterviews at a sample of special places to assess the accuracy of the information collected from the Facility Questionnaire via computer assisted telephone interview or personal visit. This evaluation will address whether classification discrepancies occur by type of special place and whether data quality differs by telephone or personal visit mode of data collection.

(E.2) Special Place Local Update of Census Addresses
This evaluation focuses on local governments' participation in the Special Place Local Update of Census Addresses. It will document changes to the address list along with operational issues that were encountered.

(E.3) Assess the Inventory Development Process for Service Based Enumerations
The purpose of this study is to assess the quality and effectiveness of the Service Based Enumeration (SBE) sites file. The quality of this file will be determined by the percentage of SBE records that were returned or could not be mailed because of incorrect addresses. We will look at deleted addresses, incomplete addresses, and added addresses. This study will also assess the addresses that could not be mailed by source to determine the relative merit of the various sources.

(E.4) Decennial Frame of Group Quarters and Sources
This study will evaluate the coverage, content, comparability, and the sources of information used to construct the group quarters frame for the decennial census (and American Community Survey), especially through comparison with the contemporary Business Register frame. This evaluation examines the feasibility and constraints to enrich or integrate these frames.

(E.5) Group Quarters Enumeration
This study will document various aspects of the group quarters enumeration. Some of the topics covered by this study include the total count of the group quarters population, the number of special places that were enumerated, and the number of group quarters that were enumerated. Additionally, the numerical distribution of group quarters per special place and of residents per group quarter will be documented.

(E.6) Service Based Enumeration

The goal of Service Based Enumeration (SBE) is to enumerate people without housing who may be missed in the traditional enumeration of housing units and group quarters. A complete enumeration of emergency shelters and soup kitchens, mobile food vans and non-sheltered outdoor locations will be conducted in late March 2000. This evaluation will document data collection completeness, last resort data collections, and whether the SBE unduplication process successfully identified individuals who were enumerated more than once. Also included in this study will be a profile of persons enumerated at SBE sites.

This page intentionally left blank.

F: Address List Development

Overview

These evaluations cover a broad spectrum of activities, both internal and external, involved with building address files and the related TIGER (geographic) database, including field operations from which address information and related map updates are gathered. The address list development category includes various evaluations of the Census Bureau's Master Address File (MAF), the TIGER database, and the Decennial Master Address File (DMAF). These include examination of the completeness and accuracy of address information in the MAF, as well as of the design of the MAF and DMAF. An evaluation of the U.S. Postal Service's Delivery Sequence File used in the MAF building process is also planned. A variety of census field and local/tribal partner operations will be evaluated to measure the impact of each operation on the MAF and the TIGER database. These include, but are not limited to: Address Listing, Block Canvassing, Update/Leave, List/Enumerate, multiple cycles of the Address List Review (also referred to as the Local Update of Census Addresses), and the New Construction Program. Combined, these field operations offer comprehensive address checks in rural and urban areas and are a primary source of address information used for MAF and TIGER database enhancement. Additional evaluations focus on the process of transferring address information to the MAF and incorporating map updates to the TIGER database from file sources and field operations.

What Will We Learn?

The findings from the address list development evaluations will provide insight into the most accurate methods for updating the MAF and the related TIGER database. This includes understanding the individual contribution of each operation as it is implemented. For each operation, we will look at the characteristics of addresses that were added, corrected, or flagged for deletion. We also will look at the geographic impact of each operation (i.e., we will examine how changes to the MAF are distributed geographically). Additionally, we will learn some things about the overall housing unit coverage in the census. Finally we will learn more about quality and coverage by examining addresses that are on the full MAF, but were not included in the census for various reasons. All of these evaluations will help inform continued MAF and TIGER database updating through the decade and also will provide insight for the 2010 Census and the American Community Survey.

Address List Development Evaluations

(F.1) Impact of the Delivery Sequence File Deliveries on the Master Address File Through Census 2000 Operations

The Delivery Sequence File (DSF) is a file of addresses produced and maintained by the U.S. Postal Service. The Census Bureau uses this file, along with the 1990 census address list and other information, to create a permanent national address list called the Master Address File (MAF). For Census 2000, the Census Bureau will use the DSF as a primary source to enhance the initial MAF for mailout/mailback areas of the country. Subsequent DSFs will be used to update the address list through April of 2000, in order to maximize the inclusion of all existing addresses in the census. This evaluation will assess the impact of each of the DSFs through Census 2000 operations by profiling the number and characteristics of housing units added to and deleted from the MAF following each delivery of the DSF.

(F.2) Address Listing Operation and its Impact on the Master Address File

For Census 2000, an Address Listing Operation was used in update/leave areas of the country to create the initial Master Address File (MAF) and provide a comprehensive update of the streets/roads and their names in the TIGER database. In this operation, census enumerators went door-to-door to identify the mailing address and physical location of every housing unit as well as the existence and name of every street and road in areas where the U.S. Postal Service does not deliver mail using house number/street name addresses. The Census Bureau used this procedure in order to create a file of good locatable addresses for Census Bureau field operations in Census 2000 as well as its future demographic surveys, including the American Community Survey. This evaluation will assess the impact of the Census 2000 Address Listing Operation on the MAF by profiling the number and characteristics of housing units added to the MAF.

(F.3) Evaluation of Address List Review 1998

The Local Update of Census Addresses (LUCA) operation (also known as Address List Review) for Census 2000 included a LUCA 98 operation that focused on mailout/ mailback areas. For this operation, local and tribal government entities were provided a Census Bureau address list containing addresses derived from the Delivery Sequence File and the 1990 Address Control File. The objective of the LUCA operations was to provide local entities the opportunity to review the Bureau's address information and related maps and then provide feedback in the form of 1) address adds, deletes and corrections and 2) street and street name adds, deletions, and corrections on the maps. The Census Bureau compared the results to the block canvassing results in mailout/mailback areas, and all discrepancies were field verified. After Census Bureau review of submissions, local and tribal entities were given the opportunity to review results and to appeal situations in which they believed the Master Address File (MAF) still was incomplete or incorrect. This evaluation will assess the number and profile of housing unit adds to the MAF, the extent of geographic clustering of these adds, and the total number and profile of housing unit deletions and corrections. The evaluation also will include information documenting the participation rates of local and tribal governments and the proportion of addresses covered by these governments.

(F.4) Block Canvassing Operation and its Impact on the Master Address File
The objective of this evaluation is to determine the extent to which the Block Canvassing Operation corrected known Master Address File (MAF) deficiencies. In 1998, the Census Bureau conducted the MAF Quality Improvement Program which measured deficiencies in the MAF as it existed prior to the Block Canvassing Operation. These deficiencies included undercoverage, overcoverage, and geocoding errors. The Block Canvassing Operation is a dependent address updating operation conducted in mailout/mailback enumeration areas. This evaluation will assess the extent to which the Block Canvassing Operation has removed the deficiencies identified in the MAF Quality Improvement Program. That is, for the MAF Quality Improvement Program sample of housing units, we will examine the changes made during Block Canvassing to see if they are consistent with our expectations from that study.

(F.5) Block Canvassing Operation
For the 1990 census, the Census Bureau conducted an operation called Precanvass to improve its address list for mailout/mailback areas. For Census 2000, a similar operation, called Block Canvassing, was implemented. As with the1990 Precanvass, this operation was conducted primarily in areas where city-style addresses are used for mail delivery; however, for Census 2000, the Block Canvassing Operation covered a larger geographic area than did the 1990 Precanvass Operation, and the scope of the operation was expanded to include map (i.e. TIGER database) updates. The objective of this evaluation is to determine the overall effect of the Block Canvassing Operation on the Master Address File (MAF) by measuring the number and characteristics of housing unit adds, deletes, and corrections to the MAF.

(F.6) Address List Review 1999
The Local Update of Census Addresses (LUCA) operation (also known as Address List Review) for Census 2000 included a LUCA 99 operation for Update/Leave areas. For LUCA 99, local and tribal government entities were provided with census housing unit block counts that were created using addresses obtained from the Address Listing Operation. Participating entities were asked to review the counts and provide feedback when they believed the number of housing unit addresses for the block should have been higher or lower. Participating governments could challenge block counts, but could not provide specific housing unit adds, corrections, or deletes. Blocks that were challenged were sent to LUCA 99 Field Verification for relisting, then returned to participating governments for another review. This evaluation will document the participation rates of those tribal and local governments that were eligible to participate, the proportion of addresses covered by those governments, the number of blocks that were challenged and went to LUCA 99 Field Verification, and the extent to which changes occurred during the field verification.

(F.7) Criteria for the Initial Decennial Master Address File Delivery
In advance of the creation of the initial Decennial Master Address File (DMAF), address information was derived from a number of files and operations, particularly the 1990 Address Control File, the Delivery Sequence Files from the U.S. Postal Service, Block Canvassing, Address Listing and the Local Update of Census Addresses operations. The status codes from

these files/operations were used to determine which addresses from the Master Address File to include in the DMAF. This evaluation will provide a profile of these addresses as well as those MAF addresses not used for the DMAF.

(F.8) The Decennial Master Address File Update Rules
A number of address list update operations occur after the delivery of the initial Decennial Master Address File (DMAF) to the printing vendor. This evaluation will assess the profile of housing units corrected, flagged for deletion, and added to the DMAF from each update.

(F.9) New Construction Adds
In this new operation, the Census Bureau will request local and tribal governments to provide information on new construction addresses (i.e., addresses for units built after the Block Canvassing operation and not accounted for by subsequent Delivery Sequence File deliveries from the U. S. Postal Service), including street and street name updates to the maps. This evaluation will document the extent of local and tribal government participation and will document what happens to these cases during Census 2000 enumeration.

(F.10) Update/Leave
The Update/Leave operation is conducted in areas where mail delivery of questionnaires would be problematic. Field staff dependently canvass their assigned area, update the address list and map, and distribute a questionnaire to each housing unit. This evaluation will document address corrections, added units, and units flagged for deletion during the operation. We also will study problem referral forms completed by enumerators for difficult listing situations (e.g., unable to obtain access, gate blocked, road washed away, no trespassing signs), to see how well these situations were followed through on and how they might have contributed to coverage errors.

(F.11) Urban Update/Leave
Urban Update/Leave is an operation that targets whole census blocks and is conducted in areas where the Census Bureau is not confident that the addressed questionnaires will be delivered to the corresponding housing units. For Census 2000, 8 of the 12 Regional Census Centers have identified blocks for this operation. The Charlotte, Kansas City, Los Angeles, and New York Regional Census Centers decided to use tool kit methods exclusively in lieu of Urban Update/Leave. This evaluation will assess the number of addresses added and deleted as a result of Urban Update/Leave and will profile the housing unit addresses as follows: single/multi-unit, P.O. Box, drop/nondrop delivery, and LUCA 98 participant/nonparticipant. It also will examine the number of Urban Update/Leave addresses that match the Delivery Sequence File and the number and profile of addresses that result in a Master Address File correction.

(F.12) Update/Enumerate
Update/Enumerate is similar to Update/Leave, except that interviewers enumerate the unit at the time of their visit rather than leaving a questionnaire to be completed and mailed back. The operation is conducted in communities with special enumeration needs and where most housing units may not have house numbers and street name addresses. These areas include some selected American Indian Reservations and the Colonias. Update/Enumerate also will be implemented in

resort areas with high concentrations of seasonally vacant housing units. Most Update/Enumerate areas are drawn from address listed areas, but some may come from block canvasses areas. This evaluation will document the number and characteristics of housing units added, corrected, and flagged for deletion in Update/Enumerate areas.

(F.13) List/Enumerate
List/Enumerate is an all-in-one operation conducted in sparsely populated areas of the country. The address list is created and the housing units are enumerated concurrently. The main objectives of this evaluation will be to profile all addresses produced by the List/Enumerate operation, as well as to specifically profile the List/Enumerate addresses that matched to the Delivery Sequence File.

(F.14) Overall Master Address File Building Process for Housing Units
The objective of this evaluation is to examine the whole series of operations that affect the Master Address File (MAF) and the corresponding TIGER database during Census 2000 and to determine their individual impact on the final census inventory of housing units. This evaluation will assess 1) the effectiveness of each component of the Census 2000 MAF building process relative to the final list of housing units in Census 2000 and 2) which MAF update operations should be retained for MAF maintenance after Census 2000 is completed. It also will measure demographic and housing characteristics by operational source.

(F.15) Quality of the Geocodes Associated With Census Addresses
The objective of this evaluation is to measure the quality of residential address geocoding in Census 2000 and to identify the source of the geocode (i.e., the TIGER database, one of the several field operations, LUCA/New Construction participants, etc.).

(F.16) Evaluation of the Block Split Operation for Tabulation Purposes
Block Split operations are conducted by the Census Bureau to provide for tabulation of data where governmental unit and statistical area boundaries do not conform to collection block boundaries. This evaluation will measure the accuracy of block splitting operations for tabulation purposes.

This page intentionally left blank.

G: Field Recruiting and Management

Overview

Prompted by the difficulties in recruiting applicants and high turnover of employees in the 1990 decennial census, the Census Bureau redesigned its recruitment, staffing, and compensation programs for Census 2000. Several new programs were developed to address the 1990 issues and to help the Census Bureau successfully recruit several million applicants, hire several hundred thousand employees, and retain this staff through the decennial census. Some of these programs include frontloading, higher pay rates, and paid advertising.

What Will We Learn?

The purpose of these evaluations is to study the effects of these new program activities upon recruitment, staffing, and retention. A contractor, for example, determined that the 1990 District Office (now LCO) pay rates were not adequately set to attract and retain staff when compared to local economic conditions of that area. The methodology to set the Census 2000 pay rates, based on this knowledge, was revised and set to a derivative of the local prevailing pay rate. The effectiveness of this higher pay rate will be evaluated, as well as other recruitment and hiring programs (such as frontloading and paid advertising). In addition, this category contains an evaluation of the Operation Control System, a system used to track work going to and from field operations. We will also learn about the overall usefulness of this system.

Recruiting and Management Evaluations

(G.1) Census 2000 Staffing Programs

This evaluation examines the effectiveness of the Census 2000 hiring programs during Nonresponse Followup (NRFU). Study questions will focus upon the effectiveness of the higher pay rate program, frontloading, paid advertising, and other areas. Some of the questions are: 1) was the Census Bureau able to adequately hire and attract staff to execute NRFU, Accuracy and Coverage Evaluation, and other various field operations; 2) were the pay rates effective in attracting and retaining staff needed for Census 2000 NRFU; and 3) did recruiting activities provide an adequate supply of applicants and replacements. A portion of this study also will examine the effectiveness of the higher pay rates on productivity and evaluate the pay model as a predictor of local economic conditions.

(G.2) Operation Control System

This evaluation examines materials, such as debriefing questionnaires, management reports, systems documentation, and cost data, to assess the effectiveness of the Operation Control System (OCS) in tracking the cost and progress of field operations. This evaluation will answer the following questions: 1) was the OCS 2000 an effective tool for tracking work going to and from field operations; 2) were the products produced (for example, listings, labels, etc.) used in the manner for which they were designed; 3) was troubleshooting necessary during production and if so, was it effective; 4) did the management reports reflect production; 5) were the reports used in managing the operations; and 6) were Field Division's overall needs met with respect to the OCS 2000?

H: Field Operations

Overview

This category includes studies of various field operations and strategies whose goals are to curb questionnaire delivery problems and obtain census data from individuals who did not respond to the census by a specified date. For example, the Local Census Office (LCO) delivery of questionnaires returned by the U. S. Postal Service "as undeliverable as addressed" is designed to increase the number of questionnaires reaching potential respondents who may not have received one otherwise. The Nonresponse Followup operation consists of sending an enumerator to collect census data from every address from which no mail, telephone, or Internet response was received. Evaluations in this category will analyze whether these operations were conducted as planned and will assess their effectiveness. Additionally, operational results will be documented for each LCO for historical purposes.

Analyses in this category also will examine our efforts to count those categorized as hard-to-enumerate. 1990 Data for Census 2000 Planning, which was previously known as the Planning Database, is composed of 1990 person and housing unit census data that are indicators of nonresponse and potential to be undercounted. This database will help the Regional Census Centers determine the placement of Questionnaire Assistance Centers and Be Counted Forms. The database will also be used by participants of our partnership program. Studies in this category will evaluate the utility of the 1990 Data for Census 2000 Planning along with the usage of Questionnaire Assistance Centers. In addition, we will evaluate our targeted enumeration methods such as blitz enumeration (use of a group of enumerators to conduct enumeration in a compressed time frame), team enumeration (two enumerators working together where safety is a concern), and the use of local facilitators (long-time neighborhood residents or church leaders who assist the enumerator in gaining entry to the neighborhood).

Because some respondents will be able to provide data without a census identification number (e.g., Be Counted and Telephone Questionnaire Assistance), it is possible that respondents will submit addresses that are not on our Master Address File. We will conduct a field verification of these types of addresses. If an enumerator verifies that the address is a valid housing unit, then it will be added to the Decennial Master Address File. We also will conduct an evaluation of the effectiveness of this operation.

What Will We Learn?

The results of these evaluations will give us an indication of how successful we were at obtaining data from nonrespondents including the hard-to-enumerate, and how to better plan these types of operations for future censuses. The evaluation of Nonresponse Followup will report proxy rates, number of partial interviews, vacant rates, and number of units enumerated during final attempt procedures, which will help us to assess whether the operation was conducted as planned. Other analyses will provide information about the quality of our enumerator training program, the usefulness of the 1990 Data for Census 2000 Planning, and a profile of Local Census Offices which will contain various descriptive statistics.

Field Operation Evaluations

(H.1) Use of 1990 Data for Census 2000 Planning
For Census 2000, the Census Bureau has developed a Graphical User Interface that will work with the data in the 1990 Data for Census 2000 Planning database to aid Regional Census Centers in planning their specific operations. This evaluation will focus on the use of these data in census planning, which includes use by partnerships and identification of special advertising campaign areas, questionnaire assistance centers, and tool kit and Be Counted sites. In addition, the study will assess the geographic distribution of tracts targeted for said operations and sites.

(H.2) Operational Analysis of Field Verification Operation for Respondent Generated Questionnaires
Respondent generated questionnaires (e.g., Be Counted, Telephone Questionnaire Assistance) or questionnaires for which an enumerator is not able to verify that the address exists are referred to as non-ID housing units. During field verification, enumerators will visit the location of these non-ID housing units and verify their existence on the ground before they are added to the Decennial Master Address File (DMAF). For Census 2000, non-ID questionnaires that are geocoded to a census block, but do not match to an address already in the MAF will be assigned for field verification. This operational analysis will attempt to answer questions such as how many units were added to the DMAF after verification and if operational problems were encountered during the implementation of field verification.

(H.3) Local Census Office Delivery of Census 2000 Mailout Questionnaires Returned by U.S. Postal Service with Undeliverable as Addressed Designation
Due to a low mail response rate and a high Undeliverable as Addressed (UAA) rate during the 1990 Census, the Census Bureau will conduct a UAA delivery operation for Census 2000 and analyze how many UAA questionnaires were designated for delivery, and how many of these were successfully redistributed by the Local Census Offices (LCOs). This evaluation also will focus on those U.S. Postal Service Sectional Centers Facilities whose delivery area covered multiple LCO borders, and will determine if delivery was successful in those areas.

(H.4) Questionnaire Assistance Centers for Census 2000
The Census Bureau will provide walk-in assistance centers where respondents can receive assistance with completing their questionnaire. Language assistance guides will be available in over 40 different languages, along with Be Counted forms that will be available in English and five other languages. This study will document various aspects of the Questionnaire Assistance Centers (QACs) such as choice of location, hours of operation, and number of employees. In addition, the frequency of use of the QACs will be analyzed.

(H.5) Nonresponse Followup for Census 2000
This operation will be conducted for all housing units in the mailout/mailback and update/leave areas for which the Census Bureau has not checked in a questionnaire by April 11, 2000. During Nonresponse Followup (NRFU), enumerators will visit each

nonresponding unit to determine the occupancy status of the unit on Census Day and to collect the appropriate data (i.e., long form or short form) for the household members. The objective of this analysis is to document various aspects of the NRFU operations. Some of the topics covered in this study include determination of NRFU workloads, identification of the demographics of those enumerated in NRFU, and documentation of the number of NRFU Simplified Enumerator Questionnaires that were partial interviews, refusals, completed via proxy respondents, or completed during final attempt procedures. The percent of NRFU units classified as occupied, vacant, or delete will be documented. Additionally, this evaluation will determine when each Local Census Office (LCO) started and completed their NRFU operation and the LCO cost of the operation.

(H.6) Operational Analysis of Non-Type of Enumeration Area Tool Kit Methods
Tool kit methods are special enumeration procedures (e.g., blitz enumeration, and the use of local facilitators) available for improving cooperation and enumeration in hard-to-enumerate areas. For this operation, the Census Bureau will assess the characteristics of areas targeted for tool kit methods based on the 1990 Data for Census 2000 Planning, and how often tool kit methods were used in areas not identified by these data.

(H.7) Evaluation of Nonresponse Followup Enumerator Training
During Census 2000, we will hire over 500,000 people to fill temporary positions. The largest number of these workers will be hired for the Nonresponse Followup (NRFU) operation. Adequate employee training will be critical to the success of NRFU. The overall objective of this evaluation is to examine the quality of the NRFU enumerator training program as well as the enumerator's state of preparedness following training. In addition, use of training materials and the adequacy of coverage of job assignments will be evaluated.

(H.8) Operational Analysis of Enumeration of Puerto Rico
Census 2000 is the first time that an update/leave mailback methodology will be used to conduct the enumeration in Puerto Rico. This evaluation will determine how many addresses were encompassed by this enumeration methodology, a profile of the addresses, and what operational problems were encountered in the field as a result of address list compilation and processing procedures. This study also will make comparisons to stateside Update/Leave data.

(H.9) Date of Reference for Respondents of Census 2000
The Census 2000 questionnaire states that the respondent should report age as of April 1, 2000. This study will document the average date of reference used by census respondents and the average date of reference by method of enumeration. This study also will document various types of discrepancies between date of birth and reported age. In addition, reported age and birth date on the census questionnaire will be compared to the same information collected by the Content Reinterview Survey.

(H.10) Local Census Office Profile

This operational summary will provide descriptive statistics at the Local Census Office (LCO) level for many census operations. For example, total housing units, average household size, and mail return rate will be among the statistics reported for each LCO.

I: Coverage Improvement

Overview

The coverage improvement evaluations examine various Census 2000 operations that are intended to improve the coverage of both housing units and people in the census. Following the mailback efforts to complete the census, a series of operations are conducted to ensure that people were counted at their correct Census Day address, confirm the status of housing units that were deleted or enumerated as vacant, and to ensure the inclusion of all persons in a household when the returned form shows discrepancies in the number of persons enumerated.

What Will We Learn?

From these evaluations we will learn about the effectiveness of these various operations as they attempt to improve census coverage. From the Nonresponse Followup operation, we will examine the potential coverage gain from identifying movers and checking to see if they were counted at their Census Day address. We will also analyze the situations where entire households were identified as having a "usual home elsewhere." For the Coverage Improvement Followup, we will examine the person and housing unit coverage gains from this operation, which determines the Census Day status of certain types of housing units (most of which are identified as deletes or coded as vacants in earlier census operations). The evaluation of the Coverage Edit Followup will measure coverage gains from this operation, which consists of contacting households whose completed forms show discrepancies regarding the number of persons enumerated, or whose completed form indicates there are more than six persons in that household. Furthermore, we will evaluate the coverage questions on the enumerator questionnaire to determine how well enumerators asked these questions and used the answers to obtain an accurate household roster.

Coverage Improvement Evaluations

(I.1) The Coverage Edit Followup for Census 2000
The Coverage Edit Followup (CEFU) is used to increase within household coverage and improve data quality in two ways. A standard questionnaire only has room for six persons, so CEFU is used to collect data on additional persons in large households. Second, it resolves discrepancies on mail return forms between the reported household size and the actual number of persons for which data are recorded on the census form. An attempt will be made to resolve all households that fail edits for these situations by using a Computer Assisted Telephone Interview. This analysis will document the workload, operational aspects, and coverage gains from conducting this operation.

(I.2) The Nonresponse Followup Whole Household Usual Home Elsewhere Probe
During the Nonresponse Followup (NRFU), List/Enumerate, and Update/ Enumerate operations, enumerators will ask respondents whether their address is a seasonal or vacation home and if the whole household has another place where they live most of the time. When respondents indicate they had a usual home elsewhere on Census Day, enumerators will record census information about this on a blank Simplified Enumerator Questionnaire (SEQ - a version of the mail return questionnaire that is easier to use for personal visit enumeration) and enumerate the current address as a vacant unit or obtain information about the people living there on Census Day. This evaluation examines how often SEQs were completed as Whole Household Usual Home Elsewhere (WHUHE), how many of these addresses were matched to an address on the Decennial Master Address File (DMAF) , how often addresses could neither be matched to the DMAF or geocoded, how often the WHUHE persons were already included on the census form for this address, and how often we found a different set of people on the census questionnaire for this address.

(I.3) Nonresponse Followup Mover Probe
In Census 2000, in-movers (households that moved there after Census Day) will be identified during the Nonresponse Followup (NRFU), List/Enumerate, and Update/Enumerate operations and will be asked if they were enumerated at their Census Day address. If a respondent does not recall completing a census form at their Census Day address, the enumerator will complete a questionnaire for the in-mover household using their Census Day address. This evaluation looks at how many of these cases occurred, and how many persons were added to the census as a result of this procedure.

(I.4) The Coverage Improvement Followup
The Coverage Improvement Followup (CIFU) universe will consist of units classified as vacant or deleted in NRFU, adds from the new construction operation, late adds from Update/Leave, blank mail returns, and lost mail returns. During CIFU, enumerators visit these units to verify the Census Day status and collect person and housing unit data as appropriate. This evaluation will document the person and housing unit coverage gain from conducting the CIFU, including the number of units that changed status from vacant to occupied or from delete to either vacant or

occupied. This study also will examine the characteristics of persons and housing units added as a result of the CIFU, start/finish dates, and the cost of the operation.

(I.5) Coverage Gain from Coverage Questions on Enumerator Completed Questionnaire
In 1990, enumerators began their interview with an explanation of who should be included as residents of the household. This procedure was changed for Census 2000. Now, enumerators begin by asking how many people were living or staying in the housing unit on Census Day. After collecting appropriate person and housing unit data, the enumerator will ask two coverage questions. The first asks about typical situations in which persons who should be included as residents tend to be missed – babies, foster children, persons away on business or vacation, roomers or housemates, and temporary residents with no other home. If someone has been missed, then he or she will be added to the form and their census information will be collected. The second question asks about typical situations in which persons who should *not* be included as residents tend to be included as such – persons away at college, in the armed forces, in a nursing home, or in a correctional facility. If someone was included on the form but should be counted elsewhere, then the enumerator will delete them from the form by marking the cancel box under their name. The purpose of this analysis is to study the effectiveness of the new coverage questions in the identification of persons who would have otherwise been missed or included in error.

This page intentionally left blank.

J: Ethnographic Studies

Overview

These evaluations will study certain aspects of coverage for various populations and attempt to identify areas where methods of collecting census data for these populations can be improved. We will examine within-household undercoverage, rostering methods (used to determine the residents of a unit), and household composition by comparing census results to Current Population Survey information for the same addresses. Another study in this category will apply social network field and analysis methods to evaluate census coverage and processes. We also will conduct ethnographic research on mobile populations and Colonias – areas lacking basic infrastructure and services along the border between the United States and Mexico.

What Will We Learn?

The comparison of the Current Population Survey and the census will give us greater insight into within household coverage errors by identifying who is missed in this survey by race, age, ethnicity, sex, and relationship. We will learn whether characteristics of the household (e.g., tenure, composition) are predictors of coverage. Other results will help us determine whether individuals can be better identified from their position in social networks (based on their interactions and transactions with others) than by comparing sets of address and person records. We will also learn how to improve procedures to enumerate mobile populations by tracing Census Day travel routes or stopover sites for a sample of such persons and determining undercounts or multiple enumerations of them in the census. We also will learn how to overcome barriers to enumerating Colonias in future censuses.

Note: Apart from the Census 2000 Evaluation Program; the Accuracy and Coverage Evaluation (A.C.E.) Survey for Census 2000 will provide a great deal of information pertaining to the coverage of various population groups. The A.C.E. Survey itself will be studied as part of the Census 2000 Evaluation Program in evaluation categories N, O, and P.

Ethnographic Studies

(J.1) Coverage, Rostering Methods and Household Composition: A Comparative Study of the Current Population Survey and Census 2000
Previous Census Bureau studies suggest a need for more analysis of the interactions between household composition, unusual living situations, and type of data collection (i.e., a census versus a survey). The census and CPS both collect household rosters and relationships information, and the CPS collects additional information on other relationships within households that can be used to study census data. This evaluation examines these within household coverage and household composition differences.

(J.2) Ethnographic Social Network Tracing
This study will use ethnographic and social network methods to study two questions: 1) What is the increased risk of coverage errors (both misses and erroneous inclusions) for persons who move once or more in a 6 month period around Census Day; and 2) Can these and similar coverage errors be reduced by identifying and tracing persons through their social networks and interactions. This would be a new approach for the Census Bureau, which traditionally has matched on non-household records (such as drivers license lists) to the census to identify persons at risk of coverage errors. The study also will examine how difficult it would be to incorporate social network methods into the census enumeration process. Various demographic, housing, household, social, and economic data will be documented for those persons who were missed or erroneously enumerated in the census.

(J.3) Comparative Ethnographic Research on Mobile Populations
In this study, a sample of selected mobile people will be traced to identify their Census Day travel routes or stopover sites. The information then will be matched and reconciled with census results. Coverage errors found in the census then will be analyzed to develop recommendations for improving procedures.

(J.4) Colonias on the U.S./Mexico Border: Barriers to Enumeration in Census 2000
Colonias are unincorporated, generally low income residential subdivisions lacking basic infrastructure and services (e.g., paved roads and public water systems) along the border between the U.S. and Mexico. In order to develop appropriate enumeration procedures and effective outreach and promotion programs for Colonias, it is necessary to better understand the unique situations and issues associated with conducting the census or other Census Bureau surveys in these areas. This research will examine the potential barriers to census enumeration in Colonias in the context of Census 2000 through participant observation, in-depth interviews, and focus groups with selected Colonia residents. Based on previous research, topics of particular interest include irregular housing, concerns regarding confidentiality, complex household structure, knowledge of English, and literacy.

K: Data Capture

Overview

The Data Capture System for Census 2000 (DCS 2000) will process more than 120 million census forms by creating a digital image of each page and interpreting the entries on each image using Optical Mark Recognition (OMR), Optical Character Recognition (OCR), or keying. These evaluations are designed to assess these components of DCS 2000, the Data Capture Audit Resolution (DCAR) process, and to measure the impact of each on data quality and on subsequent data coding operations.

What Will We Learn?

Findings from these evaluations will determine the level of accuracy at which the data capture system performed and how census data quality compares to that for capture systems used in 1990. Detailed information about the system will be collected, ranging from the number of forms processed by form type, date, and processing office, to measuring the accuracy of each of the three capture modes - OMR, OCR, and Key From Image. Operational problems and their resolution will be documented. Evaluation of the DCAR process will examine the system's ability to identify and resolve capture problems stemming from problems with response entries. Additionally, an evaluation of the interaction between the redesigned questionnaires and the new data capture system will be conducted. The impact of data capture errors on our ability to correctly assign industry and occupation codes will also be assessed.

Data Capture Evaluations

(K.1.a) The Data Capture Audit Resolution Process
This evaluation documents the results of Data Capture Audit Resolution by failure reason, form type, and Data Capture Center. Using these same categories, it also will document the number and types of changes that can be made by Audit Review clerks and the results of the Audit Count review.

(K.1.b) Quality of the Data Capture System and the Impact of Questionnaire Capture and Processing on Data Quality
This evaluation examines how the data capture system affects data quality and whether the rules for determining where cases are routed (e.g., to key from image, Data Capture Audit Resolution or audit resolution) are set appropriately. In addition, this evaluation will document and compare the data quality of each data capture method for every field on the questionnaire, as well as by form type, Data Capture Center, and racial and ethnic categories.

(K.1.c) Analysis of Data Capture System 2000 Keying Operations
This evaluation will study various aspects of the Key From Image (KFI) and Key From Paper (KFP) operations. We will document the number of questionnaires processed and production keying rates in each of these operations by form type, Data Capture Center, cluster, and date. This study also will look at the accuracy of the data captured by the KFI operation in conjunction with OCR reject rates, the content distribution of fields accepted by OCR, and those rejected and sent to KFI. Our ability to recover from KFI/KFP operational problems, the adequacy and timeliness of management reports, the cost associated with the keying operation and our ability to hire and retain keying staff also will be assessed.

(K.1.d) Synthesis of Results from K.1.a, K.1.b, and K.1.c
This report will combine and summarize the results from the following studies: The Data Capture Audit Resolution Process (K.1.a), Evaluation of the Quality of the Data Capture System and the Impact of Questionnaire Capture and Processing on Data Quality (K.1.b), and Analysis Data Capture System 2000 Keying Operations (K.1.c).

(K.2) Analysis of the Interaction Between Aspects of Questionnaire Design, Printing, and Completeness With Data Capture
This study will focus on what impact the redesigned paper questionnaires used for Census 2000 had on respondent behaviors and on the ability of the new data capture process to completely and accurately convert the questionnaire data to computer files.

(K.3) Impact of Data Capture Errors on Autocoding, Clerical Coding and Autocoding Referrals in Industry and Occupation Coding
The information provided by respondents to the industry and occupation questions on the census form must be assigned (coded) to a standard set of categories. This evaluation examines how data capture errors affect the ability of the autocoding system and clerical coders to assign correct Industry and Occupation codes.

L: Processing Systems

Overview

Once census data from all sources are captured by the Data Capture System 2000, they are stored in a file known as the Decennial Response File (DRF). Several processes then must be applied before the data can be used to produce official census counts and tabulations. One process is applied to link multiple questionnaires that were used to enumerate that same household. For example, a large family could have a mail return form with data on six members of the household and an enumerator form with data on the rest of the household. Another process is used for situations where multiple questionnaires involving different households were received for the same address. For example, one form could be for a household that moved out near Census Day, and the other form could be for the household that then moved in. A computer program known as the Primary Selection Algorithm (PSA) then is used to decide which person and housing unit data should be used for census tabulations. Following all these processes, the DRF is merged with the Decennial Master Address File (DMAF) to create the Census Unedited File (CUF), which contains the original responses for a household.

A variety of post-census activities are needed to prepare the data from the original responses to releasing the official counts and tabulations. These activities include editing and imputation, coding of write-in response items (such as race, language, industry and occupation, and place of work/migration), tabulation recoding, and data disclosure avoidance.

The Beta Site is a software testing site for Bureau of the Census application developers and is used as an integration center for Regional Census Centers (RCC) and Local Census Offices (LCO) systems, a testing center for all systems, and a support center for RCC, LCO, and the National Processing Center systems. We will examine the effectiveness of this software testing site.

What Will We Learn?

Analysis of a reinterview of multiple questionnaire addresses will determine if the PSA methodology and rules for resolving these cases accurately identified the Census Day household members. The evaluation of the DRF creation and processes will examine how well multiple forms for the same household were linked. Analysis of CUF creation will document the number of times each specific DMAF/DRF rule was applied. The Beta Site analysis will include information on whether the data collection systems were successfully integrated, and the benefits of the software testing and release process.

Processing Systems Evaluations

(L.1.) Invalid Return Detection
The objective of this evaluation is to look for large geographic areas that may have high rates of multiple Be Counted forms and forms completed by telephone per housing unit.

(L.2) Decennial Response File Stage 2 Linking and Setting of Expected Household Population
This evaluation will document how frequently census forms were linked during the Decennial Response File processing and the types of linkages that were constructed. It will also assess the accuracy of the automated process for setting the expected household size and its effects on the census population.

(L.3.a) Analysis of Primary Selection Algorithm Results (Operational Assessment)
The objective of this evaluation is to document the effects of using the Primary Selection Algorithm in resolving situations when multiple household questionnaires are received for the same address.

(L.3.b) Resolution of Multiple Census Returns Using Reinterview
The objective of this evaluation is to determine the accuracy of Primary Selection Algorithm rules for determining the Census Day residents for an address. The data will be collected using a reinterview of a sample of respondents.

(L.4) Census Unedited File Creation
This evaluation documents the results of the process of determining the final housing unit inventory. The final housing unit inventory for the census is determined during the process of creating the Census Unedited Detail File. The final housing unit inventory is created by merging information on the processed Decennial Response File with the information on the Decennial Master Address File.

(L.5) Beta Site
This evaluation will answer questions about how well the Beta Site integrated the data collection systems, and its overall utility for software testing and release.

M: Quality Assurance Evaluations

Overview

For Census 2000, the overall objective of the Quality Assurance (QA) program is to assist in producing deliverables or outputs which meet the Bureau's quality requirements. The QA program will identify when major inputs such as people, material, machinery, software, etc. do not meet quality requirements. The QA data will provide managers, supervisors or employees with information to make necessary adjustments and improvements to the system. At the end of the operational task, QA will identify and correct clusters of outputs which contain a significant number of errors.

What Will We Learn?

The QA evaluations will provide information to help determine if the QA approach used in Census 2000 is the right approach in a census environment, whether the QA operation improved the overall quality of the census, how effectively it was implemented, and how it might be improved. For example, the results of the first study will help us determine if different QA approaches should be explored for census use. For the second study, the effectiveness of variables that are used to detect enumerator falsification will be measured, and appropriate variables will be added and/or deleted from the detection process.

Quality Assurance Evaluations

(M.1) Assessment of the Quality Assurance Approach During Census 2000
For Census 2000, the overall objective of the Quality Assurance (QA) program is to assist in producing deliverables or outputs that meet the Census Bureau's quality requirements. To achieve this objective, the QA program, whenever possible, focuses on three main concepts or philosophies: prevention, improvement, and protection. The goals of this study are to document operational experiences with this approach in Census 2000; measure quality levels that were achieved, and determine if other approaches should be explored for Census 2010.

(M.2) Effectiveness of Existing Variables in the Model Used to Detect Fabrication During Reinterview, and the Identification of New Variables
The reinterview program is a quality assurance operation whose major objective is to detect enumerators who may have falsified data. This evaluation examines variables used in the fabrication model to determine if they were effective in detecting fraud; whether other variables should be added to the model; and to provide suggestions on other ways to improve this program.

N: Accuracy and Coverage Evaluation Survey Operations

Overview

The Census Bureau will conduct the Accuracy and Coverage Evaluation (A.C.E.), a nationwide sample survey, to determine the number of people and housing units missed or incorrectly counted in the census. The basic approach is to independently relist a sample of blocks, re-enumerate them during the A.C.E. survey, and then compare the results to the census data for the same blocks. The Census Bureau will use the results of the A.C.E. to correct the census counts obtained through the preceding enumeration procedures.

The studies in this category will measure how well the Census Bureau carried out different components of the A.C.E. For instance, analysis projects and evaluations will be conducted that measure the completeness of the housing unit lists used for A.C.E. interviewing, the quality of the A.C.E. person interviewing process, and the accuracy of the procedures used to match persons counted during the A.C.E. interview to those that were enumerated in the census. The success of each A.C.E. component affects the quality of the final estimates.

What Will We Learn?

The results of these A.C.E. analysis projects and evaluations will help the Census Bureau to document this coverage measurement operation and improve its procedures. For example, we will learn whether match rates were different in relisted blocks. An examination of laptop computers used during person interviews will identify errors encountered by interviewers and also will provided suggestions for how to improve the computer assisted instrument in the future. Other studies will determine how well we detect interviewer fabrication, while also looking at its effect on A.C.E.

These operational analyses and evaluations will document the A.C.E. process and give the Census Bureau greater insight into what causes error in the measurement of coverage error. Some causes of error are attributable to census questionnaire data capture. Moreover, matching errors may add to errors in the estimates of census coverage. One evaluation in this category will examine a subsample of rematched A.C.E. blocks to measure matching errors. We also will measure the effect of matching error on Dual System Estimates and undercount rates.

The evaluations in this category will help the Census Bureau to identify operational causes of error in measuring coverage and will help to minimize them when planning future censuses.

A.C.E. Survey Operations Evaluations

(N.1) Contamination of Census Data Collected in A.C.E. Blocks
This evaluation examines whether census and A.C.E. operations were kept operationally independent (a key requirement for avoiding bias in the dual-system estimates of coverage error) by comparing census results in A.C.E. and non-A.C.E. clusters.

(N.2) Analysis of Listing Future Construction and Multi-Units in Special Places
A new procedure during block relisting for A.C.E. 2000 is to include housing units under construction, and multi-unit structures within special places. We will study the effectiveness of the listing, the number of units added to A.C.E., and match rates to the census.

(N.3) Analysis of Relisted Blocks
The A.C.E. address listing operation is reviewed for high concentrations of geocoding errors, and blocks with too many errors are relisted. This analysis will examine the relisted blocks to document their characteristics and the results of matching to census listings.

(N.4) Analysis of Blocks With No Housing Unit Matching
For most blocks in the A.C.E., the matching is done in two steps: first the addresses are matched to the census address list, and then the persons are compared to the census questionnaires for matching addresses. The housing unit match step is not done for relisted blocks (see N.3) or for blocks in list/enumerate areas. The purpose of this study is to determine how this affects the person matching process.

(N.5) Analysis of Blocks Sent Directly for Housing Units Followup
The A.C.E. addresses first are computer matched to the addresses on the January 2000 version of the Decennial Master Address File and then undergo a clerical matching operation. Some blocks are sent directly for a follow-up interview when there is little perceived benefit to clerical matching. While this allows the field follow-up to begin earlier, it also may reduce the ability of that operation to resolve the match status of these units. This study will examine the effectiveness of this strategy by comparing match rates and unresolved rates for these blocks to those for blocks that did undergo clerical matching.

(N.6) Analysis of Person Interview With Unresolved Housing Unit Status
This analysis examines whether housing units with an unresolved status after the initial housing unit match are eventually resolved during the person interview and final housing unit match operations.

(N.7) Analysis on the Effects of Census Questionnaire Data Capture in A.C.E.
During the A.C.E. person matching, data capture *images* of census questionnaires are examined when the initial match results indicate the census data are insufficient for matching. This study will document how often this occurred and the effects on final match codes.

(N.8) Analysis of the Census Residence Questions Used
During A.C.E. interviewing and field followup for non-matches, persons are asked about their Census Day address and the results are used to determine where they should have been counted in the census. Persons counted at the wrong address then are classified as erroneous enumerations by the census. This study will examine the responses to these residency questions during A.C.E. and document how they affected the estimates of erroneous enumerations.

(N.9) Analysis of the Person Interview Process
This study examines the overall interviewing process. The analysis will include topics such as Computer Assisted Personal Interview instrument (i.e., laptop) performance, detection of interviewer falsification by the interviewing quality assurance process, and the impact of allowing interviewers to modify address information during matching operations.

(N.10) Falsification in A.C.E.
This evaluation examines how well the quality assurance process identified interviewers who entered false data in the A.C.E. interview and the impact of undetected false data on A.C.E. estimates.

(N.11) Extended Roster Analysis
During the census, an extended roster is used to capture names (but not data)for people in large households. A follow-up operation then collects demographic data for these people. If the follow-up does not collect the data, it will not be possible to match A.C.E. data for these people to census results and sending this case to A.C.E. field followup is pointless. This study will document how reviewing these extended rosters affected the A.C.E. person matching and followup operations.

(N.12) Matching Stages Analysis
The person matching is conducted first by computer and then undergoes three levels of clerical matching by clerks, technicians, and analysts. The goal of this analysis is to document the differences in the match codes assigned by these four different operations.

(N.13) Analysis of Unresolved Codes in Person Matching
Results from the Census 2000 Dress Rehearsal Person Followup interview indicated that there were a large volume of cases coded as unresolved when the interview was conducted with a proxy respondent. In general, proxies were able to answer whether a household lived at a given address on Census Day, but answered "Don't Know" to questions regarding a household being in a group quarter and/or having a ususal home elsewhere, which resulted in an unresolved code. The goal of this analysis is to document the coding results for specified patterns of "Don't Know" answers from proxy respondents. The coding results of proxy respondents will then be compared to those with similar patterns of answers from actual (i.e., non-proxy) respondents.

(N.14) Evaluation of Matching Error
A potential source of error in the coverage estimates are the matching operations used to classify persons as missed or erroneously enumerated in the census. This evaluation will determine the relative error associated with the matching operations and how matching error affects the Dual System Estimates.

(N.15) Outlier Analysis in the 2000 A.C.E.
In 1990, an outlier review was conducted in 104 of the blocks that contributed most to the net undercount. This review was conducted in 1991 after all operations were completed. In 2000, the outlier review is planned to be conducted before the matching is completed. Blocks will be selected for an indepth review by the analysts. Matching errors will be corrected, if they exist. In addition, the analysts will document the results of their investigation. This project will document the outlier review.

(N.16) Impact of Targeted Extended Search
This evaluation has two main purposes. The first study looks at the nature and extent of errors resulting from limiting the search area to one ring of blocks (around the sample block clusters). This is accomplished by looking at the effect of this Targeted Extended Search on the Dual System Estimates and variances for the evaluation post strata, as well as data from the production matching operation. The second study evaluates potential gains from adding a second ring of blocks to the search and match operation.

(N.17) Targeted Extended Search Block Cluster Analysis
In 1990, the search area for matching was extended to surrounding blocks for all clusters. In 2000, this only will be done for clusters deemed most likely to benefit this additional searching. This study will document the characteristics of such blocks and the effects of this strategy on final match rates.

(N.18) Effect of Late Census Data on Final Estimates
The aim of this evaluation is to determine the effect on A.C.E. adjustment factors of ignoring the small amount of census data collected after late September.

(N.19) Housing Unit and Person Coverage Analysis
This analysis provides an overall assessment of the quality of housing unit and person coverage A.C.E. operations. Some of the topics addressed in the analysis are quality of A.C.E. listing, effect of housing unit followup interviewing on the enhanced list, effectiveness of housing unit and person followup quality assurance, and noninterview rates.

(N.20) Group Quarters Analysis
In 1990, sample interviews were conducted in noninstitutional and nonmilitary group quarters. In 2000, A.C.E. sample interviews will not be conducted in any group quarters. The A.C.E. sample interviews will only be conducted in housing units, but sometimes it is difficult to determine if a place is a housing unit or a group quarters. The A.C.E. sample nonmatches in whole households will be compared to the group quarters enumerations in the census. The

purpose of this analysis is to document these matching results, identify post-strata containing a large number of group quarters, and to examine whether definitional problems led to group quarters classification errors.

(N.21) Analysis of Mobile Homes

Mobile homes were missed at a higher rate in 1990 than single family homes. Mobile homes are not a housing category for Census 2000, but they will be identified in the A.C.E. sample. This study will document the nonmatch rates for mobile homes and people in mobile homes from the A.C.E. sample matching results.

This page intentionally left blank.

O: Coverage Evaluations of the Census and of the Accuracy and Coverage Evaluation Survey

Overview

The studies in this category include a group evaluating A.C.E. coverage and a group evaluating census coverage. These studies will identify person and housing unit characteristics that are related to being missed or erroneously enumerated. Analysis in this area will also study the quality of data from proxy respondents, and the frequency and patterns of geocoding error. Furthermore, census counts and dual system estimates will be compared to demographic benchmarks to evaluate accuracy and completeness.

What Will We Learn?

Results from these evaluations will allow us to determine how complete our Master Address File was for Census 2000. Net coverage rates of housing units will be computed at the national and subnational levels along with gross omission and erroneous enumeration rates. Other studies will explain factors that contribute to housing unit coverage error. For example, we will learn whether type of address (city style versus noncity style) has an effect on housing unit coverage. In addition, there will be a study of housing unit duplication, to identify characteristics of duplicate units and their operational source.

Similarly, we will identify factors that contribute to person coverage error. For instance, studies will examine how nonmatch rates are affected by type of enumeration area (e.g., mailout/mailback, update/leave) and characteristics of blocks, households, and people. We will acquire knowledge about erroneous enumerations by determining which demographic, housing unit type, and type of enumeration variables are associated with them. Furthermore, we will conduct an analysis of measurement error, which will help us determine why people are erroneously listed in the census and the Accuracy and Coverage Evaluation.

(O.1) Type of Enumeration Area Summary
The census is conducted differently in the different types of enumeration areas. This project will document nonmatch rates and erroneous enumeration rates in the different areas. Geocoding error in the different enumeration areas also will be documented.

(O.2) Coverage of Housing Units in the Early Decennial Master Address File
The initial housing unit matching is done between the A.C.E. housing units and the housing units on the January 2000 version of the Decennial Master Address File (DMAF). This matching is conducted to link the ACE and census housing units for later processing. If an A.C.E. housing unit is linked to a census housing unit, the telephone number from a mail return questionnaire can be used to start person interviewing by phone. The results of the matching will allow an early look at the quality of the DMAF in January 2000.

(O.3) Housing Unit Coverage on the Master Address File
This evaluation assesses 1) the net coverage rate of housing units, 2) the gross omission rate of housing units, and 3) the erroneous enumeration rate of housing units. These assessments are made at the national level, smaller geographic levels, and for each post-strata. This evaluation also examines the potential impact on housing unit coverage had we excluded specific Master Address File building operations. This study is similar to the Housing Unit Coverage Study conducted in 1990.

(O.4) Analysis of Conflicting Households
During A.C.E. housing unit matching, situations are found where the census and A.C.E. listed two entirely different families. This study will document the follow-up interviewing results for these households to determine if the census was in error, the A.C.E. was in error, if the two families both live at the address, if there was misdelivery of the census form, and so on.

(O.5) Analysis of Proxy Data in the A.C.E. and in the Census
Both the census and A.C.E. sometimes must collect data from proxy respondents--persons who are not members of the household where data are needed. This study will examine match rates and erroneous enumeration rates for such cases in both the census and the A.C.E.

(O.6) P-Sample Nonmatches Analysis
This study will examine nonmatch rates for the post-strata used to form final dual-system estimates of census coverage errors. It also will examine these rates for other variables not used to form post-strata.

(O.7) Analysis of Person Coverage in Puerto Rico
The measurement of person coverage, and evaluation studies of that measurement, will be done separately for Puerto Rico. This study will document those findings and compare the results to those from the A.C.E.

(O.8) Analysis of Housing Unit Coverage in Puerto Rico
The measurement of housing coverage, and evaluation studies of that measurement, will be done separately for Puerto Rico. This study will document those findings and compare the results to those from the A.C.E.

(O.9) Geocoding Error Analysis
A housing unit and its occupants are classified as geocoding errors by the census if that housing unit is enumerated within the ACE search area and should not have been. This study will examine the frequency of geocoding error and will identify operations more prone to making such errors.

(O.10) Housing Unit Duplication in the 2000 Census
Duplication in the census is one type of erroneous enumeration. This analysis will identify duplicate housing units in Census 2000 and their characteristics. The study will also determine if duplication is more likely for one group or another (e.g. owners vs. renters). The census operations most likely to produce housing unit duplication will be identified, along with the most plausible sources of duplication.

(O.11) E-Sample Erroneous Enumeration Analysis
This study will examine erroneous enumeration rates for the post-strata used to form final dual-system estimates of census coverage errors. It also will examine these rates for other variables not used to form post-strata.

(O.12) Analysis of Nonmatches and Erroneous Enumerations Using Logistic Regression
This project looks at logistic regression as a tool to analyze the A.C.E. data. This purpose of this analysis is to build logistic regression models that relate demographic, housing unit type, and type-of-enumeration variables, to census nonmatches, A.C.E. nonmatches, and erroneous enumerations.

(O.13) Analysis of Person and Housing Unit Data Combined
For some housing units, the A.C.E. results will identify both missed and erroneously enumerated persons. This study will examine the person and housing characteristics of such cases.

(O.14) Analysis of Measurement Error
Measurement error is the term used for error in surveys due to an inability to collect the correct answer to a question. Measurement error can be attributable to the interviewer, the respondent, or the data collection instrument. The main question of this evaluation is whether or not the errors in A.C.E. residency status codes and person match codes had a significant effect on the Dual System Estimates. Other topics examined by this evaluation are the magnitude of the error attributable to the respondents or interviewers in A.C.E. Person Interview (and the A.C.E. Person Followup Interview), and the characteristics of people whose person match and residency codes were found to be incorrect in A.C.E.

(O.15) Impact of Housing Unit Coverage on Person Coverage Analysis
This analysis will include an examination of the effect that census housing unit updating operations (e.g., postal check, Local Update of Census Addresses) have on person coverage. The study also will identify characteristics of persons who were missed or erroneously enumerated due to housing unit errors, and compare them to the characteristics of those who were missed or erroneously enumerated for other reasons.

(O.16) Person Duplication in the 2000 Census
People are duplicated in the census for many different reasons. This analysis will identify the number and characteristic of duplicate persons in the 2000 Census. The study will also determine if duplication is more likely for one group or another (e.g., owners/renters). The census operations most likely to cause duplication will be identified, along with the most plausible sources of the duplication.

(O.17) Analysis of the 0-17 Age/Sex Post-Strata
Children have historically been disproportionately undercounted in the census. This study will examine coverage errors for children. It also will examine how often we encounter A.C.E. households in which all of the residents are under 16 years of age. These cases are treated as non-interviews and thus can affect the final coverage estimates for persons age 0-17. The characteristics of the people in these types of households and the household composition will be documented.

(O.18) Synthesis of What We Know About Missed Census People
The purpose of this study is to summarize and synthesize findings from A.C.E. and other sources about the causes and characteristics of persons missed or erroneously enumerated in the census.

(O.19) Analysis of Coverage of Housing Units in the Early Decennial Master Address File and Subsequent Census Coverage Improvement
The goal of this study is to assess the completeness of housing unit coverage on the early Decennial Master Address File (DMAF). We will determine which census operations contributed to undercoverage by deleting units that should have not been deleted, and which operations improved coverage by adding units not previously accounted for. We also will identify which census operations reduced housing unit duplication

(O.20) Consistency of Census Estimates with Demographic Benchmarks
This study uses independent demographic benchmarks to evaluate the accuracy of the Census 2000 counts and the completeness of coverage in Census 2000. While this approach cannot produce estimates for as many demographic groups and geographic areas as A.C.E., results can be compared to A.C.E. at aggregate levels.

(O.21) Implications of Net Census Undercount on Demographic Measures and Program Uses

This evaluation will address the effect of net census undercount on demographic measures such as growth rates, race, age and sex composition, and vital statistics rates used by a variety of programs and data users.

This page intentionally left blank.

P: Accuracy and Coverage Evaluation Survey Statistical Design and Estimation

Overview

The evaluations in this category examine the quality of Accuracy and Coverage Evaluation (A.C.E.) estimates. Analyses in this area will address the quality of Dual System Estimates (DSE) by examining estimates of variances and coefficients of variation. We will also analyze missing data procedures, compare A.C.E. results to various quality measures, and conduct a study of total error in A.C.E.

What Will We learn?

We will gain knowledge about the quality of A.C.E. estimates from the total error analysis, which will examine model and measurement error in the empirical DSE. We will also learn about the overall quality of A.C.E. by comparing its results to a synthesis of quality measurements from various coverage measurement evaluations and operational analyses. We will examine characteristics associated with missing data and the bias and uncertainty associated with the missing data procedures.

A.C.E. Survey Statistical Design and Estimation Evaluations

(P.1) Measurement of Bias and Uncertainty Associated With Application of the Missing Data Procedures
The purpose of this evaluation is to obtain a measure of the error (and sensitivity to assumptions) of missing data procedures used for the dual system estimates resulting from the A.C.E.

(P.2) Synthetic Design Research/Correlation Bias
Synthetic estimation uses a statistical model to modify coverage estimates for a particular post-strata using information from sample units outside the geographic area of interest. Because this can introduce bias, the accuracy of the Dual System Estimate (DSE) depends on the validity of the model and whether the assumptions of the synthetic model are satisfied. The purpose of this study is to determine if the assumptions were satisfied for Census 2000 A.C.E. post-strata, and to measure the effects of any biases on the DSEs.

(P.3) Variance of Dual System Estimates and Adjustment Factors
This study is designed to evaluate the quality of the Dual System Estimates (DSE) and adjustment factors by examining estimates of variances and coefficients of variation (CVs). We will compare the 2000 DSE variance estimates and CVs to the 1990 DSE variance estimates and CVs at the national-level and for various demographic variables.

(P.4) Overall Measures of A.C.E. Quality
The aim of this study is to synthesize quality measures from various coverage measurement evaluations and operational analyses to assess the overall quality of the 2000 A.C.E. Measures such as response rates, imputation rates, match rates, and correct enumeration rates by various demographic and geographic groups will be examined and, where possible, compared to the 1990 census rates.

(P.5) Total Error Analysis
The total error analysis will examine model and measurement error in the empirical Dual System Estimates. For each evaluation post-stratum, we will estimate the bias and variance in the net undercount rate for each type of nonsampling error and estimate the overall bias and variance in the net undercount rate.

Q: Organization, Budget, and Management Information System

Overview

Research in this category will document headquarters decision making processes and the impact of headquarters organizational structure on the decennial census. We plan to study the effectiveness of the Management Information System (MIS) and the Cost Model. The MIS is a data warehouse of cost and progress information for census operations that also includes an activity schedule for the decennial census. The cost model is used to formulate budgets for operations, allocate funds, and to assist in planning future census operations. Furthermore, we will conduct an evaluation that examines the role of contracting in carrying out Census 2000.

What Will We Learn?

The findings from these studies will help the Census Bureau to better manage future censuses and similar projects. These studies will document how well the MIS and the Cost Model worked in helping us manage Census 2000. We will compare the activities and recommendations of the 2000 research and development program to what was actually implemented for Census 2000 to determine which projects were most beneficial. In addition, we will examine the roles and influences of both external and internal entities on planning and implementing the census. Some of the groups that will be studied include various advisory committees, Office of the Inspector General, Congress, General Accounting Office, Census Bureau Monitoring Board, Census Bureau Executive Staff, and the Department of Commerce. Other research in this category will give us insight into the effectiveness of hiring contractors to help conduct Census 2000. We will learn how cost effective our contracts were and whether contractors are bringing in the expertise needed by the Census Bureau. We will also address whether the Census Bureau is losing "corporate knowledge" by giving contractors a major role in conducting Census 2000.

Organization, Budget, and Management Information System Evaluations

(Q.1) Management Processes and Systems of the 2000 Decennial Census
The purpose of this study is to determine how well various processes and systems worked for managing Census 2000. This analysis will include an evaluation of the Management Information System and the Cost Model. The effectiveness of decision making groups/processes (e.g., Census Operational Managers and decision memos) will be assessed. This study will also look at the organizational structure, roles, and influences of entities such as the Census Bureau Executive Staff, Department of Commerce, Inspector General, advisory committees, General Accounting Office, Census Bureau Monitoring Board, and Congress. Furthermore, a comparative study of management models will be conducted.

(Q.2) Effectiveness of the Contracting in Carrying Out the 2000 Decennial Census
The goal of this evaluation is to measure the effectiveness of contracting in Census 2000. We will look at the cost effectiveness of our contracts along with whether contractors are bringing in the expertise that is needed by the Census Bureau. This study will examine how well we managed our contracts and will determine if we are losing "corporate knowledge" by giving contractors a great deal of responsibility in this decennial census.

R: Automation of Census Processes

Overview

These studies will examine many of the major automated systems designed to support Census 2000. In general, we will assess whether the right requirements were defined for each of the systems and use this information to guide improvements needed for future censuses and surveys. The systems to be studied include:

- Telephone Questionnaire Assistance
- Internet Questionnaire Assistance
- Internet Data Collection System
- Operational Control System 2000 (OCS2000) System
- Preappointment Management System/Automated Decennial Administrative Management System (PAMS/ADAMS)
- American Fact Finder
- Data Capture
- Matching, Review, and Coding System
- Accuracy and Coverage Evaluation Survey 2000 System

What Will We Learn?

In addition to examining systems requirements, we will also assess other factors such as reliability and functionality, maintenance and security needs, and respondent acceptance. A common protocol will be designed to include general questions for the selected automated systems and debriefings will take place. We will also identify questions and concerns unique to specific systems.

Note: The plans for this evaluations category are currently being finalized. The specific evaluations will be identified in an updated version of this document.

March 29, 2000 DRAFT - Revised

Census 2000 Operational Analyses for Sampling and Estimation

A Brief Overview

Topics covered:

1 A.C.E. Sample Design
2 Long Form Sampling
3 Service Based Enumeration (SBE) Estimation
4 P- and E-Sample Missing Data
5 Dual System Estimation (DSE)
6 A.C.E. Variance Estimation
7 Long Form Estimation/Variance Estimation
8 A.C.E. Weight Trimming
9 Generalized Variances
10 Block Level Estimation
11 Unclassified Estimation
12 Housing Unit (HU) Dual System Estimation (DSE)
13 A.C.E. Targeted Extended Search

C **A.C.E. Sample Design**

The A.C.E. is a multi-stage sample design. Initially, the original ICM sample of 750,000 housing units will be selected and be sent to the field to be independently listed. This listing sample is a state-based design and will include an oversample of large block clusters, a separate sample of small block clusters in each state. In addition, a separate sample of American Indian Reservation clusters will be selected. After the sample is listed, the A.C.E. sample cluster reduction and the small block subsampling will be done. After the initial housing unit matching operation and followup, the large block subsampling will be done to reach the final A.C.E. interview sample size. The E-sample will generally overlap with the P-sample.

Questions to answer:

- What are the sample sizes in terms of clusters and expected housing units?
- What are the unbiased sampling weights after each stage of selection?
- What are the weighted population distributions by State?
- What are the sample sizes (clusters and housing units) by TEA and ACERO?
- What's the average number of interviewed A.C.E. housing units in a cluster?
- What's the sample size (clusters and housing units) by stratum before and after sampling?

Processes included:

Initial Sampling
A.C.E. Reduction
Small block subsampling
Large block subsampling
E-sample Identification

2. **Long Form Sampling**

Long form sampling uses four sampling rates based on the size (MOS) of the long form sampling entity (LFSE). MOS is an estimate of occupied housing units. Long form sampling entities are geographic and statistical areas eligible for the sampling. Sampling rates are assigned to the collection blocks as follows:

- All blocks in Puerto Rico 1-in-6
- 1-in-2 for governmental units if estimated MOS is less than 800.
- 1-in-4 for government units if MOS is 800 or more but less than 1200.
- 1-in-8 for census tracts with MOS $\geq$ 2000.
- 1-in-6 for all remaining blocks in tract with MOS less than 2000.

Questions to answer:

- What are the sizes of the universe by sampling rates?
- What are sample sizes by type of enumeration area?
- What are the response rates for long form at different geographic level?
- What are the response rates for long form by selected household type?

3. **Service Based Enumeration (SBE) Estimation**

SBE is a fundamentally different approach to counting persons without a usual home than was used in the 1990 Census. SBE counts people at facilities such as shelters, soup kitchens, mobile food vans and certain outdoor locations. SBE estimation uses the multiplicity estimator methodology for the estimation of people with no usual residence who use SBE facilities. The multiplicity estimator depends on the service usage question for shelters, soup kitchens, and mobile food vans. SBE multiplicity estimation was done in the Census 2000 Dress Rehearsal. It will be used for Census 2000 for the redistricting file but not for the apportionment counts.

We will use multiplicity estimation to estimate the number of persons that use SBE facilities and do not have a usual home.

Questions to answer:

- How many persons were enumerated as part of SBE on Be Counted Forms (BCFs) and at shelters, soup kitchens and mobile food vans, and Targeted Non-Shelter Outdoor Locations (TNSOLs) after unduplication (i.e., apportionment count)?

- What is the distribution of response and nonresponse to the usage questions by nonresponse adjustment strata? What is the average respondent usage? (Separate for shelters and soup kitchens).

- For soup kitchens, what is the distribution of the response to the shelter usage question?

- What is the multiplicity estimate and apportionment count by age/sex group? How many persons were added by the adjustment for casual users (if we use the adjustment)?

- How many replications were omitted to account for soup kitchen respondents who also used shelters?

- What are the answers to the above questions by county and state?

- For how many counties was it not appropriate to use the multiplicity estimator?

4. **P- and E-Sample Missing Data**

There are several types of missing data known to affect the Accuracy and Coverage Evaluation. These include noninterviewed households, item missing data, and unknown status for variables such as residence, match, and correct enumeration.

To compensate for missing data we have 3 procedures in place. First, a noninterview adjustment that compensates households that could not be reached for an A.C.E. interview. Second, a characteristic imputation to fill in values for person characteristics that are missing. Finally, an procedure to estimate the probability of match, residence, or correct enumeration for those persons for whom we do not have an exact figure.

Questions to answer:

- Level and degree of household nonresponse? by State? by ACERO?
- Distribution of sample sizes and adjustment factors by noninterview adjustment cell?
- Summary of sample sizes and estimated probabilities for match, enumeration and residence cells?
- Summary at the post-stratum level of effects of each individual missing data adjustment.
- Level and degree of item missing data? by State? by ACERO?
- Level and degree of missing status? by State? by ACERO?
- Distributions of imputed and non-imputed characteristics by imputation categories?
- Number and percent imputed for certain race and ethnic characteristics?

5. **Dual System Estimation (DSE)**

DSE was used for the 1990 Census Post-Enumeration survey, the 1995 and 1996 Census Tests, and the Census 2000 Dress Rehearsal. For the Dress Rehearsal, DSE was used in conjunction with raking. For Census 2000 we will use DSE for the Accuracy and Coverage Evaluation (A.C.E.); but there will be no raking.

We will use DSE as part of the Census 2000 A.C.E., to adjust for the undercount for the redistricting file. DSE will not be used for the apportionment counts.

Questions to answer:

- For each A.C.E. post-stratum what are the weighted and unweighted components of the DSE (census count, insufficient information (II), E-sample total, correct enumerations, non-mover matches and total non-movers, out-mover matches and total out-movers, total in-movers)? What are these summary statistics after collapsing age/sex? For race/origin groups? Owners and renters? High and low mail return rate? MSA/TEA group?

- What is the coverage correction factor, undercount rate and undercount by post-stratum?

- Was any collapsing of the 448 post-strata necessary due to sample size? What was the sample size (P-sample and E-sample) by post-strata?

- What is the consistency between the E-sample and P-sample responses for each of the post-stratification variables for not imputed and total persons?

6. **A.C.E. Variance Estimation**

For the 2000 Census, the methodology that will be used to calculate the variances will be a stratified jacknife. This methodology will reflect the double sampling, TES, missing data and DSE.

The variance estimation summaries will focus on the different components of the variance estimate. We will conduct a quantitative analysis of the contribution of each variance component (such as the variance that results from imputation of missing data, i.e. , correct enumeration probability and p-sample matching probability) to the overall variance estimate of the dual system estimate.

Questions to answer:

- How do observed coefficients of variation compare to expected coefficients of variation for selected demographic groups? by post-stratum? By State? By Congressional District?

- What=s the contribution of imputing match probability to the overall variance estimates? By post-strata?

- How do the erroneous enumerations contribute to the overall variance estimate? By post-strata?

- How do the various stages of sampling contribute to the overall variance estimate? By post-strata?

7. **Long Form Estimation/ Variance Estimation**

Long form was used in the1990 Census. However, the long form was not used in 1995 and 1996 Census tests. The long form was used in Census 2000 dress rehearsal. For dress rehearsal estimation is used as an operational test and will not produce official long form data estimates. The long form estimation will use a weighting approach including a raking methodology for the dress rehearsal.

Long form estimation for Census 2000 will use a weighting approach which will utilize raking methodology. Raking will use marginal controls from the census data which are not corrected for coverage error. Then coverage correction factors will be applied at the person level to the results of raking.

We will review operational tallies which provide on overview of the results from implementation of the weighting and variance estimation.

Questions to answer:

- What are the estimates of householders before applying coverage correction factors?
- What estimates of householders after applying coverage correction factors?
- What are actual variances for long form estimates at different levels of geography?
- What are the generalized variances for long form entities at different level of geography?

8. **A.C.E. Weight Trimming**

Due to large block and small block cluster subsampling and oversampling of difficult to enumerate blocks and inconsistent blocks, there will be variation in the unbiased weights for the A.C.E. sample design. If this variation is too large, some weight trimming may be implemented. For the Dress Rehearsal, a small amount of weight truncation was implemented with no increase in other weights to account for the truncation (not necessary for dual system estimation). The A.C.E. design is not finalized. However, some weight trimming (or truncation) may be needed. Research continues to determine how to truncate weights if needed.

Questions to answer:

- How do the dual system estimates and variances compare using trimmed weights and not trimming the weights?
- What are the weight distributions by trimming Cells before and after trimming?
- What are the weighted non-matches and non-match rates before and after trimming by trimming Cells? and by Region/MSA/TEA areas?
- What are the weighted erroneous enumerations and erroneous enumeration rates before and after trimming by trimming Cells? and by Region/MSA/TEA areas?
- What are the weighted E-sample and P-sample person estimates before and after trimming by State?
- What are the E-sample and P-sample American Indians on AIR estimates before and after trimming.

9. **Generalized Variances**

For small geographic areas such as blocks, direct variance estimates are smoothed by fitting a standard GATT curve. The parameters from these models are provided to data users for calculation of standard errors. The methodology is equivalent to what was used in dress rehearsal.

The generalized variance summaries for A.C.E. data will focus on how well the generalized variance function model approximated the estimated variances at various geographic levels and characteristic estimates.

The generalized variance research for Long Form will focus on use of the design factor to approximate the variances of sample estimates at various levels of geography, particularly census tracts and block groups.

Questions to answer:

- How well the fit of the weighted GATT Curve model approximate the estimated variances at all levels of geography?
- Which level of geography of the data, that is fit into the model, estimates the other levels of geography the best?
- How well the generalized design factors approximate the estimated variances?
- Differences in the design factors for metropolitan statistical area (MSA) versus non MSA?
- Summary of effect of removal of outliers?

10. **Block Level Estimation**

Block level estimation for the 1990 Census Post-Enumeration survey and the Census 2000 Dress Rehearsal assumed that within poststrata the estimated coverage factor applied to all small areas (synthetic assumption). Controlled rounding was used to create whole person records in order to correct for undercount or overcount.

We will use the synthetic assumption and controlled rounding for block level estimation for the Census 2000 Accuracy and Coverage Evaluation (A.C.E.).

Questions to answer:

- What are the rounded and unrounded counts by post-stratum for each block (prepare a block file)?

- Using this file what are the rounded and unrounded counts by post-stratum at the county, state, and national level?

- At the county, state, post-stratum, and national level what are the number of undercount or over count persons records created by synthetic estimation?

- What is the distribution of the relative and absolute effect of synthetic estimation on block totals?

11. **Unclassified Estimation**

Unclassified units are housing units with unknown status (occupied, vacant or nonexistent) and occupied housing units with unknown population count. The missing status or population count must be estimated prior to Population Division=s Edit & Imputation. For the Census 2000 Dress Rehearsal unclassified estimation was part of the estimation for Nonresponse follow-up and Undeliverable as Addresses Vacant sampling. Since this sampling will not be part of Census 2000, a separate estimation for unclassified housing units is necessary.

The nearest-neighbor hot deck imputation method will be used for Unclassified Estimation for Census 2000.

Questions to answer:

- What is the distribution of classified and unclassified units by occupied, vacant , and delete after the completion of unclassified estimation? By LCO, County, and State? By donor/donee group?

- What is the average household size for classified and unclassified occupied units after the completion of unclassified estimation? By LCO, County, and State? By donor/donee group?

- What percentage of housing units are classified and unclassified prior to unclassified estimation? By LCO, County, and State? By donor/donee group?

- Based on the A.C.E. post-stratum of the head of household what are the number of occupied classified units and their average household size by post-stratum? By A.C.E. block clusters and non A.C.E. block cluster?

- Based on the A.C.E. post-stratum of the head of household what are the number of occupied unclassified units and their average household size after unclassified estimation by post-stratum?

- By A.C.E. block clusters and non A.C.E. block cluster?

- What are the number of persons in classified and unclassified (imputed by unclassified estimation) by A.C.E. post-stratum. By A.C.E. block clusters and non A.C.E. block cluster?

12. **Housing Unit (HU) Dual System Estimation (DSE)**

HU DSE was used to support evaluation of the 1990 Census using data from the 1990 Census Post-Enumeration Survey. For the Census 2000 Dress Rehearsal, HU DSE was planned to support housing unit long form weighting as well as for evaluation. However, HU DSE was canceled for the Dress Rehearsal.

We will use HU DSE as part of the Census 2000 Accuracy and Coverage Evaluation (A.C.E.) to support housing unit long form weighting and to evaluate HU coverage.

Questions to answer:

- For each A.C.E. housing unit post-stratum what are the weighted and unweighted components of the DSE (census count, E-sample total, correct enumerations, matches and total)? What are the resulting, erroneous enumeration, and match rates by post-stratum?

- What is the coverage correction factor, undercount rate and undercount by post-stratum?

- What is the consistency between the E-sample and P-sample responses for each of the post-stratification variables?

13. **A.C.E. Targeted Extended Search**

The Census 2000 A.C.E. will implement a targeted surrounding block search operation, hereinafter, Targeted Extended Search or simply TES. The rationale is to develop a criteria by which to identify block clusters that will benefit the most from surrounding block search. In 1990, the majority of the matches found in surrounding blocks were in addresses that were incorrectly geocoded in a surrounding block. Thus, the targeting criterion or criteria will be based on information or factors which give evidence of census geocoding error. For instance; block clusters with high rates of A.C.E. housing units without a census match (coded CI) and/or census geocoding error (coded GE) will be included in the TES. The goal of this operation is to increase both the matching and correct enumeration rates compared to a design that limits the search to the A.C.E. block cluster.

Questions to be answered:

- Summary of TES block clusters by certainty strata, TES sampling universe and out of scope?

- Summary of TES block clusters by State and A.C.E sampling stratum?

- What % of gecoding and P-Sample nonmatches are accounted for in the TES blocks based on initial housing unit information? Include distribution of gecoding and nonmatches for A.C.E. block clusters.

- What=s the effect of the TES on the DSE for total population? By poststrata? By race/Hispanic Origin group?

- What is the effect of TES on each Post-stratum=s match and correct enumeration rate?

- What=s the effect of the TES operation on the reliability of the DSE=s? By poststrata? By race/Hispanic Origin group?

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 6

March 22, 2000

Prepared by: Maria Urrutia and Annette Quinlan

The sixth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on March 22, 2000 at 10:30. The agenda for the meeting was an overview of A.C.E. 2000 Evaluations.

Persons in attendance:

Kenneth Prewitt
William Barron
Nancy Potok
Paula Schneider
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
Susan Miskura
Tommy Wright
Donna Kostanich
Louisa Miller
Raj Singh
Rita Petroni
Carolee Bush
Maria Urrutia
Annette Quinlan

I. A.C.E. 2000 Evaluations

The A.C.E. 2000 evaluations are arranged into two basic categories: (1) operational summaries and (2) A.C.E. evaluations. These were discussed with the ESCAP.

1. Operational Summaries

These are evaluations that will provide real time results of key A.C.E. Sampling

and Estimation activities. These evaluations result from the verification process, and as such will be available for the decision process on determining whether to use A.C.E. to adjust the redistricting data required by PL-94-171. Donna Kostanich distributed a brief overview of Census 2000 Operational Analyses for Sampling and Estimation that will be conducted by DSSD staff. This overview included a short description of key A.C.E. processes and included a summary of the measures that will be produced for each activity. Comments or additional suggestions are welcome and should be sent to Donna Kostanich by May 5.

The highlights of these evaluations included A.C.E. Sample Design, P- and E-Sample Missing Data, Dual System Estimation (DSE), A.C.E. Targeted Extended Search, and A.C.E. Weight Trimming. The discussion on A.C.E. weight trimming raised a sensitive issue. The potential use of weight trimming may conflict with our plans to completely prespecify the A.C.E. methodology. It was decided that the ESCAP will review and approve any usage of weight trimming for the A.C.E. This issue will be revisited at a future ESCAP meeting.

2. A.C.E. Evaluations

Ruth Ann Killion distributed study plans for three sets of A.C.E. Evaluations. They described A.C.E. Operations Evaluations, Coverage Evaluations for A.C.E., and A.C.E. Statistical Design and Estimation Evaluations. These are long term evaluations based on additional research, such as re-interview studies, that would be used to assess the overall accuracy of the A.C.E. As an attachment to the study plans, background information for the Total Error Model study evaluations was provided. All handouts are included with these minutes.

The A.C.E. Evaluations conducted by PRED staff will occur after Census and A.C.E. processing operations have been completed. The ESCAP requested that Ruth Ann Killion and her staff review the evaluations to determine whether any preliminary findings would be available to assist in the review of the A.C.E. results prior to releasing redistricting data.

There was discussion of the concept of the Total Error Model and the key evaluations that will be used to construct it. In short, the Total Error Model presents an overall quantification of the sampling and nonsampling errors associated with the A.C.E. and the Census data adjusted based on the A.C.E. results. The Total Error Model incorporates and combines the individual evaluation components to produce an overall measurement of accuracy. The results of the Total Error Model Evaluation will not be available until late 2001.

II. Next Meeting

The next meeting scheduled for Wednesday April 12, 2000 will discuss A.C.E. Post-stratification.

ESCAP Committee

Kenneth Prewitt
William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson, Chair
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura

cc:

Teresa Angueira
Bill Bell
Debbie Bolton
Genny Burns
Carolee Bush
Gerald Gates
Ed Gore
Dave Hubble
Donna Kostanich
Ellen Lee
Charlene Leggieri
Don Malec
Betsy Martin
Catherine Miller
Fay Nash
Sally Obenski
Miguel Perez
Ed Pike
Magdalena Ramos
Gregg Robinson
Raj Singh
Maria Urrutia
Signe Wetrogen
David Whitford
Henry Woltman
Tommy Wright

Kathleen P Zveare
04/10/2000 03:29 PM

To: Margaret A Applekamp/DIR/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Elizabeth Centrella/DSSD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Angela Frazier/DMD/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Jane F Green/DSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC
cc:
Subject: 4/12 ESCAP Meeting

The April 12 ESCAP Meeting will be held in Rm. 2412/3 from 10:30-12. The agenda is as follows:

Post-Stratification - Donna Kostanich/Howard Hogan

UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

April 19, 2000

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES #Q-24

MEMORANDUM FOR Howard Hogan
Chief, Decennial Statistical Studies Division

From: Donna Kostanich
Assistant Division Chief, Sampling and Estimation
Decennial Statistical Studies Division

Prepared by: Richard Griffin
Dawn Haines
Estimation Staff

Subject: Accuracy and Coverage Evaluation Survey: Final Post-stratification Plan for Dual System Estimation

I. INTRODUCTION

The goal of post-stratification is to group together people who have similar coverage by the census. A common assumption is that people who share similar housing, similar language, similar cultural attitudes, and similar education would also share similar census coverage. Tenure, race and ethnic origin often serve as a marker for these similarities.

This memorandum presents the final post-stratification plan for the Accuracy and Coverage Evaluation (A.C.E.) Survey including Puerto Rico. The plan for Census 2000 A.C.E. is summarized in Section III. The detailed definitions of the poststratification variables and the race and Hispanic origin groups are given in Sections IV. and V., respectively.

II. BACKGROUND

The 2000 A.C.E. is different from the 1990 Post Enumeration Survey (PES). The A.C.E. will have approximately twice the sample size of the PES. This larger sample size allows for the formation of more post-strata and more post-strata have the advantage of reducing correlation bias. Additionally in 2000 multiple responses to the race question will be permitted; whereas in 1990 only one race could be selected.

The 1990 PES had 357 post-strata defined by a cross-classification of 51 post-stratum groups by seven age/sex groups. The 357 design started with a cross-classification of seven variables: age, sex, race, Hispanic origin, tenure, urbanicity, and region. There were 840 cells in the cross-classification. Collapsing was necessary in order to produce post-strata with sufficient sample for reliable Dual System Estimation (DSE). The attachment shows the 51 post-stratum groups for the 1990 PES after collapsing and the seven age/sex groups.

Race and Hispanic origin were the most important variables. After collapsing, five race/Hispanic origin post-strata were maintained: Non-Hispanic White or Other, Black, Hispanic White or Other, Asian and Pacific Islander, and Reservation Indians. Off-reservation American Indians were placed in either the Non-Hispanic White or Other group or the Hispanic White or Other group depending on whether they were of Hispanic origin. Within each of these race/Hispanic origin post-strata, seven age/sex categories were maintained.

The other variables were collapsed in the following order: region, urbanicity, then tenure, if necessary. For American Indians residing on reservations, all these variables were collapsed. For Asian and Pacific Islanders, region and urbanicity were collapsed and tenure maintained. For the Black and Hispanic White or Other groups, region was collapsed for two levels of urbanicity. For Non-Hispanic White or Other, the full cross-classification of region, urbanicity and tenure were maintained.

The 1990 PES for Puerto Rico had 21 post-strata defined by a cross-classification of 3 Place Type categories and seven age/sex categories. The place types were central city areas in Metropolitan Statistical Areas, non-central cities in Metropolitan Statistical Areas, and areas outside of Metropolitan Statistical Areas. The seven age/sex categories were the same as those used for the U.S. These 1990 post-stratification groups for Puerto Rico are also given in the attachment.

III. CENSUS 2000 A.C.E. POST-STRATIFICATION PLAN

For the Census 2000 A.C.E. we will retain most of the 1990 PES post-stratification variables and we will include several additional variables. The 2000 A.C.E. post-strata will be defined by nine variables: age, sex, race, Hispanic origin, tenure, region, Metropolitan Statistical Area size, Type of Enumeration Area, and tract level return rate. The Metropolitan Statistical Area size variable is replacing the urbanicity variable which will not be available until the summer of 2001. Type of Enumeration Area and the tract return rate are two new features of the 2000 A.C.E. post-stratification. The mailout/mailback areas will be differentiated from other types of enumeration areas. Tracts will be classified by high or low return rate. Additionally, multiple responses to the race question will be reflected in the race and Hispanic origin groupings.

Table 1a shows the 64 post-stratum groups for the Census 2000 A.C.E.. Within each post-stratum group there will be seven age/sex groups (Table 1c). Thus, there is a maximum of $64 \times 7 = 448$ post-strata, and of course there will be fewer if further collapsing is necessary. The post-stratification plan was chosen to reduce correlation bias without having an adverse effect on the variance of the Dual System Estimator.

For the Census 2000 A.C.E. in Puerto Rico, post-strata will be defined by five variables: age, sex, tenure, Metropolitan Statistical Area, and tract-level return rate. The variable region is not applicable for Puerto Rico. Further, there is only one Type of Enumeration Area (Update/Leave) in Puerto Rico, so this variable is not utilized. Table 1b shows the 12 post-stratum groups used in the Puerto Rico Census 2000 A.C.E. Survey. Within each post-stratum group, the seven age/sex groups in Table 1c are utilized. Thus, there is a maximum of $12 \times 7 = 84$ post-strata, and of course there will be fewer if further collapsing is necessary.

Tables 1a and 1b show the 64 and 12 post-stratum groups for the U.S. and Puerto Rico, respectively. Table 1c presents the seven age/sex groups which are used for both the U.S. and Puerto Rico. Subsequent sections of this memorandum provide a detailed description of the post-stratification domains and variables, including any alternative definitions for Puerto Rico. An extensive explanation of the domains is presented in Section V.

Table 1a: Census 2000 A.C.E. - 64 Post-Stratum Groups (U.S.)

Race/Hispanic Origin Domain Number*		Tenure	MSA/TEA	High Return Rate				Low Return Rate			
				N	M	S	W	N	M	S	W
Domain 7 (Non-Hispanic White or "Some other race")		Owner	Large MSA MO/MB	1	2	3	4	5	6	7	8
			Medium MSA MO/MB	9	10	11	12	13	14	15	16
			Small MSA & Non-MSA MO/MB	17	18	19	20	21	22	23	24
			All Other TEAs	25	26	27	28	29	30	31	32
		Non-owner	Large MSA MO/MB	33				34			
			Medium MSA MO/MB	35				36			
			Small MSA & Non-MSA MO/MB	37				38			
			All Other TEAs	39				40			
Domain 4 (Non-Hispanic Black)		Owner	Large MSA MO/MB	41				42			
			Medium MSA MO/MB								
			Small MSA & Non-MSA MO/MB	43				44			
			All Other TEAs								
		Non-owner	Large MSA MO/MB	45				46			
			Medium MSA MO/MB								
			Small MSA & Non-MSA MO/MB	47				48			
			All Other TEAs								
Domain 3 (Hispanic)		Owner	Large MSA MO/MB	49				50			
			Medium MSA MO/MB								
			Small MSA & Non-MSA MO/MB	51				52			
			All Other TEAs								
		Non-owner	Large MSA MO/MB	53				54			
			Medium MSA MO/MB								
			Small MSA & Non-MSA MO/MB	55				56			
			All Other TEAs								
Domain 5 (Native Hawaiian or Pacific Islander)		Owner		57							
		Non-owner		58							
Domain 6 (Non-Hispanic Asian)		Owner		59							
		Non-owner		60							
American Indian or Alaska Native	**Domain 1** (On Reservation)	Owner		61							
		Non-owner		62							
	Domain 2 (Off Reservation)	Owner		63							
		Non-owner		64							

- For Census 2000 persons can self identify with more than one race group. For post-stratification, persons are included in a single Race/Hispanic Origin domain. This does not change a person's actual response and all persons will be tabulated based on their actual response in the census. An extensive explanation of the domains is presented in Section V.

Table 1b: Census 2000 A.C.E. - 12 Post-Stratum Groups (Puerto Rico)

Tenure	MSA	High Return Rate	Low Return Rate
Owner	San Juan CMSA	1	2
	Other MSA	3	4
	Non-MSA	5	6
Non-owner	San Juan CMSA	7	8
	Other MSA	9	10
	Non-MSA	11	12

Table 1c: Census 2000 A.C.E. - 7 Age/Sex Groups (U.S. & Puerto Rico)

	Male	Female
Under 18	A	
18 to 29	B	C
30 to 49	D	E
50+	F	G

Key:

Return Rate: Tract-level variable measuring the proportion of occupied housing units in the mailback universe which returned a census questionnaire. Low return rate tracts are those tracts whose return rate is less than or equal to the 25th percentile return rate.

MSA: Metropolitan Statistical Area or Consolidated Metropolitan Statistical Area, as defined by the Office of Management and Budget (OMB), will be referred to as MSA.

TEA: Type of Enumeration Area.

MO/MB: Mailout/Mailback Type of Enumeration Area.

N, M, S, W: Refers to region - Northeast, Midwest, South, West.

"Some other race": One of six possible major race categories obtained from the census questionnaire.

Further details on the variable definitions are included in the following sections.

IV. CENSUS 2000 POST-STRATIFICATION VARIABLES

A. Post-stratification Variables

A.C.E. post-stratification will use the following variables:

- Race/Hispanic Origin - seven categories (omitted for Puerto Rico)
- Age/Sex - seven categories
- Tenure - two categories
- Metropolitan Statistical Area (MSA) by Type of Enumeration Area (TEA) - four categories (three categories for Puerto Rico)
- Return Rate - two categories
- Region - four categories (omitted for Puerto Rico)

The seven Race/Hispanic Origin domains are:

- American Indian or Alaska Native on Reservations
- Off-Reservation American Indian or Alaska Native
- Hispanic
- Non-Hispanic Black
- Native Hawaiian or Pacific Islander
- Non-Hispanic Asian
- Non-Hispanic White or "Some other race"

See Section V. for further details on the Race/Hispanic Origin domains. Inclusion in a Race/Hispanic Origin domain is complicated as it depends on several variables and whether there are multiple race responses. In addition, inclusion in a Race/Hispanic Origin domain **does not** change a persons Race/Hispanic Origin response. All Census 2000 tabulations will be based on the actual responses. For example, a person who responds as American Indian on a reservation and Black will be placed in the first Race/Hispanic Origin domain (Group 1) for post-stratification purposes but will be tabulated in the census as American Indian/Black.

The seven Age/Sex categories are:

- Under 18
- 18 - 29 Male
- 18 - 29 Female
- 30 - 49 Male
- 30 - 49 Female
- 50+ Male
- 50+ Female

The two Tenure categories are:

- Owner

- Non-owner

The four MSA/TEA categories are:

- Large MSA Mailout/ Mailback (MO/MB)
- Medium MSA MO/MB
- Small MSA or Non-MSA MO/MB
- All other TEAs

MSA/CMSA FIPS codes, as defined by the Office of Management and Budget (OMB), will be used for post-stratification. For simplification, MSA/CMSA will herein be referred to as MSA. Large MSA consists of the ten largest MSAs based on unadjusted, Census 2000 total population counts including the population in Group Quarters. Medium MSAs are those (besides the largest 10) which have at least 500,000 total population. Small MSAs are those with a total population size strictly less than 500,000. For post-stratification purposes, MO/MB areas are contrasted with the non-MO/MB areas.

For Puerto Rico there are three MSA categories. The TEA portion of this variable is nonexistent since all of Puerto Rico is Update/Leave. The three MSA categories are:

- San Juan CMSA (San Juan-Caguas-Arecibo CMSA)
- Other MSA (Aguadilla, Mayaguez, and Ponce MSAs)
- Non-MSA

The two Return Rate categories are:

- High
- Low

Return rate is a tract-level variable measuring the proportion of occupied housing units in the mailback universe which returned a census questionnaire. Low (high) return rate tracts are those tracts whose return rate is less than or equal to (greater than) the 25^{th} percentile return rate. Separate 25^{th} percentile cut-off values will be formed for the six applicable Race/Hispanic Origin by Tenure groups. Persons in List/Enumerate, Rural Update/Enumerate, and Urban Update/Enumerate TEAs are automatically placed in the High category. For Puerto Rico, distinct 25^{th} percentile return rate cut-off values will be formed for each Tenure category.

The four Region categories are:

- Northeast
- Midwest
- South
- West

B. Pre-collapsing

All Race/Hispanic Origin, Age/Sex, and Tenure categories for the U.S. will initially be maintained. The pre-collapsing plan for Region, MSA/TEA and Return Rate varies as follows:

- Non-Hispanic White or "Some other race" Owners: No collapsing
- Non-Hispanic White or "Some other race" Non-owners: Eliminate Region
- Non-Hispanic Black: Eliminate Region and partial collapsing of the MSA/TEA variable within Return Rate and Tenure categories
- Hispanic: Eliminate Region and partial collapsing of the MSA/TEA variable within Return Rate and Tenure categories
- Native Hawaiian or Pacific Islander: Eliminate the Region, Return Rate and MSA/TEA variables (Retain Tenure and Age/Sex only)
- Non-Hispanic Asian: Eliminate the Region, Return Rate and MSA/TEA variables (Retain Tenure and Age/Sex only)
- American Indian or Alaska Native on Reservations: Eliminate the Region, Return Rate and MSA/TEA variables (Retain Tenure and Age/Sex only)
- Off-Reservation American Indian or Alaska Native: Eliminate the Region, Return Rate and MSA/TEA variables (Retain Tenure and Age/Sex only)

For Puerto Rico, all 84 post-strata defined by MSA, Tenure, Return Rate, and Age/Sex will initially be maintained. Thus, there will be no pre-collapsing for Puerto Rico.

C. Post-collapsing

Depending on the actual A.C.E. sample sizes, additional collapsing may be necessary. The collapsing procedure is hierarchical which requires a pre-defined collapsing order. Given the pre-collapsing plan which yielded 448 post-strata, not much post-collapsing is anticipated. However, an extensive post-collapsing strategy is presented for completeness and to satisfy the requirement of pre-specification.

Note that collapsing does not necessarily imply elimination of a variable. Collapsing can refer to a reduction in the number of categories for a variable. For both the U.S. and Puerto Rico, a post-stratum is deemed too small if it contains fewer than 100 P Sample persons. The following general outline describes the post-collapsing hierarchy which is applied to both the U.S. and Puerto Rico. Any differences in definitions for Puerto Rico are noted.

If any of the 448 U.S. or 84 Puerto Rico post-strata are too small, Age/Sex will be collapsed first. This means that within any of the 64 U.S. (or 12 Puerto Rico) post-stratum groups, the seven Age/Sex categories defined in Table 1c will be reduced to the following three categories: Under 18, 18+ Male, and 18+ Female.

If some post-strata are still too small and require collapsing, Region will be collapsed next, if applicable. This collapsing applies only to the Non-Hispanic White or "Some other race" domain since the variable Region is only included in their post-stratification definition. In this case, all levels of Region (Northeast, Midwest, South, West) will be combined to eliminate the variable.

Next, the four-level MSA/TEA variable in the U.S. will be collapsed, if necessary, into the following two groups:

- Large and Medium MSA MO/MB
- Small MSA and Non-MSA MO/MB and All Other TEAs

For Puerto Rico, the three-level MSA variable will be collapsed, if necessary, into the following two groups:

- San Juan CMSA
- Other MSA and Non-MSA

If further collapsing is necessary, Return Rate is the next variable to collapse. High and Low Return Rate categories are combined to eliminate the variable.

Further collapsing involves the variable MSA/TEA in the U.S. (MSA in Puerto Rico). If necessary, the two groups defined above would be combined together to eliminate the variable MSA/TEA for the U.S. (MSA in Puerto Rico) completely.

The next variable to collapse is Tenure. Owner and Non-owner categories are combined to eliminate the variable entirely, if necessary.

If collapsing is still needed, the three remaining Age/Sex post-strata will be combined together to eliminate the Age/Sex variable completely.

In the event that there are not at least 100 P Sample persons in a Race/Hispanic Origin domain, all persons in that domain will be combined with Domain 7, which includes Non-Hispanic White and "Some other race."

V. RACE AND HISPANIC ORIGIN CLASSIFICATIONS

The Census 2000 questionnaire has 15 possible race responses. The 15 responses are collapsed into six major race groups as shown below. Races which are included in the major groups are shown in parentheses. Persons self-identifying with a single race essentially place themselves into one of these six categories.

- White
- Black (Black, African American, Negro)
- American Indian or Alaska Native
- Asian (Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Other Asian)
- Native Hawaiian or Pacific Islander (Native Hawaiian, Guamanian or Chamorro, Samoan, Other Pacific Islander)
- "Some other race"

For the first time in census history, persons will be able to respond to more than one race category. Allowing persons to self-identify with multiple races results in many more than six race groups. In fact, after collapsing race to the six major groups, there are $2^6-1 = 63$ possible race combinations. It is necessary to subtract the 1 in this equation since each individual is assumed to have a race.

The race variable defined above is often cross-classified with the Hispanic origin variable to define post-strata. The Hispanic origin variable consists of two responses, No and Yes. Categories which are included in the Yes response are shown in parentheses.

- No, not Spanish/Hispanic/Latino
- Yes (Mexican, Mexican American, Chicano, Puerto Rican, Cuban, Other Spanish/Hispanic/Latino)

Combining the race and Hispanic origin variables yields $63 \times 2 = 126$ possible Race/Hispanic Origin groups. It is important to note that any post-stratification plan of interest cannot support 126 Race/Hispanic Origin groups. As a solution, each of the 126 Race/Hispanic Origin response possibilities are assigned to one of seven Race/Hispanic Origin domains. The seven Race/Hispanic Origin domains are defined as follows:

- American Indian or Alaska Native on Reservations
- Off-Reservation American Indian or Alaska Native
- Hispanic
- Non-Hispanic Black
- Native Hawaiian or Pacific Islander
- Non-Hispanic Asian
- Non-Hispanic White or "Some other race"

Note that missing race and Hispanic origin data are imputed. Rules for classifying the 126 race and Hispanic origin combinations into one of the seven Race/Hispanic Origin domains are now presented.

Many of the decisions on how to classify multiple race persons are based on cultural, linguistic, and sociological factors which are known to affect coverage and are not necessarily data-driven.

A hierarchy is used to assign persons to a Race/Hispanic Origin domain. The Race/Hispanic Origin designation occurs in the following order: American Indian or Alaska Native on Reservations, Off-Reservation American Indian or Alaska Native, Hispanic, Non-Hispanic Black, Native Hawaiian or Pacific Islander, Non-Hispanic Asian, and Non-Hispanic White or "Some other race." All census data are tabulated using the race and Hispanic origin categories selected by census respondents.

For the following tables, Indian Country (IC) is a block-level variable that indicates whether a collection block is (wholly/partially) inside an American Indian reservation/trust land, Tribal Jurisdiction Statistical Area (TJSA), Tribal Designated Statistical Area (TDSA), or Alaska Native Village Statistical Area (ANVSA).

Tables 2 and 3 display the assignment of Race/Hispanic Origin domains. Table 2 applies to Hispanic persons while Table 3 applies to non-Hispanic persons. The first six rows of Tables 2 and 3 correspond to a single race response. The remaining portion of the tables addresses the assignment of multiple race responses to a single Race/Hispanic Origin domain. Although a person may be associated with multiple race responses, each person is included in only one of the seven Race/Hispanic Origin domains. All persons with a common number are assigned to the same Race/Hispanic Origin domain. Following is a verbal description of who is included in each Race/Hispanic Origin domain and their associated domain number.

Domain 1 (Includes American Indian or Alaska Native on Reservations): This domain includes any person living on a reservation marking American Indian or Alaska Native either as their single race or as one of many races, regardless of their Hispanic origin.

Domain 2 (Includes Off-Reservation American Indian or Alaska Native): This domain includes any person living in IC but not on a reservation who marks American Indian or Alaska Native either as their single race or as one of many races, regardless of their Hispanic origin. This domain also includes any non-Hispanic person not living in IC who marks American Indian or Alaska Native as their single race.

Domain 3 (Includes Hispanic): This domain includes all Hispanic persons who are not included in Domains 1 or 2. All Hispanic persons who self-identify with three or more races (excluding American Indian or Alaska Native in IC) are included in Domain 3. The only exception to this rule occurs when a Hispanic person lives in the state of Hawaii and classifies themselves as Native Hawaiian or Pacific Islander, regardless of whether they identify with a single or multiple race. All Hispanic persons satisfying this condition are re-classified into Domain 5.

Domain 4 (Includes Non-Hispanic Black): This domain includes any non-Hispanic person who marks Black as their only race. It also includes the combination of Black and American Indian or Alaska Native not in IC. In addition, people who mark Black and another single race group (Native Hawaiian or Pacific Islander, Asian, White, or "Some other race") are included in Domain 4. The only exception to this rule occurs when a Non-Hispanic Black person lives in the state of Hawaii and classifies themselves

as Native Hawaiian or Pacific Islander. All Non-Hispanic Black persons satisfying this condition are re-classified into Domain 5.

Domain 5 (Includes Native Hawaiian or Pacific Islander): This domain includes any person marking the single race Native Hawaiian or Pacific Islander. It also includes the combination of Native Hawaiian or Pacific Islander and American Indian or Alaska Native not in IC. Also included is the combination of Native Hawaiian or Pacific Islander with Asian. All persons living in the state of Hawaii who classify themselves as Native Hawaiian or Pacific Islander, regardless of their Hispanic origin and whether they identify with a single or multiple race, are also included in Domain 5.

Domain 6 (Includes Non-Hispanic Asian): This domain includes any non-Hispanic person marking Asian as their single race. If a person self-identifies with Asian and American Indian or Alaska Native not in IC, they are included in Domain 6.

Domain 7 (Includes Non-Hispanic White or "Some other race"): Non-Hispanic White or Non-Hispanic "Some other race" persons are included Domain 7. Non-Hispanic persons who self-identify with American Indian or Alaska Native not in IC and are White or "Some other race" are classified into Domain 7. If a Native Hawaiian or Pacific Islander response is combined with a White or "Some other race" response, they also are included in Domain 7. A person who self-identifies with Asian and White or Asian and "Some other race" is also included in this domain.
Finally, all non-Hispanic persons who self-identify with three or more races (excluding American Indian or Alaska Native in IC) are included in Domain 7. The only exception to this rule occurs when a Non-Hispanic White or Non-Hispanic "Some other race" person lives in Hawaii and classifies themselves as Native Hawaiian or Pacific Islander, regardless of whether they identify with other races. Persons who satisfy this criteria are re-classified into Domain 5.

Table 2: Census 2000 A.C.E. Post-stratification Domains for Hispanic

			Indian Country (IC)	
		Not in IC	Not On Res.	On Res.
Single race:				
American Indian or Alaska Native		3	2	1
Black		3	3	3
Native Hawaiian or Pacific Islander		3*	3	3
Asian		3	3	3
White		3	3	3
"Some other race"		3	3	3
American Indian or Alaska Native and:	Black	3	2	1
	Native Hawaiian or Pacific Islander	3*	2	1
	Asian	3	2	1
	White	3	2	1
	"Some other race"	3	2	1
Black and:	Native Hawaiian or Pacific Islander	3*	3	3
	Asian	3	3	3
	White	3	3	3
	"Some other race"	3	3	3
Native Hawaiian or Pacific Islander and:	Asian	3*	3	3
	White	3*	3	3
	"Some other race"	3*	3	3
Asian and:	White	3	3	3
	"Some other race"	3	3	3
American Indian or Alaska Native and:	Two or More Races	3*	2	1
All Else**		3*	3	3

* All persons living in the state of Hawaii who classify themselves as Native Hawaiian or Pacific Islander, regardless of their Hispanic origin and whether they identify with a single or multiple race, are included in Domain 5, which includes Native Hawaiian or Pacific Islander.

** All Else encompasses all remaining combinations which exclude American Indian or Alaska Native.

Table 3: Census 2000 A.C.E. Post-stratification Domains for Non-Hispanic

			Indian Country (IC)	
		Not in IC	Not On Res.	On Res.
Single race:				
American Indian or Alaska Native		2	2	1
Black		4	4	4
Native Hawaiian or Pacific Islander		5	5	5
Asian		6	6	6
White		7	7	7
"Some other race"		7	7	7
American Indian or Alaska Native and:	Black	4	2	1
	Native Hawaiian or Pacific Islander	5	2	1
	Asian	6	2	1
	White	7	2	1
	"Some other race"	7	2	1
Black and:	Native Hawaiian or Pacific Islander	4*	4	4
	Asian	4	4	4
	White	4	4	4
	"Some other race"	4	4	4
Native Hawaiian or Pacific Islander and:	Asian	5	5	5
	White	7*	7	7
	"Some other race"	7*	7	7
Asian and:	White	7	7	7
	"Some other race"	7	7	7
American Indian or Alaska Native and:	Two or More Races	7*	2	1
All Else**		7*	7	7

* All persons living in the state of Hawaii who classify themselves as Native Hawaiian or Pacific Islander, regardless of their Hispanic origin and whether they identify with a single or multiple race, are included in Domain 5, which includes Native Hawaiian or Pacific Islander.

** All Else encompasses all remaining combinations which exclude American Indian or Alaska Native.

ATTACHMENT: 1990 PES Post-Stratification

This attachment provides a brief summary of the 1990 PES post-stratification for the U.S. and Puerto Rico. Included below are the 51 post-stratum groups for the U.S. and the three post-stratum groups for Puerto Rico. Each of these post-stratum groups are further subdivided into the same seven age/sex groups.

Table 4a: 1990 PES 357 Design - 51 Post-Stratum Groups (U.S.)

Race/Hispanic Origin	Tenure	Urbanicity	N	M	S	W
Non-Hispanic White or Other	Owner	Large Urbanized Areas	1	2	3	4
		Other Urban	5	6	7	8
		Non-Urban	9	10	11	12
	Non-owner	Large Urbanized Areas	13	14	15	16
		Other Urban	17	18	19	20
		Non-Urban	21	22	23	24
Black	Owner	Large Urbanized Areas	25	26	27	28
		Other Urban	29			
		Non-Urban	30			
	Non-owner	Large Urbanized Areas	31	32	33	34
		Other Urban	35			
		Non-Urban	36			
Hispanic White or Other	Owner	Large Urbanized Areas	37	38	39	40
		Other Urban	41			
		Non-Urban	42			
	Non-owner	Large Urbanized Areas	43	44	45	46
		Other Urban	47			
		Non-Urban	48			
Asian or Pacific Islander	Owner		49			
	Non-owner		50			
Reservation Indians			51			

Table 4b: 1990 PES - 3 Post-Stratum Groups (Puerto Rico)

Place Type	
Central City in an MSA/PMSA	1
Non-central City in an MSA/PMSA	2
Not in an MSA/PMSA	3

Table 4c: 1990 PES - 7 Age/Sex Groups (U.S. & Puerto Rico)

	Male	Female
Under 18	A	
18 to 29	B	C
30 to 49	D	E
50+	F	G

Key:

MSA: Metropolitan Statistical Area, as defined by the Office of Management and Budget (OMB), will be referred to as MSA.

PMSA: Primary Metropolitan Statistical Area, as defined by the Office of Management and Budget (OMB), will be referred to as PMSA.

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 7

April 12, 2000

Prepared by: Maria Urrutia and Annette Quinlan

The seventh meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on April 12, 2000 at 10:30. The agenda for the meeting was A.C.E. Post-Stratification.

Persons in attendance:

Kenneth Prewitt
William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
John Thompson
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura
Tommy Wright
Raj Singh
Gregg Robinson
Dawn Haines
Maria Urrutia
Annette Quinlan

I. Overview of A.C.E. Post-Stratification

Raj Singh and Dawn Haines discussed and distributed a draft memorandum describing the final post-stratification plan for A.C.E. Dual System Estimation. The memorandum details the definitions of the post-strata variables that will be used in the 2000 design.

Howard Hogan summarized how he had incorporated comments from a previous ESCAP meeting on post-stratification into this memo, which has been finalized and is attached.

Howard indicated the main unresolved issue is how to collapse post-strata in the event of small A.C.E. sample sizes. This decision has not been finalized but the basic methodology will be completed by late April and the final specifications will be completed by June. All post-stratification issues and decisions will be made before the appropriate data are available.

The US post-stratification design for 2000 A.C.E. will contain a maximum of 448 post-strata, as compared to 357 post-strata for the 1990 post-stratification design. The major differences from the 1990 design are as follows: (1) the 2000 A.C.E. design includes the new variable mail return rate, (2) region is included only for Non-Hispanic White or "Some Other Race" Owners, (3) the 1990 PES urbanicity variable has been redefined by combining Metropolitan Statistical Area (MSA) by Type of Enumeration Area (TEA), (4) TEA is also a new variable for 2000 A.C.E., and (5) the document describes in detail the treatment of multiple race responses for creating post-stratification domains.

The Puerto Rico A.C.E. post-stratification plan has a maximum of 84 post-strata. The variables Tenure and Return Rate have both been added since 1990. A post-collapsing plan for Puerto Rico will also be defined before data are available.

II. Next Meeting

The next meeting scheduled for Wednesday April 26, 2000 will discuss A.C.E. Weight Trimming.

ESCAP Committee

Kenneth Prewitt
William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson, Chair
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura

cc:

Teresa Angueira
Bill Bell
Debbie Bolton
Genny Burns
Carolee Bush
Gerald Gates
Ed Gore
Dave Hubble
Donna Kostanich
Ellen Lee
Charlene Leggieri
Don Malec
Betsy Martin
Catherine Miller
Fay Nash
Sally Obenski
Miguel Perez
Ed Pike
Magdalena Ramos
Gregg Robinson
Raj Singh
Maria Urrutia
Signe Wetrogen
David Whitford
Henry Woltman
Tommy Wright

Kathleen P Zveare
04/25/2000 03:21 PM

To: Margaret A Applekamp/DIR/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Elizabeth Centrella/DSSD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Angela Frazier/DMD/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Jane F Green/DSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC

cc:

Subject: Agenda for 4/26 ESCAP Meeting

The agenda for tomorrow's ESCAP meeting is:

Weight Trimming - Donna Kostanich

Time: 10:30-12:00

Room: 2412/3

Thoughts on A.C.E. Weight Trimming — DRAFT 4/26/00

Goal: Reduce the contribution of variance due to outlier clusters dominating a post-stratum's coverage correction factor.

Key Dates:

June, 2000 Determine final weight trimming plans.

Dec. 8, 2000 Apply and verify the weights to Missing Data Files.

Assumptions:

S Cluster level trimming

S Identify clusters to downweight and proportionately upweight the remaining clusters.

S Identify outlier clusters separately for American Indian Reservations, rest of the U.S., and Puerto Rico.

S Distribute weights within sampling stratum if feasible.

S Sampling Staff will write the programs to identify clusters and do the weight trimming. This requires running parallel systems to verify the results. The weights will be transmitted to programmers electronically to be applied to Missing Data files.

Three options:

1. Trim weights based on total weighted housing unit estimates of block clusters
   - implement trimming methodology: Oct. 21 to Nov. 30
   - separately for P & E samples
   - does not reflect any matching results
   - does not reflect TES

2. Trim weights using the initial housing unit match results
   - implement trimming methodology: Oct. 21 to Nov. 30
   - use housing unit match results as proxy for person matching
   - will not reflect changes to census since the January DMAF
   - could misidentify clusters to downweight
   - does not reflect TES, but could if willing to make a guess at effect of TES

3. Trim weights using the person match results
   - implement trimming methodology: Dec. 1 to Dec. 5
   - reflects impact of TES
   - relies on person matching ending on Nov. 30
   - can use HU match results as a contingency

April 26, 2000 Draft

Potentially Influential Cluster

TEA:	Urban Update/Leave
1990 Housing Unit Count:	217
Keyed and Valid Independent Listing Count:	192
January DMAF	1153
Housing Unit Matching Results	
Matches	192
Erroneous Enumerations	961
1990 Demographic/Tenure Distributions	
Black Renters	97.0%
Hispanic Renters	1.7%
Black Owners	0.8%
Other Renters	0.5%
Final P-sample Weight:	383.130

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 8

April 26, 2000

Prepared by: Maria Urrutia and Annette Quinlan

The eighth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on April 26, 2000 at 10:30. The agenda for the meeting was A.C.E. Weight Trimming.

Persons in attendance:

William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Susan Miskura
Donna Kostanich
Raj Singh
Gregg Robinson
Signe Wetrogan
Carolee Bush
Maria Urrutia
Annette Quinlan

I. A.C.E. Weight Trimming

Donna Kostanich presented options for A.C.E. weight trimming. The major goal of weight trimming would be to reduce the effect on the A.C.E. estimates due to outlier clusters dominating a post-stratum's coverage correction factor. The assumptions and requirements for weight trimming were discussed, as were alternatives. To facilitate understanding the potential effects of weight trimming, an example of an influential cluster was distributed and discussed. The example and options for weight trimming are attached.

The ESCAP discussed three options for implementing weight trimming:

1) Identify block clusters for weight trimming based on total weighted housing unit estimates of block clusters from the initial housing unit phase. For example, a cluster with a large weighted estimate of housing units from the initial phase would be identified for weight trimming.
2) Identify block clusters for weight trimming using the results from the initial housing unit match that occurred in April 2000. For example, these block clusters would be identified if they had a large number of non-matching housing units.
3) Identify block clusters for weight trimming based on the results of the person match that will be completed in November 2000. For example, these block clusters would be identified if they had a large number of non-matched person records.

The timing for alternatives 1 and 2 would occur from 10/21/00 to 11/30/00 since they do not require the person match results. Since these alternatives do not include the interviewing results they would not identify all clusters that may require weight trimming. Given that, we decided that alternative 3 would be the best option to use for weight trimming. We noted, however, that there were timing concerns because this option will occur from 12/1/00 to 12/5/00, allowing five days for implementation. Therefore, there is a risk of extending the A.C.E. schedule if this process requires more than five days to complete, including the necessary review process.

It was also decided that the weight trimming process, including the criteria, would be identified by June. It is critical that this process and criteria be pre-specified and publicly available. Before any weight trimming would be implemented, the ESCAP will review to ensure that pre-specified criteria are met.

II. Next Meeting

The next meeting scheduled for Wednesday May 24, 2000 will discuss Telephone Interviewing and Synthetic Estimation.

Kathleen P Zveare
05/23/2000 02:32 PM

To: Margaret A Applekamp/DIR/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Elizabeth Centrella/DSSD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Angela Frazier/DMD/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Jane F Green/DSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC
cc:
Subject: Agenda for 5/24 ESCAP Meeting

The agenda for the May 24 ESCAP meeting scheduled from 10:30-12 in Rm. 2412/3 is as follows:

1. Telephone Interviewing
2. Synthetic Estimation

OVERVIEW OF SYNTHETIC ESTIMATION

Dawn Haines

May 22, 2000

Goal: Obtain an integer number of persons for each post-stratum within each tabulation block, representing either overcounts or undercounts.

- Synthetic estimation begins after Dual System Estimates are calculated for each post-stratum.

- The procedure involves carrying down and rounding from post-strata to tabulation blocks.

- This occurs at the following levels:

  - State
  - County
  - Tract
  - Block

- Finally, the number of records to replicate for each post-stratum within each tabulation block is determined.

May 22, 2000 **CONTROLLED ROUNDING EXAMPLE**

Dawn Haines

Suppose our nation is made up of the following i = 5 post-strata: White, Black, AIAN, API, and Other. Dual System Estimates (DSE_i), census counts (C_i), and coverage correction factors (CCF_i) are given for each post-stratum i. DSE and census totals over post-strata are presented.

Table 1: Post-stratum Information

i	White	Black	AIAN	API	Other	Total
DSE_i	3,733,740.22	433,730.80	102,864.39	80,277.83	115,871.70	4,466,484.94
C_i	3,809,939	409,180	94,371	79,483	110,354	4,503,327
CCF_i	0.98	1.06	1.09	1.01	1.05	

The coverage correction factor for post-stratum i, CCF_i, is formed by dividing the DSE for post-stratum i by its census count, denoted

$$CCF_i = \frac{DSE_i}{C_i}.$$

Table 2: Control-rounded Dual System Estimates

	White	Black	AIAN	API	Other	Total
DSE_i^R	3,733,740	433,731	102,864	80,278	115,872	4,466,485

Each Level of Carrying down requires 4 steps:

1) Census Counts: $C_{i,s}$

2) Synthetic: $\hat{N}_{i,s}^{S} = C_{i,s} \times CCF_i$

3) Adjusted Synthetic: $\hat{N}_{i,s}^{AS} = \hat{N}_{i,s}^{S} \times \frac{DSE_i^R}{DSE_i}$

(Note that the ratio of the rounded to the unrounded DSE is replaced by the ratio of rounded to unrounded Synthetic for all levels of carrying down except for the 1st level State.)

4) Rounded Synthetic: $\hat{N}_{i,s}^{RS}$ (Control rounding of Adjusted Synthetic.)

3 States in U.S.

Table 3: State-level Census Counts $C_{i,s}$

	White	Black	AIAN	API	Other	Total
02 (AK)	415,492	22,451	85,698	19,728	6,674	550,043
09 (CT)	2,859,353	274,269	6,654	50,698	96,142	3,287,116
10 (DE)	535,094	112,460	2,019	9,057	7,538	666,168
Total	3,809,939	409,180	94,371	79,483	110,354	4,503,327

Table 4: State-level Synthetic Estimates

$\hat{N}^{S}_{i,s} = C_{i,s} \times CCF_i$ **(9147.57 = 9057 × 1.01)**

	White	Black	AIAN	API	Other	Total
02 (AK)	407,182.16	23,798.06	93,410.82	19,925.28	7,007.70	551,324.02
09 (CT)	2,802,165.94	290,725.14	7,252.86	51,204.98	100,949.10	3,252,298.02
10 (DE)	524,392.12	119,207.60	2,200.71	**9,147.57**	7,914.90	662,862.90
Total	3,733,740.22	433,730.80	102,864.39	80,277.83	115,871.70	4,466,484.94

Table 5: State-level Adjusted Synthetic Estimates

$\hat{N}^{AS}_{i,s} = \hat{N}^{S}_{i,s} \times \frac{DSE^{R}_{i}}{DSE_i}$ **(9147.59 = 9147.57 ×** $\frac{80278}{80277.83}$ **)**

	White	Black	AIAN	API	Other	Total
02 (AK)	407,182.14	23,798.07	93,410.47	19,925.32	7,007.72	551,323.71
09 (CT)	2,802,165.78	290,725.27	7,252.83	51,205.09	100,949.36	3,252,298.33
10 (DE)	524,392.09	119,207.66	2,200.70	**9,147.59**	7,914.92	662,862.96
Total	3,733,740.00	433,731.00	102,864.00	80,278.00	115,872.00	4,466,485.00

Table 6: State-level Control-rounded Synthetic Estimates $\hat{N}^{RS}_{i,s}$

	White	Black	AIAN	API	Other	Total
02 (AK)	407,182	23,798	93,411	19,925	7,008	551,324
09 (CT)	2,802,166	290,725	7,253	51,205	100,949	3,252,298

10 (DE)	524,392	119,208	2,200	9,148	7,915	662,863
Total	3,733,740	433,731	102,864	80,278	115,872	4,466,485

3 Counties in Delaware

Table 7: County-level Census Counts $C_{i,c}$

DE	White	Black	AIAN	API	Other	Total
001	87,300	20,631	614	**1,420**	1,028	110,993
003	355,399	72,834	760	7,048	5,905	441,946
005	92,395	18,995	645	589	605	113,229
Total	535,094	112,460	2,019	9,057	7,538	666,168

Table 8: County-level Synthetic Estimates

$\hat{N}^{S}_{i,c} = C_{i,c} \times CCF_i$ **(1434.20 = 1420 × 1.01)**

DE	White	Black	AIAN	API	Other	Total
001	85,554.00	21,868.86	669.26	**1,434.20**	1,079.40	110,605.72
003	348,291.02	77,204.04	828.40	7,118.48	6,200.25	439,642.19
005	90,547.10	20,134.70	703.05	594.89	635.25	112,614.99
Total	524,392.12	119,207.60	2,200.71	**9,147.57**	7,914.90	662,862.90

Table 9: County-level Adjusted Synthetic Estimates

$\hat{N}^{AS}_{i,c} = \hat{N}^{S}_{i,c} \times \frac{\hat{N}^{RS}_{i,s}}{\hat{N}^{S}_{i,s}}$ **(** $1434.27 = 1434.20 \times \frac{9148}{9147.57}$ **)**

DE	White	Black	AIAN	API	Other	Total
001	85,553.98	21,868.93	669.04	**1,434.27**	1,079.41	110,605.64
003	348,290.94	77,204.30	828.13	7,118.81	6,200.33	439,642.52
005	90,547.08	20,134.77	702.82	594.92	635.26	112,614.85
Total	524,392.00	119,208.00	2,200.00	**9,148.00**	7,915.00	662,863.00

Table 10: County-level Control-rounded Adjusted Synthetic Estimates $\hat{N}^{RS}_{i,c}$

DE	White	Black	AIAN	API	Other	Total
001	85,554	21,869	669	**1,434**	1,080	110,606
003	348,291	77,204	828	7,119	6,200	439,642
005	90,547	20,135	703	595	635	112,615
Total	524,392	119,208	2,200	**9,148**	7,915	662,863

4 Tracts in County 1 in Delaware

Table 11: Tract-level Census Counts $C_{i,t}$

Co. 1, DE	White	Black	AIAN	API	Other	Total
tract 1	29,004	6,854	216	422	239	36,735
tract 2	5,408	9,315	173	**314**	411	15,621
tract 3	37,816	3,298	175	477	113	41,879
tract 4	15,072	1,164	50	207	265	16,758
Total	87,300	20,631	614	1,420	1,028	110,993

Table 12: Tract-level Synthetic Estimates

$\hat{N}^{S}_{i,t} = C_{i,t} \times CCF_i$ (**317.14 = 314 × 1.01**)

Co. 1, DE	White	Black	AIAN	API	Other	Total
tract 1	28,423.92	7,265.24	235.44	426.22	250.95	36,601.77
tract 2	5,299.84	9,873.90	188.57	**317.14**	431.55	16,111.00
tract 3	37,059.68	3,495.88	190.75	481.77	118.65	41,346.73
tract 4	14,770.56	1,233.84	54.50	209.07	278.25	16,546.22
Total	85,554.00	21,868.86	669.26	**1,434.20**	1,079.40	110,605.72

Table 13: Tract-level Adjusted Synthetic Estimates

$\hat{N}^{AS}_{i,t} = \hat{N}^{S}_{i,t} \times \frac{\hat{N}^{RS}_{i,c}}{\hat{N}^{S}_{i,c}}$ (**317.10 = 317.14 ×** $\frac{1434}{1434.20}$)

Co. 1, DE	White	Black	AIAN	API	Other	Total
tract 1	28,423.92	7,265.29	235.35	426.16	251.09	36,601.81
tract 2	5,299.84	9,873.96	188.50	**317.10**	431.79	16,111.19
tract 3	37,059.68	3,495.90	190.68	481.70	118.72	41,346.68
tract 4	14,770.56	1,233.85	54.48	209.04	278.40	16,546.33
Total	85,554.00	21,869.00	669.00	**1,434.00**	1,080.00	110,606.00

Table 14: Tract-level Control-rounded Adjusted Synthetic Estimates $\hat{N}^{RS}_{i,t}$

Co. 1, DE	White	Black	AIAN	API	Other	Total
tract 1	28,424	7,265	236	426	251	36,602
tract 2	5,300	9,874	188	317	432	16,111
tract 3	37,059	3,496	191	482	119	41,347

tract 4	14,771	1,234	54	209	278	16,546
Total	85,554	21,869	669	1,434	1,080	110,606

3 Blocks in Tract 2 in County 1 in Delaware

Table 15: Block-level Census Counts $C_{i,b}$

tract 2, Co. 1,DE	White	Black	AIAN	API	Other	Total
block 1	1,785	2,508	32	150	128	4,603
block 2	1,429	4,283	71	64	245	6,092
block 3	2,194	2,524	70	**100**	38	4,926
Total	5,408	9,315	173	314	411	15,621

Table 16: Block-level Synthetic Estimates

$\hat{N}^{S}_{i,b} = C_{i,b} \times CCF_i$ (***101.00 = 100 × 1.01***)

tract 2 Co. 1, DE	White	Black	AIAN	API	Other	Total
block 1	1,749.30	2,658.48	34.88	151.50	134.40	4,728.56
block 2	1,400.42	4,539.98	77.39	64.64	257.25	6,339.68
block 3	2,150.12	2,675.44	76.30	***101.00***	39.90	5,042.76
Total	5,299.84	9,873.90	188.57	317.14	431.55	16,111.00

Table 17: Block-level Adjusted Synthetic Estimates

$\hat{N}^{AS}_{i,b} = \hat{N}^{S}_{i,b} \times \frac{\hat{N}^{RS}_{i,t}}{\hat{N}^{S}_{i,t}}$ (***100.96 = 101.00*** $\times \frac{317}{317.14}$)

tract 2 Co. 1, DE	White	Black	AIAN	API	Other	Total
block 1	1,749.35	2,658.51	34.77	151.43	134.54	4,728.61
block 2	1,400.46	4,540.03	77.16	64.61	257.52	6,339.77
block 3	2,150.18	2,675.47	76.07	***100.96***	39.94	5,042.62
Total	5,300.00	9,874.00	188.00	317.00	432.00	16,111.00

Table 18: Block-level Control-rounded Adjusted Synthetic Estimates $\hat{N}^{RS}_{i,b}$

tract 2 Co. 1, DE	White	Black	AIAN	API	Other	Total
block 1	1,749	2,659	35	151	135	4,729
block 2	1,401	4,540	77	65	257	6,340

block 3	2,150	2,675	76	101	40	5,042
Total	5,300	9,874	188	317	432	16,111

Number of Records to Create

Table 19: Difference Between Block-level Control-rounded Adjusted Synthetic Estimates and Census Counts

$$\hat{N}_{i,b}^{RS} - C_{i,b}$$

tract 2 in Co. 1, DE	White	Black	AIAN	API	Other	Total
block 1	- 36	151	3	1	7	126
block 2	- 28	257	6	1	12	248
block 3	- 44	151	6	1	2	116
Total	- 108	559	15	3	21	490

ESCAP Committee

Kenneth Prewitt
William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson, Chair
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura

cc:

Teresa Angueira
Bill Bell
Debbie Bolton
Genny Burns
Carolee Bush
Gerald Gates
Ed Gore
Dave Hubble
Donna Kostanich
Ellen Lee
Charlene Leggieri
Don Malec
Betsy Martin
Catherine Miller
Fay Nash
Sally Obenski
Miguel Perez
Ed Pike
Magdalena Ramos
Gregg Robinson
Raj Singh
Maria Urrutia
Signe Wetrogen
David Whitford
Henry Woltman
Tommy Wright

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 9

May 24, 2000

Prepared by: Maria Urrutia and Annette Quinlan

The ninth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on May 24, 2000 at 10:30. The agenda for the meeting was A.C.E. Telephone Interviewing and A.C.E. Synthetic Estimation.

Persons in attendance:

Kenneth Prewitt
William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura
Donna Kostanich
Raj Singh
Tommy Wright
Carolee Bush
Maria Urrutia
Annette Quinlan

I. A.C.E. Telephone Interviewing

Howard Hogan discussed the goals and concerns of A.C.E. telephone interviewing. Cases are selected for the telephone phase only if the A.C.E. housing unit was independently listed in the A.C.E., was matched to the census during the initial phase of the A.C.E. Housing Unit Matching operation, and returned their Census 2000 questionnaire with a valid telephone number.

The benefits of conducting the telephone interviews include:

(1) Full operational test of the CAPI instrument and its control and support system earlier in the process.
(2) A reduced number of movers between Census Day and interview day.
(3) The interviews may be more accurate because they are occurring closer to Census Day.
(4) Training supervisory staff on the CAPI instrument to provide an opportunity for them to become more familiar with the instrument and its functions before the person visit interviews are conducted.

The potential concerns of conducting telephone interviews are:

(1) The possible mode effect on the completeness and the data quality of the interview. For example, is there a difference on the completeness of the interview when it is conducted over the phone as compared to a personal visit.
(2) The possible lack of independence between the census and the A.C.E. There is a possibility that the respondent can remember how they answered similar questions during NRFU of late mail returns or other later census operations.

These potential concerns are not anticipated to have an appreciable impact on the estimates. The mode effect will not impact the estimates because (1) the same instrument that is used for telephoning is also used for the personal visit interviews, (2) if there is any resistence, the interview goes to the field for those households which do not want to respond by telephone, and (3) the telephone interviews represent the more cooperative households. The independence concern will not impact the estimates because (1) we have already matched the A.C.E. address list to the Census and we only conduct the telephone interviews at matched households and (2) the telephone interviews are only conducted in areas where there is little risk of mail delivery problems. Staff will be evaluating the results of the A.C.E. telephone interview operation.

II. Synthetic Estimation

Donna Kostanich described synthetic, or indirect, estimation and provided an example, which is attached. The goal of synthetic estimation is to carry down coverage correction factors from the DSE to the block level. We use a control rounding procedure, as we do for long form estimation, to ensure that corrections are made in the form of an integer.

The statistical correction that results from the A.C.E. is carried down to census blocks by applying the coverage correction factors within each A.C.E. post-stratum. The goal in constructing post-strata is to form groupings of the population that capture differences in the probabilities of being included in the census and the A.C.E. In effect, the inclusion probabilities

are more similar for individuals within the same post-stratum than for individuals in different post-strata. The coverage correction factors are calculated for each post-stratum, based on a representative sample of the post-stratum, and thus reflect the net coverage of all people within the post-stratum. This is the underlying basis for applying this factor to the data records within the corresponding post-stratum to produce statistically corrected block totals which serve as the basis for Census 2000 tabulations.

The accuracy of the estimates that result from the application of the coverage correction factors depends on the degree to which the net coverage for areas or groups within a post-stratum is similar to the coverage correction factor that was developed for that post-stratum. The coverage correction factor is measured for the post-stratum based on a representative sample, and thus represents the net coverage for the post-stratum. Clearly, within the post-stratum, some degree of variation is expected from the measured coverage correction factor, and this variation will most likely be relatively greater for small areas. Thus, it is inevitable that the A.C.E. will result in the population in some blocks being overestimated and the population in other blocks being underestimated. The A.C.E. statistical correction was never intended nor expected to produce unqualified improvement in the smallest geographic areas, like blocks. That the A.C.E. does not produce improvement for every single block, however, is no reason to forego the benefits that will flow from the use of corrected census population counts at geographic levels of significance to data users. The Census Bureau expects that the A.C.E. estimates will produce better data for aggregations - such as states, congressional districts, counties, and cities - that are the basic areas for which census data are used.

The controlled rounding program, as described above, integerizes unrounded synthetic estimates. The rounding occurs in stages from (1) state, (2) county, (3) tract, and then by (4) block level. At each stage the controlled rounded estimates will differ by less than one from the unrounded estimator. The software is being double programmed to verify the results.

Bob Fay discussed the 1990 PES synthetic estimation assumptions, and his analysis of the effects of synthetic estimation on the analysis of the accuracy of the 1990 PES.

Application of synthetic estimation inevitably results in some degree of heterogeneity bias in estimates for states, counties, and other geographic areas of interest. The total error model of Mary Mulry and Bruce Spencer did not attempt to account for synthetic estimation bias, basically because the 1990 PES data were too thin to provide any reliable direct measures of this source of error.

At the time, there were competing hypotheses or possibilities:

(1) Heterogeneity, although omitted, might be so small as to be negligible.
(2) Heterogeneity might be large and, because it was omitted from the model, the error in the

adjusted figures might be much larger than estimated. Therefore, the loss function analysis might incorrectly favor adjustment.

Bob Fay attempted a brief summary of an empirical study presented in a 1993 paper coauthored by John Thompson. The analysis focused on the possible effect of heterogeneity on the loss function analysis at the state level. By taking census variables, such as mail response rates, poverty, and unemployment, and scaling them to the approximate level of the percent undercount, they constructed eight artificial populations. Unlike the PES, the artificial populations were essentially unaffected by sampling variance, since the variables were measured either by the whole census or the long form. In other words, heterogeneity bias could be studied for the artificial populations without the limitations imposed by the sampling error in the 1990 PES data. The loss function analysis, mimicking the 1990 PES by omitting any allowance for heterogeneity, could then be compared to the actual losses with and without adjustment, including the effect of heterogeneity in the calculation. In seven out of eight of the artificial populations, the results indicated the following:

> Heterogeneity, although a potentially significant source of error, led the loss function to understate the error in the unadjusted census counts by about as much as, or even more than, the error in the adjusted counts. Hence, although it omitted an important source of error, the loss function analysis generally could be trusted when it showed an advantage to the adjusted counts over the unadjusted. The one exception was the artificial population based on unemployment rate.

Bob noted that, resources permitting, it would be helpful to replicate this sort of study with 2000 data at some point.

III. Next Meeting

The next meeting scheduled for Wednesday June 28, 2000 will discuss Missing Data and Correlation Bias.

ESCAP Committee

Kenneth Prewitt
William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson, Chair
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura

cc:

Teresa Angueira
Bill Bell
Debbie Bolton
Genny Burns
Carolee Bush
Gerald Gates
Ed Gore
Dave Hubble
Donna Kostanich
Ellen Lee
Charlene Leggieri
Don Malec
Betsy Martin
Catherine Miller
Fay Nash
Sally Obenski
Miguel Perez
Ed Pike
Magdalena Ramos
Gregg Robinson
Raj Singh
Maria Urrutia
Signe Wetrogen
David Whitford
Henry Woltman
Tommy Wright

Kathleen P Zveare
06/26/2000 01:14 PM

To: Margaret A Applekamp/DIR/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Elizabeth Centrella/DSSD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Angela Frazier/DMD/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Jane F Green/DSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC
cc: Patrick J Cantwell/DSSD/HQ/BOC@BOC, J Gregory Robinson/POP/HQ/BOC@BOC
Subject: Agenda for 6/28 ESCAP Meeting

Below is the agenda for the June 28 ESCAP Meeting scheduled from 10:30-12 in Rm. 2412/3:

1. Missing Data---Pat Cantwell

2. Overview of Correlation Bias---Raj Singh

3. How Do We Measure Correlation Bias Using DA---Greg Robinson.

Description of Before-Followup (BFU) Groups for the E-Sample

BFU Group 1. *Matches needing FU.* This group contains all E-Sample persons who match to a P-Sample person, but are sent to followup. This situation occurs when a person matches, but the P-Sample person to whom he or she matches has unresolved residence status.

BFU Group 2. *Possible matches.* This group contains all E-Sample persons with a match code of "possible match".

BFU Group 3. *Partial HH nonmatches.* This group contains all E-Sample persons unresolved before followup who did not match to a P-Sample person but who were living in a household where at least one person did match.

BFU Group 4. *Whole HH nonmatches (where HU matched); not conflicting HHs.* This group contains all E-Sample persons unresolved before followup who lived in a household where no persons matched but the housing unit did match to an A.C.E. housing unit. Note that this group does not contain persons from conflicting households*.

BFU Group 5. *Nonmatches from conflicting HHs; HU not in regular NRFU.* This group contains all E-Sample persons unresolved before followup who lived in a household where no person matched, the housing unit was **not** in regular nonresponse followup, and the household is a conflicting household*.

BFU Group 6. *Nonmatches from conflicting HHs; HU in regular NRFU.* This group contains all E-Sample persons unresolved before followup who lived in a household where no person matched, the housing unit **was** in regular nonresponse followup, and the household is a conflicting household*.

BFU Group 7. *Whole HH nonmatches; HU did not match during HU matching.* This group contains all E-Sample persons unresolved before followup who lived in a household where no person matched and the census housing unit did not match to an A.C.E. housing unit.

BFU Group 8. *Persons resolved before FU.* This group contains all E-Sample persons whose enumeration status was resolved before followup. The following people are included:

- Matches not needing followup.
- Duplicates of another E-Sample person
- Persons erroneously enumerated due to geocoding error.
- Persons identified as fictitious before followup.

BFU Group 9. *Persons with insufficient information for matching.* This group includes all persons who do not have a full name and at least 2 person characteristics.

*Conflicting households are those in which the census household has all different persons from the matching A.C.E. household. All persons in the census household and the A.C.E. household are nonmatches (no possible matches). Note that persons in both the census household and the A.C.E.

household can have insufficient information for matching in a conflicting household.

Description of Before-Followup (BFU) Groups P-Sample

BFU Group 1. *Matches needing FU.* This group contains all P-Sample persons who match to an E-Sample person, but are sent to followup. This situation occurs when a person matches, but has unresolved residence status.

BFU Group 2. *Possible matches.* This group contains all P-Sample persons with a match code of "possible match".

BFU Group 3. *Partial HH nonmatches.* This group contains all P-Sample persons needing followup who did not match to an E-Sample person but who were living in a household where at least one person did match.

BFU Group 4. *Whole HH nonmatches needing FU (not conflicting HHs).* This group contains all P-Sample persons needing followup who lived in a household where no persons matched but the household is **not** conflicting*.

BFU Group 5. *Nonmatches from conflicting HHs needing FU.* This group contains all P-Sample persons needing followup who lived in a household where no person matched and the household is a conflicting* household.

BFU Group 6. *Persons resolved before FU.* This group contains all P-Sample persons whose residence status was resolved before followup. The following people are included:

- Matches not needing followup.
- Nonmatches where the data were collected from a household member (i.e., a nonproxy interview), the household was not conflicting*, and the residence status is known.
- Partial household nonmatches with a code of "NC".
- Duplicates of another P-Sample person.

BFU Group 7. *Persons with insufficient information for matching.* This group includes all persons who do not have a full name and at least 2 person characteristics.

*Conflicting households are those in which the census household has all different persons from the matching A.C.E. household. All persons in the census household and the A.C.E. household are nonmatches (no possible matches). Note that persons in both the census household and the A.C.E. household can have insufficient information for matching in a conflicting household.

Missing Data for the Accuracy and Coverage Evaluation

June 28, 2000

1. Background: Types of missing data

- Noninterviews (P Sample only)
  - --- census day
  - --- A.C.E. interview day

- Item nonresponse (P Sample only)
  only for tenure, race, hispanic origin, age category, sex

- Unresolved status
  - --- correct enumeration (E Sample)
  - --- match (P Sample)
  - --- residence (P Sample)

2. Noninterviews (housing-unit level)

compute two adjustments of sampling weights at housing-unit level (see Attachment 1):

- one for census day, applied to person non-movers and out-movers
- one for A.C.E. interview day, applied to person in-movers

3. Item nonresponse

characteristic imputation (see Attachment 2)

tenure, race, hispanic origin: nearest-neighbor hot deck

age, sex: impute from distribution of characteristic,
conditioned on certain variables

4. Unresolved status

- choice of imputation cells for the U.S. (see Attachment 3)
  criteria
  - --- variables that discriminate well
  - --- minimum expected frequencies of resolved cases per cell
  - --- no collapsing

  research

--- 1990 PES, Hudson and Clarke
--- 1998, Dress Rehearsal, Malec

! Puerto Rico (see Attachment 3)
--- start with the same cells
--- remove race, number of imputes
--- maintain expected frequencies

! procedures used in the 1990s
1990 PES: logistic regression modeling; Belin, Diffendal, Fay
1995, 1996, 1998: increasing use of imputation cell estimation
--- 1995, for residence status
--- 1996, for match status
--- 1998 (Dress Rehearsal), for all three statuses

! original possibilities for using logistic regression in the 2000 A.C.E.
--- Belin's program with coding changes
--- vendor software package
--- a new program written for the A.C.E.

! reasons for using imputation cell estimation in production for the A.C.E.
--- logistic regression options not acceptable
--- differences between logistic regression and imputation cell estimation
appear to be small, based on 1990 PES, 1995 and 1996

5. Verification

a) verifying that the programs are running correctly
--- production program, all components (except for characteristic imputation)
already double programmed
--- independent program being written

b) operational analyses
hundreds of computations and tables for review
statistical analyses done automatically, with unusual results raising flags

c) detailed analyses
additional tables also available

Adjustment for Noninterviews: The Procedure **Attachment 1**
Page 1 of 2

We define an interview and a noninterview. For the given reference date, that is, separately for census day and A.C.E. interview day,

> interview: a housing unit is an interview if there is at least one person (with name and at least two demographic characteristics) who possibly or definitely was a resident of the housing unit on the given reference date;
>
> noninterview: an occupied housing unit that is not an interview is a noninterview.

(An example on the next page illustrates these definitions and the procedure below.)

Procedure

1. Assign all occupied housing units to noninterview adjustment cells: block cluster × type of basic address (single-family home, apartment, or other). See note below on collapsing.

2. Assign *every* occupied unit a census-day interview status (interview or noninterview) *and* an A.C.E. interview-day interview status.

3. The noninterview adjustment factor is the ratio of the *weighted* number of housing-unit interviews and noninterviews to the weighted number of housing-unit interviews.

4. Compute the adjustment for census day; apply it to the person weights of non-movers and out-movers in interviewed housing units.

5. Compute the adjustment for A.C.E. interview day; apply it to the person weights of in-movers in interviewed housing units.

Notes:

- Although interview status is a housing-unit characteristic, mover status is a person characteristic. People in the same household can have different mover statuses.

- Vacant housing units do not contribute to the noninterview adjustment.

- People in noninterviewed housing units do not contribute to the components of DSE.

- If the *unweighted* number of interviewed housing units in a cell is less than half the number of

noninterviewed units, the cell is collapsed. Rules for collapsing are in Q-25.

Suppose a block cluster has nine housing units, all of the same type of basic address, for example, all single family homes, as depicted below.

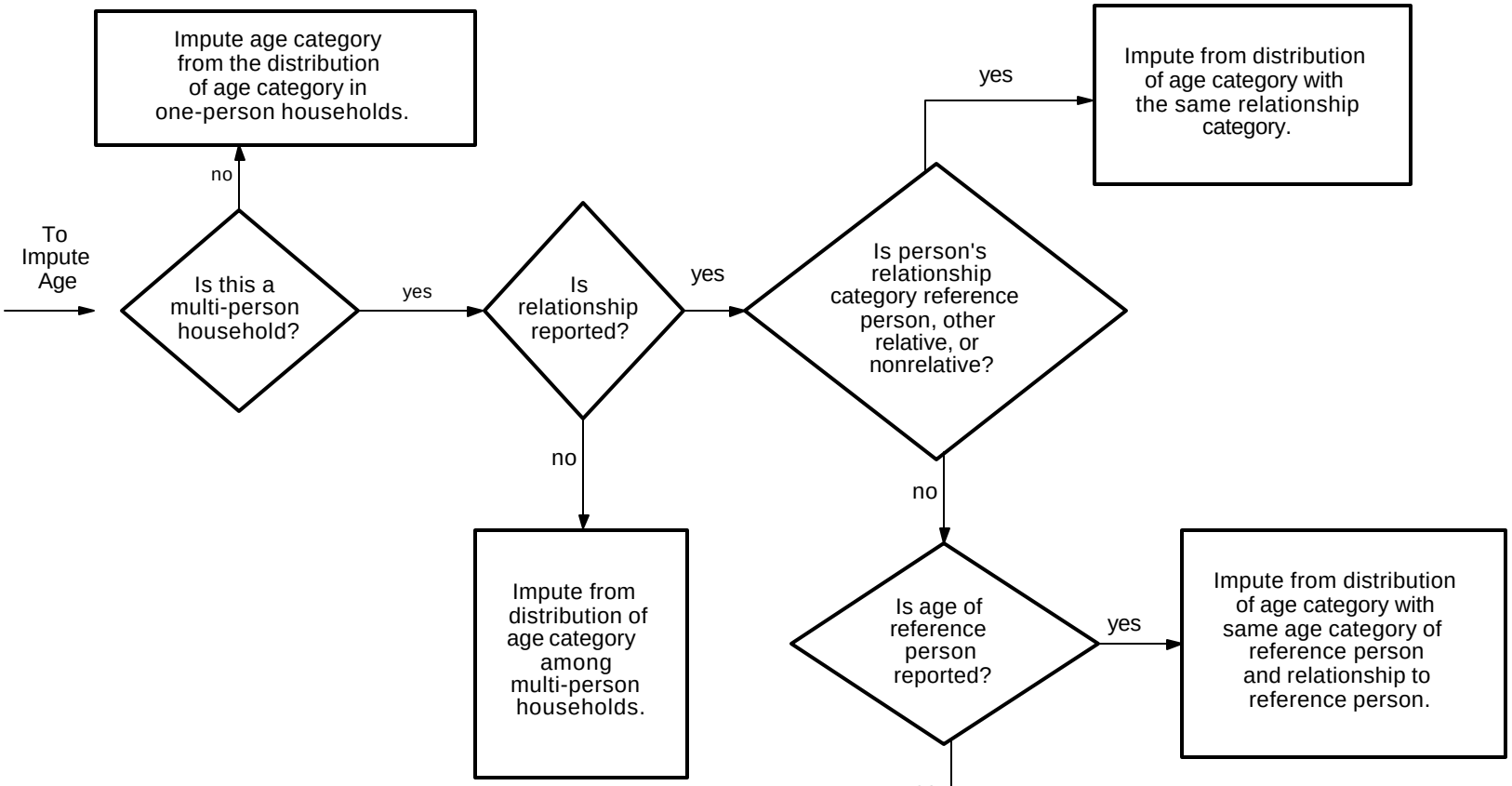
Housing Unit	Weight	Actual Situation	Status of (and Information from) A.C.E. Interview	Census-Day Status [1]	A.C.E. Interview-Day Status [1]
1	100	Resident on 4/1/00 and at time of A.C.E. intvw.	Interviewed in A.C.E.	I	I
2	100	Resident on 4/1 and at time of A.C.E. intvw.	Neighbor (proxy) interviewed in A.C.E.	I	I
3	100	Resident on 4/1 and at time of A.C.E. intvw.	No one interviewed in A.C.E.	NI	NI
4	100	Vacant on 4/1, resident at time of A.C.E. intvw.	Interviewed in A.C.E., knows of 4/1 status	Vacant	I
5	100	Vacant on 4/1, resident at time of A.C.E. intvw.	Intvw'd in A.C.E., no knowledge of 4/1 status	NI	I
6	100	Vacant on 4/1, resident at time of A.C.E. intvw.	No one interviewed in A.C.E.	NI	NI
7	100	Resident on 4/1, vacant at time of A.C.E. intvw.	Information obtained from proxy	I	Vacant
8	100	Resident on 4/1, vacant at time of A.C.E. intvw.	No info on 4/1 status; Census staff determines vacant at time of A.C.E.	NI	Vacant
9	100	Resident on 4/1, different resident at time of A.C.E.	Interviewed in A.C.E., knows of 4/1status	I	I

[1] Interview Status: I = interview, NI = noninterview

In this cluster × TBA (noninterview cell), to people in interviewed housing units, apply the following noninterview adjustments:

(1) to the person weights of non-movers and out-movers,
census-day NI adjustment = 800 / 400 = 2

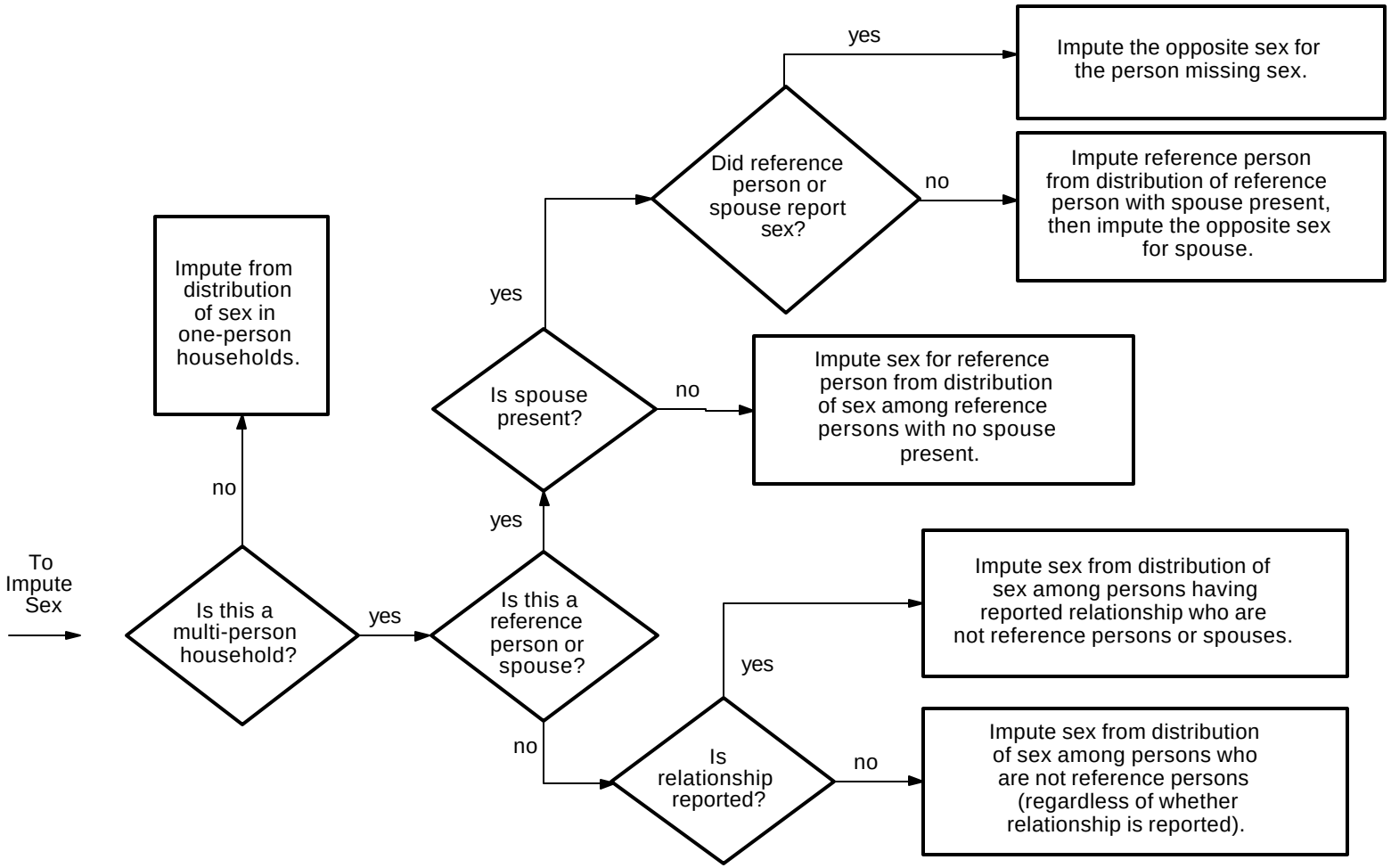
(2) to the person weights of in-movers,
A.C.E. interview-day NI adjustment = 700 / 500 = 1.4



Age is collapsed into the following categories for the purpose of imputation:
0 - 17, 18 - 29, 30 - 49, 50 and over.

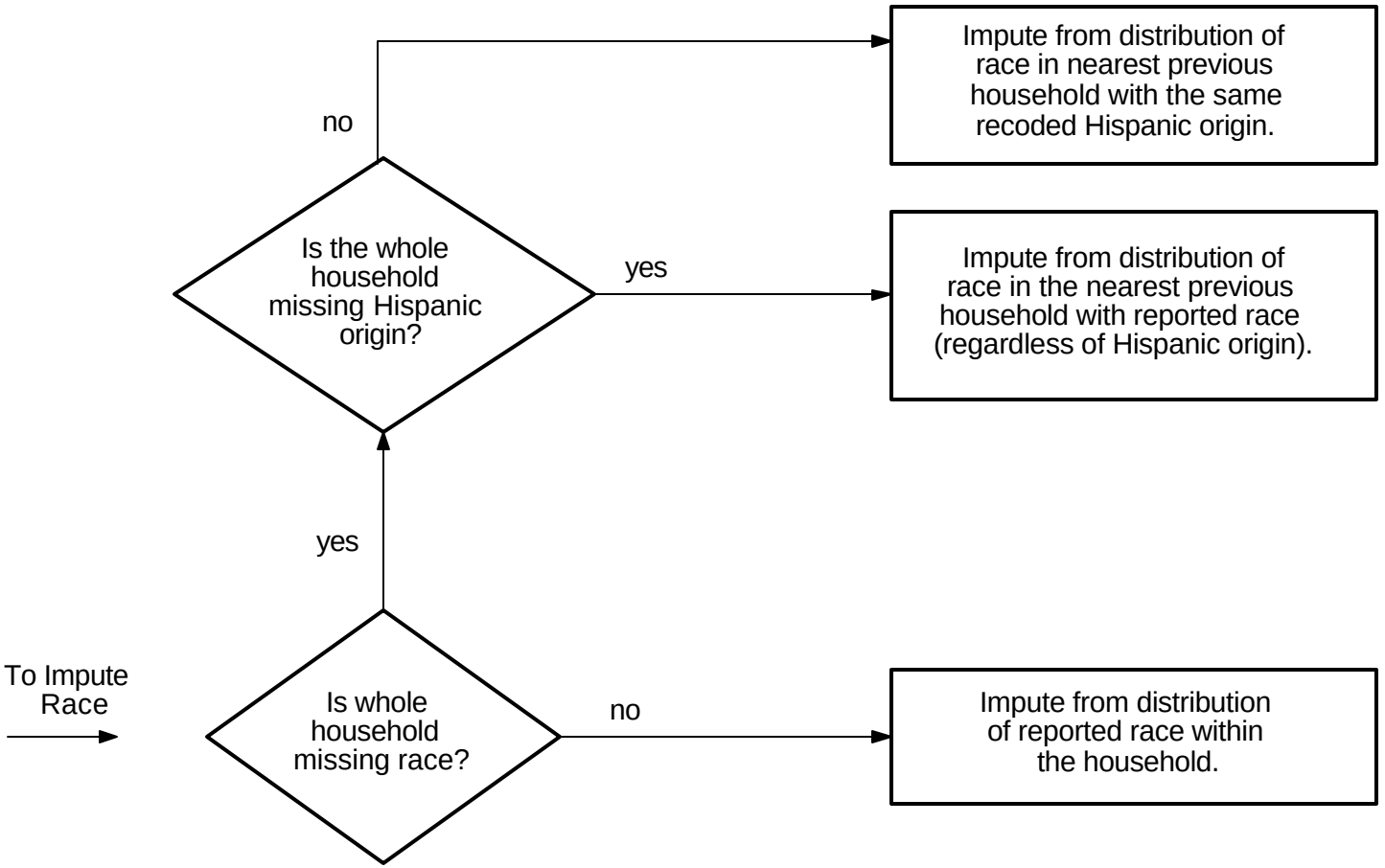
Relationship is collapsed into the following seven categories for the purpose of age imputation:
reference person, spouse, child, sibling, parent, other relative, and nonrelative.

Unless otherwise stated, all distributions are among multi-person households



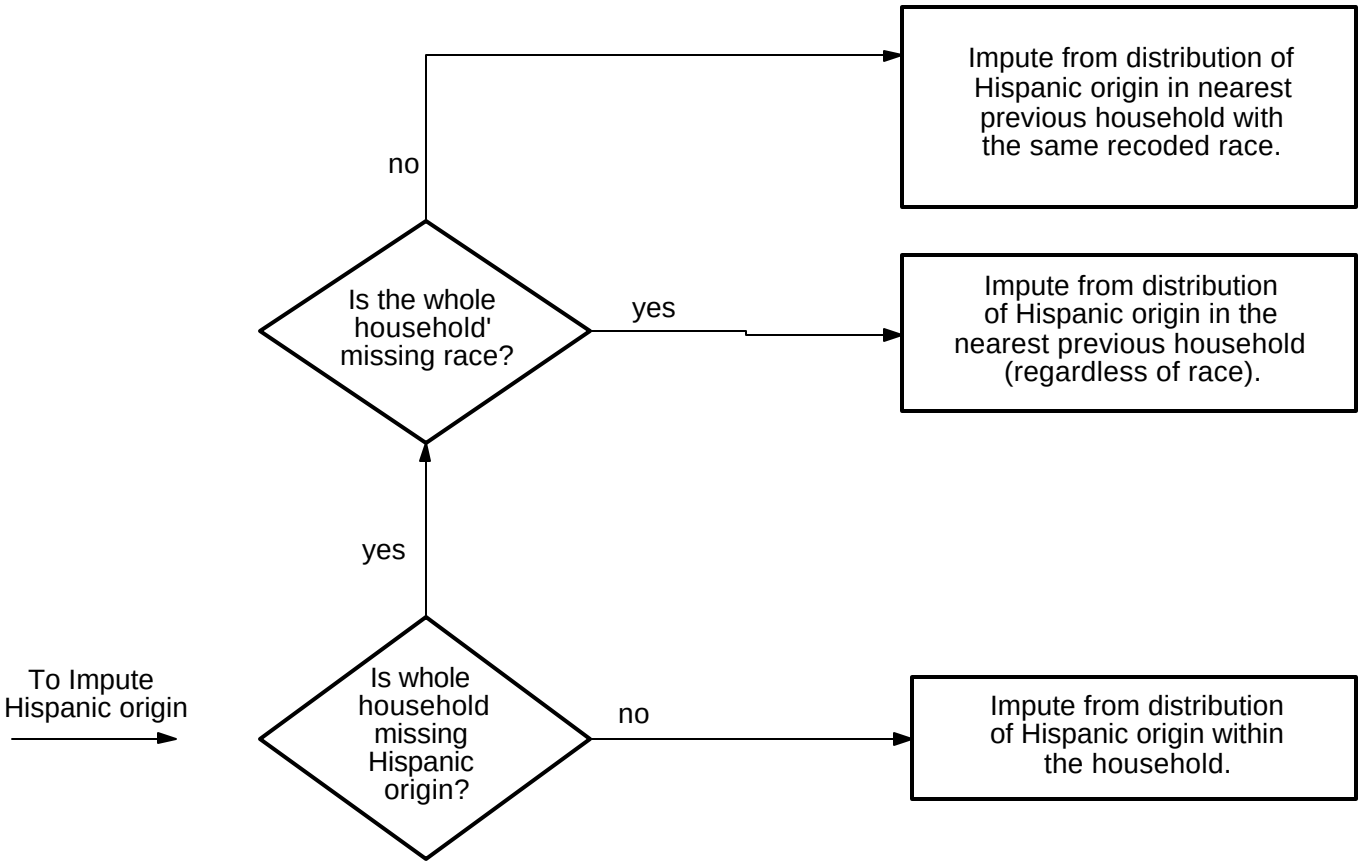
Relationship is collapsed into the following seven categories for the purpose of sex imputation: reference person, spouse, child, sibling, parent, other relative, and nonrelative.

Unless otherwise stated, all distributions are among multi-person households



Note: We can impute any of the 63 possible values of race.

A household's recoded Hispanic origin is the Hispanic origin of the first person on the household roster with reported Hispanic origin. Recoded Hispanic origin has three categories: Hispanic, Non-Hispanic, and missing (indicating the whole household is missing Hispanic origin).



Note: We collapse Hispanic origin to Non-Hispanic/Hispanic for imputation purposes.

A household's recoded race is the race of the first person on the household roster with reported race. Recoded race has four categories: (1) white only, (2) other or both white and other, (3) all reported races not covered by (1) or (2), and (4) missing (indicating that the whole household is missing race). Note that "other" here refers to a race written in by the respondent.

Imputation Cells for Unresolved Cases

Enumeration Status (E Sample)

United States

BFU Group	0 Imputes		1+ Imputes		Total Projected Frequency [1]
1 = Matches needing FU					4,600 (0.6%)
2 = Possible matches					4,600 (0.6%)
3 = Partial HH nonmatches	V3a [2]	V3b	V3a	V3b	35,000 (4.7%)
4 = Whole HH nonmatches (where HU matched); not conflicting HHs	nonhispanic white	other			48,000 (6.4%)
5 = Nonmatches from conflicting HHs; HU not in regular NRFU					
6 = Nonmatches from conflicting HHs; HU in regular NRFU					
7 = Whole HH nonmatches; HU did not match during HU matching	nonhispanic white	other			25,000 (3.4%)
8 = Persons resolved before FU	nonhispanic white	other			611,000 (81.5%)
9 = Persons with insufficient information for matching	(no unresolved cases)				21,600 (2.9%)
Total	600,000 (75 - 85 %)		150,000 (15% - 25 %) [3]		750,000 (100%)

[1] Based on the Dress Rehearsal (DR).

[2] V3a = age 18-29, relationship is child of HHer; V3b = other. In the DR about 12% of BFU Group 3 cases fell into the V3a cell.

[3] Based roughly on prior tests (15% in the Dress Rehearsal w/o Menominee; 20% in 1995; higher in Chicago, 1996).

Puerto Rico

BFU Group	Total Projected Frequency
1 = Matches needing FU	225 (0.6%)
2 = Possible matches	225 (0.6%)
3 = Partial HH nonmatches	1,750 (4.7%)
4 = Whole HH nonmatches (where HU matched); not conflicting HHs	2,400 (6.4%) (in 4, 5, and 6 combined)

5 = Nonmatches from conflicting HHs; HU not in regular NRFU	
6 = Nonmatches from conflicting HHs; HU in regular NRFU	
7 = Whole HH nonmatches; HU did not match during HU matching	1,275 (3.4%)
8 = Persons resolved before FU	30,550 (81.5%)
9 = Persons with insufficient information for matching	1,100 (2.9%) (no unresolved cases)
Total	37,500 (100%)

Imputation Cells for Unresolved Cases

Match Status (P Sample)

Note 1: In the Dress Rehearsal, about 96% of all unresolved matches were people with insufficient information for matching. These cases are *not* followed up. 90% or more of them did not have a valid name; their imputation rates were higher than other cases for race (27.3%), hispanic origin (21.6%), sex (26.2%), age (45.6%), and probably tenure. (Data from Menominee are excluded from these rates.)

Note 2: In both tables, to get a rough idea of the frequencies, it is assumed that all characteristics are independent; we realize that the characteristics may be strongly correlated. Further, the proportions and correlations for Puerto Rico are likely to be different from those of the U.S.

United States

Mover Status	Address Code				Total
	1 = HU Match from initial matching		2 = HU Nonmatch or conflicting household		
Non-mover	0 imputes 601,000	1+ imputes 19,000 [1]	0 imputes 90,000	1+ imputes 3,000 [1]	713,000 (95%)
Mover	0 imputes 32,000	1+ imputes 1,000 [1]	5,000		38,000 (5% [2])
Total	653,000 (87%)		98,000 (13% [3])		750,000

[1] Frequency of 3% - 4% suggested by data from the Dress Rehearsal.

[2] Suggested by Dress Rehearsal data; may be higher due to added time before A.C.E. field operations.

[3] Suggested by current (2000 A.C.E.) HU matching, early results.

Puerto Rico

Mover Status	Address Code		Total
	1 = HU Match from initial matching	2 = HU Nonmatch or conflicting household	
Non-mover	31,000	4,625	35,625 (95%)
Mover	1,625	250	1,875 (5% [2])
Total	32,625 (87%)	4,875 (13% [3])	37,500

Imputation Cells for Unresolved Cases

Residence Status (P Sample)

Note 1: Residence status was treated differently in the 1990 PES, because mover procedure B was used.

Note 2: In the U.S. and Puerto Rico, for people with insufficient information for matching (BFU Group 7): within tenure (× race groups, for the U.S.), derive the weighted residence proportion across BFU Groups 1 - 5; that is, discount cases in Group 6 (those resolved before follow-up)

United States

BFU Group	Owner				Non-Owner				Total Projected Frequency [1]
	Nonhispanic White		Other		Nonhispanic White		Other		
1 = Matches needing FU									4,800 (0.6%)
2 = Possible matches									4,800 (0.6%)
3 = Partial HH nonmatches	V3a	V3b	V3a	V3b	V3a	V3b	V3a	V3b	35,600 (4.7%)
4 = Whole HH nonmatches needing FU (not conflicting HHs)									24,400 (3.3%)
5 = Nonmatches from conflicting HHs needing FU									
6 = Persons resolved before FU									673,000 (89.7%)
7 = Persons with insufficient information for matching (note: no resolved cases)	weighted average over BFU groups 1 - 5		weighted average over BFU groups 1 - 5		weighted average over BFU groups 1 - 5		weighted average over BFU groups 1 - 5		7,600 (1.0%)
Total									750,000 (100%)

[1] Based very roughly on Dress Rehearsal sites, Sacramento and South Carolina

Puerto Rico

BFU Group	Owner	Non-Owner	Total Projected Frequency [1]
1 = Matches needing FU			225 (0.6%)
2 = Possible matches			225 (0.6%)
3 = Partial HH nonmatches			1,750 (4.7%)
4 = Whole HH nonmatches needing FU (not conflicting HHs)			1,250 (3.3%)
5 = Nonmatches from conflicting HHs needing FU			

6 = Persons resolved before FU			33,650 (89.7%)
7 = Persons with insufficient information for matching (note: no resolved cases)	wgt. avg. over BFU groups 1 - 5	wgt. avg. over BFU groups 1 - 5	375 (1.0%)
Total			37,500 (100%)

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 10

June 28, 2000

Prepared by: Maria Urrutia and Annette Quinlan

The tenth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on June 28, 2000 at 10:30. The agenda for the meeting was A.C.E. missing data.

Persons in attendance:

Kenneth Prewitt
Nancy Potok
Paula Schneider
Cynthia Clark
John Thompson
Jay Waite
Bob Fay
Sally Obenski
John Long
Susan Miskura
Raj Singh
Tommy Wright
Patrick Cantwell
Rita Petroni
Gregg Robinson
Carolee Bush
Maria Urrutia
Annette Quinlan

I. A.C.E. Missing Data

This was the detailed presentation to the ESCAP on missing data. The missing data procedures have been finalized and the purpose of the meeting was to update the ESCAP.
Pat Cantwell presented the results and a summary document is attached. The detailed missing data procedures may be found in the DSSD Memorandum Series Chapter Q-25.

The highpoints of the missing data discussion were as follows:

- Missing data may occur in three areas of the A.C.E.: noninterviews, item nonresponse, and status.

- They are addressed using the following basic methodology:

  (1) Noninterviews are handled through two weighting adjustments, one applied to in-movers and the other to out-movers and non-movers.

  (2) Item nonresponse is addressed through imputation. We use a hot deck approach to impute for tenure, race, and Hispanic origin. Age category and sex are imputed using distributions based on responses.

  (3) Three types of status can remain unresolved even after all person DSE follow-up is complete:

  a) Enumeration status for E-Sample persons - whether the person was correctly or erroneously enumerated in the census.
  b) Match status for P-Sample persons - whether the person matched to someone enumerated in the census.
  c) Residence status for P-Sample people - whether the person was a resident at that address on Census Day.

  For people with unresolved status, we use an imputation cell procedure whereby resolved and unresolved persons are allocated to cells according to their operational or other characteristics. Unresolved persons are given a probability of Enumeration, Match, or Residence status equal to the weighted proportion among the resolved cases in the cell.

  In 1990 we used a logistic regression modeling approach to address cases with unresolved status. This approach also resulted in the assignment of probabilities for the categories described above. We made this change because the cell method will perform adequately for the assignment of probabilities, and offers the advantages of operational efficiency and more straightforward validation.

II. Next Meeting

The next meeting scheduled for Wednesday July 12, 2000 will discuss Correlation Bias.

ESCAP Committee

William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson, Chair
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura

cc:

Kenneth Prewitt
Teresa Angueira
Bill Bell
Debbie Bolton
Genny Burns
Carolee Bush
Gerald Gates
Ed Gore
Dave Hubble
Donna Kostanich
Ellen Lee
Charlene Leggieri
Don Malec
Betsy Martin
Catherine Miller
Fay Nash
Sally Obenski
Miguel Perez
Ed Pike
Magdalena Ramos
Gregg Robinson
Raj Singh
Maria Urrutia
Signe Wetrogen
David Whitford
Henry Woltman
Tommy Wright

Kathleen P Zveare
07/11/2000 02:39 PM

To: Margaret A Applekamp/DIR/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Elizabeth Centrella/DSSD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Angela Frazier/DMD/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Jane F Green/DSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC
cc: J Gregory Robinson/POP/HQ/BOC@BOC, Rita J Petroni/PRED/HQ/BOC@BOC
Subject: Agenda for 7/12 ESCAP Meeting

The agenda for the July 12 ESCAP Meeting is as follows:

1. Overview of Correlation Bias---Howard Hogan

2. How Do We Measure Correlation Bias Using DA---Greg Robinson

3. Correlation Bias for Evaluation Purposes--Rita Petroni

Presentation on Demographic Analysis and Measurement of "Correlation Bias"

July 12, 2000

- **The Method of Demographic Analysis**

  Examples of components
  Historical record of DA coverage measurements

- **Differences of DA and Survey coverage results that point to "correlation bias"**

  Differences in net undercount rates
  Differences in sex ratios

- **Challenges for DA coverage measurement in Census 2000**

  Measurement of uncertainty
  Accommodation to new race question (mark one or more)
  Greater focus on sex ratios

What is Demographic Analysis?

, Population (<65)= Births (since 1935)
- Deaths (to persons born after 1935)

+Immigrants (born after 1935)
- Emigrants (born after 1935)

, Population (65+)= Medicare Count
+ Estimated unenrolled

Table 1: Illustrative Values of DA Components for the Estimated U.S. Resident Population, April 1, 2000 (Numbers in Millions)

Component	Age in 2000				
	All Ages	Under 15	15-44	45-64	65+
Total	280.0	60.0	117.0	68.1	35.0
Under age 65:					
Births	235.0	59.0	112.0	64.0	-
Deaths	-14.0	-0.7	-5.0	-8.3	-
Immigrants	29.0	2.0	12.0	15.0	-
Emigrants	-5.0	-0.3	-2.0	-2.7	-
Ages 65+:					
Medicare	35.0	-	-	-	35.0

Figure 1. Percent Net Undercount by Race
1940-1990

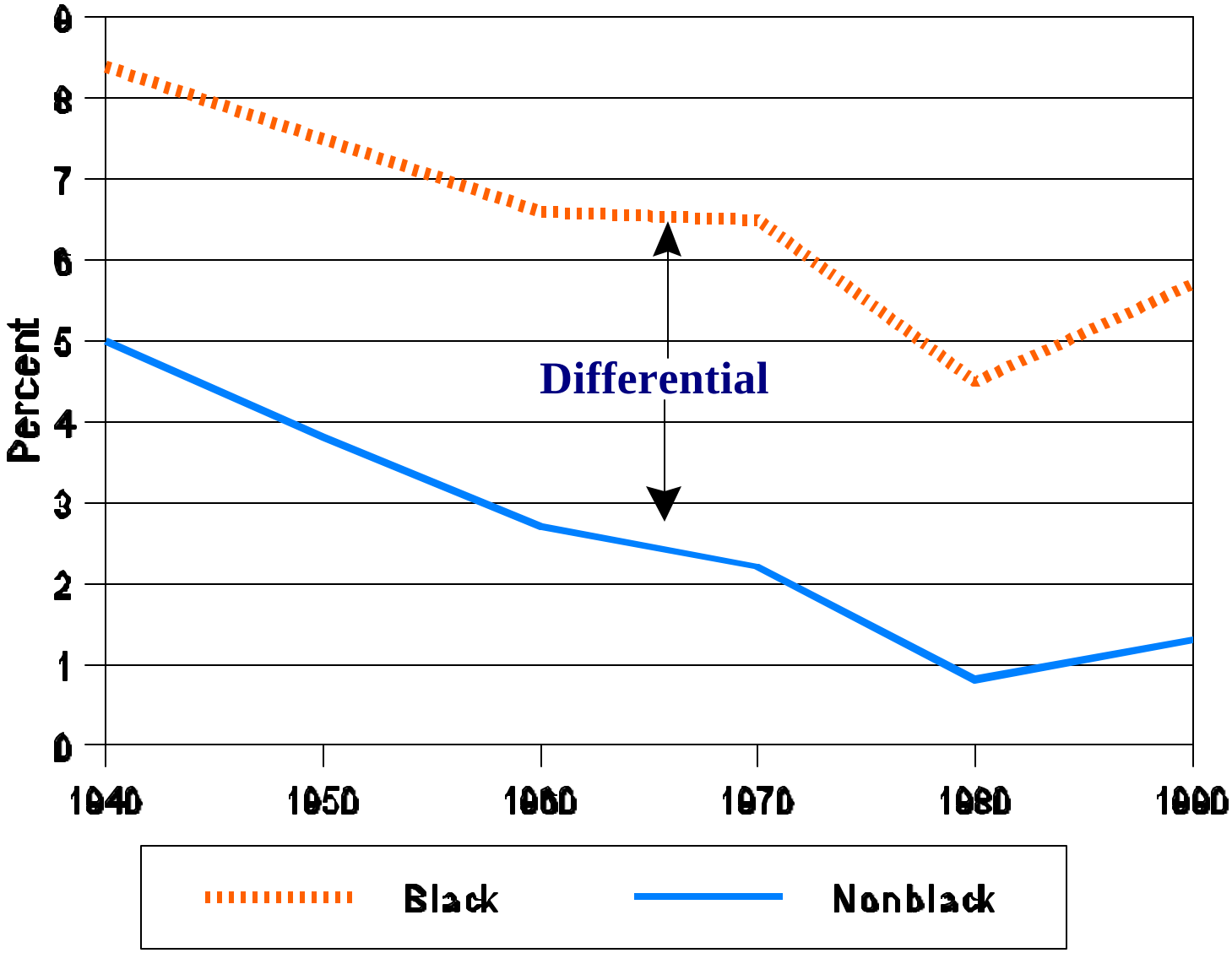
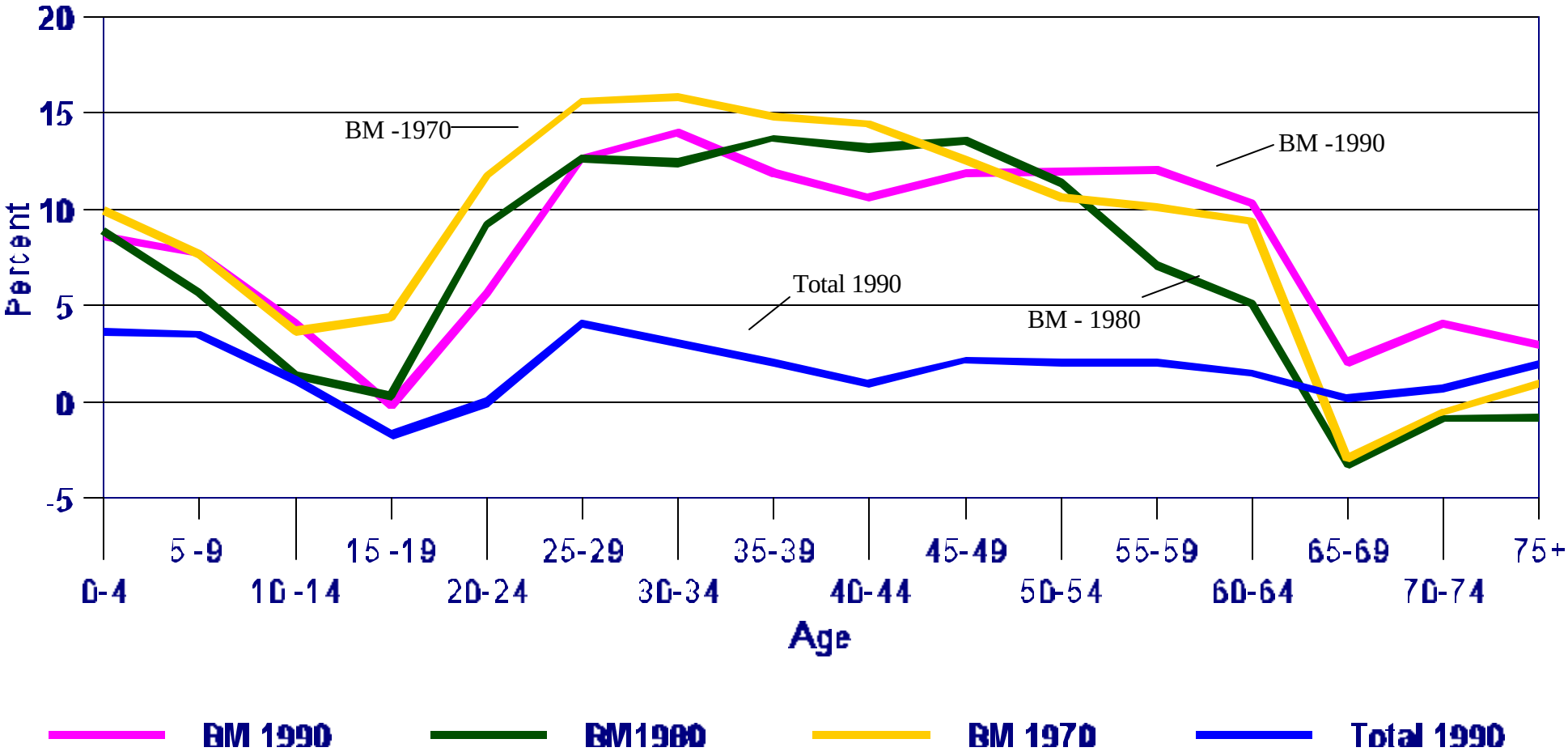
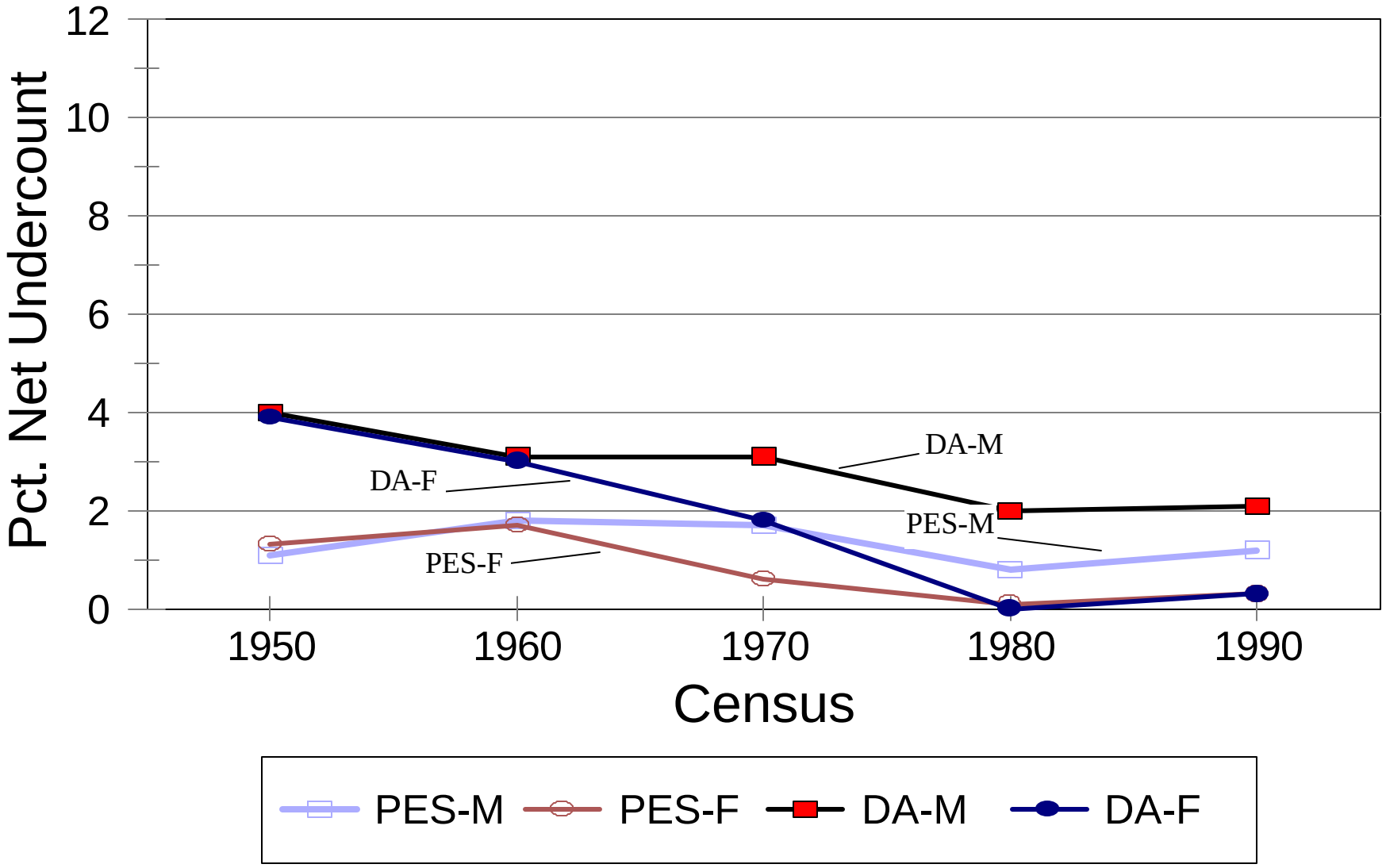


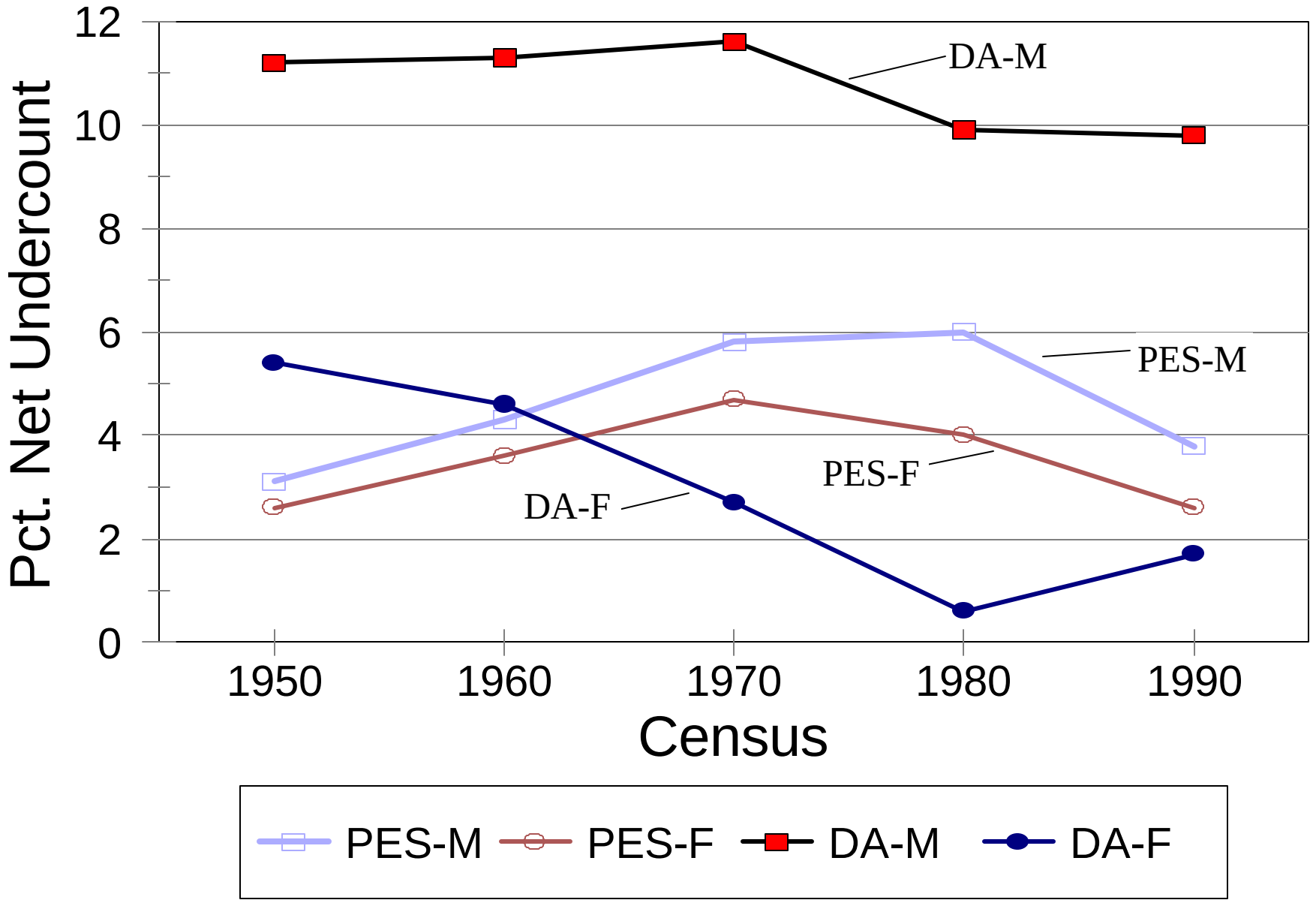
Figure 2. Percent Net Undercount: Black Males 1970 - 1990, and Total, 1990



DA and Survey Estimates of Net Undercount for Nonblack Adults, by Sex

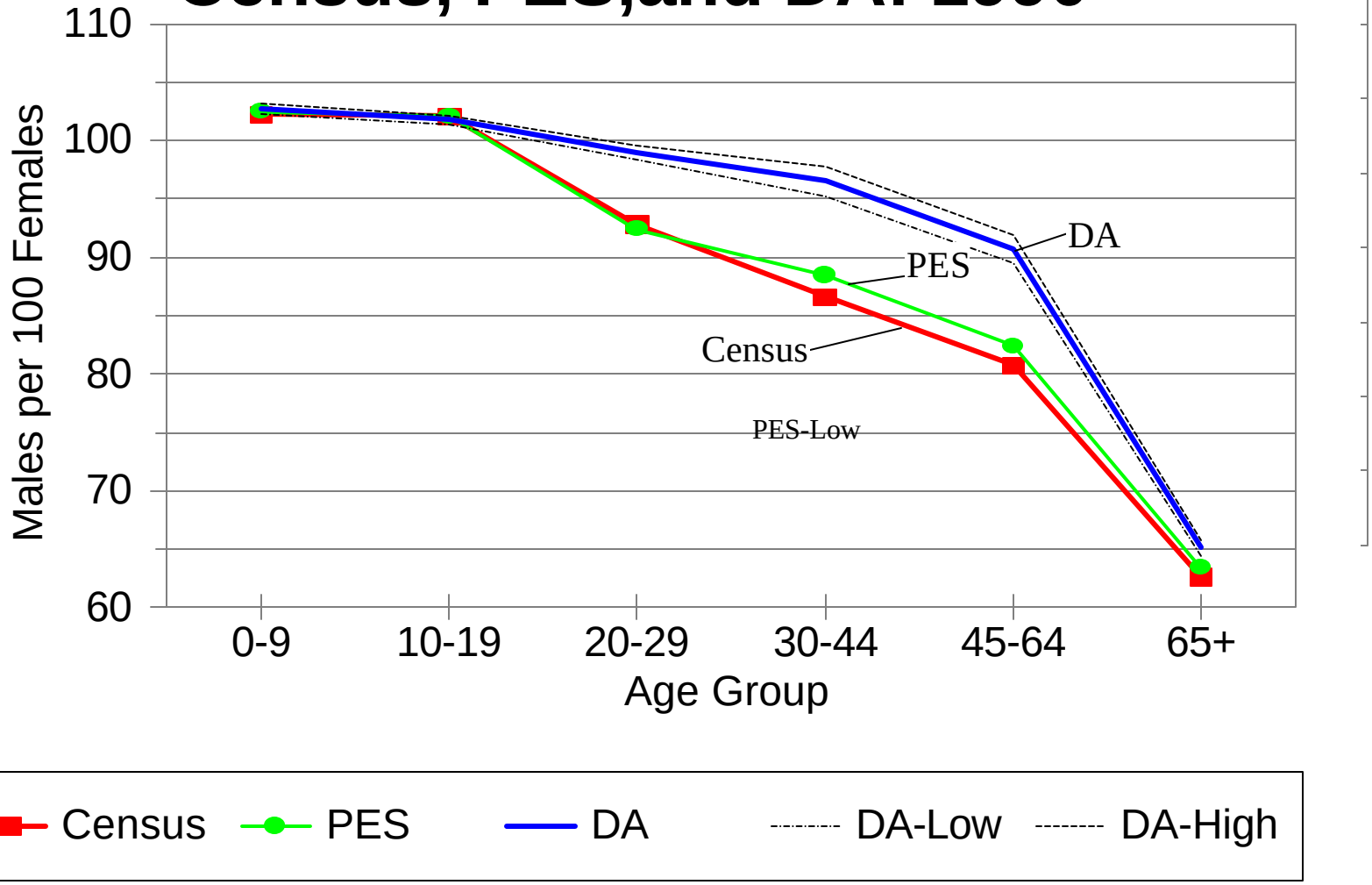


DA and Survey Estimates of Net Undercount for Black Adults, by Sex

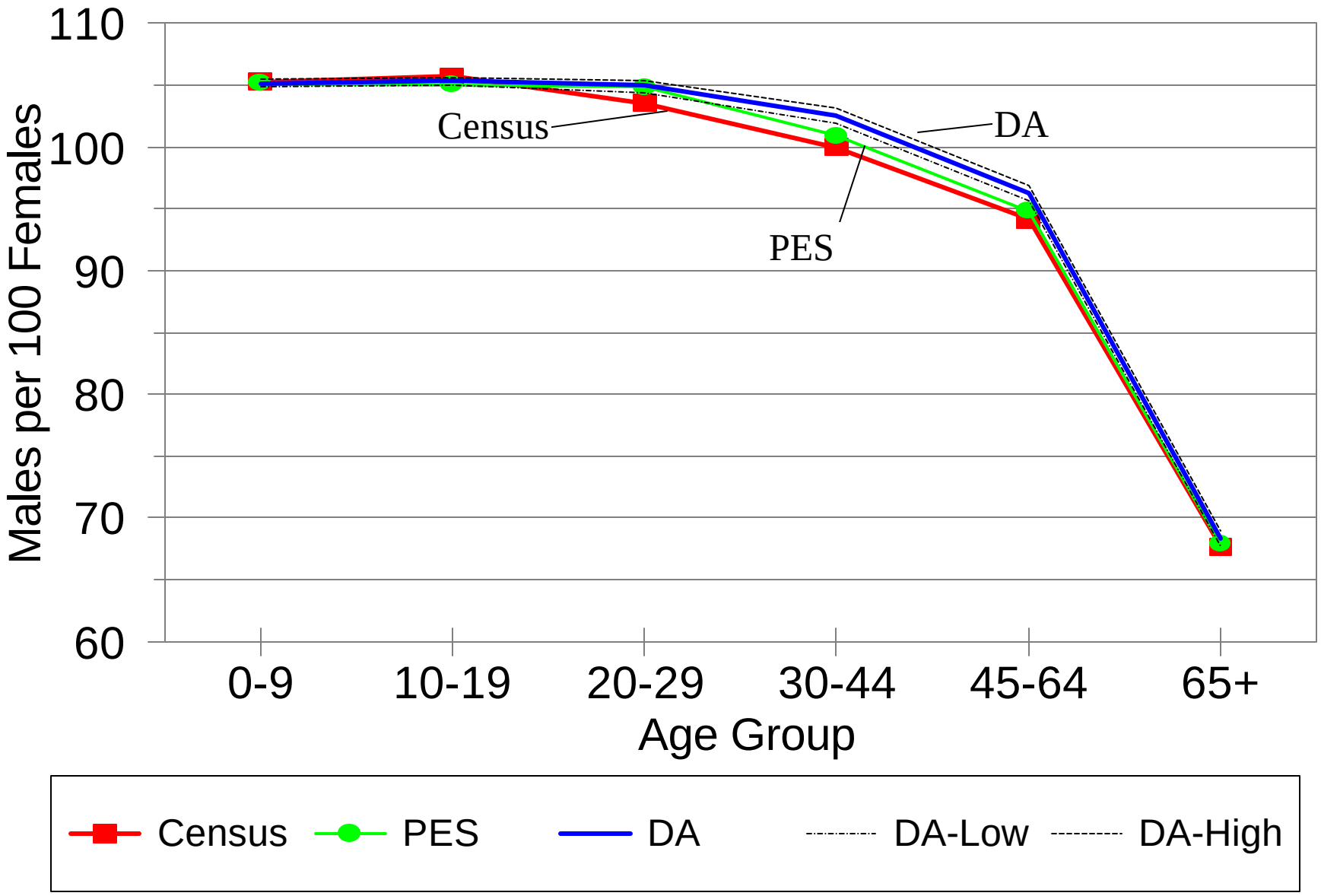


1990 Undercount Uncertaintv Intervals

Comparison of Sex Ratios for Blacks Census, PES,and DA: 1990



Comparison of Sex Ratios for Nonblacks Census, PES,and DA: 1990



Sex Ratios for Black Adults
Census, Surveys, and DA: 1960-1990

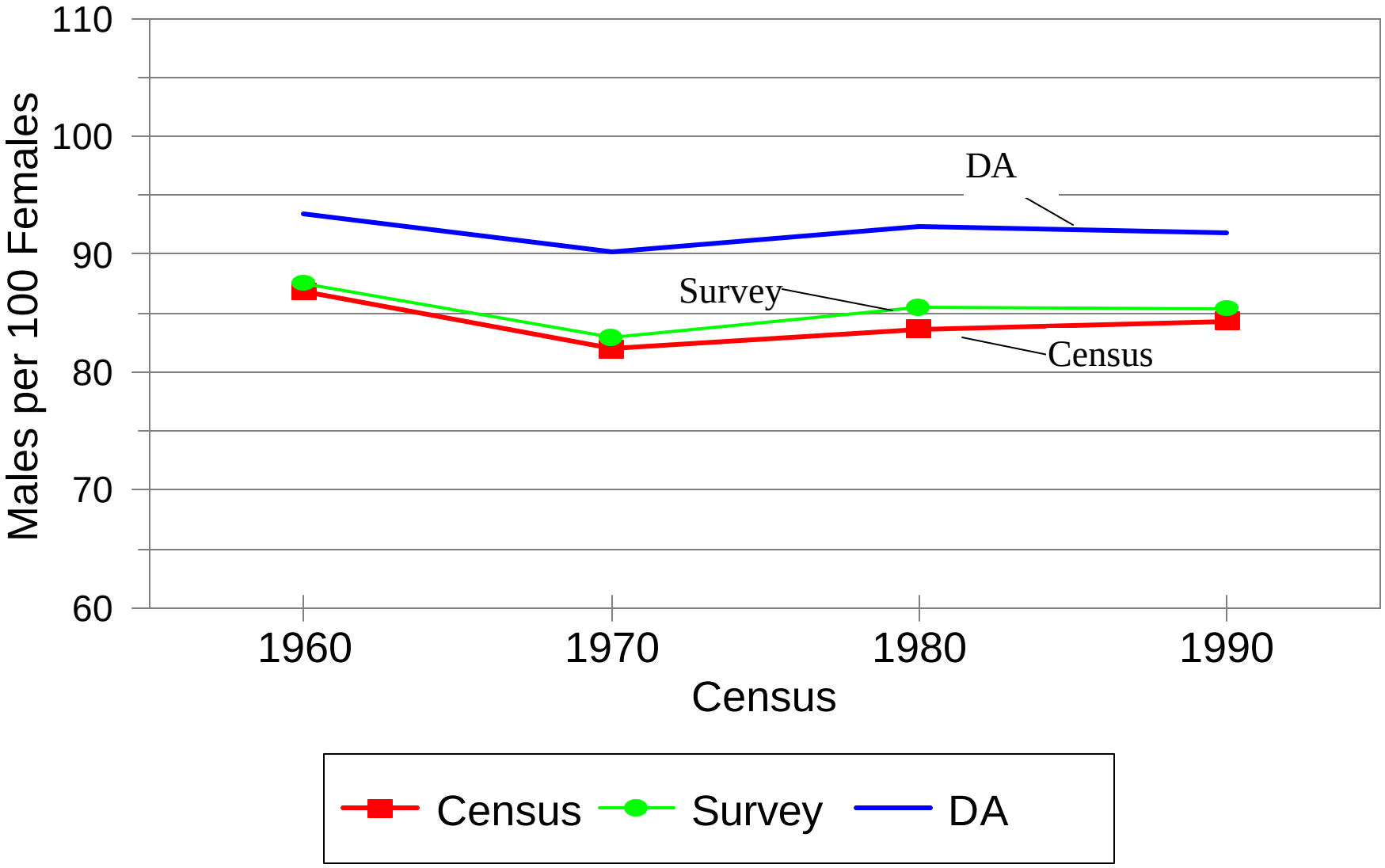
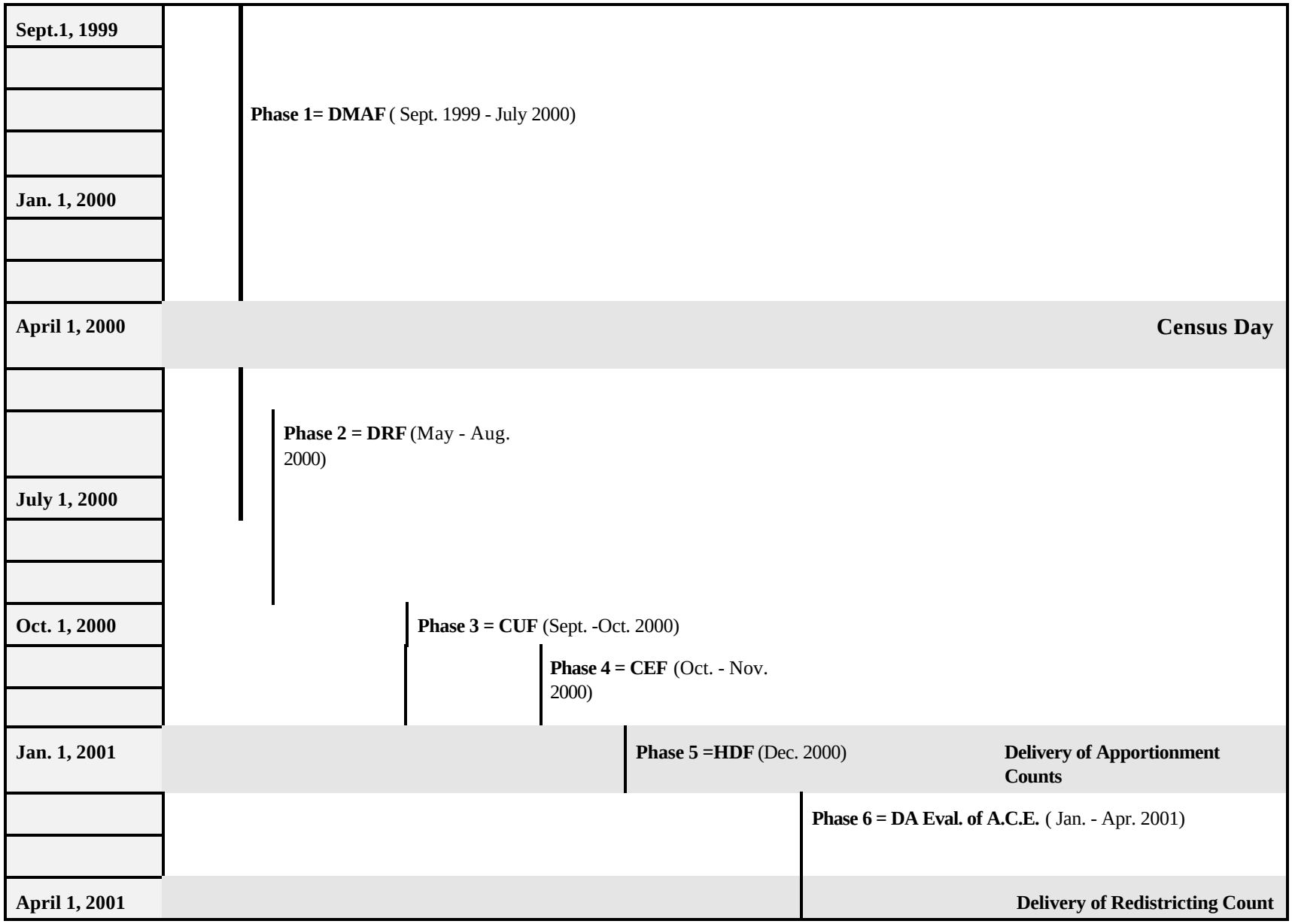


Figure 10
Sequence of Activities for Census 2000 Demographic Analysis Program

Time Line



November 3, 1997

MEMORANDUM FOR: Distribution List

FROM: J. Gregory Robinson
Chief, Population Analysis and Evaluation Staff

SUBJECT: **The Differential Undercount of Adult Black Men: Is it a Myth?**

A major goal of Census 2000 is to reduce the differential undercount. The most widely recognized differential is the chronic high undercount of adult black men. For the past six census the undercount rate of adult black males has ranged from 10 to 12 percent. These rates are approximately 8 percentage points higher than the overall undercount.

This statement is made from the perspective of Demographic Analysis (DA). Past surveys, including the 1990 PES, have detected coverage differences but not nearly to the extent of DA. If we want a "one number census" that is demographically sound, then steps must be taken to ensure that the survey estimates of undercount approximate the levels suggested by DA.

The Problem

We plan to use Dual System Estimates (DSE) as a vehicle for eliminating differential undercounts in the 2000 census. However, historical survey estimates give little assurance that the large undercounting of Black men will be observed or remedied. As demonstrated below, the survey approach has never measured an undercount of adult black men that is significantly higher (both statistically and realistically) than that of black women or black children.

Table 1. **Percent Net Undercount: Survey Estimates**

		Black	Black	Black Adults	
Year	Total	Total	Children	Male	Female
1990	1.6	4.4	7.0	**3.8**	2.6
1980	1.4	6.0	7.6	**6.0**	4.0
1970	2.3	6.2	7.8	**5.8**	4.7
1960	1.8	3.6	3.1	**4.3**	3.6
1950	1.4	3.2	3.9	**3.1**	2.6

Source: Appendix Table 1

The 1990 PES is a case in point. While the total Black undercount of 4.4 percent was significantly higher than the overall total of 1.6 percent, the Black undercount was most pronounced for children (7.0 percent)--not adult men. The adult Black male (3.8) and female (2.6) percent undercounts were

lower and the male-female difference is not statistically significant.

Where did the media and other observers arrive at the conclusion that black men are severely undercounted? The answer is its simply the profile of net undercount measured by demographic analysis:

Table 2. **Percent Net Undercount: DA Estimates**

		Black	Black	Black Adults	
Year	Total	Total	Children	Male	Female
1990	1.8	5.7	5.9	**9.8**	1.7
1980	1.2	4.5	3.7	**9.9**	0.6
1970	2.7	6.5	5.9	**11.6**	2.7
1960	3.1	6.6	5.0	**11.3**	4.6
1950	4.1	7.5	6.4	**11.2**	5.4
1940	5.4	8.4	7.5	**12.0**	6.5

Source: Appendix Table 1

DA shows a disproportionately high undercount for adult black males--9.8 percent to 12.0 percent nationally (!) in every census since 1940. The undercount rate of children is also relatively high, but only one-half the size of the black male undercount. The net undercounts of black women are moderate in comparison.

There are two reasons why the survey estimates show such a different pattern for adults. The persistent understatement of black men is attributable to the "correlation bias" problem. That is, many persons missed by the census are not being picked up in the survey interview, leading to an understatement of the measured undercount. The overstatement of the net undercount of black women (relative to DA) is less well-known and understood.

Appendix Table 1 provides additional detail on the net undercount estimates for Nonblacks. The DA and survey estimates do not differ as much as for Blacks. In fact, the two approaches measure the same small undercount for adult nonblack females. However, the survey results consistently understate the undercount of adult nonblack men (relative to DA).

A Final Point: Apart from compensating for correlation bias, DA provides a basis to "smooth" age-sex anomalies in the survey estimates. The most obvious example is where the 1990 PES measured a larger undercount of Black women aged 18-29 (5.5 percent) than Black men (3.6 percent). That result is completely contradicted by the 1990 DA estimates (Black female = 2.9, males = 7.7) and the historical record. The detailed age structure of the PES estimates also failed consistency standards.

The Solution: Incorporate Demographic Analysis (DA)

Demographic analysis is the standard for describing historical trends in coverage and differentials by age, sex, and race. The current census 2000 plan--that relies on survey estimation alone for ICM--runs the risk of failing to reduce the differential undercount in a demographically consistent manner. In particular, DA could expose the failure of the 2000 census to reduce the adult Black male undercount. The incorporation of DA into the ICM process can help ensure our goal of a "one number census that is right the first time". If DA is not used, we must spell out the specific improvements to the survey methodology that will render moot the need for DA.

cc:

Schneider (Dir)	Killion (DSSD)	Wright (SRD)	Long (POP)
Thompson	Vacca	Singh	Wetrogan
Bounpane	Whitford	Bell	Miller
Marx	Waltman	Isaki	Del Pinal
Waite	Griffin	Petroni	Spencer
Fay	Schindler	Weiler (FLD)	Hollmann
Mckenney	Haines	Blass	Das Gupta
			West
Hogan (SVSD)			Word
			Robinson

APPENDIX TABLE 1: COMPARISON OF ESTIMATES OF PERCENT NET UNDERCOUNT BASED ON SURVEY AND DEMOGRAPHIC MEASUREMENT APPROACHES: 1940 - 1990

Census	Coverage Evaluation Program	Total			Black			Nonblack		
		Both Sexes	Black	Non-black	Children	Adults Male	Adults Female	Children	Adults Male	Adults Female
		(1)	(2)	(3)	(4)	(5)	(6)		(7)	(8)
A. Survey approach										
1990	Post Enumeration Survey	1.6	4.4	1.2	7.0	3.8	2.6	2.5	1.2	0.3
1980	Post Enumeration Program	1.4	6.0	0.8	7.6	6.0	4.0	1.5	0.8	0.1
1970	CPS/Census Match	2.3	6.3	1.8	7.8	5.8	4.7	2.1	1.7	0.6
1960	Survey Coverage Study	1.8	3.6	1.6	3.1	4.3	3.6	1.2	1.8	1.7
1950	Post Enumeration Survey	1.4	3.2	1.2	3.9	3.1	2.6	1.1	1.1	1.3
1940	N/A									
B. Demographic approach										
1990	Demographic Analysis	1.8	5.7	1.3	5.9	9.8	1.7	1.6	2.1	0.3
1980	Demographic Analysis	1.2	4.5	0.8	3.7	9.9	0.6	0.3	2.0	0.0
1970	Demographic Analysis	2.7	6.5	2.2	5.9	11.6	2.7	1.8	3.1	1.8
1960	Demographic Analysis	3.1	6.6	2.7	5.0	11.3	4.6	2.0	3.1	3.0
1950	Demographic Analysis	4.1	7.5	3.8	6.4	11.2	5.4	3.3	4.0	3.9
1940	Demographic Analysis	5.4	8.4	5.0	7.5	12.0	6.3	6.3	4.9	4.7
C. Method difference (= A - B)										
1990	Survey minus DA	-0.2	-1.3	-0.1	1.1	-6.0	0.9	0.9	-0.9	0.0
1980	Survey minus DA	0.2	1.5	0.0	3.9	-3.9	3.4	1.2	-1.2	0.1
1970	Survey minus DA	-0.4	-0.2	-0.4	1.9	-5.8	2.0	0.3	-1.4	-0.2
1960	Survey minus DA	-1.3	-3.0	-1.1	-1.9	-7.0	-1.0	-0.8	-1.3	-0.3
1950	Survey minus DA	-2.7	-4.3	-2.6	-2.5	-8.1	-2.8	-2.2	-2.9	-2.6

Note:

Demographic analysis estimates represent percent net undercoverage for all years. Estimates for 1940-80 represent revised estimates that are consistent with the methodology and components used to produce the 1990 demographic estimates.

Survey-based estimates represent percent net undercoverage for all years. Survey estimates for 1960 and 1950 refer to Black-and-other-races instead of Black. Estimates for 1980 represent a composite of 9 sets of estimates (sets 14-20, 14-9, and 14-8 are excluded). Approximate sample sizes of estimates: 1990 - 144,000 interviewed households; 1980 - 84,000 CPS households for P-sample and 110,000 census households for E-sample; 1970 - 45,000 interviewed households; 1960 - about 35,000 households; 1950 - about 25,000 households.

Adult - DA estimates refer to population 18 and over in 1990 and 20 and over in 1950-1980; survey estimates for 1990 (PES) refer to population 18 and over in 1990, 20 and over in 1980 (PEP), and 15 and over in 1950-1970.

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 11

July 12, 2000

Prepared by: Maria Urrutia and Annette Quinlan

The eleventh meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on July 12, 2000 at 10:30. The agenda for the meeting was A.C.E. correlation bias.

Persons in attendance:

Kenneth Prewitt
William Barron
Nancy Potok
Paula Schneider
Nancy Gordon
John Thompson
Jay Waite
Howard Hogan
John Long
Susan Miskura
Donna Kostanich
Fay Nash
Tommy Wright
David Hubble
Rita Petroni
Gregg Robinson
Carolee Bush
Maria Urrutia
Annette Quinlan

I. A.C.E. Correlation Bias

The purpose of this meeting was to describe and discuss correlation bias with the ESCAP. Howard Hogan provided a general discussion of correlation bias and Gregg Robinson discussed methods to measure correlation bias based on Demographic Analysis.

Howard Hogan defined the two causes of correlation bias, causal dependence and heterogeneity, the handout is attached. Causal dependence occurs when the event of being included in the census affects a person's chance of being included in the A.C.E., or vice versa. Correlation bias also results from heterogeneity of the probabilities of being included in either the census or the coverage measurement surveys. The theoretical assumptions underlying the DSE do not require that the initial census and the coverage measurement survey have the same probability of including people. However, DSE does assume that there are not groups of the population within post-strata that have different inclusion probabilities for both the initial census and the coverage measurement survey. When this situation occurs the DSE will be subject to correlation bias and will understate the "true" population totals. The attachment describes examples of this phenomenon.

Gregg Robinson provided an overview of the Demographic Analysis (DA) program, which included a description of the DA method, the major finding of DA measurements of net undercount trends and differentials from previous censuses, and how the coverage patterns based on DA estimates compare to coverage patterns measured by previous census coverage measurement surveys (such as the 1990 PES) and how DA provides a measurement of correlation bias. Differences in the DA and DSEs in 1990 and 1980 for some age-sex groups (such as adult Black men between the ages of 18 and 29) provide measures of "correlation bias."

The method of DA relies on aggregate administrative records, which are essentially independent of the census. The DA estimates for the population under 65 years of age in 2000 (born after 1935) are based on the compilation of historical data or estimates of births, deaths, immigrants, and emigrants. Administrative Medicare data are used to estimate the population 65 and over. Limitations of the DA estimates were also discussed, including problems with estimating some of the components (e.g., undocumented immigrants), the inability to provide coverage estimates for detailed race/ethnic groups, and how DA only provides estimates at the national level.

DA has been used over the years to describe historical trends in coverage differentials by age, sex, and race (Black, Nonblack). It has provided a consistent tracking system by which the percent undercount rates can be compared from decade to decade. DA has measured a persistent and disproportionate undercount of adult Black men and Black children in the censuses of 1940 to 1990. The net undercount of adult Black men during the 1940 - 1970 censuses exceeded 10 percent nationally and during the 1980 and 1990 censuses was approximately 10 percent nationally.

A table and figures were also presented which compared the DA and coverage measurement survey results of each census since 1950. While the two methods have been in close agreement regarding the overall net undercount in the most recent censuses (e.g, the PES measured a net undercount of 1.6 percent in 1990; DA measured a slightly higher rate of 1.8

percent), certain differences emerge in the comparison of estimates dis-aggregated by age, sex, and race. In particular, the survey net undercount estimates for adult Black men are substantially lower than the corresponding DA estimates. The difference remains even after accounting for the "uncertainty" in the measured undercounts.

The Demographic Analysis estimates demonstrate that the coverage measurement survey results have consistently understated the undercount of adult Black men. This is the empirical evidence of "correlation bias". DA sex ratios have been shown to be less subject to uncertainty than the DA "point" estimates themselves. Comparison of DA and coverage measurement survey-based sex ratios (ratio of males to females) for adult Blacks further confirm correlation bias.

II. Next Meeting

The next meeting scheduled for Wednesday July 26, 2000 will discuss plans for how correlation bias will be estimated for A.C.E. evaluation purposes.

ESCAP Committee

William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson, Chair
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura

cc:

Kenneth Prewitt
Teresa Angueira
Bill Bell
Debbie Bolton
Genny Burns
Carolee Bush
Gerald Gates
Ed Gore
Dave Hubble
Donna Kostanich
Ellen Lee
Charlene Leggieri
Don Malec
Betsy Martin
Catherine Miller
Fay Nash
Sally Obenski
Miguel Perez
Ed Pike
Magdalena Ramos
Gregg Robinson
Raj Singh
Maria Urrutia
Signe Wetrogen
David Whitford
Henry Woltman
Tommy Wright

Kathleen P Zveare
07/25/2000 10:53 AM

To: Margaret A Applekamp/DIR/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Elizabeth Centrella/DSSD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Angela Frazier/DMD/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Jane F Green/DSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC
cc: Rita J Petroni/PRED/HQ/BOC@BOC, Deborah A Fenstermaker/DSSD/HQ/BOC@BOC
Subject: Agenda for 7/12 ESCAP Meeting

The agenda for the July 26 ESCAP Meeting scheduled from 10:30-12 in Rm. 2412/3 is as follows:

1. Correlation Bias for Evaluation Purposes--Rita Petroni

2. A.C.E. Analysis--Howard Hogan/Debbie Fenstermaker

DRAFT 7/26/00

Summary of A.C.E. Quality Indicators

For discussion purposes: Please note that this plan is in the process of being developed. Changes and refinements should be assumed.

1. Estimation Quality Indicators

   - Census 2000 counts, A.C.E. estimates with SEs, and A.C.E. undercount rates with SEs for total pop and seven race/origin domains
   - DA undercounts and sex ratios from 1940 to 2000
   - Poststratum-level details about the components of the DSE: weighted and unweighted matches, correct enumerations, P- and E-sample population estimates, movers, census counts, census IIs, sample sizes, DSEs, undercounts, CVs
   - Census counts, undercounts, and CVs for aggregated areas/groups, such as states
   - Comparisons with 1990 PES where appropriate

2. Missing Data Quality Indicators

   - Comparison of weighted P-sample, weighted E-sample, and census item missing data rates for poststratification factors: race, Hispanic origin, age, sex, tenure
   - Weighted estimate of noninterviewed P-sample housing units, and weight distribution of interviewed P-sample housing units before and after noninterview adjustment
   - Distribution of P-sample residency status before and after residence status imputation
   - Distribution of P-sample match status and match rates before and after match status imputation
   - Distribution of E-sample enumeration status and CE rates before and after enumeration status imputation

3. A.C.E. Interviewing Quality Indicators

   - Overall interview results: number of interviews, noninterviews, and non-occupied housing units for interview day and Census Day
   - Detailed interview results: weighted and unweighted number and percentages of household-member interviews, proxies, refusals, other noninterviews, vacants, and non-housing units
   - Distribution of interviews by mode (telephone or personal visit) and mover status
   - NRCO results by census return rate categories, TEA, and MSA status
   - Detailed information about timing of phone and personal visit interview phases by LCO and cumulative completion rate over time by ACERO

4. Person Match Quality Indicators

- After follow-up person match results: matches, A.C.E. non-matches, A.C.E. out-of-scope, census CEs, census EEs, unresolved
- Distribution of match results by race/origin domains

5. P-Sample and E-Sample Weight Quality Indicators

- Trimmed clusters with their net errors before and after trimming
- Distribution of P-sample and E-sample cluster and housing unit weight variation
- Influential clusters identified using a jackknife procedure

6. Sample Design Quality Indicators

- Comparison of weighted P-sample, weighted E-sample, and census housing units and persons
- Comparison of weighted P-sample, weighted E-sample, and census distribution of post-stratum factors
- Comparison of occupancy/vacancy rates for weighed P sample, weighted E sample, and census

7. Other

- Degree of consistency of P-sample and E-sample responses for post-stratification data items, with and without imputed data

Estimation of Correlation Bias for Evaluations

1. What is Correlation Bias?

Dual System Estimation is subject to correlation bias. Correlation bias occurs because people missed in the census enumeration are also more likely to be missed in the A.C.E.

2. What was Done to Measure Correlation Bias in 1990?

For the 1990 Coverage Measurement Evaluations, correlation bias was measured by comparing Post Enumeration Survey (PES) estimates of population size with estimates derived from combining results with demographic analysis (DA) sex ratios according to a method developed by Bell (1993). Sex ratios rather than population totals were used to minimize the effect of errors in demographic analysis.

Bell's approach to estimating correlation bias at the national level:

1. uses the usual dual system estimates (DSE) for females (i.e. the approach assumes no correlation bias).

2. selects a model for males that produces alternative postratum DSEs allowing for some dependence between the census and the coverage measurement survey.

3. estimates the dependence by controlling the alternative male DSEs to reproduce DA sex ratios, assuming female DSEs are correct, when aggregated to the national level.

The national level of correlation bias was then taken to be the difference between the resulting DSE for males and the usual DSE for males. No adjustments were made for "other" sources of bias (Mulry, 1991). The correlation bias was then distributed to the PES poststrata proportional to the estimate of the number of males in the fourth cell of the DSE for the poststratum (CAPE, 1992).

This approach provided the best subnational indications of correlation bias that were available, but it does have several limitations. First, it assumes that demographic analysis sex ratios are accurate and that there is no correlation bias for females. Rough evidence from demographic analysis totals for females in 1980 and 1990 do not refute this latter assumption, but this does not prove that correlation bias for females is entirely absent within poststrata. Also, the suitability of any alternative male DSE for this procedure depends on how will its underlying assumption conforms with reality. Unfortunately, this is uncheckable from our data. A fourth limitation is the occurrence of negative estimates of census counts less the estimates of erroneous enumerations (Bell, 1991). When the estimate is negative, no amount of the estimated people missed due to correlation bias is allocated to that post-stratum (CAPE, 1992). A final potential limitation was not adjusting DSEs for other biases (i.e. measurement bias, contamination bias, ratio estimator bias). This potentially leads to an underestimate of correlation bias.

Spencer (2000c) looked at how unmeasured correlation bias impacts loss function analysis that compares adjusted and unadjusted census counts. He found that if the unmeasured correlation bias is positively correlated with the undercount, the fact that there is unmeasured correlation bias will tend to make the census look unduly favorable relative to the DSE. If unmeasured correlation bias is uncorrelated with the undercount, then there is no systematic favoring toward either the census or the DSE in comparisons of relative accuracy. 1990 PES results and participant observer studies suggest that it is plausible that correlation bias is correlated with the undercount.

3. Models and Assumptions

The 2x2 table used for DSE is:

Census	PES		
	In	Out	Total
In	x_{11}	x_{12}	x_{1+}
Out	x_{21}	$\hat{x}_{22}$	$\hat{x}_{2}$
Total	x_{+1}	$\hat{x}_{2}$	$\hat{x}_{..}$

For the usual DSE, we assume no correlation bias or that:

$$\frac{\Pr[\text{In PES} \mid \text{In Census}]}{\Pr[\text{In PES} \mid \text{Not In Census}]} = 1.$$

This is equivalent to the assumption that:

$$\frac{x_{11}\hat{x}_{22}}{x_{12}x_{21}} \quad 1.$$

So Bell's approach assumed for females that:

$$\frac{F_{11}\hat{F}_{22}}{F_{12}F_{21}} \quad 1.$$

For males, he considered four DSE models based on these assumptions:

1. $$\frac{M_{11}\hat{M}_{22}}{M_{12}M_{21}} = \frac{\Pr[\text{In PES}|\text{ In Census}]}{\Pr[\text{In PES}|\text{ Not In Census}]} = C$$

2. $$\frac{M_{11}(M_{21}\ \hat{M}_{22})}{M_1\ M_{21}}\ C$$

3. $$\frac{\hat{M}_{22}\,\hat{F}_{..}}{\hat{M}_{..}\,\hat{F}_{22}}\ C$$

4. $$\frac{M_{21}\,\hat{M}_{..}}{M_1\ (M_{21}\ \hat{M}_{22})} = \frac{\Pr[\text{In PES}|\text{Not In Census}]}{\Pr[\text{In Census}]} = C$$

For each model, C was assumed constant within age/sex groups for males.

4. What Does the Census Bureau Plan to Do for Census 2000?

For the 2000 Total Error Model, we again expect to use demographic analysis to obtain national levels of correlation bias and a modeling approach to distribute the correlation bias across poststrata.

The Bureau will re-evaluate whether to use DA data and knowledge of other DSE biases to obtain national levels of correlation bias. We are considering a method based on DA totals and a method based on sex ratios which treats correlation bias for females as negligible. For both methods we are also considering whether to make adjustments for contamination, ratio estimator, and measurement error biases (Spencer, 2000a).

To assist in determination of which estimates of correlation bias to use, we plan to develop point estimates of other biases in DSE estimates of males and females by black and non-black and given resources we may estimate some selected variances.

Measurement error biases will be obtained from an evaluation sample of about 2300 clusters - over

twice the number of evaluation clusters used for the 1990 evaluations.

We are considering various modeling approaches to distribute correlation bias to poststrata. In addition to Bell's approach (1993), we are considering approaches developed by Elliott and Little (1999) and by Haberman, Jiang, and Spencer (1998).

Elliot and Little improve the application of Bell's models. They propose some general principals for aiding the choice among the alternative models. Using these principles, they choose a model and imbed it within a more comprehensive Bayesian model for counts in poststrata of the population. Through judicious choice of parameterization and prior distributions, their Bayesian model eliminates negative cell estimates and reduces outlying predictions of undercount rates. In addition, their model detects, through posterior predictive distributions, strata in which large negative raw cell estimates may be due to bias rather than variance. Their model also allows direct control over the inter-strata sex ratio variation through the variance parameter. Additionally, their method can be extended to provide estimates of precision that incorporate uncertainty in the estimates from demographic analysis and other sources.

The Haberman et.al. approach is similar to the approach of Bell. It is applicable when the capture probabilities can vary from individual to individual according to a logit model, whereas Bell's models apply to the post-stratification model, in which capture probabilities are assumed constant. The two approaches both use sex ratios and in effect constrain the adjustment factors to match the sex ratios. The females can be assumed to have zero correlation bias, or the DA totals can be used to constrain the adjustment factors for both females and males to agree with the DA totals.

The modeling approach we use will be determined by consideration of the strength of each approach and resources available to implement the approaches.

5. References

Bell, W. R. (1991), "Use of Alternative Dual System Estimation to Measure Correlation Bias," 1990 Coverage Studies and Evaluation Memorandum Series #0-3, Project P13, July 10, 1991.

Bell, W.R. (1993), "Using Information from Demographic Analysis in Post-Enumeration Survey Estimates," Journal of the American Statistical Association, 88, 1106-1118.

CAPE (1992), "Assessment of Accuracy of Adjusted Versus Unadjusted 1990 Census Base for Use in Intercensal Estimates," Report of the Committee on Adjustment of Postcensal Estimates, Bureau of the Census, Department of Commerce, August 7, 1992.

Elliott, M. R. and Little, R. J. A. (1999), "A Bayesian Approach to Combining Information from a

Census, A Coverage Measurement Survey and Demographic Analysis," to appear in the Journal of the American Statistical Association.

Haberman, S. J., Jiang, W. and Spencer, B. D. (1998), "Activity 7: Develop Methodology for Evaluating Model-Based Estimates of the Population Size for States," Final Report under contract 50-YABC-2-66023 for the Bureau of the Census. Chicago: National Opinion Research Center.

Mulry, M.H. (1991), "1990 Post Enumeration Survey Evaluation Project P16: Total Error in PES Estimates for Evaluation Post Strata," 1990 Coverage Studies and Evaluation Memorandum Series #R-6, July 11, 1991. Washington, DC.: Bureau of the Census.

Spencer, B. D. (2000a), "Final Report on Correlation Bias Methodology," Under contract number 50-YABC-7-66020 for the Bureau of the Census. Abt Associates, Inc. and Spencer Statistics, Inc., April 20, 2000.

Spencer, B. D. (2000b), " Total Error Model for Census 2000: How Components of Error Can Be Estimated from the Bureau's Planned Evaluation Studies", Final Report under contract number 50-YABC-7-66020 for the Bureau of the Census. Abt Associates, Inc. and Spencer Statistics, Inc., revised May 17, 2000.

Spencer, B. D. (2000c), " A Strategy for Analyzing Whether Adjustment of Census 2000 Will Improve Redistricting," draft 1.0, June 8, 2000.

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 12

July 26, 2000

Prepared by: Maria Urrutia and Annette Quinlan

The twelfth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on July 26, 2000 at 10:30. The agenda for the meeting was A.C.E. correlation bias for evaluation purposes and A.C.E. quality analysis.

Committee Attendees:

William Barron
Nancy Potok
Cynthia Clark
Nancy Gordon
John Thompson
Howard Hogan
John Long
Susan Miskura

Other Attendees:

Kenneth Prewitt
Donna Kostanich
Fay Nash
Raj Singh
Tommy Wright
David Hubble
Rita Petroni
Gregg Robinson
Maria Urrutia
Annette Quinlan

I. A.C.E. Estimation of Correlation Bias for Evaluations

Rita Petroni presented and distributed the correlation bias estimation plans for evaluation

purposes. The handout is attached. Rita first reviewed the method that was used to produce correlation bias estimates for the 1990 PES evaluation. She then described four models which were considered in 1990 to produce the sub-national evaluation estimates and totals, and described plans for Census 2000. Demographic Analysis correlation bias estimates are produced at a national level. The challenge for evaluating the effects of correlation bias is to produce estimates for sub-national areas in order to evaluate the A.C.E. The sub-national correlation bias estimates will not be available until mid 2002 for the A.C.E. evaluations.

In 1990, we used Demographic Analysis (DA) sex ratios rather than DA totals to dampen concerns about uncertainty in DA estimates for population totals. Gregg Robinson noted in the previous ESCAP meeting that sex ratios were less subject to uncertainty. In 1990 the following approach was used by Bell, as stated in the handout, to obtain estimates of correlation bias at the national level.

> Use the DSE for females (which assumes no correlation bias) and select a model for males that produces alternative post stratum DSEs allowing for some dependence between the census and the coverage measurement survey. Then estimate the dependence by controlling the alternative male DSEs to reproduce DA sex ratios, assuming female DSEs are correct, when aggregated to the national level (Bell 1993). The national level of correlation bias was then taken to be the difference between the resulting DSE for males and the usual DSE for males. No adjustments were made for "other" sources of bias (Mulry, 1991). The correlation bias was then distributed to the PES post strata proportional to the estimate of the number of males in the fourth cell of the DSE for the post stratum (CAPE, 1992).

The fourth cell represents persons missed in both the census and the A.C.E. For post-strata with negative cell estimates, no correlation bias was assigned. The negative values resulted from the use of "unbiased" estimators of the fourth cell with high sampling variance. One of the strengths of the DSE is that it does not require direct estimation of the fourth cell.

Rita talked about alternatives we are considering for 2000 which include the four 1990 approaches and an approach developed by Elliott and Little. The four 1990 approaches included:

> The Fixed Odds Ratios Model which assumes that the odds of appearing in the A.C.E. given that an individual was enumerated in the census relative to the odds of appearing in the A.C.E. given that an individual was not enumerated in the census are arbitrary for males and equal to 1 for females in a one-stratum design (Wolter 1990).
>
> The Fixed Relative Risk Model which assumes a constant relative risk for enumeration

in the census and A.C.E. for males and independence for females (Bell 1993).

The Fixed Sex Ratio Model which assumes that the sex ratio in the fourth cell is constant across strata (Bell 1993).

The Generalized Behavioral Response Model which assumes the probability of being included in the A.C.E. given that an individual was not enumerated in the census divided by the probability of being included in the census is constant across the post-strata for males (Bell 1993).

The Fixed Odds Ratio Model was used for the analysis in 1990. The approach developed by Elliott and Little improves the 1990 models. Using some general principles, they choose a model and embed it with a more comprehensive Bayesian model. Rita also discussed an alternative approach from Haberman, Jiang, and Spencer which is similar to that which was used in 1990 but does not assume that capture probabilities are constant for individuals. We concluded the discussion by noting that Rita and staff need to conduct additional exploration and conduct technical consultation with outside experts prior to selecting the methodology that will be used for 2000. We would consider implementing sensitivity analysis as a method of looking at how much variance is possible among the various model assumptions.

II. Targeted Extended Search - Change to Procedure

The Targeted Extended Search (TES) is an operation that is designed to reduce high variances that are caused by large-scale geocoding error in a subset of clusters. Geocoding error causes both erroneous enumerations and nonmatches on the census side and ACE side. During the TES operation, the blocks surrounding a cluster are searched for nonmatched and erroneously enumerated people.

A needed change has been discovered in the way housing units added to the initial census were treated in the TES procedure. For TES clusters, the census housing unit will be treated as a correct enumeration if it is located in a surrounding block. When needed, cases are sent to the field for map spotting. We have found problems with our TES procedure resulting from housing units added to the initial census since January. These housing units were not included in the TES operation. Our planning up to this point was based on an expectation that this would not adversely affect the A.C.E. processes or estimates. More recent analysis has indicated that we must include these housing units in our estimation process. We will do this by developing an imputation process to assign enumeration status for these units. We will impute correct enumeration status for census geocoding errors where we do not know if it is in the surrounding block or some other block further away.

We will finalize this modification to our estimation before any data are available for review.

III. A.C.E. Quality Analysis

Howard Hogan has been directing the development of a proposal for data that will be presented to the ESCAP for its deliberation in deciding whether to use statistical correction for redistricting data. Howard presented the preliminary quality indicators that have been identified for the analysis of the initial census and the A.C.E. results. The handout is attached.

IV. Next Meeting

The next meeting scheduled for Wednesday August 9, 2000 will continue to discuss the process for evaluating the initial census and the A.C.E. results.

ESCAP Committee

William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson, Chair
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura

cc:

Kenneth Prewitt
Teresa Angueira
Bill Bell
Debbie Bolton
Genny Burns
Carolee Bush
Gerald Gates
Ed Gore
Dave Hubble
Donna Kostanich
Ellen Lee
Charlene Leggieri
Don Malec
Betsy Martin
Catherine Miller
Fay Nash
Sally Obenski
Miguel Perez
Ed Pike
Magdalena Ramos
Gregg Robinson
Raj Singh
Maria Urrutia
Signe Wetrogan
David Whitford
Henry Woltman
Tommy Wright

There was no agenda developed or used for the August 9, 2000 meeting.

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 13

August 9, 2000

Prepared by: Maria Urrutia and Annette Quinlan

The thirteenth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on August 9, 2000 at 10:30. The agenda for the meeting was the A.C.E. quality indicator analysis preparation.

Committee Attendees:

William Barron
Bob Fay
John Thompson
Howard Hogan
Ruth Ann Killion
Susan Miskura

Other Attendees:

Kenneth Prewitt
Donna Kostanich
Louisa Miller
Raj Singh
David Whitford
Debbie Fenstermaker
Michael Ikeda
Maria Urrutia
Annette Quinlan

I. A.C.E. Quality Indicator Analysis

Due to staff schedules, the full ESCAP could not meet. Rather than cancel the meeting, Howard Hogan conducted an informal discussion on the progress in developing the set of A.C.E. quality indicators that will be assessed as part of the Committee's deliberations. This

discussion will continue at the next ESCAP meeting.

II. Next Meeting

The next meeting scheduled for Wednesday August 23, 2000 will discuss how 1990 data will be used to inform the 2000 decision process.

ESCAP Committee

William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson, Chair
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura

cc:

Kenneth Prewitt
Teresa Angueira
Bill Bell
Debbie Bolton
Genny Burns
Carolee Bush
Gerald Gates
Ed Gore
Dave Hubble
Donna Kostanich
Ellen Lee
Charlene Leggieri
Don Malec
Betsy Martin
Catherine Miller
Fay Nash
Sally Obenski
Miguel Perez
Ed Pike
Magdalena Ramos
Gregg Robinson
Raj Singh
Maria Urrutia
Signe Wetrogan
David Whitford
Henry Woltman
Tommy Wright

There was no agenda developed or used for the August 23, 2000 meeting.

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 14

August 23, 2000

Prepared by: Maria Urrutia and Annette Quinlan

The fourteenth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on August 23, 2000 at 10:30. The agenda for the meeting was to continue the discussion on how census data will be used to inform the 2000 decision process.

Committee Attendees:

William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
John Long
Susan Miskura

Other Attendees:

Donna Kostanich
Raj Singh
Tommy Wright
Debbie Fenstermaker
David Hubble
Carolee Bush
Nick Birnbaum
Maria Urrutia
Annette Quinlan

I. A.C.E. Quality Analysis

The ESCAP continued the discussion of the A.C.E. measures of quality from the previous meeting. The ESCAP also discussed the presentation of these materials for the meeting sponsored by the National Academy of Sciences scheduled for October 2, 2000. Howard Hogan and Census Bureau staff will prepare reports on how we plan to analyze data for the following areas:

- Assessing Results from the Accuracy and Coverage Evaluation
- Overall Quality Indicators
- Decomposition of Dual System Estimation (DSE) Components
- Missing Data Results
- Person Matching and Follow-up Results
- Person Interviewing Results
- Dual System Estimation Results
- Demographic Analysis Results
- Variance Estimates by Size of Geographic Area
- Consistency of Post-Stratification Variables
- Quality of Census Processes
- Synthetic Assumptions
- Correlation Bias Results

We also discussed various methods by which the overall data can be synthesized. This discussion will continue at the next ESCAP meeting.

II. Next Meeting

The next meeting scheduled for Wednesday September 13, 2000 will continue the discussion on the preparation for the October 2 meeting with NAS.

ESCAP Committee

William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson, Chair
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long
Susan Miskura

cc:

Kenneth Prewitt
Teresa Angueira
Bill Bell
Debbie Bolton
Genny Burns
Carolee Bush
Gerald Gates
Ed Gore
Dave Hubble
Donna Kostanich
Ellen Lee
Charlene Leggieri
Don Malec
Betsy Martin
Catherine Miller
Fay Nash
Sally Obenski
Miguel Perez
Ed Pike
Magdalena Ramos
Gregg Robinson
Raj Singh
Maria Urrutia
Signe Wetrogan
David Whitford
Henry Woltman
Tommy Wright

Annette M Quinlan
09/12/2000 08:05 AM

To: Kathleen P Zveare/DMD/HQ/BOC@BOC
cc: Maria E Urrutia/DMD/HQ/BOC@BOC
Subject: Agenda for ESCAP Meeting #15

Agenda for ESCAP Meeting #15, September 13, 2000

We will continue the discussion on preparation for October 2 meeting with NAS.

DRAFT

September 8, 2000

Outline of
Assessing the Results from the
Accuracy and Coverage Evaluation

Howard Hogan

1. The purpose of this document

2. Review of A.C.E. operations
    - 2.1 Were the steps between processing and estimation properly carried out?
    - 2.2 Were the A.C.E. operations well conducted and well controlled?

3. Review of the measures of A.C.E. quality
    - 3.1 Individual Components of A.C.E. quality
        - 3.1.1 What is the level of A.C.E. Sampling?
        - 3.1.2 Consisted reporting of Census Day residence
        - 3.1.3 Matching Error
        - 3.1.4 A.C.E. Fabrications
        - 3.1.5 Missing Data
        - 3.1.6 Balancing error
        - 3.1.7 Error in measuring erroneous enumerations
        - 3.1.8 Correlation Bias
        - 3.1.9 Synthetic Bias and Synthetic Variability
        - 3.1.10 Other measurement and technical errors
            - Technical Ratio bias
            - Contamination error
            - Inconsistent Poststratification
    - 3.2 Synthesizing the components of A.C.E. Quality
        - 3.2.1 How do the individual components of A.C.E. quality combine to affect the accuracy of the population estimates?
        - 3.2.2 How does these accuracy of the A.C.E. compare to the accuracy of the uncorrected census?

4. Comparison with historical patterns and independent benchmarks
    - 4.1 Comparison with Demographic Analysis and Demographic Projections
    - 4.2 Comparison with historical patterns
    - 4.3 External measures of Census Quality
    - 4.4 Census Quality Assurance indicators
    - 4.5 Other reports of Census Quality

5. Forming an overall Assessment

Data and Analysis to Inform the ESCAP Decision

1. Introduction.

2. Review of the Quality of the Uncorrected Census.

   2.1 **Comparison with Demographic Analysis and Demographic Estimates**.
   2.2 **Direct Measures of Census Quality**.

3. Review of A.C.E. Operations.

   3.1 **Proper Execution of the Steps Between Processing and Estimation.**
   3.2 **Conduct and Control of the A.C.E. Operations.**

4. Review of Measures of A.C.E. Quality.

   4.1 **Individual Components of A.C.E. Quality**.

      4.1.1 *Sampling Variance*.
      4.1.2 *Consistent Reporting of Census Day Residence*.
      4.1.3 *Matching Error*.
      4.1.4 *A.C.E. Fabrications*.
      4.1.5 *Missing Data.*
      4.1.6 *Balancing Error.*
      4.1.7 *Erroneous Enumerations.*
      4.1.8 *Correlation Bias.*
      4.1.9 *Synthetic Bias and Synthetic Variability.*
      4.1.10 *Other Measurement and Technical Errors.*

   4.2 **Synthesizing A.C.E. Quality.**

      4.2.1 *Combining the Components of A.C.E. Quality to Assess Accuracy*.
      4.2.2 *Comparing the Accuracy of the A.C.E. to the Accuracy of the Uncorrected Census*.

5.0 Forming an Overall Assessment

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 15

September 13, 2000

Prepared by: Nick Birnbaum.

The fifteenth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on September 13, 2000 at 10:30. The meeting did not run the allotted hour and a half because several ESCAP members had to depart early due to other commitments.

The agenda for the meeting was to continue the discussions from previous meetings regarding drafts of the documents to be presented to the National Academy of Sciences (NAS) panel on October 2.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
John Thompson
Jay Waite
Bob Fay
Howard Hogan

Other Attendees:

Kenneth Prewitt
Donna Kostanich
Raj Singh
Tommy Wright
Debbie Fenstermaker
Roxie Jones
Louisa Miller
Gregg Robinson
Nick Birnbaum
Kathleen Styles
Maria Urrutia
Annette Quinlan

I. Comparison of A.C.E. and Census Quality

Howard Hogan presented an outline of an overview document that he will draft to summarize the various analyses that the set of documents will contain. An introductory section of the overview document will define the purpose of the series of documents. Second, the analysis reports relating to the A.C.E. operations and the transitional steps will be reviewed to determine how well the specific operations and steps were executed and documented. Third, a discussion of the reports relating to measures of A.C.E. quality will be presented and include a determination of the validity of each component. Next, there will be an examination of the reports comparing the initial census data to historical patterns and independent benchmarks to determine how well the initial census was conducted. The last section will discuss the analysis report relating to the formulation of an overall assessment.

During the discussion, ESCAP members emphasized the fact that their task is to examine the quality of both the initial census counts and the A.C.E., analyze the deficiencies in both, and determine if the results of the A.C.E. could be used to improve the accuracy of the initial counts. That is, the Committee has not pre-judged the superior accuracy of the A.C.E.-based results; these results would only be applied if they can correct the deficiencies in the initial counts. Howard Hogan agreed that this point would be clearly articulated in the overview document, and that the individual analysis reports would reflect the fact that the Committee would be examining the assessments of both the initial counts and the A.C.E. results.

Attached are the outline that was discussed at the meeting and a revised version. Comments on the analysis reports were to be sent to John Thompson and Howard Hogan before c.o.b. September 20.

II. Next Meeting

The next meeting, scheduled for Wednesday September 27, 2000, will continue discussions of the documents to be presented at the NAS panel workshop.

Attachments

Kathleen P Zveare
09/27/2000 09:01 AM

To: Margaret A Applekamp/DIR/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Elizabeth Centrella/DSD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Angela Frazier/DMD/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Jane F Green/DSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC
cc:
Subject: Agenda for Today's ESCAP Meeting

The agenda for today's ESCAP meeting from 10:30-12 in Rm. 2412/3 is as follows:

Continue preparations for the October 2 National Academy of Science meeting.

Decennial Statistical Studies Division
Census 2000 Procedures and Operations Memorandum Series

Chapter B: A.C.E. Review

Chapter Code	Subject
B-1	Data and Analysis to Inform the ESCAP Recommendation
B-2	Overall Census and A.C.E. Quality Indicators
B-3	Quality of Census 2000 Processes
B-4	Accuracy and Coverage Evaluation Survey: Demographic Analysis Results
B-5	Accuracy and Coverage Evaluation Survey: Person Interviewing Results
B-6	Accuracy and Coverage Evaluation Survey: Person Matching and Followup Results
B-7	Accuracy and Coverage Evaluation Survey: Missing Data Results
B-8	Accuracy and Coverage Evaluation Survey: Decomposition of Dual System Estimate Components
B-9	Accuracy and Coverage Evaluation Survey: Dual System Estimation Results
B-10	Accuracy and Coverage Evaluation Survey: Consistency of Post-Stratification Variables
B-11	Accuracy and Coverage Evaluation Survey: Variance Estimates by Size of Geographic Area
B-12	Accuracy and Coverage Evaluation Survey: Correlation Bias

Chapter Code	Subject
B-13	Accuracy and Coverage Evaluation Survey: Comparing Accuracy
B-14	Accuracy and Coverage Evaluation Survey: Synthetic Assumptions
B-15	Accuracy and Coverage Evaluation Survey: Contributions of Service Based Enumeration Multiplicity Estimation to Corrected Census Results
B-16	Demographic Full Count Review Report

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 16

September 27, 2000

Prepared by: Nick Birnbaum.

The sixteenth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on September 27, 2000 at 10:30.

The agenda for the meeting was to continue the discussions from previous meetings regarding the documents to be presented to the National Academy of Sciences (NAS) Panel on October 2.

Committee Attendees:

William Barron
Nancy Potok
Cynthia Clark
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Susan Miskura
Ruth Ann Killion

Other Attendees:

Kenneth Prewitt	Donna Kostanich
Tommy Wright	Debbie Fenstermaker
Roxie Jones	Louisa Miller
Gregg Robinson	Nick Birnbaum
Kathleen Styles	Maria Urrutia
Annette Quinlan	Carolee Bush

I. Preparations for the October 2 NAS Panel Workshop

John Thompson opened the meeting and asked Dr. Prewitt to briefly discuss his planned opening remarks for the NAS Panel meeting. Dr. Prewitt discussed how he planned to address, among other things, the rationale for not making partial or preliminary data publicly available during the ESCAP decision-making process:

- To avoid confusion
- To avoid the appearance of political manipulation
- The ESCAP needs to deliberate without external scrutiny, pressure, or influence.

The data will be made available after the decision has been made. To date, the Census Bureau, as a statistical agency, has exercised more transparency regarding its statistical programs than other statistical agencies would perhaps be comfortable with providing. There has been a great deal of pre-specification in order to support this transparency. However, this pre-specification is not always optimal. It limits changes to the methodology, in response to unanticipated circumstances, that would result in improvements to the data.

There was some discussion regarding the timing of the issuance of the Panel's report. It was noted that although we might receive some informal feedback as a result of our presentation at the October 2 workshop, the workshop would not result in formal recommendations from the Panel.

Howard Hogan then reviewed his overview document for the NAS presentation, as this would be the basis for leading the discussion at the workshop. With regard to analysis report #13, Howard stated that it was important to emphasize that the loss function analyses would not be determinative – that they would be one component of the Committee's assessment of both the initial counts and the A.C.E. results. Howard indicated that it was important to reiterate that some data that are relevant to the loss function analysis would not be available within the time frame for the Committee to make a recommendation, and would have to be estimated based on 1990 data. The specifics of the estimation were deferred for future discussion.

[Note: Howard Hogan's overview document and the rest of the final prototype analysis reports provided to the NAS Panel on October 2, 2000, are available upon request. For the convenience of the reader, an index of those documents is attached.]

In wrapping up the meeting, a couple of administrative issues were discussed. It was suggested that the points contained in the section entitled "Forming an Overall Assessment," which is the very last section of the overview document, be discussed at the beginning of the meeting, in the

event that time constraints did not enable Howard to walk through the entirety of his overview document.

It was announced that revised versions of 14 of the 16 documents would be distributed later today and that revised versions of the remaining two documents would be distributed tomorrow. Also, a meeting agenda, a document index, and directions to the National Academy would be distributed to Committee members.

II. Next Meeting

The next meeting, scheduled for October 11, 2000, will address the effect of late census data on Dual System Estimation and examine the dual system estimate variances from the 1990 PES.

Attachment

Kathleen P Zveare
10/10/2000 10:27 AM

To: Margaret A Applekamp/DIR/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Elizabeth Centrella/DSD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Angela Frazier/DMD/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Jane F Green/DSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC
cc:
Subject: Agenda for 10/11 ESCAP Meeting

The agenda for the October 11 ESCAP Meeting scheduled from 10:30-12 in Rm. 2412/3 is as follows:

1. Update on Census Processes - Jay Waite

2. NAS Reports - Howard Hogan

Dual System Estimate. For a given post-stratum, the formula for the dual system estimate is as follows:

$$DSE = (C - II)\left(\frac{CE}{N_e}\right)\left(\frac{N_p}{M}\right) \quad (1)$$

where

C	=	the census count;
II	=	the number of census people with insufficient information;
CE	=	the estimated number of correct enumerations from the E Sample;
N_e	=	the estimated number of people from the E Sample;
N_p	=	the estimated total population from the P Sample;
M	=	the estimated number of persons from the P-sample population who match to the Census.

Persons in Group Quarters and the Remote Alaska type of enumeration area are excluded from the 2000 A.C.E., and thus from the above numbers. For the 1990 PES, persons in the Remote Alaska type of enumeration area were excluded while persons from Group Quarters were included in these numbers.

The 2000 A.C.E. and the 1990 PES differ procedurally in their treatment of movers; that is persons whose location at the time of the survey interview differ from their location on Census Day. See the section on Movers.

Coverage Correction Factor. The coverage correction factor (CCF) is a measure of correction to assess the degree of net overcount or net undercount of the household population within the Census. The coverage correction factor (CCF) for a post-stratum is the ratio of the DSE over the census count.

$$CCF = \frac{DSE}{C} \quad (2)$$

For example, a coverage correction factor of 1.05 would imply that for every 100 people within the given post-stratum, there is a net undercount of five persons.

Table B-1: 2000 A.C.E. Results--Total Race/Hispanic Origin Domain

Total Domain 1
(AI on Res) Domain 2
(AI off Res) Domain 3
(Hispanic) Domain 4
(Black) Domain 5
(NH or PI) Domain 6
(Asian) Domain 7
(Wt or Oth) Total

Census Counts
Data-Defined Persons (DD) 987,654,321 This number is for display purposes only.

Insufficient Information (II)
Total Persons (C)

P Sample
Nonmover Sample Size
Inmover Sample Size
Outmover Sample Size
Weighted Nonmovers (N_n)
Weighted Inmovers (N_i)
Weighted Outmovers (N_o)
Weighted Nonmover Matches (M_n)
Weighted Outmover Matches (M_o)
Weighted P-Sample Persons (N_p)
Weighted P-Sample Matches (M)

E Sample
E-Sample Size
Correct Enumeration Sample Size
Weighted E-Sample Persons (N_e)
Weighted Correct Enumerations (CE)

Estimates
Dual System Estimate (DSE)
Standard Error (SE)
Coefficient of Variation (CV) (%)

Coverage Correction Factor
Standard Error (SE)
Coefficient of Variation (CV) (%)

Net Undercount Percent (UC) (%)
Standard Error

Table 4. Estimates for Revised Post-Strata Groups

	Percent undercount					Standard errors				
	All	*NE*	*S*	*MW*	*W*	*All*	*NE*	*S*	*MW*	*W*
Non-Hispanic White and Other										
Owner										
Large Urbanized Areas		−2.13	.68	−.26	−.34		1.08	.71	.39	.65
Other Urban		−1.08	.52	−.10	.62		.49	.42	.40	.58
Non-Urban		−.54	.18	−.71	.29		.70	.69	1.18	.69
Nonowner										
Large Urbanized Areas		1.16	2.56	2.33	3.18		1.39	1.48	1.61	1.62
Other Urban		3.41	3.20	1.23	4.49		1.51	1.74	1.09	1.34
Non-Urban		6.52	6.23	2.85	6.08		4.20	1.71	1.51	1.81
Black										
Owner										
Large Urbanized Areas		1.63	2.16	.81	6.10		1.91	.90	.87	1.91
Other Urban	1.34					.98				
Non-Urban	3.52					1.90				
Nonowner										
Large Urbanized Areas		8.37	6.27	5.99	9.96		1.61	1.90	1.68	2.72
Other Urban	4.15					1.18				
Non-Urban	4.62					5.33				
Non-Black Hispanic										
Owner										
Large Urbanized Areas		.67	2.53	−4.33	2.89		4.45	.90	2.58	.87
Other Urban	.94					1.64				
Non-Urban	2.73					2.69				
Nonowner										
Large Urbanized Areas		6.72	9.34	6.64	5.91		3.51	2.59	3.26	1.84
Other Urban	6.60					2.74				
Non-Urban	15.80					5.01				
Asian and Pacific Islander										
Owner	−1.45					1.50				
Nonowner	6.96					2.52				
Reservation Indians	12.22					4.73				

Table 5. Counts and Undercount Rates by State: Adjustment and Revised

Name	Census count*	Adjustment estimate			Revised estimate		
		Count*	Rate	SE	Count*	Rate	SE
Alabama	4,041	4,146	2.5	.4	4,113	1.8	.3
Alaska	550	561	1.9	.4	561	2.0	.4
Arizona	3,665	3,790	3.3	.5	3,754	2.4	.5
Arkansas	2,351	2,403	2.2	.4	2,392	1.7	.3
California	29,760	30,888	3.7	.4	30,595	2.7	.4
Colorado	3,294	3,376	2.4	.5	3,364	2.1	.4
Connecticut	3,287	3,306	.6	.6	3,308	.6	.4
Delaware	666	687	3.0	.4	678	1.8	.4
DC	607	639	5.0	.5	628	3.4	.9
Florida	12,938	13,278	2.6	.4	13,197	2.0	.4
Georgia	6,478	6,633	2.3	.4	6,619	2.1	.3
Hawaii	1,108	1,136	2.5	.5	1,129	1.9	.8
Idaho	1,007	1,035	2.8	.5	1,029	2.2	.4
Illinois	11,431	11,592	1.4	.4	11,544	1.0	.4
Indiana	5,544	5,586	.7	.4	5,572	.5	.4
Iowa	2,777	2,807	1.1	.5	2,788	.4	.4
Kansas	2,478	2,506	1.2	.4	2,495	.7	.4
Kentucky	3,685	3,768	2.2	.4	3,746	1.6	.4
Louisiana	4,220	4,332	2.6	.4	4,314	2.2	.4
Maine	1,228	1,240	1.0	.6	1,237	.7	.6
Maryland	4,781	4,869	1.8	.4	4,882	2.1	.4
Massachusetts	6,016	6,039	.4	.5	6,045	.5	.5
Michigan	9,295	9,404	1.2	.4	9,361	.7	.4
Minnesota	4,375	4,419	1.0	.4	4,394	.4	.4
Mississippi	2,573	2,632	2.2	.4	2,629	2.1	.4
Missouri	5,117	5,184	1.3	.4	5,149	.6	.4
Montana	799	822	2.8	.5	818	2.4	.5
Nebraska	1,578	1,595	1.0	.4	1,589	.6	.4
Nevada	1,202	1,232	2.4	.5	1,231	2.3	.4
New Hampshire	1,109	1,116	.6	.5	1,119	.8	.5
New Jersey	7,730	7,836	1.4	.5	7,774	.6	.6
New Mexico	1,515	1,586	4.5	.5	1,563	3.1	.5
New York	17,990	18,304	1.7	.5	18,262	1.5	.6
North Carolina	6,629	6,815	2.7	.4	6,753	1.8	.3
North Dakota	639	648	1.4	.5	643	.7	.5
Ohio	10,847	10,933	.8	.4	10,922	.7	.4
Oklahoma	3,146	3,214	2.1	.4	3,203	1.8	.3
Oregon	2,842	2,898	1.9	.4	2,896	1.9	.4
Pennsylvania	11,882	11,957	.6	.5	11,917	.3	.5
Rhode Island	1,003	1,006	.3	.6	1,005	.1	.6
South Carolina	3,487	3,590	2.9	.4	3,559	2.0	.4
South Dakota	696	707	1.5	.5	703	1.0	.5
Tennessee	4,877	5,012	2.7	.4	4,964	1.7	.3
Texas	16,987	17,551	3.2	.4	17,470	2.8	.4
Utah	1,723	1,757	1.9	.5	1,753	1.7	.5
Vermont	563	571	1.4	.7	569	1.1	.8
Virginia	6,187	6,353	2.6	.4	6,314	2.0	.4
Washington	4,867	4,987	2.4	.4	4,958	1.8	.4
West Virginia	1,793	1,842	2.6	.4	1,819	1.4	.4
Wisconsin	4,892	4,924	.7	.4	4,922	.6	.4
Wyoming	454	466	2.7	.5	464	2.2	.4

* All counts in thousands.

Appendix: Table A.1. Adjustment Dual-System Estimates

Percent Undercount by Post-Stratum Group

	Direct				Smoothed			
North East	All Other	Black	Hispanic	Asian	All Other	Black	Hispanic	Asian
New England								
Central Cities	-1.74	5.69			-1.16	4.25		
Non Central City MSA	0.61				0.19			
Other Places 10,000+	0.54	5.88 *			0.59	5.39 *		
Other areas	1.68				1.79			
Middle Atlantic								
New York City CC's								
Non-owner	2.06	6.44	4.00	9.47	0.87	7.76	1.73	10.50
Owner	-2.64	-2.86			-0.23	-0.15		
Other Large MSA Central city								
Non-owner	-6.41	10.78	9.91		-0.37	7.74	2.01	
Owner	-2.93	2.66			-0.19	-0.03		
Central cities of Small MSA	2.05	17.92			0.07	9.34		
Non Central City in NYC PMSA	5.03	5.63			0.42	6.73		
Non Central City in Other Large MSA	-0.80				0.36			
Non Central City in Small MSA	-0.78	5.88 *			-0.09	5.39 *		
Other Places 10,000+	1.36				0.41			
Other areas	0.43				0.70			
South								
South Atlantic								
Large MSA Central city								
Non-owner	11.49	10.46			5.00	9.33		
Owner	1.09	1.68	2.77		1.72	0.95	4.92	
Central cities of Small MSA	2.84	4.93			2.74	4.00		
Non Central City in Large MSA	0.93	4.17	13.79		0.44	1.97	5.13	
Non Central City in Small MSA	3.50	0.27			2.80	3.59		
Other Places 10,000+	1.23	-1.71			1.51	1.60		
Other areas	3.25	5.68			2.71	2.64		
East South Central								
Large MSA Central city								
Non-owner	2.17	6.46			4.80	5.81		
Owner	3.19				2.56			
Central cities of Small MSA	0.90				2.58			
Non Central City in MSA	1.42	4.82			2.31	2.26		
Other Places 10,000+	-6.02				1.84			
Other areas	-0.95				1.65			
West South Central								
Houston,Dallas, Ft.Worth CC's								
Non-owner	6.24				4.60			
Owner	0.56	8.09	8.96		1.49	6.64	7.11	
Other Large MSA Central city								
Non-owner	1.34				3.23			
Owner	-1.16	4.54	3.18		0.69	4.82	3.76	
Central cities of Small MSA	-3.16				2.48			
Non Central City in MSA	2.07				2.28			
Other Places 10,000+	1.19	1.66	2.36		1.25	2.28	5.11	
Other areas	1.72				1.96			

	Direct				Smoothed			
	All Other	Black	Hispanic	Asian	All Other	Black	Hispanic	Asian
Midwest								
East North Central								
Chicago Detroit CC's								
Non-owner	2.76	6.76	0.38		5.17	5.77	-1.61	
Owner	-0.05	0.42			1.12	1.98		
Other Large MSA Central city								
Non-owner	1.56	4.03			1.04	4.49		
Owner	-1.24	7.09			-0.15	0.64		
Central cities of Small MSA	1.76	4.61			2.09	5.44		
Non Central City in Large MSA	0.84				0.59			
Non Central City in Small MSA	0.96	3.99 *			0.64	4.66 *		
Other Places 10,000+	0.42				0.20			
Other areas	-1.64				-0.99			
West North Central								
Large MSA Central city								
Non-owner	5.20	5.47			2.47	5.44		
Owner	-0.53				-0.33			
Central cities of Small MSA	1.82	4.85			1.90	7.23		
Non Central City inLarge MSA	1.09				0.71			
Non Central City in Small MSA	0.22	3.99 *			1.64	4.66 *		
Other Places 10,000+	0.83				0.75			
Other areas	0.78				0.31			
West								
Mountain								
Large MSA Central city								
Non-owner	4.65	1.48			5.03	4.61		
Owner	1.24				0.98			
Central cities of Small MSA	2.88				1.52			
Non Central City in MSA	0.60	7.39 *			0.75	7.80 *		
Other Places 10,000+	1.22				1.45			
Other areas	3.00				3.22			
Pacific								
Non-owner								
Los Angeles/Long Beach CC's	6.44	7.38	10.14	6.29	4.75	6.83	7.87	6.50
Other Large MSA Central city	3.73				3.72			
Central cities of Small MSA								
Owner								
Los Angeles/Long Beach CC's	-0.35	8.36	2.01	3.10	1.39	7.86	1.95	4.80
Other Large MSA Central city	1.39				1.39			
Central cities of Small MSA								
Central cities of Small MSA	0.56				0.95			
Non Central City in Large MSA/PMSA	1.05	14.32	5.65	0.82	0.17	16.37	6.94	0.79
Non Central City in Small MSA	2.90				3.15			
Other Places 10,000+	1.38	7.39 *		-3.22	1.89	7.80 *		0.18
Other areas	3.15				1.92			
Reservation Indian				12.72				12.72

NOTE: Bold indicates cell is significantly different from zero at 90% level. Boxes show which cells were combined to form post-strata. Asians are included in All Other when not separately shown. * indicates that the cell is combined with another non-adjacent cell.

[*Received January 1992. Revised November 1992.*]

REFERENCES

Belin, T., and Diffendal, G. (1991), "Results From the Handling of Unresolved Enumeration Status, Missing Characteristic Data, and Noninterviews in the 1990 Post-Enumeration Survey," STSD Decennial Census Memorandum Series V-112, *U.S. Bureau of the Census*, Washington, DC.

Furnival, G. M., and Wilson, R. W. (1974), "Regression by Leaps and Bounds," *Technometrics*, 16, 499–512.

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 17

October 11, 2000

Prepared by: Maria Urrutia and Annette Quinlan

The seventeenth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on October 11, 2000 at 10:30. Howard Hogan discussed the potential effects of late census data on the Dual System Estimation (DSE) process and the DSE variances from 1990.

Committee Attendees:

William Barron
Nancy Potok
Paula Schneider
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
John Long
Susan Miskura

Other Attendees:

Kenneth Prewitt
Donna Kostanich
Raj Singh
Tommy Wright
Sally Obenski
Kathleen Styles
Maria Urrutia
Annette Quinlan

I. Census Processes

Jay Waite updated the ESCAP on the status of the Hundred percent Census Unedited File (HCUF) creation. At the time of the meeting, over half of the HCUF state files had been approved.

In the most recent Executive State Of the Census (ESOC) report there was a reference to the possibility of duplicated addresses in the census. Jay briefly described this situation and summarized what actions are being taken to address this issue.

II. Dual System Estimation

Howard Hogan discussed the potential effect of late census data on the Dual System Estimates (DSEs). Late census data are treated, during matching and estimation, in the same manner as any whole person substitutions in the census. Since it would be impossible to match such people, late census data and whole person substitutions are not included in the matching operations or the calculation of the DSE. However, the late census data are included with the final census counts used in calculating the Coverage Correction Factors (CCFs); that is, they are included since the CCF is the ratio, for a given post-stratum, of the DSE to the final census count, including late census data.

Howard then discussed the table which will be prepared to summarize the A.C.E. results for ESCAP deliberation purposes (attached). The table will now include an additional variable to distinguish between the census counts before late census data and after late census data.

III. 1990 Estimates and Variances

Howard Hogan discussed the direct and smoothed estimation results and variances from the 1990 Post Enumeration Survey. This was to provide background information to the Committee for them to develop an understanding of what occurred in 1990.

In 1990, two models were used for estimating the total population, a direct model and a smoothed model. The smoothed model has been eliminated for the 2000 A.C.E. The direct and smoothed percent undercount rates for 1990 are summarized in the attached tables. The smoothing made some of the estimates more reasonable and resulted in less variation.

In 1990, when the decision was made to drop plans for smoothing the estimates, it was also decided to decrease the number of post-strata and thereby increase the number of people in each post-stratum so as to reduce the variance of the estimates. This history led to a discussion

of the relative size of each post-stratum. It was decided that a useful tool in aiding the ESCAP deliberation process would be a supplementary table showing the sizes of the post-strata. This would show how large the groups were and the potential impact on the estimated coverage errors that will be calculated for the post-strata.

The ESCAP also discussed the decisions that would be addressed. It was decided that the ESCAP would focus initially on the decision regarding adjustment of the redistricting data. The ESCAP would then determine whether additional decisions are necessary.

IV. Next Meeting

The next meeting scheduled for Wednesday October 25, 2000 will discuss loss functions.

Attachments

There was no agenda developed or used for the October 25, 2000 meeting.

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 18

October 25, 2000

Prepared by: Nick Birnbaum.

The eighteenth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on October 25, 2000 at 10:30. The agenda for the meeting was to familiarize the Committee members with the purpose of loss functions and the complexities inherent in this type of analysis. Refer to the DSSD Census 2000 Procedures and Operations Memorandum Series B-13 for a detailed discussion of the methodology to be used for conducting these analyses.

Committee Attendees:

William Barron
Nancy Potok
Nancy Gordon
Paula Schneider
Cynthia Clark
John Thompson
Ruth Ann Killion
Bob Fay
Howard Hogan
Susan Miskura
John Long

Other Attendees:

Kenneth Prewitt
Donna Kostanich
Raj Singh
Tommy Wright
Nick Birnbaum
Kathleen Styles
Maria Urrutia
Annette Quinlan
Carolee Bush

I. Brief Introduction to Loss Function Analysis

John Thompson began the meeting with a brief discussion of loss functions, summarizing what they are, how they are used, and providing an historical context (that is, the role they played in the 1990 adjustment decision process).

As has been mentioned earlier, the Census Bureau will base its decision on a comprehensive set of data and analyses -- the loss function analysis is one tool among many the Committee will utilize for examining the accuracy of the initial counts and the A.C.E. results.

At this point, John turned the meeting over to Howard Hogan, who provided the Committee with a brief A.C.E. operational update before giving his presentation on loss function analysis.

II. Loss Functions - The Fundamentals

Howard explained that loss functions are used to compare two sets of counts or share distributions -- unadjusted versus adjusted -- to determine which set is closer to the "true" count or share distribution. That is, loss functions involve comparing the census errors to the coverage measurement survey errors to determine which has the smaller "loss" when compared to the "true" counts or shares. Since the "true" count or share distribution can never be known, one has to rely on an estimated truth (a target number or share distribution) to perform the loss function analysis. Estimates of the "true" population and "true" population shares (or proportions) are produced for states and sub-state areas, depending on the desired level of analysis. These estimated "truths" have variances and biases associated with them, making the loss function analysis more complex. John briefly discussed how this comparison problem was addressed in 1990.

The input to the loss function analysis is based on a total error model used to estimate the net effect of sampling and non-sampling error in the initial census and the A.C.E. The components of the total error model are derived from the Census Bureau's evaluation studies providing various measures of sampling and non-sampling error. Because the Census Bureau will not complete some of its evaluations until late 2001 or 2002, complete information on the components of the total error model will not be available within the time frame for producing the ESCAP recommendation. Consequently, the comparison of accuracy between the adjusted and unadjusted Census 2000 population data will be modeled from 1990 census components. The methodology for developing the components of the Census 2000 total error model will be discussed in more detail at future ESCAP meetings.

III. Next Meeting

The agenda for the next meeting, scheduled for November 8, 2000, is to examine preliminary demographic analysis estimates.

Kathleen P Zveare
11/02/2000 01:51 PM

To: Margaret A Applekamp/DIR/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Angela Frazier/DMD/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Jane F Green/DSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC
cc:
Subject: Agenda for 11/8 ESCAP Meeting

The agenda for the November 8 ESCAP Meeting scheduled from 10:30-12 in Rm. 2412/3 is as follows:

Demographic Analysis - Greg Robinson

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 19

November 8, 2000

Prepared by: Nick Birnbaum.

The nineteenth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on November 8, 2000 at 10:30. The agenda for the meeting was to familiarize the Committee members with the techniques of demographic analysis (DA) and to provide preliminary data on the DA estimated resident population as of April 1, 2000.

Committee Attendees:

William Barron
Nancy Potok
Paula Schneider
John Thompson
Jay Waite
Ruth Ann Killion
Howard Hogan
Susan Miskura

Other Attendees:

Raj Singh
Tommy Wright
Gregg Robinson
Signe Wetrogan
Roxie Jones
Nick Birnbaum
Kathleen Styles
Maria Urrutia
Annette Quinlan
Carolee Bush

I. Demographic Analysis -- What It Is And How It Is Used

Gregg Robinson began his presentation by explaining demographic analysis (DA). The purpose of the meeting was to explain DA to the Committee members in preparation for their analysis of the Census 2000 data.

DA represents a macro-level approach to measuring net undercount, where analytic estimates are constructed based on various types of demographic data essentially independent of the census, such as administrative statistics on births, deaths, and immigration; estimates of emigration and undocumented migration, and Medicare data. The difference between the DA estimated population and the census count provides an estimate of net census undercount or overcount.

DA uses the following demographic accounting equation:

For the population under 65:

> Births (since 1935) minus Deaths (to persons born after 1935) plus Immigrants (born after 1935) minus Emigrants (born after 1935).

For the population 65 and over:

> Medicare Count plus
> Estimated unenrolled.

The immigration and emigration components are developed from the following sub-components:

<u>Immigration components</u>:

> Legal immigrants
> Net migration from Puerto Rico
> Net migration of temporary residents
> Net migration of Federal civilian employees
> Net movement of Armed Forces overseas
> Net undocumented migration

<u>Emigration components</u>:

> Foreign-born emigration
> Native emigration

The techniques of DA allow for refinement of the estimates: for example, adjustments for imperfections in the component data, such as those relating to birth registration completeness, or the constant evaluation of time series of DA coverage estimates and the underlying component data (for internal and longitudinal consistency).

Gregg presented a preliminary DA estimate for the U.S. Resident Population as of April 1, 2000, by DA component and broad age group. He also presented data on the change in components from April 1, 1990 to April 1, 2000. Gregg then walked the Committee through the preliminary component estimates for April 1, 2000, explaining how they were developed and the uncertainties associated with them. The official April 1, 2000 DA population estimates will be available to the Committee in sufficient time to allow for review and analysis of these data to properly inform its recommendation to the Director.

Gregg also presented historical data on DA estimates of net undercount by race and age, and DA estimates of sex ratios compared with the census and coverage measurement surveys.

Finally, there was some discussion of the assignment of race for birth data in which one parent is Black and the other Non-Black, and its impact on demographic analysis estimates of the Black and Non-Black populations.

II. Next Meeting

The agenda for the next meeting, scheduled for November 22, 2000, is to examine results from the A.C.E. Before Follow-up Person Matching.

Kathleen P Porter
11/20/2000 03:09 PM

To: Margaret A Applekamp/DIR/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Angela Frazier/DMD/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Jane F Green/DSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Marvin D Raines/DIR/HQ/BOC@BOC, Mary E Williams/DIR/HQ/BOC@BOC
cc: Kirsten K West/POP/HQ/BOC@BOC, Danny R Childers/DSSD/HQ/BOC@BOC
Subject: Agenda for Nov. 22 ESCAP Meeting

The agenda for the November 22 ESCAP Meeting scheduled from 10:30-12 in Rm. 2412/3 is as follows:

1. How the Bureau will release results of demographic analysis - John Long

2. Person duplication and A.C.E. non-interview rates - Dan Childers

3. Demographic benchmark analysis of housing units - Kirsten West

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting #20

November 22, 2000

Prepared by: Nick Birnbaum.

The twentieth meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on November 22, 2000 at 10:30.

The agenda for the meeting was to discuss how the Census Bureau will release the results of demographic analysis, and to examine results from A.C.E. Before Follow-up Person Matching and preliminary demographic analysis benchmarks of housing units.

Committee Attendees:

Paula Schneider
Nancy Potok
Cynthia Clark
John Thompson
Jay Waite
Bob Fay
Howard Hogan
John Long
Susan Miskura
Ruth Ann Killion

Other Attendees:

Kenneth Prewitt
Donna Kostanich
Dan Childers
Debbie Fenstermaker
Kathleen Styles
Maria Urrutia
Signe Wetrogan
Alan Tupek
Jason Devine
Marvin Raines
Tommy Wright
David Whitford
Nick Birnbaum
Carolee Bush
Annette Quinlan
Raj Singh
Kirsten West

I. Availability of Demographic Analysis (DA) Estimates

John Long spoke briefly about the availability of demographic analysis estimates with April 1, 2000, as the reference date. Census level (not adjusted for 1990 undercount) population estimates for April 1, 2000, are currently available to the public (on the Census Bureau's web site). These data are at the national level by age, sex, race, and Hispanic origin.

Additionally, users, if so inclined, could produce national level adjusted population estimates for April 1, 2000, by applying the 1990 PES or the 1990 demographic analysis undercount rates to the census level population estimates mentioned above. These undercount rates are also available on the Census Bureau's web site.

The official demographic analysis population data for April 1, 2000, will be released at the time the Committee delivers its report to the Director, as part of the documentation underlying its analyses and recommendation regarding the use of the adjusted data for redistricting purposes.

II. Demographic Benchmark Analysis of Housing Unit Estimates

Kirsten West and Jason Devine presented preliminary demographic benchmark analysis results for housing units. The demographic benchmark is the housing unit estimate for July 1999 projected to April 1, 2000. The analysis is done for the nation and for groups of counties by type of enumeration area (TEA). For the nation and mailout/mailback counties, the preliminary census file housing unit count was slightly lower than the benchmark. For counties that are solely update/leave, the preliminary census file housing unit count mildly exceeded the benchmark, but it was noted that there is greater uncertainty associated with the DA housing unit estimates for these areas.

III. Results from A.C.E. Before Follow-up (BFU) Person Matching

Dan Childers began his presentation on the results from the BFU Person Matching by reviewing data on P-sample noninterview rates (unweighted data). He also discussed data showing percentages of person and housing unit duplication in the E-sample (unweighted data). Dan will continue his presentation on the BFU Person Matching results at the next meeting.

IV. Next Meeting

The agenda for the next meeting, scheduled for November 30, 2000, is to continue discussions of the results from BFU Person Matching.

fyi
----- Forwarded by Maria E Urrutia/DMD/HQ/BOC on 01/29/2001 03:04 PM -----
Kathleen P Porter
11/28/2000 02:01 PM

To: Angela Frazier/DMD/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Danny R Childers/DSSD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, Margaret A Applekamp/DIR/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Marvin D Raines/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Mary E Williams/DIR/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Kathleen P Porter/DMD/HQ/BOC@BOC
cc:
Subject: Re: Agenda for Nov. 30 ESCAP Meeting

PLEASE REMEMBER TO BRING THE FOLDER THAT YOU RECEIVED AT THE LAST MEETING.

THANKS.

Kathleen P Porter
11/28/2000 10:49 AM

To: Margaret A Applekamp/DIR/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Robert E Fay

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 21

November 30, 2000

Prepared by: Maria Urrutia and Annette Quinlan

The twenty first meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on November 30, 2000 at 2:30. The agenda for the meeting was to continue the discussion of the A.C.E. Before Followup Match results.

Committee Attendees:

William Barron
Nancy Potok
Paula Schneider
Cynthia Clark
Nancy Gordon
John Thompson
Jay Waite
Bob Fay
Howard Hogan
Ruth Ann Killion
John Long

Other Attendees:

Kenneth Prewitt	Debbie Fenstermaker
Marvin Raines	Roxie Jones
Donna Kostanich	Kathleen Styles
Raj Singh	Maria Urrutia
David Whitford	Annette Quinlan
Danny Childers	

III/DIR/HQ/BOC@BOC, Angela Frazier/DMD/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Sue A Kent/DMD/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Marvin D Raines/DIR/HQ/BOC, Mary E Williams/DIR/HQ/BOC, Danny R Childers/DSSD/HQ/BOC

cc:

Subject: Agenda for Nov. 30 ESCAP Meeting

* * * * * * * * * * PLEASE NOTE THE DATE AND TIME OF THIS MEETING * * * * * * * * * * * *

The agenda for the November 30 ESCAP Meeting scheduled from 2:30-4:00 in Rm. 2412/3 is as follows:

A.C.E. Before Follow-Up Match Results (continued) - Dan Childers

I. Before Followup Match Results

Danny Childers continued the Before Followup (BFU) matching results discussion from the previous ESCAP meeting (November 22). All results discussed are the preliminary, unweighted results and don't include the Targeted Extended Search results. These data were discussed in the context of how they compared to 1990 results, and as indications of gross trends that might result in the final DSE data.

The preliminary BFU matching results for the P and E samples by different geographic areas, including Type of Enumeration Area (TEA) and region, were presented to the ESCAP. These data were discussed, and it was noted that, for the P-sample, they compared reasonably with analogous 1990 results.

The preliminary E-sample BFU erroneous enumeration results were also examined, but only final 1990 results were available, so a full comparison could not be made. It was possible to compare the geocoding error component with 1990 results. This comparison indicated that the level of geocoding error for 2000 was likely to be larger than in 1990. The effect of greater geocoding error was discussed, and it was noted that geocoding errors, if clustered, would result in increased variances. Therefore, we will carefully consider the estimates of variance that will be produced for the DSEs.

Another finding for the E-sample was the low percentage of people counted at the wrong location. Howard Hogan and his staff are examining potential causes for this finding, and will report back at a subsequent meeting.

The ESCAP also considered the BFU E- and P-sample results in terms of how they might predict final coverage errors. The data did not allow for any quantitative assessments, but did indicate that the minority post-strata were likely to have higher undercount rates than the non-minority post-strata. However, this comparison was very rough and could not be put into quantitative measurements.

II. Next Meeting

The next meeting scheduled for Wednesday December 6, 2000 will discuss Service Based Enumeration quality indicators.

Preston J Waite
Sent by: Sue A Kent
12/05/2000 03:36 PM

To: Angela Frazier/DMD/HQ/BOC@BOC, Annette M Quinlan/DMD/HQ/BOC@BOC, Barbara E Hotchkiss/DSD/HQ/BOC@BOC, Betty Ann Saucier/DIR/HQ/BOC@BOC, Carnelle E Sligh/PRED/HQ/BOC@BOC, Carolee Bush/DMD/HQ/BOC@BOC, Cynthia Z F Clark/DIR/HQ/BOC@BOC, Danny R Childers/DSSD/HQ/BOC@BOC, Donna L Kostanich/DSSD/HQ/BOC@BOC, Ellen Lee/DIR/HQ/BOC@BOC, Geneva A Burns/DMD/HQ/BOC@BOC, Hazel V Beaton/SRD/HQ/BOC@BOC, Howard R Hogan/DSSD/HQ/BOC@BOC, Jeannette D Greene/DIR/HQ/BOC@BOC, John F Long/POP/HQ/BOC@BOC, John H Thompson/DMD/HQ/BOC@BOC, Kathleen M Styles/DMD/HQ/BOC@BOC, Kathleen P Porter/DMD/HQ/BOC@BOC, Kenneth Prewitt/DIR/HQ/BOC@BOC, Linda A Hiner/DSSD/HQ/BOC@BOC, Lois M Kline/POP/HQ/BOC@BOC, Margaret A Applekamp/DIR/HQ/BOC@BOC, Maria E Urrutia/DMD/HQ/BOC@BOC, Marvin D Raines/DIR/HQ/BOC@BOC, Mary A Cochran/DIR/HQ/BOC@BOC, Mary E Williams/DIR/HQ/BOC@BOC, Nancy A Potok/DIR/HQ/BOC@BOC, Nancy M Gordon/DSD/HQ/BOC@BOC, Nicholas I Birnbaum/DMD/HQ/BOC@BOC, Patricia E Curran/DIR/HQ/BOC@BOC, Paula J Schneider/DIR/HQ/BOC@BOC, Phyllis A Bonnette/DIR/HQ/BOC@BOC, Preston J Waite/DMD/HQ/BOC@BOC, Rajendra P Singh/DSSD/HQ/BOC@BOC, Robert E Fay III/DIR/HQ/BOC@BOC, Ruth Ann Killion/PRED/HQ/BOC@BOC, Susan Miskura/DMD/HQ/BOC@BOC, Tommy Wright/SRD/HQ/BOC@BOC, William G Barron Jr/DIR/HQ/BOC@BOC
cc: Kathleen P Porter/DMD/HQ/BOC@BOC

Subject: 12-6 ESCAP Meeting Agenda (Noon - 1:30 p.m.)

The agenda for tomorrow's ESCAP Meeting (12:00 - 1:30 p.m.), in Room 2412/3 follows:

Preview of SBE Quality Indicators - Rick Griffin

Attached is an overview memorandum on SBE for your review.

UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

December 5, 2000

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES #Q-36

MEMORANDUM FOR Howard Hogan
Chief, Decennial Statistical Studies Division

From: Donna Kostanich *(signed 12/5/2000)*
Assistant Division Chief, Sampling and Estimation
Decennial Statistical Studies Division

Prepared by: Felipe Kohn
Estimation Staff
Decennial Statistical Studies Division

Subject: Census 2000 Service Based Enumeration: Overview of Multiplicity Estimation

I. INTRODUCTION

A key component of Census 2000 is the enumeration of persons with no usual residence. These persons had the opportunity to be enumerated during Service Based Enumeration (SBE) at shelters, soup kitchens, mobile food vans and Targeted Non-sheltered Outdoor Locations (TNSOLS). In addition, persons with no usual residence had the opportunity to be enumerated by completing Be Counted Forms (BCF). For the uncorrected Census count, the Census Bureau will include the persons actually enumerated (after unduplication) at shelters, soup kitchens, mobile food vans, TNSOLS and on BCF forms indicating no usual residence as the official count.

For the corrected Census count the Bureau plans to use the multiplicity estimator (based on the service usage questions asked in the questionnaires) to estimate the number of persons without usual residence who use shelters, soup kitchens and mobile food vans (SBE facilities). In addition, persons enumerated in TNSOLS and persons without usual residence that filed BCFs will augment the estimate. Since each of the SBE facilities is enumerated on only one day, the multiplicity estimator uses frequency of facility usage responses to estimate the number of persons using services but not on the day of enumeration. The multiplicity estimator is a statistical technique based on the service usage questions asked in the Individual Census Report (ICR) used in shelters, and the Individual Census Questionnaire (ICQ) used in soup kitchens and other facilities.

This memorandum provides an overview of the SBE estimation procedures. For more detail see the memorandum in Chapter Q of this series, Subject: Census 2000 Service-Based Enumeration: Computer Specifications for Multiplicity Estimation. Multiplicity Estimation is done independently for each county.

II ENUMERATION

Census Bureau enumerators visited shelters to collect information on the 27th of March and the following day went to collect information in soup kitchens and mobile food vans.

For the Census 2000 SBE enumeration, the Bureau used four forms: two for the population found in shelters and two for the population found in soup kitchens and mobile food vans. Questions were asked about shelter usage and soup kitchen or mobile food van usage. The responses to the shelter usage question and the soup kitchen usage question whenever the respondent answers "No" to the shelter usage question will be used for production estimation for the SBE population in the 2000 Census.

Shelter Usage

For the shelter population, the Bureau used the ICR short form as well as long form versions. In these questionnaires Question 2 d asks the following shelter question: "Including tonight, how many nights during the past 7 nights did you stay in a SHELTER?". The ICQ short and long versions given to persons in soup kitchens and using mobile food vans also ask clients about their shelter usage. Question 10 in the short version and question 9 in the long version ask "Including last night, during the past seven nights did you stay in a SHELTER?"(see summary table below). In the ICR questionnaire the respondent is asked about his/her usual place of residence. This question is on the ICR because this form is used for other data collection operations. However, for SBE we assume that all persons staying in a shelter should have the shelter's address as their usual place of residence regardless of their response to the usual place of residence question.

Soup Kitchen or Mobile Food Van Usage

The soup kitchen or mobile food van usage question is limited to the ICQ questionnaire short and long versions. Question 9 in the short and Question 8 in the long ask the following question: "Including today, how many days during the past seven days did you receive a meal from a SOUP KITCHEN or MOBILE FOOD VAN?"(see summary table below).

Each person is also asked if he/she has a usual residence. Any person who responds that he/she has a usual residence is removed from the file.

The ICR questionnaire is used for the TNSOL enumeration, but persons enumerated there are not asked the shelter usage question or the usual place of residence question.
In addition, persons without usual residence could fill and return a BCF. The BCF does not include any service usage question.

Table 1: Summary of Service Based Enumeration

Enumerated Population	**Date**	**Questionnaire Used**	**Usage Questions**
Shelter	March 27th	ICR Short & Long	"Including tonight, how many nights during the past seven nights did you stay in a shelter?"
Soup Kitchen or Mobile Food Van	March 28th	ICQ Short & Long	"Including today how many days during the past seven days did you receive a meal from a SOUP KITCHEN or MOBILE FOOD VAN?" "Including last night, during the past seven nights did you stay in a SHELTER?"
Targeted Non-Sheltered Outdoors Locations (TNSOLS)	March 31st	ICR Short	Not Asked

III. RULES OF UNDUPLICATION

The persons without usual residence that use service facilities are very transient by definition. Since the enumeration of shelters was done the 27th of March, and the enumeration of soup kitchens was done the following day, it is expected that at least some of the persons were enumerated more than once. In the paragraph below, we describe the rule we used for ensuring one (and only one) record for each person.

If a person is enumerated in a shelter and in a soup kitchen or food van, then when unduplicating the shelter questionnaire should be kept. If a person is enumerated in a service

facility and in a TNSOL, when unduplicating the form collected in the service facility should be kept. Finally, if a person is enumerated in SBE and on a BCF with no usual residence, the SBE questionnaire will be kept (see summary table below).

Table 2: Summary of Unduplications

Duplicate Enumerations		**Questionnaire Kept**
Shelter	Soup Kitchen	Shelter
Shelter	TNSOL	Shelter
Shelter	BCF	Shelter
Soup Kitchen	TNSOL	Soup Kitchen
Soup Kitchen	BCF	Soup Kitchen
TNSOL	BCF	TNSOL

IV. IMPUTATION

If a questionnaire has some missing demographic data the Population Division imputes that missing data item in the same way as in housing questionnaires. Imputation of missing response to the usage questions on SBE questionnaires is done as part of SBE estimation prior to multiplicity estimation. For each state, the mean usage of respondents will be imputed on the input file (Hundred percent Census Edited File, HCEF) for non-respondents within age/sex/type of facility (shelter or soup kitchen/mobile food van) groups prior to multiplicity estimation. Imputation cells with less than 10 respondents will be collapsed. For soup kitchen and mobile food van questionnaires, no imputation of the soup kitchen/mobile food van question is necessary if the response to the shelter usage question is "YES".

V. MULTIPLICITY ESTIMATION

A. If the usage question used in estimation had to be imputed for all persons in a shelter, soup kitchen, or mobile food van, then all persons at the SBE site are given a weight of one (i.e., no multiplicity estimation for the site).

B. Assigning Weights

1. Shelter Persons

After the unduplication of enumerated persons in shelters, the initial weight for all shelter person records will be calculated using the formula below:

$$W_j = \frac{7}{A_j}$$ where A_j is the response or imputed value for the ICR shelter usage question for the j-th person enumerated in a shelter .

2. Soup Kitchens and Mobile Food Vans persons

For all ICQ records (after unduplication) if their response to the shelter usage question is "YES" they are given a weight of zero for the multiplicity estimator. These persons are accounted for in the multiplicity factor applied to shelter respondents.

If the response to the shelter question is "NO" or there is "no response", then after the unduplication of enumerated persons in soup kitchens and mobile food vans, the initial weight for all soup kitchen and mobile food van records will be calculated using the formula below:

$$W_j = \frac{7}{B_j} where,$$ B_j is the response or imputed value for the ICQ soup kitchen usage response for the j-th person enumerated in a soup kitchen or mobile food van.

3. TNSOLS and BCF with no usual residence have an effective weight of one and are not subject to the multiplicity estimation process.

Thus, the multiplicity estimator of persons without usual residence that use services has the formula shown below:

$$\hat{X} = \sum_{j=1}^{n} \frac{7}{A_j} + \sum_{j=1}^{m} \frac{7}{B_j} + X_{TNSOL} + X_{BCF}$$

Where, after unduplication, n is the number of persons enumerated at shelters and m is the number of persons enumerated at soup kitchens or mobile food vans who did not answer "YES" to the shelter usage question.

VI. RECORD CREATION

The output file from multiplicity estimation will be the HEDF with SBE persons WITHOUT WEIGHTS. Thus, the multiplicity weights described in Section V. are reduced by one and control rounded to integers (the sum of the integer weights will be within 1 of the sum of the unrounded weights) to produce rounded replication weights. Any person who matched to a person enumerated at a TNSOL or BCF with no usual residence is given a replication weight of zero (i.e., they represent only themselves and no additional people) to account for other TNSOL or BCF persons who would have matched if SBE was done on a different day(s) of the week. For the final HEDF all SBE enumerated persons after unduplication remain on the file and are duplicated a number of times equal to these rounded replication weights.

In addition we do not want to remove persons from the file who have a weight of zero in the multiplicity estimator since they have taken the time to supply responses to demographic data. These are the persons enumerated at soup kitchens and mobile food vans who responded "YES" to the shelter usage question. If the multiplicity estimator turns out to be less than the uncorrected count this is indicative of considerable response bias. Thus, we use the following rule for each county.

Count 1: The number of persons added to the uncorrected count due to multiplicity estimation.

Count 2: The number of persons enumerated at soup kitchens and mobile food vans who responded "YES" to the shelter usage question (these are the persons given a weight of zero for the multiplicity estimate).

If, Count 2 is greater than Count 1 then multiplicity estimation will not be done for the county (since the multiplicity estimation process would produce a count lower than the uncorrected count) . The final count for SBE for the corrected Census count file will be the same as for the uncorrected Census count (i.e., the number of persons after unduplication enumerated at shelters, soup kitchens, mobile food vans, TNSOLS, or BCFs without usual residence).

If, Count 1 is greater than or equal to Count 2, a sample of size s = Count 2 of the persons with rounded replication weights not equal to zero will be selected allowing persons to be selected more than once with probability proportional to their replication weights. Each selected person will have their rounded replication weight reduced by 1 for each time they are selected in the sample (i.e., a person selected twice will have their rounded replication weight reduced by 2). Here we are reducing the number of duplications so that we can leave persons with a weight of zero on the file. For each person with a weight of zero, we reduce the number of duplications by one.

VII. REFERENCES

Sirken, M.G. (1970), "Households Surveys with Multiplicity", Journal of the American Statistical Association, Volume 63, 257-266.

Sirken, M.G. (1972), "Stratified Sample Surveys with Multiplicity", Journal of the American Statistical Association, Volume 67, 224-227.

Sirken, M.G. (1972), "Variance Components of Multiplicity Estimators", Biometrics, Volume 28, 869-873.

Sirken, M.G. and Leoy, P.S. (1974), "Multiplicity Estimation of Proportions based on Ratios of Random Variables"

Thompson, S.K., (1992), "Sampling", John Wiley and Sons, Inc.

cc: Census 2000 Procedures and Operations Memorandum Distribution List

Minutes of the Executive Steering Committee on
Accuracy and Coverage Evaluation (A.C.E.) Policy (ESCAP) Meeting # 22
December 6, 2000

Prepared by: Nick Birnbaum.

The twenty-second meeting of the Executive Steering Committee on Accuracy and Coverage Evaluation Policy was held on December 6, 2000 at 12:00.

The agenda for the meeting was to discuss preliminary data from the Service Based Enumeration.

Committee Attendees:

Nancy Potok
Paula Schneider
Cynthia Clark
John Thompson
Jay Waite
Bob Fay
William Barron
Nancy Gordon
Susan Miskura
John Long
Ruth Ann Killion

Other Attendees:

Marvin Raines
Donna Kostanich
Raj Singh
Debbie Fenstermaker
Roxie Jones
Nick Birnbaum
Carolee Bush
Kathleen Styles
Maria Urrutia
Annette Quinlan
Rick Griffin
Annetta Clark Smith
Denise Smith

Felipe Kohn

I. Discussion of Service Based Enumeration Procedures and Preliminary Data

DSSD staff provided the Committee with some background information regarding the Service Based Enumeration, particularly the use of the multiplicity estimator. The source of this information is DSSD Census 2000 Procedures and Operations Memorandum Series #Q-36, "Census 2000 Service Based Enumeration: Overview of Multiplicity Estimation." This document provides an overview of the SBE estimation procedures and is attached.

DSSD has examined both preliminary data from the actual counts from the SBE and related operations including Targeted Non-sheltered Outdoor Locations and Be Counted Forms indicating no usual residence, and the estimates using the multiplicity estimator for the SBE facilities. The actual counts (unweighted) will be included in the apportionment tabulations.

DSSD staff explained how the estimator is computed and walked through the shelter and soup kitchen usage questions. Respondents' answers are weighted to reflect the fact that the enumeration of SBE facilities only takes place on one day and one has to account, through estimation, for those service users who didn't utilize the services on that particular day. Preliminary SBE data were then presented to the Committee.

DSSD staff expressed concerns about the response patterns to the usage questions. Consequently, the Committee requested more detailed information regarding the usage patterns.

The Committee also reviewed unweighted Census 2000 SBE data in comparison to similar data from the 1990 S-Night operation.

II. Next Meeting

The next meeting, to be held on December 13, will examine more detailed demographic analysis estimates.

Attachment

At this time, additional ESCAP agenda, handouts, and minutes are not available. These documents will be available in the near future.