MASTER FILE

December 31, 2002

DSSD A.C.E. REVISION II MEMORANDUM SERIES #PP-12

PRED CENSUS AND SURVEY MEASUREMENT STAFF MEMORANDUM SERIES: CSM-A.C.E. REVISION II-08R

MEMORANDUM FOR:	Donna Kostanich Chair, A.C.E. Revision II Planning and Management Group Decennial Studies Statistics Division
From:	Mary Mulry *signed 12/31/02* Chair, A.C.E. Revision II Quality Indicators Group Statistical Research Division
Through:	David Hubble *signed 12/31/02* Assistant Division Chief, Evaluations Planning, Research, and Evaluation Division (PRED)
Prepared by:	Katie Bench Mathematical Statistician, PRED
Subject:	A.C.E. Revision II: Study Plan for P-sample Match Rate Corrected for Error Due to Inconsistent Poststratification Variables

This document contains the A.C.E. Revision II study plan for the "P-sample Match Rate Corrected for Error Due to Inconsistent Poststratification Variables" evaluation.

Contact Katie Bench at katie.bench@census.gov or at 301-457-4306 if you have any questions or comments.

cc: DSSD A.C.E. REVISION II MEMORANDUM SERIES Distribution List
     Ruth Ann Killion   (PRED)
     David Hubble   (PRED)

P-sample match rate corrected for error due to
inconsistent poststratification variables
Katie Bench

I. Background

An inconsistent poststratification bias in the dual system estimate (DSE) occurs when the coverage factors for gross undercount are derived for poststrata based on the P-sample and then applied to the poststrata based on census enumerations, or what we call E-sample poststrata (Haberman and Spencer, 2001). This inconsistency bias in the DSE was computed by Shelby Haberman for the original A.C.E. estimates, and used to produce a corrected P-sample population size. Inconsistency bias also affects estimates of the P-sample match rate. So, to measure the impact of this bias on the A.C.E. Revision II estimate, this evaluation will calculate a P-sample match rate corrected for the error due to inconsistent poststratification variables, and compare the A.C.E. Revision II estimate to the A.C.E. Revision II estimate adjusted for inconsistency bias.

II. Questions to be Answered

How much do inconsistent poststratification variables contribute to error in the estimation of the P-sample match rate, and hence the DSE?

III. Project Description and Methodology

A. Objectives

In this evaluation we will estimate the P-sample match rate corrected for error due to inconsistent poststratification variables, using a program written by Shelby Haberman. The corrected P-sample match rate will then be used to estimate a corrected DSE, and the bias in the DSE due to inconsistent poststratification variables. Variances for the corrected P-sample match rate, DSE, and associated biases will also be calculated using a simple jackknife estimation.

B. Statistical Methodologies

To estimate the P-sample match rate corrected for error due to inconsistent poststratification, we will run a modification of Shelby Haberman's program to produce the A.C.E. sample's corrected P-sample match rate. This will be done for the new A.C.E. P-sample poststrata. This corrected P-sample match rate will then be used to calculate poststrata DSEs corrected for the inconsistent poststratification variables as specified in the attachment "P-sample match rate corrected for error due to inconsistent poststratification variables," by Mary Mulry dated 11/22/02.

1

Shelby Haberman's program does not estimate variances for the corrected P-sample match rate. So, we will estimate variances for the corrected DSE and inconsistent poststratification bias using a jackknife estimation method. To estimate these variances, we first form 32 replicates. Second, we calculate the corrected P-sample match rate, the uncorrected P-sample match rate, and all other P-sample or E-sample components of the DSE for the full A.C.E. Revision II and the 32 replicates. Third, we calculate the variances using the full sample estimates and the estimates for the 32 replicates.

The inconsistent poststratification bias in the DSE will be calculated by A.C.E. Revision II full P-sample by full E-sample poststrata.

IV. Data Requirements

   A. Sources

   • Ztest2.dat updated with the new poststrata for the A.C.E. sample. Ztest2.dat contains linked P-sample and E-sample records
   • Production Sample Design File
   • Missing data output files which include:
      ‣ A.C.E. Revision II Estimation E-sample File
      ‣ A.C.E. Revision II Estimation P-sample File
      ‣ A.C.E. production P-sample File with new AGE imputation
      ‣ A.C.E. production E-sample File (unchanged from production)
   • Estimation and identification of duplicate persons in Census 2000 output files
   • File containing insufficient information persons, late add persons, and the Census count including late adds by poststrata.

   B. Output

   • Table of statistically significant differences between the A.C.E. Revision II DSE and the corrected A.C.E. Revision II DSE (inconsistency bias). The table will be broken out by A.C.E. Revision II full P-sample by full E-sample poststrata.

V. Division Responsibilities

   A. DSSD

      • Provide PRED with the needed input files.
      • Provide PRED with the A.C.E. Revision II DSE program used in Doug Olson's variance system.

B. PRED

- Update the file ztest2.dat with the new A.C.E. Revision II review sample P-sample poststrata.
- Revise Shelby's program to compute inconsistency for the A.C.E. sample using the new P-sample poststrata.
- Write and run programs to form r replicates, and create input files for Shelby's program.
- Run Shelby's program to compute inconsistency bias.
- Write and run programs to obtain the A.C.E. Revision II DSEs, corrected A.C.E. Revision II DSEs, and associated variances. These programs will incorporate a revised version of the A.C.E. Revision II DSE program used in Doug Olson's variance system.
- Write any necessary specifications.
- Perform the analysis.

It should be noted that significant work on the responsibilities listed above will begin after all A.C.E. Revision II plans are finalized.

VI. Milestone Schedule

ACTIVITY	DIVISION
Update Shelby Haberman's inconsistency bias program with new P-sample poststrata	PRED
Update the linked production A.C.E. P-sample and E-sample file (ztest2.dat) with the new P-sample poststrata	PRED
Update the program that creates input files for Shelby Haberman's program for each replicate	PRED
Receive input files necessary to calculate the A.C.E. Revision II DSEs	DSSD
Run the program that create input files for Shelby Haberman's program for each replicate	PRED
Run Shelby's program to compute inconsistency bias for each replicate	PRED
Write and run programs to obtain A.C.E. Revision II DSEs, corrected DSEs, and associated variances. These programs will incorporate a revised version of the A.C.E. Revision II DSE program used in Doug Olson's variance system.	PRED
Perform Analysis	PRED
Prepare Report	PRED

VII. Limitations

The variances will be based on only 32 replicates; instead, of on a replicate for each A.C.E. cluster.

VIII. References

Haberman, Shelby J., and Spencer, Bruce D. (2001), "Estimation of Inconsistent Poststratification in the 2000 A.C.E.," August 29, 2001.

Mulry, Mary. (2002), "P-Sample match rate corrected for error due to inconsistent poststratification variables," draft November 22, 2002.

P-Sample match rate corrected for error
due to inconsistent poststratification variables

Mary Mulry

Inconsistency in the E-sample and P-sample reporting of the characteristics used in defining the poststrata may create a bias in the dual system estimate (DSE). This bias affects the estimation of the P-sample match rate.

The analysis for the A.C.E. Revision II will follow a similar investigation for the original A.C.E. The basic approach is to estimate the inconsistency in the poststratification variables using the matches and then assume that the rates also held for the nonmatches. The models used for the inconsistency of the original A.C.E. poststrata ("Estimation of Inconsistent Poststratification in the 2000 A. C. E.", by Shelby J. Haberman and Bruce D. Spencer, 12/17/01) were fitted in two steps, first (I) models for inconsistency of basic variables, and then (ii) derivation of inconsistency probabilities for poststratification given the inconsistency probabilities of the basic variables. The inconsistency probabilities led to an estimate of the bias in the P-sample match rate that was used to estimate the bias in the DSE.

The approach we are taking for the A.C.E. Revision II DSE is to calculate the proportions for the poststrata for the A.C.E. Sample. The proportions will not be applied in calculations of the double sampling adjustments based on the A.C.E. Revision II Sample, a subsample of the A.C.E. Sample. We assume the models in (I) and (ii) have been revised to reflect revisions to the variables used in the P-sample poststratification.

We examine the error due to inconsistent poststratification variables with the current formulation of the match rate for the calculation of the A.C.E. Revision II presented in "Chapter 6: A.C.E. Revision II Estimation, draft 9/19/02" by William Bell, Richard Griffin, Donna Kostanich, and Eric Schindler. We will use the same definitions of variables as found in the draft of Chapter 6.

Match rate for A.C.E. Revision II

The match rate for poststratum j for the calculation of the A.C.E. Revision II from Equation (6) of the draft of Chapter 6 is the following:

$$r_{Mj} = \frac{M_{nm,j}^{ND} f_{j'}^{Mnm} + \dfrac{M_{om,j} f_{j'}^{Mom}}{P_{om,j} f_{j'}^{Pom}} [P_{im,j} f_{j'}^{Pim} + g_j \sum_{s\in j} W_{P,s}\, p_s(1-h_s)PRres_{P,s}] + \sum_{s\in j} W_{P,s}\, p_s h_s PRres_{P,s} PRm_{P,s}}{P_{nm,j}^{ND} f_{j'}^{Pnm} + [P_{im,j} f_{j'}^{Pim} + g_j \sum_{s\in j} W_{P,s}\, p_s\,(1-h_s)PRres_{P,s}] + \sum_{s\in j} W_{P,s}\, p_s\, h_s\, PRres_{P,s}}$$

where

$P_{nm,j}^{ND}$ = P-sample nonmovers without a census duplicate in poststratum j

$M_{nm,j}^{ND}$ = P-sample nonmover matches without a census duplicate in poststratum j

$P_{om,j}$ = P-sample outmovers in poststratum j

$M_{om,j}$ = P-sample outmover matches in poststratum j

$P_{im,j}$ = P-sample inmovers in poststratum j

$f_{j'}^{MG}$ = double sampling adjustment for P-sample group G without census duplicates, where G = nm or om, in A.C.E. Revision II Sample poststratum j'. The A.C.E. Revision II Sample poststrata are collapsed A.C.E. sample poststrata.

$f_{j'}^{PG}$ = double sampling adjustment for matches in group G without census duplicates, where G = nm, om, or im, in A.C.E. Revision II Sample poststratum j'

p_s = probability that person s has a census duplicate outside the search area

h_s = probability that person s with a census duplicate outside the search area is retained after unduplication (see Appendix 6.1 of draft Chapter 6)

$W_{P,s}$ = P-sample weight for person s. The weight is assumed to include the probability of residence in draft Chapter 6, but that formulation needs to be reconsidered.

$PRm_{P,s}$ = probability person s with a census duplicate was matched in production

g_j = estimated proportion of P-sample persons in poststratum j with census duplicates outside the search area who are not retained as resident nonmovers by the duplicate study because they should have been coded as inmovers.

For ease of discussion, we rewrite the match rate for poststratum j as

$$
r_{M,j} = \frac{M_{nm,j}^{ND} f_{j'}^{Mnm} + \dfrac{M_{om,j} f_{j'}^{Mom}}{P_{om,j} f_{j'}^{Pom}} \left[P_{im,j} f_{j'}^{Pim} + P_{nm-im,j}^{D} \right] + M_{nm,j}^{D}}{P_{nm,j}^{ND} f_{j'}^{Pnm} + P_{im,j} f_{j'}^{Pim} + P_{nm-im,j}^{D} + P_{nm,j}^{D}}
$$

where

$$
P_{nm,j}^{D} = \sum_{s \in j} W_{P,s} \, p_s \, h_s \, PRres_{P,s} = \text{P-sample nonmovers with census duplicates in poststratum j}
$$

$$P^{D}_{nm-im,j} \; = \; g_j \sum_{s \in j} W_{P,s} \; p_s \; (1-h_s) \; PRres_{P,s} = \text{P-sample nonmovers with census duplicates in}$$

poststratum j who are not retained as nonmovers by the duplicate study because they should have been coded as inmovers.

$$M^{D}_{nm,j} \; = \; \sum_{s \in j} W_{P,s} \; p_s \; h_s \; PRres_{P,s} \; PRm_{P,s} = \text{P-sample nonmover matches with census duplicates in}$$

poststratum j

Correcting match rate for inconsistent poststratification variables

To define a match rate corrected for inconsistent poststratification, we need to define

$\hat{f}_G(j,k) \; = \;$ the proportion of group G persons enumerated in P-sample poststratum k who belong to P-sample poststratum j, based on their E-sample poststratification variables. The estimation of this proportion is based on the matched P-sample persons in group G. In this application, group G may be nonmovers, outmovers, or inmovers.

Next we need to define the following:

$$P^{ND}_{nm,j,I} \; = \; \sum_{k} \hat{f}_{nm}(j,k) \; P^{ND}_{nm,k}$$

$$M^{ND}_{nm,j,I} \; = \; \sum_{k} \hat{f}_{nm}(j,k) \; M^{ND}_{nm,k}$$

$$P_{G,j,I} \; = \; \sum_{k} \hat{f}_{G}(j,k) \; P_{G,k}, \text{ for G = im or om}$$

$$M_{om,j,I} \; = \; \sum_{k} \hat{f}_{om}(j,k) \; M_{om,k}$$

$$P^{D}_{G,j,I} \; = \; \sum_{k} \hat{f}_{nm}(j,k) \; P^{D}_{G,j} \; , \text{ for G = nm or nm-im}$$

$$M^{D}_{nm,j,I} \; = \; \sum_{k} \hat{f}_{nm}(j,k) \; M^{D}_{nm,j}$$

Then we define the match rate corrected for inconsistent poststratification variables, assuming no other errors are present, by the following:

$$r_{M,j,I} = \frac{M^{ND}_{nm,j,I}\, f^{Mnm}_{j'} + \dfrac{M_{om,j,I}\, f^{Mom}_{j'}}{P_{om,j,I}\, f^{Pom}_{j'}}\,[\, P_{im,j,I}\, f^{Pim}_{j'} + P^{D}_{nm-im,j,I}\,] + M^{D}_{nm,j,I}}{P^{ND}_{nm,j,I}\, f^{Pnm}_{j'} + P_{im,j,I}\, f^{Pim}_{j'} + P^{D}_{nm-im,j,I} + P^{D}_{nm,j,I}}$$

Calculation of error in the A.C.E. Revision II

Recall the general form of the DSE from Equation (2) in draft Chapter 6 is

$$DSE = Cen\,(1 - r_{II})\,\frac{r_{CE}}{r_{M}}$$

where

$r_{II} = (II + LA)\,/\,Cen$ for the given post-stratum, with II as the census imputations, LA as the late adds, and Cen the census count including the late adds.

r_{CE} = the correct enumeration rate for the given poststratum. The formula for the correct enumeration rate that incorporates a correction for the double sampling and census duplicates is given in Equation (5) in the draft of Chapter 6.

The A.C.E. Revision II estimate for estimation cell ij formed by the intersection of E-sample poststratum I and P-sample poststratum j is

$$DSE_{ij} = Cen\,(1 - r_{II,ij})\,\frac{r_{CE,i}}{r_{M,j}}\,.$$

Then the A.C.E. Revision II estimate that incorporates a correction for the error due to inconsistent poststratification variables for the estimation cell ij is given by

$$DSE_{ij,I} = Cen\,(1 - r_{II,ij})\,\frac{r_{CE,i}}{r_{M,j,I}}$$

The bias due to inconsistent poststratification variables is

$$DSE_{ij} - DSE_{ij,I}$$