UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

MASTER FILE

December 31, 2002

DSSD A.C.E. REVISION II MEMORANDUM SERIES #PP-57

MEMORANDUM FOR: Donna L. Kostanich

From: Michael Beaghen M. B.
Chair, A.C.E. Revision II Missing Data Subgroup

Prepared by: Michael Beaghen, Robert Sands
Decennial Statistical Studies Division

Subject: A.C.E. Revision II: Results from the Imputation for Unresolved Enumeration, Residency and Match Status

This memorandum documents the results of the assignment of probabilities of enumeration status to unresolved E-sample people and of probability of Census Day residency and/or match status to unresolved P-sample people in the Accuracy and Coverage Evaluation Revision II. Contact Michael Beaghen at Michael.A.Beaghen@Census.gov or at 301-763-9258 if you have any questions or comments.

A.C.E. Revision II
December 31, 2002

A.C.E. Revision II: Results from the Imputation for Unresolved Enumeration, Residency and Match Status

Michael Beaghen and
Robert Sands

Decennial Statistical
Studies Division

USCENSUSBUREAU

Helping You Make Informed Decisions

Introduction

This memorandum documents the results of the assignment of probabilities of enumeration status to unresolved E-sample people and of probability of Census Day residency and/or match status to unresolved P-sample people in the Accuracy and Coverage Evaluation Revision II. It also discusses the assignment of probabilities to conflicting cases. The methodology is specified in Beaghen & Sands (2002).

In the Accuracy and Coverage Evaluation (A.C.E.)[1], P-sample people with unresolved Census Day residency or match status occurred in one of two ways. First, the A.C.E. person interview (PI) may not have provided sufficient information for match and followup. Second, the A.C.E. person followup (PFU) may not have collected adequate information to allow us to determine a person's Census Day residency status or their match status. A.C.E. E-sample people with unresolved enumeration status likewise arose in this second manner; the PFU did not collect adequate information to determine the person's enumeration status. In the A.C.E. Revision II unresolved cases also arose because of the Evaluation Followup (EFU).

The A.C.E. Revision II assignment of probability of correct enumeration, residency and match status used the method of defining imputation cells and donor pools. See Ikeda & McGrath (2001) for how this was done in the A.C.E. It is necessary to define some key terms used in this document:

imputation cell:	a group of people, both resolved and unresolved, who are similar in some way relevant to their enumeration, residency or match status;
recipients:	the unresolved people associated with an imputation cell;
donors:	the resolved people associated with an imputation cell.

The proportion of donors who are correctly enumerated, residents or matched, is the probability of correct enumeration, residency or match that we assign to the unresolved people in the cell (for an illustration of the method see **Example 1** below).

[1]In this document A.C.E. refers to the original A.C.E. that generated the March 2001 estimates of census coverage.

Imputation for Revision II[2] P-Sample People with Insufficient Information for Match and Followup

The Revision II P-sample people with insufficient information for match and followup tended to be the same people who had insufficient information for match and followup in the A.C.E., except for some rare cases with coding changes. People who had insufficient information in the A.C.E. were not sent to EFU. There were about three million weighted people with insufficient information for match and followup in both the A.C.E. and the Revision II samples.

In the A.C.E., P-sample people with insufficient information for match and followup were assigned a probability of Census Day residency equal to the residency rate of P-sample people who went to PFU. In the A.C.E. Revision II we improved upon this by defining finer imputation cells that took into account whether or not the housing unit was matched, non-matched, or had a conflicting household (a conflicting household was said to exist when the P-sample household had no people in common with the E-sample household). See **Table 5** for the number of recipients or unresolved cases, the number of donors who are residents and the total number of donors for each cell.

The probability of match was assigned based on the overall match rate, divided into groups based on mover status and housing unit match status as was done in the A.C.E., and additionally on conflicting household status. See **Table 6** for the number of recipients or unresolved cases, the number of donors who are matched and the number of donors who are residents for each cell.

Imputation for P-Sample and E-Sample People with Incomplete or Ambiguous Followup

In contrast to P-sample people with insufficient information, the residency status for Revision II P-sample people and the correct enumeration status for Revision II E-sample people often changed from the A.C.E. to the Revision II coding because the Revision II coding processed not just the original information from the PFU, but also the new information from the EFU. Thus while the EFU information resolved many cases that were unresolved in the A.C.E. on account of the PFU.

EFU cases with incomplete or ambiguous information were a new source of unresolved cases in the Revision II coding. There were about the same weighted number of E-sample unresolved cases in the Revision II as in the A.C.E., more than six million. About half of the six million Revision II E-sample unresolved were new unresolved cases resulting from EFU information; about half were the same people as in the A.C.E.; note that the EFU information allowed us to

[2]Please note that the A.C.E. Revision II sample that was used to correct for measurement error is known as the Revision II sample and the people selected in that sample are known as Revision II people. Likewise, the results of the A.C.E. Revision II coding are known as the Revision II coding.

resolve about three million people unresolved in the A.C.E. In contrast, the Revision II coding generated substantially more P-sample unresolved cases than the A.C.E., 4.6 million versus 2.7 million. We saw this increase because all the Revision II P-sample except those with insufficient information went to EFU, including whole households of non-matched people who had not gone to PFU. These people were assumed in the A.C.E. to be resolved and could have become unresolved because of the EFU.

Originally the A.C.E. missing data plan based the imputation cells on information obtained before any followup was conducted. An ad hoc fix to the A.C.E. missing data methodology was effected by using information from the person followup (Cantwell & Childers, 2001). Based on the PFU keyed data we created the after followup groups for 'potential fictitious' and 'lived elsewhere on Census Day'. The new cells used information highly relevant to resident or enumeration status. Further, they showed greater discrimination in assigning probabilities of correct enumeration and residency. In the A.C.E. Revision II we entirely abandoned the before followup imputation cells and defined our cells based on after followup information. This change was the single most important improvement in the A.C.E. Revision II missing data methodology.

To define the after followup groups we employed the keyed responses to the PFU and EFU questionnaire check boxes and the 'why' codes. Why codes were clerically applied codes that took into account both the responses in the questionnaire checkboxes and the handwritten notes (Adams & Krejsa, 2002). Using the keyed results and the why codes we identified the following:

- unresolved cases with the same history, i.e., the recipient or imputation cells;
- the resolved followup cases that shared that history up to the point of being unresolved, i.e., the donor pool.

We defined PFU after followup groups for those cases that were unresolved as a result of the PFU, and EFU after followup groups for those cases unresolved on account of the EFU. It was necessary to define separate groups for the PFU and EFU because their interviews and questionnaires were different. However, the same after followup groups were employed for the P-sample and E-sample unresolved cases, as the PFU and EFU questions about Census Day residency were the same as the EFU and PFU questions about enumeration status.

It often happened that both the PFU interview and the EFU interview were unresolved. In that case in order to assign a cell for imputation we chose the unresolved interview that was more informative. When both interviews had the same level of information we usually chose the EFU over the PFU because we believed the EFU questionnaire questions were more sharply defined.

At this point it may help to give an example of an after followup group.

Example 1
One cell of unresolved E-sample people or recipients was defined as people with evidence from the EFU interview that they had moved in since Census Day, or moved out before Census Day, though the EFU interview did not provide the address they moved to or from. We could not determine the enumeration status of these people since we did not know whether the Census Day address was in the A.C.E. cluster. The corresponding donor pool consisted of those resolved people who indicated in the followup that they moved in after Census Day, or moved out before Census Day; these were generally people who provided the mover address in the EFU.

In **Table 1** this is the cell 'Moved In after Census Day or Moved out before Census Day' (EMO). Note that there were 1,537,389 recipients. Of the 1,701,178 resolved people who moved in before Census Day, or moved out after Census Day, 472,549 were correctly enumerated. Thus we assign the recipients a probability of correct enumeration of 0.27778.

We had an analogous after followup group for people unresolved because they indicated they moved in after Census Day or moved out before Census Day on the person followup interview. This cell is found in **Table 2** (PMO).

Table 1 shows the E-sample EFU cells; **Table 2** shows the E-sample PFU cells; **Table 3** shows the P-sample EFU cells; and **Table 4** shows the P-sample PFU cells.

Revision II E-Sample and P-Sample Conflicting Coding Cases

When the A.C.E. person followup (PFU) and the evaluation followup (EFU) interviews had contradictory information and we could not determine which was correct, the Revision II coding assigned the case a code of conflicting (conflicting coding is not to be confused with conflicting households, which was described earlier). All cases found to be conflicting in the Revision II automated coding were sent to analysts for clerical review. By examining the handwritten notes of interviewers, the analysts could often determine which of the interviews was the better and appropriately assign a code. There were some cases where the interviews appeared to be of equal quality, such as when both respondents were household members or both respondents were of equal caliber proxy. For these conflicting cases, the interviews seemed equally likely to be correct based on the expertise of the analysts. Therefore, probabilities of 0.5 were assigned both for correct enumeration status of Revision II E-sample conflicting cases and for Census Day residency status of Revision II P-sample conflicting cases. It should be noted that the recoding of the Revision II samples resulted in considerably less conflicting cases than the PFU/EFU Review sample. The PFU/EFU Review sample had about 2.6 million weighted people (Adams & Krejsa, 2001) in contrast to only about 100,000 weighted people in the Revision II samples (Adams & Krejsa, 2002).

REFERENCES

Adams, T. & Krejsa, E., 2001: *ESCAP II: Results of the Person Followup and Evaluation Followup Forms Review,* Executive Steering Committee for A.C.E. Policy II (ESCAP II) Report 24.

Adams, T. & Krejsa, E., 2002: *A.C.E. Revision II Measurement Subgroup Documentation,* DSSD A.C.E. Revision II Memorandum Series #PP-6.

Beaghen, M. & Sands, D. 2002: *A.C.E. Revision II: Specifications for the Assignment of Probability of Enumeration Status, Census Day Residency and Match Status,* A.C.E. Revision II Memorandum Series #PP-23.

Cantwell, P., & Childers, M., 2001: *Accuracy and Coverage Evaluation Survey: A Change to the Imputation Cells to Address Unresolved Resident and Enumeration Status,* DSSD Census 2000 Procedures and Operations Memorandum Series Chapter Q-44.

Ikeda, M., & McGrath, D., 2001: *Accuracy and Coverage Evaluation Survey: Specifications for the Missing Data Procedures; Revision of Q-25,* DSSD Census 2000 Procedures and Operations Memorandum Q-62.

Table 1: E-sample EFU Cells

Cell	Cell Description	Number of Recipients	Number of Donor Correct Enumerations	Total Number of Donor Enumerations	Proportion Correct
EKR[1]	Potential Fictitious	258,195.7[2]	805,729.52 (553)[3]	929,467.01 (675)	0.86687
EMO	Mover In After Census Day or Moved Out before Census Day	1,537,389	472,548.92 (318)	1,701,177.79 (996)	0.27778
EMP	Mover Status Unresolved	155,432.3	4,344,918.87 (1,606)	4,491,184.84 (1,721)	0.96743
EIB	Non-interview: Non-Conflicting Household	493,726.2	226,383,614 (50,994)	230,559,049.48 (53,958)	0.98189
EIC	Non-interview: Conflicting Household	55,301.7	1,523,652.89 (1,378)	1,829,650.09 (1,597)	0.83276
E2B	Non-interview - blank form: Non-Conflicting Household	329,535.2	247,281,151.92 (58,541)	253,385,774.68 (62,965)	0.97591
E2C	Non-interview - blank form: Conflicting Household	17,242.4	1,933,578.18 (1,795)	2,617,556.99 (2,525)	0.73870

[1]The three letter cell names are included to identify the cells with the definitions in Beaghen & Sands (2002).

[2]The weights of these counts reflect the Revision II sampling and the Targeted Extended Search sampling.

[3]In parenthesis are the unweighted counts; each person with a non-zero weight counts as one.

Table 1: E-sample EFU Cells

Cell	Cell Description	Number of Recipients	Number of Donor Correct Enumerations	Total Number of Donor Enumerations	Proportion Correct
ENL	Never Lived Here	216,696.8	1,677,690.73 (617.5)[4]	2,516,994.46 (1,171)	0.66655
EOR	Other Residence - Don't know which residency was Census Day	377,386.9	7,405,391.65 (1,631)	8,844,744.28 (2,506)	0.83726
EKB	Didn't Answer Other Residence Questions: Non-Conflicting Household	1,966,332	212,000,021.41 (45,961.5)	214,431,816.54 (47,817)	0.98866
EKC	Didn't Answer Other Residence Questions: Conflicting Household	160,395.8	1,324,904.89 (1,199.5)	1,493,070.18 (1,318)	0.88737
EAD	Didn't Answer Other Residence Questions: Other Residence - Didn't give address	45,543.43	531,151.53 (198)	1,874,182.50 (974)	0.28340

[4]Conflicting people count as one half a correct enumeration.

Table 2: E-sample PFU Cells

Cell	Cell Description	Number of Recipients	Number of Donor Correct Enumerations	Total Number of Donor Enumerations	Proportion Correct
PKR[5]	Potential Fictitious	4,166.39[6]	184,481.36 (185)[7]	308,772.33 (321)	0.59747
PMO	Mover In After Census Day or Moved Out before Census Day	166,981.1	89,041.52 (104.5)	573,138.76 (416)	0.15536
PN2	Non-interview - 'Did Not Live Here'	85,232.62	1,131,659.57 (931.5)	3,333,957.28 (2,407)	0.33943
PNI	Non-interview - 'Lived Here'	10,464.51	31,209,023.55 (29,139.5)	32,855,352.15 (30,964)	0.94989
PN3	Non-interview - DK/ref 'Lived Here'	8,663.304	153,065.70 (107)	215,987.45 (162)	0.70868
PN4	Noninterview - blank 'Lived Here'	31,861.69	2,442,652.43 (2,093)	3,243,196.14 (3,029)	0.75316
POR	Other Residence - Don't know which residency was Census Day	33,784.64	1,001,087.32 (685.5)	2,033,553.60 (1,401)	0.49228

[5]The three letter cell names are included to identify the cells with the definitions in Beaghen & Sands (2002).

[6]The weights of these counts reflect the Revision II sampling and the Targeted Extended Search sampling.

[7]In parenthesis are the unweighted counts; each person with a non-zero weight counts as one.

Table 2: E-sample PFU Cells

Cell	Cell Description	Number of Recipients	Number of Donor Correct Enumerations	Total Number of Donor Enumerations	Proportion Correct
POK	Didn't Answer Other Residence Questions	470,829.90	30,115,957.84 (27,873)	31,498,796.39 (29,471)	0.95610
PAD	Other Residence - Didn't give address	14,221.12	151,885.08 (109.5)	956,734.15 (675)	0.15875

Table 3: P-sample EFU Cells

Cell	Cell Description	Number of Recipients	Number of Donor Residents	Total Number of Donor Residents and Nonresidents	Proportion Residents
EKR[8]	Potential Fictitious	151,069.5[9]	904,879.23 (412)[10]	936,637.74 (447)	0.96609
EMO	Mover In After Census Day or Moved Out before Census Day	1,653,063	364,045.97 (125)	1,569,764.09 (731)	0.23191
EMP	Mover Status Unresolved	155,377.6	3,917,388.91 (1,153)	4,055,181.28 (1,237)	0.96602
EIB	Noninterview: Conflicting Household	60,756.87	1,513,202.02 (1,498)	2,005,783.54 (1,915)	0.75442
EIC	Noninterview: Non-Conflicting Household, PFU	81,576.77	7,577,687.33 (6,030)	9,329,510.61 (7,234)	0.81223
EID	Noninterview: Non-Conflicting, No PFU	345,598.2	217,366,749.65 (28,492)	218,941,159.07 (28,816)	0.99281
E2B	Noninterview - blank form: Conflicting Household	8,049.89	1,795,011.74 (1,792)	2,645,598.19 (2,538)	0.67849

[8]The three letter cell names are included to identify the cells with the definitions in Beaghen & Sands (2002).

[9]The weights of these counts reflect the Revision II sampling and the Targeted Extended Search sampling, though not the non-interview adjustment.

[10]In parenthesis are the unweighted counts; each person with a non-zero weight counts as one.

Table 3: P-sample EFU Cells

Cell	Cell Description	Number of Recipients	Number of Donor Residents	Total Number of Donor Residents and Nonresidents	Proportion Residents
E2C	Noninterview - blank form: Non-Conflicting Household, PFU	36,212.83	10,368,040.0 (8,336)	12,743,183.28 (10,008)	0.81361
E2D	Noninterview - blank form: Non-Conflicting, No PFU	237,366.4	235,189,934.63 (31,750)	237,816,995.71 (32,314)	0.98895
ENL	Never Lived Here	230,434.3	1,545,528.60 (435.5)	2,323,139.27 (800)	0.66528
EOR	Other Residence - Don't know which residency was Census Day	423,376.1	7,068,943.66 (1,089)	8,544,598.88 (1,784)	0.82730
EKF	Didn't Answer Other Residence Questions: PFU	390,959.9	7,742,010.95 (6,484.5)	8,611,769.78 (7,137)	0.89900
EKU	Didn't Answer Other Residence Questions: No PFU	433,048.6	204,532,556.33 926,091)	205,332,178.56 (26,241)	0.99611
EAD	Other Residence - Didn't give address	56,366.33	467,233.79 (130.5)	1,860,984.32 (787)	0.25107

Table 4: P-sample PFU Cells

Cell	Cell Description	Number of Recipients	Number of Donor Residents	Total Number of Donor Residents and Nonresidents	Proportion Residents
PKR[1]	Potential Fictitious	11,807.56[12]	56,998.54 (78)[13]	166,260.98 (162)	0.34283
PMO	Mover In After Census Day Moved or Out before Census Day	127,713.9	26,422.15 (23)	468,443.75 (335)	0.05640
PN2	Noninterview - 'Did Not Live Here'	129,883.3	442,518.47 (350.5)	2,476,383.88 (1,611)	0.17870
PNI	Noninterview - 'Lived Here'	7,810.71	10,053,748.47 (9,432)	10,995,966.27 (10,235)	0.91431
PN3	Noninterview - DK/ref 'Lived Here'	11,626.56	56,998.57 (54)	116,031.88 (106)	0.49123
PN4	Noninterview - blank 'Lived Here'	28,382.97	2,036,760.76 (1,714.5)	2,519,011.65 (2,159)	0.80856
POR	Other Residence - Don't know which residency was Census Day	34,409.83	479,948.96 (356.5)	1,290,664.80 (981)	0.37186

[11]The three letter cell names are included to identify the cells with the definitions in Beaghen & Sands (2002).

[12]The weights of these counts reflect the Revision II sampling and the Targeted Extended Search sampling, though not the non-interview adjustment.

[13]In parenthesis are the unweighted counts; each person with a non-zero weight counts as one.

Table 4: P-sample PFU Cells

Cell	Cell Description	Number of Recipients	Number of Donor Residents	Total Number of Donor Residents and Nonresidents	Proportion Residents
POK	Didn't Answer Other Residence Questions	278,749.7	9,753,562.52 (9,154)	10,562,568.29 (9,847)	0.92341
PAD	Other Residence - Didn't give address	37,641.17	98,562.95 (62)	698,129.27 (488)	0.14118

Table 5: P-sample Insufficient Information for Match and Followup - Cells for Probability of Residency

Cell	Cell Description	Number of Recipients	Number of Donor Residents	Total Donor Residents and Nonresident	Proportion Residents
PKC[14]	Insufficient information for match and followup: Conflicting Household	177,124[15]	1795011.74 (1,792)[16]	2,645,598.19 (2,538)	0.67849
PKH	Insufficient information for match and followup: Matched Housing Unit	2,675,959	9253499.95 (7,053.5)	11,404,116.90 (8,405)	0.81142
PKN	Insufficient information for match and followup: Non-Matched Housing Unit	230,764.4	1541514.59 (2,705.5)	2,057,678.58 (3,168)	0.74915
PPM	Possible matches	26,639.94	1672552.11 (1,335)	1,758,639.97 (1,407)	0.95105

[14]The three letter cell names are included to identify the cells with the definitions in Beaghen & Sands (2002).

[15]The weights of these counts reflect the Revision II sampling and the Targeted Extended Search sampling, though not the non-interview adjustment.

[16]In parenthesis are the unweighted counts; each person with a non-zero weight counts as one.

Table 6: P-sample Insufficient Information and Possible Match - Cells for Match Probability

Cell	Cell Description	Number of Recipients	Number of Donor Matched Residents	Total Donor Residents	Proportion Matches
MCF[17]	Conflicting Household	185,677.7[18]	2,316,763.18 (217)[19]	260,260.16 (2,323)	0.11234
MHN	HU Match, Non-mover	1,347,498	214,976,518.08 (28,595)	227,274,404.3 (34,938)	0.94589
MHU	HU Match, Mover	1,346,547	4,672,696.79 (1,871)	5,860,982.16 (2,698)	0.79725
MNN	HU Non-match, Non-mover	86,851.52	14,064,082.68 (2,896)	17,358,866.88 (5,859)	0.81020
MMN	HU Non-match, Mover	143,912.9	362,304.25 (289)	620,138.16 (618)	0.58423

[17]The three letter cell names are included to identify the cells with the definitions in Beaghen & Sands (2002).

[18]The weights of these counts reflect the Revision II sampling and the Targeted Extended Search sampling, though not the non-interview adjustment.

[19]In parenthesis are the unweighted counts; each person with a non-zero weight counts as one.