



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

December 31, 2002

**MASTER FILE**

DSSD A.C.E. REVISION II MEMORANDUM SERIES # PP- 45

PRED CENSUS AND SURVEY MEASUREMENT STAFF MEMORANDUM SERIES: CSM-  
A.C.E. REVISION II -R5R

MEMORANDUM FOR: Donna Kostanich  
Chair, A.C.E. Revision II Planning and Management Group

From: Mary Mulry (*signed 12/31/02*) *MM*  
Chair, A.C.E. Revision II Quality Indicators Subgroup

Prepared by: Tamara S. Adams and Elizabeth A. Krejsa  
A.C.E. Revision II Measurement Subgroup

Subject: At-Risk Codes Evaluation

This evaluation reports the effect of using automated enumeration and residence status coding for the A.C.E. Revision II on the dual system estimates.

For any questions regarding this document, please contact Tamara Adams or Elizabeth Krejsa.

cc: DSSD A.C.E. Revision II Memorandum Series Distribution List  
Ruth Ann Killion (PRED)  
David Hubble (PRED)

# Evaluation of At-Risk Codes

Tamara S. Adams  
Elizabeth A. Krejsa

---

Demographic Statistical Methods Division  
Planning, Research, and Evaluation Division

U S C E N S U S B U R E A U

*Helping You Make Informed Decisions*

## EXECUTIVE SUMMARY

### **What is the primary goal of the At-Risk Codes evaluation?**

Since the evaluations of the Accuracy and Coverage Evaluation Survey (A.C.E.) found errors in the assignment of enumeration and residence status of both the E-sample and the P-sample, the revision of the A.C.E. included recoding a subsample of the A.C.E. sample and using the results in a double sampling ratio adjustment. The recoding operation for the A.C.E. Revision II assigned some of the E-sample enumeration status codes and some of the P-sample residence status codes by a computer algorithm and the rest by analysts at the National Processing Center (NPC). The primary goal of the At-Risk Codes evaluation is to estimate the potential error in the A.C.E. Revision II dual system estimates (DSEs) due to the automated assignment of enumeration and residence status for some of the cases. Since the only portions of the DSE that involve the revision coding are the double-sampling ratios, we will concentrate our analysis on those factors.

### **What did the At-Risk Codes evaluation find in the E-sample?**

Twelve of the 31 E-sample double-sampling ratios have significant differences between the A.C.E. Revision II and those with the at-risk adjustment. The differences in the double-sampling ratios range from  $-0.0025$  ( $se=0.0016$ ) to  $0.0009$  ( $se=0.00004$ ). The largest increase in a poststratum due to the at-risk adjustment is an additional 8,743 (0.035%) correct enumerations; the largest decrease in a poststratum due to the at-risk adjustment is a decrease of 10,798 (0.246%) correct enumerations.

### **What did the At-Risk Codes evaluation find in the P-sample?**

In the P-sample, none of the double-sampling ratios with the at-risk adjustments were significantly different from the A.C.E. Revision II double-sampling adjustments. The differences ranged from  $-0.0009$  ( $se=0.04$ ) to  $0.0006$  ( $se=0.03$ ). The largest increase in residents due to the at-risk adjustment is an increase of 7,988 (0.031%) additional residents in a poststratum; the largest decrease in residents due to the at-risk adjustment is a decrease of 8,342 (0.093%) residents in a poststratum.

### **What are the implications for the A.C.E. Revision II dual system estimates?**

Based on the data, we see very small differences in every poststratum in the double-sampling ratios. Since the double-sampling ratios are the only portions of the DSE that are affected by revision coding, we expect to see small differences in the A.C.E. Revision II dual system estimates.

## 1. BACKGROUND

Since the evaluations of the Accuracy and Coverage Evaluation Survey (A.C.E.) found errors in the assignment of enumeration and residence status of both the E-sample and the P-sample, the revision of the A.C.E. included recoding a subsample of the A.C.E. sample and using the results in a double sampling ratio adjustment. The recoding operation for the A.C.E. Revision II assigned some of the E-sample enumeration status codes and some of the P-sample residence status codes by a computer algorithm and the rest by analysts at the National Processing Center (NPC). The primary goal of the At-Risk Codes evaluation is to estimate the potential error in the A.C.E. Revision II dual system estimates (DSEs) due to the automated assignment of enumeration and residence status for some of the cases.

### *1.1 A.C.E. Revision II Background*

In the fall of 2001, the results of coding the data collected in the Evaluation Followup (EFU) showed a significant increase in erroneous enumerations in the E-sample and nonresidents in the P-sample that were not detected in the coding of the A.C.E. The EFU was conducted for a subsample of the A.C.E. sample and asked more probing questions about Census Day residence than had been asked during the CAPI interview and Production Followup (PFU) for the A.C.E. An additional review (called the PFU/EFU Review) of 17,522 PFU and EFU E-sample cases was conducted by the analysts, the most skilled coders, at the NPC. This review confirmed that the A.C.E. had underestimated the number of erroneous enumerations (Adams and Krejsa, 2001).

These errors in the A.C.E. needed to be corrected for in the A.C.E. Thus, the A.C.E. Revision II required more coding by the analysts at the NPC who had coded cases during the PFU/EFU Review. Since A.C.E. Revision II had to provide accurate subpopulation estimates, the subsample had to be larger than the one used for the PFU/EFU Review. Recoding the entire A.C.E. sample was not possible because the EFU collected data in only 2,259 out of the 11,303 A.C.E. sample clusters. Even clerically recoding the approximately 70,000 E-sample cases and 52,000 P-sample cases in the EFU sample was not feasible because of time constraints.

### *1.2 Using the Keyed Data in A.C.E. Revision II*

Fortunately, both the PFU and EFU questionnaires had been keyed and were available in electronic form for the A.C.E. Revision II process. A new strategy evolved to combine automated coding and clerical coding to provide high quality data in the time allotted. The plan restricted the clerical review to the more difficult cases and automated the assignment of codes to the more straightforward cases.

Initially an automated algorithm assigned an enumeration status code (or residence status code) and a why code which described the reason for the code assigned. The detailed codes can be summarized by the following broad groupings:

- No followup
- Noninterview
- Geocoding issues

- Mover issues
- Other residence issues
- Group quarter issues
- Died before census day or born after census day
- Lived there, no unusual living situations noted

A three-step process was followed to assign final codes to each case:

- Validation – Determine for each why code category if the automated enumeration status coding is of high quality by assessing the level of agreement between the automated codes and the PFU/EFU Review codes, for cases that were coded by both procedures.
- Targeting – Target only those why code categories that have automated enumeration status codes with low levels of agreement with the PFU/EFU Review data.
- Clerical Coding – Clerically recode only cases in the targeted why code categories. The clerical recoding took advantage of handwritten interviewer comments (Adams and Krejsa, 2002).

This strategy reduced the clerical workload to 23,988 people, a workload that could be completed in the allotted time. Most cases that received codes during PFU/EFU Review retained these codes and were not sent for a second clerical coding. (Adams and Krejsa 2002) Table 1 shows the number of cases that received automated codes and clerical codes in the E-sample and the P-sample.

**Table 1. Final Coding of Cases in A.C.E. Revision II**

	<b>E-sample</b>	<b>P-sample</b>
Cases not sent to Clerical*	39,509	31,528
Cases sent to Clerical		
PFU/EFU Review	15,678	7,035
A.C.E. Revision II Clerical	14,131	14,108
Cases without Forms to Review		
In A.C.E. Revision II Sample (duplicates, insufficient information for matching and followup, cases without EFU, others)	7,323	8,654
Not in A.C.E. Revision II Sample	90,477	106,422

\*=At-risk cases

Note that matches are included in both the E- and P-sample counts

Cases that received automated coding are called the “at-risk cases.” The automated codes are believed to have a higher risk of error than the clerically assigned codes. The At-Risk Codes evaluation attempts to estimate the potential error in the “at-risk cases” by examining the error in the automated codes for cases in the PFU/EFU Review sample.

## 2. METHODS

### 2.1 Error Factors

To assess the potential error in the DSE due to the at-risk cases, we used the error rates observed in the PFU/EFU Review to derive estimated error factors for the at-risk cases. The underlying assumption for this approach is that the at-risk cases have the same error factor as the cases in their keyed code category<sup>1</sup> that were in the PFU/EFU Review. We used the following approach to calculate the potential error in the DSE:

- Create Donor Cells – These are cases in a given combined keyed code category<sup>2</sup> in the PFU/EFU Review.
- Calculate Error Factors – We calculated the error factor for each combined keyed code category in the PFU/EFU Review. The error factor is how much error we could incur by accepting that category without further review.
- Calculate Average Unresolved Probabilities – To calculate error factors for unresolved cases, we used the same correct enumeration probability<sup>3</sup> for all cases with that why code category.

### *2.1.1. Creating Donor Cells*

For each at-risk case, the computer code was categorized as to why that case received the code it did (called a why code). For example, an E-sample person was coded a correct enumeration (CE) and given a why code of ‘Lived Here’ because he/she lived at the followup address and had no other residence and was not in a group quarters on Census Day. Each person in the PFU/EFU Review was also coded using the computer algorithm in the same way but benefited from a second clerical review. Such a review can use notes on the form that cannot be coded by a computer and is not subject to keying error like the computer data.

Using these why codes, donor cells for error factors were formed by combining the PFU computer why code and the EFU computer why code. Some combined key code categories were collapsed to form larger donor cells. The combined keyed codes were formed to calculate the error factor that accounts for the coding of both forms and the final result. Each combination resulted in a best enumeration status for E-sample cases or best residence status for P-sample cases. See Appendix A for details on the formation of donor cells.

### *2.1.2 Calculating Error Factors*

Error factors were calculated by determining how often the keyed enumeration status was different from the final status in the PFU/EFU Review. We consider the “error” to be the deviation of that case from its A.C.E. Revision II correct enumeration probability.

---

<sup>1</sup> A “keyed code category” consists of all cases within a given why code and match code grouping (see Appendix A for details). For instance, all E-sample persons who are erroneous enumerations because they lived in a dorm are within one “keyed code category”.

<sup>2</sup> A “combined keyed code category” consists of all cases that have a keyed code category for PFU and a keyed code category for EFU. For instance, all E-sample persons who, in EFU, are erroneous enumerations because they lived in a dorm, and who, in PFU, are correct enumerations because they had no other residences, moving, or group quarters are once combined keyed code category.

<sup>3</sup> Note that throughout this report we use the E-sample for discussion’s sake. The P-sample is analogous, where enumeration status is equivalent to residence status.

We calculated an error factor for each of the combined keyed code categories as follows (A.C.E. Revision II is abbreviated as RevII):

$$error\ factor = \frac{\sum_{PFU / EFU\ Rev, combined\ category} rewt * ceprobi_{Final\ RevII}}{\sum_{PFU / EFU\ Rev, combined\ category} rewt * ceprobi_{Bestkey}} \quad (1)$$

where:

$ceprobi_{FinalRevII}$ <sup>4</sup>=the probability of correct enumeration for that case’s A.C.E. Revision II final code. The probability is usually 1 or 0, but for unresolved cases it can be within that range.

$ceprobi_{Bestkey}$ =the probability of correct enumeration for that case’s best key code. This is always the same for every case within the summation. See section 2.1.3 for issues with unresolved cases. For erroneous enumeration cases which have a  $ceprobi$  of zero, we take this value to be 1, since we cannot divide by 0.

Table 2 shows these error factors for the E-sample, Table 3 for the P-sample.

**Table 2. E-sample Error Factors**

Why Code Combination Group Description	Keyed Enumeration Status	Error factor
A -Group Quarters	Erroneous	.01779
B - No Knowledgeable Respondent	Unresolved	1.00606
C - Lived Here on both forms	Correct	.99685
D - Lived Here on EFU, no PFU form	Correct	.99921
E - Other Residence/Lived Here, on both forms	Correct	.99659
F - Other Residence Lived Here, on one form	Correct	.99993
G - Partial Information on both forms	Unresolved	.99435
H - PFU Lived Here, EFU Not Enough Information	Correct	.99692
I - No PFU form, EFU Not Enough Information, A.C.E. override	Correct	.99651
J - Other Residence unknown, both forms	Unresolved	1.00056

For at-risk cases with a keyed erroneous enumeration status (group A), the A.C.E. Revision II  $ceprob$  is zero because they have zero probability of being a correct enumeration. The error factor above of .01779 means that the keyed coding underestimated the correct enumeration probability by that amount. This error factor was determined when cases in the PFU/EFU Review that were also assigned an erroneous enumeration status according to the keyed data were clerically reviewed. 1.779 percent of those cases were determined to be correct enumerations.

For correct enumeration cases (groups C, D, E, F, H, and I), the A.C.E. Revision II correct enumeration probability is 1. The error factor for these cases ranges from .99659 to .99993.

<sup>4</sup> Note: we use the  $ceprobi$ , as opposed to  $ceprob$ .  $ceprob$  captures the adjustment for duplicates in the non-E-sample which we do not want to capture.  $ceprobi$  does not capture it.

This means that the keyed coding overestimated the correct enumeration probability by 1 minus these factors. The error factors were determined when some cases in the PFU/EFU Review which were coded a correct enumeration using the keyed data, were actually erroneous enumerations as determined by a clerical review.

The unresolved cases (groups B and J) are discussed in section 2.1.3.

**Table 3. P-sample Error Factors**

Why Code Combination Group	Keyed Residence Status	Error factor
M – Nonresident	Nonresident	.50321
N – Not Enough Information	Unresolved	.99415
O – Lived Here/Other Residence both forms	Resident	.99684
P – Lived here, one form blank	Resident	.99929
Q – No PFU and EFU other residence lived here	Resident	.99993
R – PFU resident, EFU not enough information	Resident	.9967
S – PFU not enough information, EFU lived here	Resident	.98324
T – No PFU, EFU unresolved address information	Unresolved	1.14457

For at-risk cases with a keyed nonresident residence status (group M), the A.C.E. Revision II residence probability is zero because they have zero probability of being a resident. The error factor above of .50321 means that the keyed coding underestimated the residence probability by that amount. The error factor was determined when cases in the PFU/EFU Review that were also assigned a nonresident residence status according to the keyed data were clerically reviewed. Over 50 percent of those cases were determined to be residents. This high error factor is likely due to some cases being coded unresolved and imputed at a high rate of residence and others being coded residents due to a misclassification of assisted living facilities as nursing homes.

For residents (groups O, P, Q, R, and S), the A.C.E. Revision II residence probability is 1. The error factor for these cases ranges from .98324 to .99993. This means that the keyed coding overestimated the residence probability by 1 minus these factors. The error factors were determined when some cases in the PFU/EFU Review which were coded a resident using the keyed data, were actually nonresidents as determined by a clerical review.

The unresolved cases (groups N and T) are discussed in section 2.1.3.

### *2.1.3 Cases with Unresolved Codes from Keying*

Unresolved cases were not assigned correct enumeration probabilities during the computer coding. At-risk cases were assigned correct enumeration probabilities in the final missing data process (see Beaghen 2002 for details). PFU/EFU Review cases were assigned correct enumeration probabilities based on the result of the clerical review, not the computer coding. Ideally, to calculate an error factor, a correct enumeration probability would have been calculated based on the keyed data alone and the deviation from that probability would be determined. Instead, if a PFU/EFU Review case remained unresolved the correct enumeration probability used is the one calculated by the missing data process. For cases that were coded unresolved using keyed data but were determined to be correct or erroneous by the clerical



review, we used an average probability of correct enumeration as calculated from the PFU/EFU review sample:

$$ceprob_{ue} = \frac{\sum rewt * ceprobi_{Final RevII}}{\sum rewt} \quad (2)$$

PFU / EFUreview,withUEcodeinwhycodecat

$ceprob_{ue}$  is the average A.C.E. Revision II probability for the **final** why code group because the final why code was used to calculate  $ceprobi_{FinalRevII}$ . Using this  $ceprob_{ue}$  we calculated the error factor as:

$$error\ factor_{ue} = \frac{\sum rewt * ceprobi_{Final RevII}}{\sum rewt * ceprobi_{Final RevII} + \sum rewt * ceprob_{UE}} \quad (3)$$

PFU / EFU Rev.combined category

PFU / EFU Rev,combined category, where final code=UE

PFU / EFU Rev,combined category, where final code=CE or EE

The average correct enumeration probabilities for the E-sample unresolved groups were calculated from cases in the PFU/EFU Review with a final A.C.E. Revision II code of unresolved and a why code combination based on the final why codes equivalent to donor groups B, G, and J. Table 4 shows the average correct enumeration probabilities for these groups.

**Table 4. E-sample Average Correct Enumeration Probabilities**

Final Why Code Combination Group	Cell Size	$ceprob_{ue}$
B - No Knowledgeable Respondent	212	.8818
G - Partial Information on both forms	259	.8276
J - Other Residence unknown, both forms	767	.9816

Using the above average correct enumeration probabilities, the error factors were calculated according to formula 3. The error factors for these cases ranged from .99435 to 1.00606 (as shown in Table 2). An error factor greater than 1 indicates that the correct enumeration probability was underestimated. The correct enumeration probability for at-risk cases in group B ranged from .33943 to .956098, in group G from .27777 to .98865, and in group J from .88736 to .998659. Therefore, since most of the correct enumeration probabilities within a group are less than their respective error rates, we can say that the A.C.E. Revision II process underestimated the correct enumeration probabilities.

The average residence probabilities for P-sample unresolved groups were calculated from cases in the PFU/EFU Review with a final A.C.E. Revision II code of unresolved and a why code combination based on the final why codes equivalent to donor groups N and T. Table 5 shows the average residence probabilities (rprob) for these groups.

**Table 5. P-sample Average Residence Probabilities**

Final Why Code Combination Group	Cell Size	rprob <sub>ue</sub>
N - Not Enough Information	1171	.9614
T - No PFU, EFU unresolved address information	15	.8044

Using the above residence probabilities, the error factors were calculated according to formula 3. At-risk cases in group N have residence probabilities ranging from .01787 to .96609. The error factor for group N (as shown in Table 3) is .99415. At-risk cases in group T have residence probabilities ranging from 0 to .99281. The error factor for group T (as shown in Table 3) is 1.14457. An error factor greater than 1 means that the residence probability was underestimated.

#### 2.1.4 Applying the Error Factors

If we examine the A.C.E. Revision II DSE formula (see Appendix B), we see that the only terms affected by the A.C.E. Revision II coding are the double-sampling ratios,  $f_s$ , which are applied to those cases without links to beyond the search area. We recalculated the  $f$  terms as follows:

- We assigned recipient cells (see Appendix A) to the cases that were coded using the keyed data for A.C.E. Revision II. These recipient cells were assigned analogous to the donor cells.
- We applied the error factors. The error factor calculated from equation (1) is equivalent to the correct enumeration probability if we had not used the keyed data. So, we next calculated the  $f$  terms. For example, for the E-sample we calculated the  $f_i$  term as follows:

$$f_{1,i',ar} = \frac{CE_{i'}^{ND*}}{CE_{i'}^{ND}} = \frac{\sum_{\text{all cases in RevII}} cephrob * finalwgt}{CE_{i'}^{ND}}$$

where  $cephrob$  is the A.C.E. Revision II probability of correct enumeration for cases with a clerical review; for at-risk cases it is  $error\ factor * cephrob$  for correct and unresolved cases and  $error\ factor * (1 - cephrob)$  for erroneous cases. P-sample  $f$  terms are calculated similarly; see Appendix B for details.

#### 2.2 Sampling error assessment

We used a non-stratified, delete-a-group jackknife to calculate the standard errors on the  $f$  factors shown in the results section.

### 3. LIMITS

We made assumptions about the P-sample error rate based on the E-sample error rates. We used the E-sample PFU/EFU Review cases to calculate error factors for P-sample groups because the P-sample was not represented in the PFU/EFU Review sample.

To calculate error factors for unresolved cases, we needed to use the same correct enumeration probability for all cases with that why code category. Therefore, we assumed that the correct enumeration probability assigned during A.C.E. Revision II is not affected by the keyed data coding.

### 4. RESULTS

#### 4.1 *E-sample Double-Sampling Ratios*

Table 6 below shows the E-sample double-sampling ratio adjustment applied to the non-duplicate e-sample correct enumerations ( $f_1$ ) for each poststratum. We present the A.C.E. Revision II  $f_1$ , followed by the  $f_1$ , which includes the error factors for at-risk cases, and the difference between the two. Significant differences are shown with a \*. Standard errors for the estimates are available in Appendix C. The  $f_1$  term is multiplied by the weighted estimate of non-duplicate link correct enumerations ( $CE^{ND}$ ) in the A.C.E. Revision II DSE, cumulated to the revision poststratum level. Given the difference in the  $f_1$  term, the effect of the at-risk cases in the  $CE^{ND} * f_1$  term of the A.C.E. Revision II DSE is presented in the last column.

**Table 6. E-sample Double-Sampling Ratios by Poststratum Group**

#	Poststratum Group Description	A.C.E.			Lower Bound	Upper Bound	CE <sup>ND</sup>	Effect of At-Risk
		II f <sub>1</sub>	At-Risk f <sub>1</sub>	Difference				
1	All proxy cases	0.97670	0.97917	-0.00246	-0.00559	0.00066	4,382,552	-10,798
2	American Indians on Reservations (AIR) who are Non-proxy	0.97450	0.97360	0.00090*	0.00082	0.00097	363,750	326
<b>Non-Proxy/non-AIR</b>								
3	Nuclear member 0-9 M&F	0.99682	0.99655	0.00027*	0.00007	0.00046	27,593,801	7,323
4	Nuclear member 10-17 M&F	0.99294	0.99266	0.00029*	0.00011	0.00046	23,226,971	6,622
5	Nuclear member 18-29 M	0.99175	0.99188	-0.00013	-0.00095	0.00069	5,543,037	-745
6	Nuclear member 18-29 F	0.99331	0.99445	-0.00114	-0.00309	0.00080	7,883,434	-9,025
7	Nuclear member 30-49 M	0.99420	0.99384	0.00035*	0.00025	0.00045	24,727,916	8,743
8	Nuclear member 30-49 F	0.99657	0.99631	0.00026*	0.00011	0.00041	28,638,422	7,461
9	Nuclear member 50+ M	0.99189	0.99172	0.00017	-0.00007	0.00041	22,732,599	3,916
10	Nuclear member 50+ F	0.99592	0.99588	0.00004	-0.00028	0.00036	22,092,197	881
11	Other member 0-9 M&F	0.98178	0.98142	0.00035	-0.00017	0.00088	3,246,796	1,142
12	Other member 10-17 M&F	0.98826	0.98788	0.00037	-0.00017	0.00091	2,110,004	783
13	Other member 18-29 M	0.97940	0.97936	0.00004	-0.00038	0.00046	10,536,263	405
14	Other member 18-29 F	0.97507	0.97470	0.00037*	0.00014	0.00061	8,557,076	3,199
15	Other member 30-49 M	0.97777	0.97818	-0.00041	-0.00165	0.00083	7,378,013	-3,029
16	Other member 30-49 F	0.98937	0.98933	0.00004	-0.00064	0.00072	5,372,144	217
17	Other member 50+ M	0.98894	0.98856	0.00038*	0.00020	0.00057	3,543,650	1,351
18	Other member 50+ F	0.98688	0.98666	0.00022	-0.00003	0.00047	6,516,082	1,419
19	Nuclear member 0-17 M&F	0.99493	0.99465	0.00027*	0.00014	0.00041	5,120,570	1,408
20	Nuclear member 18-29 M&F	0.99267	0.99340	-0.00073	-0.00208	0.00061	750,233	-550
21	Nuclear member 30-49 M&F	0.99549	0.99518	0.00030*	0.00020	0.00041	4,780,216	1,448
22	Nuclear member 50+ M&F	0.99390	0.99379	0.00011	-0.00014	0.00035	3,159,331	336
23	Other member 0-17 M&F	0.98451	0.98415	0.00036	-0.00002	0.00074	422,593	152
24	Other member 18-29 M&F	0.97743	0.97724	0.00019	-0.00006	0.00045	1,751,582	335
25	Other member 30-49 M&F	0.98282	0.98304	-0.00021	-0.00098	0.00055	1,129,658	-242
26	Other member 50+ M&F	0.98754	0.98727	0.00027*	0.00007	0.00047	909,399	246
27	Nuclear member 0-29 M&F	0.99451	0.99442	0.00009	-0.00022	0.00040	880,251	78
28	Nuclear member 30+ M&F	0.99477	0.99456	0.00021*	0.00008	0.00035	1,080,427	232
29	Other member 0-29 M&F	0.97895	0.97872	0.00023	-0.00001	0.00046	952,728	217
30	Other member 30+ M&F	0.98540	0.98535	0.00005	-0.00032	0.00042	10,622,018	535
31	Nuclear member 0+ M&F	0.99467	0.99450	0.00016*	0.00001	0.00032	n/a	n/a
32	Other member 0+ M&F	0.98269	0.98256	0.00012	-0.00012	0.00037	691,140	86

Although twelve of the poststratum differences are significant the potential effect in these poststrata is very small.

#### 4.2 P-sample Double-Sampling Ratios

Table 7 below shows the P-sample double-sampling ratio adjustment applied to the in-movers ( $f_5$ ) for each poststratum. We present the A.C.E. Revision II  $f_5$ , followed by the  $f_5$ , which includes the error factors for at-risk cases, and the difference between the two. The  $f_5$  term is multiplied by the weighted estimate of in-movers ( $P_{IM}$ ) in the A.C.E. Revision II DSE, cumulated to the revision poststratum level. Given the difference in the  $f_5$  term, the effect of the at-risk cases in the  $P_{IM} * f_5$  term of the A.C.E. Revision II DSE is presented in the last column.

**Table 7. P-sample Inmover Double-Sampling Ratios by Poststratum Group**

#	Poststratum Group Description	A.C.E.			Lower Bound	Upper Bound	$P_{IM}$	Effect of At-Risk
		Revision II $f_5$	At-Risk $f_5$	Difference				
1	Owner	0.88696	0.88687	0.00009	-0.00784	0.00802	5,431,577	491
2	Non-Owner	1.00994	1.00993	0.00001	-0.00103	0.00105	8,138,569	95

Table 8 below shows the P-sample double-sampling ratio adjustment applied to the outmover matches ( $f_3$ ) for each poststratum. We present the A.C.E. Revision II  $f_3$ , followed by the  $f_3$ , which includes the error factors for at-risk cases, and the difference between the two. The  $f_3$  term is multiplied by the weighted estimate of outmover matches ( $M_{OM}$ ) in the A.C.E. Revision II DSE, cumulated to the revision poststratum level. Given the difference in the  $f_3$  term, the effect of the at-risk cases in the  $M_{OM} * f_3$  term of the A.C.E. Revision II DSE is presented in the last column.

**Table 8. P-sample Outmover Matches Double-Sampling Ratios by Poststratum Group**

#	Poststratum Group Description	A.C.E.			Lower Bound	Upper Bound	$M_{OM}$	Effect of At-Risk
		Revision II $f_3$	At-Risk $f_3$	Difference				
3	Owner	0.78901	0.78875	0.00025	-0.02192	0.02243	2,562,607	648
4	Non-Owner	0.81818	0.81788	0.00030	-0.02569	0.02629	4,285,674	1,270

Table 9 below shows the P-sample double-sampling ratio adjustment applied to the outmover residents ( $f_4$ ) for each poststratum. We present the A.C.E. Revision II  $f_4$ , followed by the  $f_4$ , which includes the error factors for at-risk cases, and the difference between the two. The  $f_4$  term is multiplied by the weighted estimate of outmover residents ( $P_{OM}$ ) in the A.C.E. Revision II DSE, cumulated to the revision poststratum level. Given the difference in the  $f_4$  term, the effect of the at-risk cases in the  $P_{OM} * f_4$  term of the A.C.E. Revision II DSE is presented in the last column.

**Table 9. P-sample Outmover Residents Double-Sampling Ratios by Poststratum Group**

#	Poststratum Group Description	A.C.E.			Lower Bound	Upper Bound	$P_{OM}$	Effect of At-Risk
		Revision II $f_4$	At-Risk $f_4$	Difference				
3	Owner	0.76287	0.76249	0.00038	-0.03283	0.03359	3,176,542	1,203
4	Non-Owner	0.83096	0.83050	0.00046	-0.04024	0.04117	5,665,657	2,630

Table 10 below shows the P-sample double-sampling ratio adjustment applied to the nonmover matches ( $f_2$ ) for each poststratum. We present the A.C.E. Revision II  $f_2$ , followed by the  $f_2$ , which includes the error factors for at-risk cases, and the difference between the two. The  $f_2$  term is multiplied by the weighted estimate of nonmover matches who are not duplicate links ( $M_{nm}^{ND}$ ) in the A.C.E. Revision II DSE, cumulated to the revision poststratum level. Given the difference in the  $f_2$  term, the effect of the at-risk cases in the  $M_{nm}^{ND} * f_2$  term of the A.C.E. Revision II DSE is presented in the last column.

**Table 10. P-sample Nonmover Matches Double-Sampling Ratios by Poststratum Group**

#	Poststratum Group Description	A.C.E. Revision II $f_2$	At-Risk $f_2$	Difference	Lower Bound	Upper Bound	$M_{nm}^{ND}$	Effect of At-Risk
5	American Indians on Reservations (AIR)	0.99567	0.99504	0.00063	-0.05502	0.05629	326,562	207
6	Non-AIR/Owner 0-9 M&F	1.00141	1.00116	0.00025	-0.02124	0.02173	21,210,912	5,198
7	Non-AIR/Owner 10-17 M&F	1.00114	1.00090	0.00023	-0.02036	0.02083	20,086,401	4,716
8	Non-AIR/Owner 18-29 M	1.00344	1.00319	0.00025	-0.02178	0.02228	8,296,407	2,084
9	Non-AIR/Owner 18-29 F	1.00064	1.00037	0.00027	-0.02371	0.02425	8,284,367	2,265
10	Non-AIR/Owner 30-49 M	1.00424	1.00397	0.00027	-0.02324	0.02378	24,719,866	6,627
11	Non-AIR/Owner 30-49 F	1.00325	1.00305	0.00020	-0.01754	0.01795	27,372,047	5,538
12	Non-AIR/Owner 50+ M	1.00446	1.00424	0.00022	-0.01941	0.01986	24,012,868	5,375
13	Non-AIR/Owner 50+ F	1.00246	1.00224	0.00022	-0.01889	0.01932	28,469,122	6,202
14	Non-AIR/Non-Owner 0-9 M&F	1.00600	1.00566	0.00034	-0.02912	0.02979	10,689,763	3,590
15	Non-AIR/Non-Owner 10-17 M&F	0.99989	0.99953	0.00036	-0.03103	0.03174	6,865,092	2,457
16	Non-AIR/Non-Owner 18-29 M	1.01411	1.01361	0.00050	-0.04327	0.04427	6,702,914	3,346
17	Non-AIR/Non-Owner 18-29 F	1.00866	1.00814	0.00051	-0.04456	0.04559	7,479,345	3,844
18	Non-AIR/Non-Owner 30-49 M	1.00313	1.00268	0.00044	-0.03833	0.03921	9,121,797	4,033
19	Non-AIR/Non-Owner 30-49 F	0.99992	0.99950	0.00042	-0.03635	0.03719	9,740,097	4,084
20	Non-AIR/Non-Owner 50+ M	0.99607	0.99568	0.00040	-0.03445	0.03525	4,320,957	1,717
21	Non-AIR/Non-Owner 50+ F	1.00025	0.99992	0.00033	-0.02826	0.02891	6,325,607	2,062
22	Non-AIR/Owner 18+ M	1.00422	1.00397	0.00025	-0.02140	0.02189	930,272	230
23	Non-AIR/Owner 18+ F	1.00258	1.00236	0.00022	-0.01889	0.01933	1,073,217	234
24	Non-AIR/Non-Owner 18+ M	1.00493	1.00448	0.00045	-0.03897	0.03987	313,817	141
25	Non-AIR/Non-Owner 18+ F	1.00267	1.00224	0.00042	-0.03662	0.03747	357,207	151

Table 11 below shows the P-sample double-sampling ratio adjustment applied to the nonmover residents ( $f_6$ ) for each poststratum. We present the A.C.E. Revision II  $f_6$ , followed by the  $f_6$ , which includes the error factors for at-risk cases, and the difference between the two. The  $f_6$  term is multiplied by the weighted estimate of nonmover residents who are not duplicate links ( $P_{nm}^{ND}$ ) in the A.C.E. Revision II DSE, cumulated to the revision poststratum level. Given the

difference in the  $f_6$  term, the effect of the at-risk cases in the  $P_{nm}^{ND} * f_6$  term of the A.C.E. Revision II DSE is presented in the last column.

**Table 11. P-sample Nonmover Residents Double-Sampling Ratios by Poststratum Group**

#	Poststratum Group Description	A.C.E. Revision II $f_6$	At-Risk $f_6$	Difference	Lower Bound	Upper Bound	$P_{nm}^{ND}$	Effect of At-Risk
5	American Indians on Reservations (AIR)	1.00058	1.00057	0.00001	-0.00143	0.00146	364,634	5
6	Non-AIR/Owner 0-9 M&F	1.00125	1.00095	0.00030	-0.02562	0.02621	22,618,695	6,683
7	Non-AIR/Owner 10-17 M&F	1.00073	1.00048	0.00025	-0.02144	0.02193	21,040,524	5,202
8	Non-AIR/Owner 18-29 M	0.99695	0.99788	-0.00093	-0.08224	0.08039	8,997,016	-8,342
9	Non-AIR/Owner 18-29 F	1.00487	1.00489	-0.00002	-0.00221	0.00216	8,903,635	-216
10	Non-AIR/Owner 30-49 M	1.00258	1.00227	0.00031	-0.02661	0.02723	26,022,990	7,988
11	Non-AIR/Owner 30-49 F	1.00302	1.00278	0.00024	-0.02060	0.02107	28,558,930	6,785
12	Non-AIR/Owner 50+ M	1.00446	1.00419	0.00027	-0.02322	0.02376	25,015,085	6,701
13	Non-AIR/Owner 50+ F	1.00330	1.00305	0.00026	-0.02216	0.02268	29,598,783	7,567
14	Non-AIR/Non-Owner 0-9 M&F	1.00464	1.00419	0.00046	-0.03956	0.04047	12,027,720	5,488
15	Non-AIR/Non-Owner 10-17 M&F	1.00764	1.00715	0.00049	-0.04224	0.04321	7,643,656	3,724
16	Non-AIR/Non-Owner 18-29 M	1.02601	1.02636	-0.00034	-0.03051	0.02982	7,852,335	-2,696
17	Non-AIR/Non-Owner 18-29 F	1.01453	1.01397	0.00056	-0.04868	0.04980	8,390,818	4,711
18	Non-AIR/Non-Owner 30-49 M	0.99797	0.99744	0.00053	-0.04618	0.04724	10,366,998	5,522
19	Non-AIR/Non-Owner 30-49 F	0.99810	0.99756	0.00053	-0.04629	0.04736	10,701,972	5,714
20	Non-AIR/Non-Owner 50+ M	0.99861	0.99823	0.00038	-0.03259	0.03335	4,796,173	1,803
21	Non-AIR/Non-Owner 50+ F	1.00498	1.00463	0.00036	-0.03095	0.03166	6,833,517	2,439
22	Non-AIR/Owner 18+ M	1.00253	1.00243	0.00011	-0.00920	0.00941	997,915	106
23	Non-AIR/Owner 18+ F	1.00338	1.00317	0.00021	-0.01839	0.01881	1,130,400	240
24	Non-AIR/Non-Owner 18+ M	1.00736	1.00715	0.00021	-0.01809	0.01851	392,801	82
25	Non-AIR/Non-Owner 18+ F	1.00507	1.00457	0.00050	-0.04292	0.04391	417,234	207

There are no significant differences in any of the P-sample poststratum differences<sup>1</sup>

## 5. CONCLUSIONS

As we can see from the above tables, 12 of the 31 of the E-sample double-sampling ratios differ significantly between A.C.E. Revision II and the at-risk adjustment. Those with significant differences, although statistically significant, are quite small. In the P-sample, none of the double-sampling ratios with the at-risk adjustments were significantly different from the A.C.E. Revision II double-sampling adjustments.

The largest increase in a poststratum due to the at-risk adjustment is an additional 8,743 (0.035%) correct enumerations; the largest decrease in a poststratum due to the at-risk adjustment is a decrease of 10,798 (0.246%) correct enumerations. There are some poststrata that have a slightly larger percent increase in correct enumerations. The largest increase in residents due to the at-risk adjustment is an increase of 7,988 (0.031%) additional residents in a poststratum; the largest decrease in residents due to the at-risk adjustment is a decrease of 8,342 (0.093%) residents in a poststratum. There are some poststrata that have a slightly larger percent increase in residents.

We can therefore conclude that augmenting the clerical coding procedures with automated coding introduced very small errors into the A.C.E. Revision II DSEs and saved significant time and resources in the A.C.E. Revision II process.

## **6. REFERENCES**

Adams, Tamara and Krejsa, Elizabeth (2001). "ESCAP II: Results of the Person Followup and Evaluation Followup Forms Review." ESCAP Report No. 24. U.S. Census Bureau.

Adams, Tamara and Krejsa, Elizabeth (2002). "A.C.E. Revision II Measurement Subgroup Documentation." DSSD A.C.E. Revision II Memorandum Series #PP- 6.

Beaghen, Michael (2002). "A.C.E. Revision II: Specifications for the Assignment of Probability of Enumeration Status, Census Day Residency and Match Status." DSSD A.C.E. Revision II Memorandum Series #PP-23.

Mulry, Mary (2002). "A.C.E. Revision II: Design and Methodology, Chapter 7: Assessing the Estimates" DSSD A.C.E. Revision II Memorandum Series #PP-30.



## Appendix A - Why Code Categories and Combined Groups

### E-sample

The recipient cells contain cases where the final measurement code was determined by the computer using keyed data (i.e. the at-risk cases). To create recipient cells, we determined all the combinations of PFU why codes and EFU why codes used in coding. Table A1 below shows those combinations and the number of cases per cell.

**Table A1. E-sample Recipient Cells**

<b>Keyed Why Code Combination (PFUKY EFUKY)</b>	<b>Description of Why Codes from each form</b>	<b>Keyed Enumeration Status</b>	<b>Cell Size</b>
GQ DO	PFU = followup person (FUP) was in a group quarters, EFU= FUP was in a dorm	EE	28
GQ GQ	PFU and EFU = FUP was in a group quarters	EE	3
GQ ORDK	PFU = FUP was in a group quarters, EFU = respondent didn't know if FUP had another residence	EE	1
KR KR	PFU and EFU = no knowledgeable respondent could be located	UE	10
NI KR	PFU = noninterview, EFU = no knowledgeable respondent could be located	UE	16
ORDK KR	PFU = respondent didn't know if FUP had another residence, EFU = no knowledgeable respondent could be located	UE	43
LH LH	PFU and EFU = FUP lived at the followup address	CE	18,957
NF LH	PFU = no followup, EFU = FUP lived at the followup address	CE	17,294
OP LH	PFU = FUP had another residence but stayed here most of the time, EFU= lived here	CE	217
LH OR1	PFU = lived here, EFU = FUP had another residence but stayed here most of the time	CE	441
OP OR1	PFU and EFU = FUP had another residence but stayed here most of the time	CE	85
NF OR1	PFU = no followup, EFU = FUP had another residence but stayed here most of the time	CE	342
NI ORDK	PFU = noninterview, EFU= respondent didn't know if FUP had another residence	UE	22
ORDK MICD	PFU = respondent didn't know if FUP had another residence, EFU = moved in but don't know when	UE	7
NI MICD	PFU = noninterview, EFU = moved in but don't know when	UE	1
KR ORDK	PFU = no knowledgeable respondent could be located, EFU = respondent didn't know if FUP had another residence	UE	14
LH ORDK	PFU = lived here, EFU = respondent didn't know if FUP had another residence	CE	843
OP ORDK	PFU = FUP had another residence but stayed here most of the time, EFU = respondent didn't know if FUP had another residence	CE	15
NF KR	PFU = no followup, EFU = no knowledgeable respondent could be located	CE	175

**Table A1. E-sample Recipient Cells Con't**

Keyed Why Code Combination (PFUKY EFUKY)	Description of Why Codes from each form	Keyed Enumeration Status	Cell Size
NF ORDK	PFU = no followup, EFU= respondent didn't know if FUP had another residence	CE	623
ORDK ORDK	PFU and EFU = respondent didn't know if FUP had another residence	UE	342

A few combinations were excluded from the evaluation. They include any case with a final code of KE and any usual home elsewhere (UHE) case. These types of cases were excluded because they were determined to be correct and not contain error. This accounts for 22 unweighted cases. Two additional cases were excluded because the final code was not as expected.

The donor groups contain cases where the final measurement code was determined by the PFU/EFU Review<sup>5</sup>. These cases were also computer coded using keyed data. To create donor groups we,

1. used the keyed why code combinations from the recipient cells,
2. determined the number of PFU/EFU Review cases with the same keyed why code combination, and
3. collapsed combinations together to form groups of at least 30 donor cases; the combination must have the same enumeration status to be collapsed.

**Table A2. E-sample Donor Groups – PFU/EFU Review Cases**

Keyed Why Code Combination (PFUKY EFUKY)	Keyed Enumeration Status	Donor Cell Size	Collapsed Donor Group	Collapsed Donor Group Description
GQ DO	EE	66	A	Group Quarters, Erroneous
GQ GQ	EE	7	A	Group Quarters, Erroneous
GQ ORDK	EE	8	A	Group Quarters, Erroneous
KR KR	UE	6	B	Not Enough Information, Unresolved
NI KR	UE	8	B	Not Enough Information, Unresolved
ORDK KR	UE	34	B	Not Enough Information, Unresolved
LH LH	CE	1,044	C	Lived Here – 2 forms, Correct
NF LH	CE	6,071	D	Lived Here – 1 form, Correct
OP LH	CE	19	E	Other Residence, Lived Here - Correct
LH OR1	CE	38	E	Other Residence, Lived Here – Correct
OP OR1	CE	9	E	Other Residence, Lived Here – Correct
NF OR1	CE	164	F	Other Residence – 1 form, Correct
NI ORDK	UE	30	G	Partial Information, Unresolved
ORDK MICD	UE	24	G	Partial Information, Unresolved
NI MICD	UE	2	G	Partial Information, Unresolved
KR ORDK	UE	10	G	Partial Information, Unresolved
LH ORDK	CE	116	H	Lived Here, Not Enough Information – Correct
OP ORDK	CE	2	H	Lived Here, Not Enough Information – Correct

<sup>5</sup> Some PFU/EFU Review cases were sent back for additional clerical review in the A.C.E. Revision II Clerical Coding operation.

<b>Keyed Why Code Combination (PFUKY EFUKY)</b>	<b>Keyed Enumeration Status</b>	<b>Donor Cell Size</b>	<b>Collapsed Donor Group</b>	<b>Collapsed Donor Group Description</b>
NF KR	CE	24	I	Not Enough Information, Production override - Correct
NF ORDK	CE	247	I	Not Enough Information, Production override - Correct
ORDK ORDK	UE	303	J	Other Residence Unknown both forms, Unresolved

We then collapsed the recipient cells into these groups. This process yields the following combined groups:

<b>Combined Group</b>	<b>Donors (PFU/EFU Review cases)</b>	<b>Recipients (At-Risk cases)</b>
A – Group Quarters	81	32
B – Not Enough Information on both forms	48	69
C – Lived Here on both forms	1,044	18,957
D – Lived Here on EFU, no PFU form	6,071	17,294
E – Other Residence Lived Here, on both forms	66	743
F – Other Residence Lived Here, on one form	164	342
G – Partial Information	66	44
H– PFU Lived Here, EFU Not Enough Information	118	858
I – Not Enough Information, A.C.E. override, no PFU form	271	800
J - Other Residence unknown, both forms	303	342
Total	8,232	39,481

## **P-sample**

The P-sample recipient cells contain cases where the final measurement code was determined by the computer using keyed data (i.e. the at-risk cases). To create recipient cells, we determined all the combinations of PFU why codes and EFU why codes used in coding, as was done for the E-sample. Table A4 below shows those combinations and the number of cases per cell.

<b>Keyed Why Code Combination (PFUKY EFUKY)</b>	<b>Description of Why Codes from each form</b>	<b>Keyed Residence Status</b>	<b>Cell Size</b>
Blank DO	PFU = no followup, EFU= FUP in a dorm	NN	7
Blank GQ	PFU = no followup, EFU=FUP in a group quarters	NN	5
Blank MS	PFU = no followup, EFU=FUP in a military group quarters	NN	3
Blank NH	PFU = no followup, EFU=FUP in a nursing home	NN	1
Blank OR2	PFU = no followup, EFU = FUP has another residence and stays there most of the time	NN	7

**Table A4. P-sample Recipient Cells Con't**

<b>Keyed Why Code Combination (PFUKY EFUKY)</b>	<b>Description of Why Codes from each form</b>	<b>Keyed Residence Status</b>	<b>Cell Size</b>
GQ Blank	PFU = FUP in a group quarters, EFU = no followup	NN	6
GQ DO	PFU = FUP in a group quarters, EFU = FUP in a dorm	NN	29
GQ GQ	PFU = FUP in a group quarters, EFU = FUP in a dorm	NN	4
GQ ORDK	PFU = FUP in a group quarters, EFU = respondent didn't know if FUP had another residence	NN	1
GQ OR2	PFU = FUP in a group quarters, EFU = FUP has another residence and stays there most of the time	NN	1
NI OR2	PFU = noninterview, EFU = FUP has another residence and stays there most of the time	NN	2
ORDK GQ	PFU = respondent didn't know if FUP had another residence, EFU = FUP in a group quarters	NN	1
ORDK OR2	PFU = respondent didn't know if FUP had another residence, EFU = FUP has another residence and stays there most of the time	NN	2
ORDK ORDK	PFU = respondent didn't know if FUP had another residence, EFU = respondent didn't know if FUP had another residence	NU	139
ORDK blank	PFU = respondent didn't know if FUP had another residence, EFU = no followup	NU	55
KR Blank	PFU = no knowledgeable respondent, EFU = no followup	NU	1
KR ORDK	PFU = no knowledgeable respondent, EFU = respondent didn't know if FUP had another residence	NU	6
KR KR	PFU = no knowledgeable respondent, EFU = no knowledgeable respondent	NU	2
NI Blank	PFU = noninterview, EFU = no followup	NU	7
NI ORDK	PFU = noninterview, EFU = respondent didn't know if FUP had another residence	NU	11
NI KR	PFU = noninterview, EFU = no knowledgeable respondent	NU	18
ORDK KR	PFU = respondent didn't know if FUP had another residence, EFU = no knowledgeable respondent	NU	25
LH LH	PFU = FUP lived here, EFU = FUP lived here	NR	5,091
LH OR1	PFU = FUP lived here, EFU = FUP had another residence but stayed here most of the time	NR	142
OP LH	PFU = FUP had another residence but stayed here most of the time, EFU = FUP lived here	NR	94
OP OR1	PFU = FUP had another residence but stayed here most of the time, EFU = FUP had another residence but stayed here most of the time	NR	33
Blank LH	PFU = no followup, EFU = FUP lived here	NR	22,224
LH blank	PFU = FUP lived here, EFU = no followup	NR	1,292
OP Blank	PFU = FUP had another residence but stayed here most of the time, EFU = no followup	NR	34

<b>Keyed Why Code Combination (PFUKY EFUKY)</b>	<b>Description of Why Codes from each form</b>	<b>Keyed Residence Status</b>	<b>Cell Size</b>
Blank OR1	PFU = no followup, EFU = FUP had another residence but stayed here most of the time	NR	509
Blank ORDK	PFU = no followup, EFU = respondent didn't know if FUP had another residence	NR	970
Blank KR	PFU = no followup, EFU = no knowledgeable respondent	NR	204
LH ORDK	PFU = FUP lived here, EFU = respondent didn't know if FUP had another residence	NR	183
OP ORDK	PFU = FUP has another residence and stayed here most of the time, EFU = respondent didn't know if FUP had another residence	NR	3
ORDK LH	PFU = respondent didn't know if FUP had another residence, EFU = FUP lived here	NR	233
NI LH	PFU = noninterview, EFU = FUP lived here	NR	75
NI OR1	PFU = noninterview, EFU = FUP had another residence but stayed here most of the time	NR	3
ORDK OR1	PFU = respondent didn't know if FUP had another residence, EFU = FUP had another residence but lived here most of the time	NR	9
Blank AD	PFU = no followup, EFU = FUP has another residence and stays there most of the time, respondent didn't provide a valid address	NU	3
Blank DF	PFU = no followup, EFU = FUP has another residence, respondent didn't know which place FUP stayed at most of the time	NU	1
Blank GB	PFU = no followup, EFU = geocoding section was left blank	NU	1

The donor groups contain cases where the final measurement code was determined by the PFU/EFU Review<sup>6</sup>. These cases were also computer coded using keyed data. To create donor groups we,

1. used the keyed why code combinations from the recipient cells,
2. determined the number of E-sample PFU/EFU Review cases with the same keyed why code combination, and
3. collapsed combinations together to form groups of at least 30 donor cases; the combination must have the same enumeration status to be collapsed. Since enumeration status has a counterpart in residence status we've converted the terminology: a correct enumeration to a resident; an erroneous enumeration to a nonresident, and unresolved remains unresolved.

<sup>6</sup> Some PFU/EFU Review cases were sent back for additional clerical review in the A.C.E. Revision II Clerical Coding operation.

**Table A5. P-sample Donor Groups – PFU/EFU Review Cases**

<b>Keyed Why Code Combination (PFUKY EFUKY)</b>	<b>Keyed Residence Status</b>	<b>Donor Cell Size</b>	<b>Collapsed Donor Group</b>	<b>Collapsed Donor Group Description</b>
Blank DO	NN	25	M	Nonresident
Blank GQ	NN	8	M	Nonresident
Blank MS	NN	7	M	Nonresident
Blank NH	NN	27	M	Nonresident
Blank OR2	NN	23	M	Nonresident
GQ Blank	NN	5	M	Nonresident
GQ DO	NN	66	M	Nonresident
GQ GQ	NN	7	M	Nonresident
GQ JBPb	NN	8	M	Nonresident
GQ OR2	NN	2	M	Nonresident
NI OR2	NN	5	M	Nonresident
ORDK GQ	NN	2	M	Nonresident
ORDK OR2	NN	4	M	Nonresident
ORDK ORDK	NU	303	N	Not Enough Information, Unresolved
ORDK blank	NU	19	N	Not Enough Information, Unresolved
KR Blank	NU	4	N	Not Enough Information, Unresolved
KR ORDK	NU	10	N	Not Enough Information, Unresolved
KR KR	NU	6	N	Not Enough Information, Unresolved
NI Blank	NU	1	N	Not Enough Information, Unresolved
NI ORDK	NU	30	N	Not Enough Information, Unresolved
NI KR	NU	8	N	Not Enough Information, Unresolved
ORDK KR	NU	34	N	Not Enough Information, Unresolved
LH LH	NR	1,045	O	Lived Here/Other residence both forms
LH OR1	NR	39	O	Lived Here/Other residence both forms
OP LH	NR	19	O	Lived Here/Other residence both forms
OP OR1	NR	9	O	Lived Here/Other residence both forms
Blank LH	NR	5,902	P	Lived Here, One form blank
LH blank	NR	30	P	Lived Here, One form blank
OP Blank	NR	0	P	Lived Here, One form blank
Blank OR1	NR	161	Q	PFU blank, EFU = other residence lived here
Blank ORDK	NR	224	R	EFU not enough information, PFU resident
Blank KR	NR	24	R	EFU not enough information, PFU resident
LH ORDK	NR	116	R	EFU not enough information, PFU resident
OP ORDK	NR	2	R	EFU not enough information, PFU resident
ORDK LH	NR	1,181	S	PFU not enough information, EFU lived here
NI LH	NR	159	S	PFU not enough information, EFU lived here
NI OR1	NR	14	S	PFU not enough information, EFU lived here
ORDK OR1	NR	41	S	PFU not enough information, EFU lived here
Blank AD	NU	11	T	PFU blank, EFU unresolved address information
Blank DF	NU	32	T	PFU blank, EFU unresolved address information
Blank GB	NU	0	T	PFU blank, EFU unresolved address information

Based on the E-sample donor group collapsing, we then collapsed the P-sample recipient cells into these groups. This process yields the following combined groups:

**Table A6. Combined Group Donor and Recipient Counts**

<b>Combined Group</b>	<b>Donors (PFU/EFU Review cases)</b>	<b>Recipients (At-Risk cases)</b>
M – Nonresident	189	69
N – Not Enough Information	415	264
O – Lived Here/Other Residence both forms	1,112	5,360
P – Lived here, one form blank	5,932	23,550
Q – No PFU and EFU other residence lived here	161	509
R – PFU resident, EFU not enough information	366	1,360
S – PFU not enough information, EFU lived here	1,395	320
T – No PFU, EFU unresolved address information	43	5
Total	9,613	31,437

## Appendix B – A.C.E. Revision II DSE

$$\text{Re DSE } C_{ij} = (Cen_{ij} - H_{ij} - LA_{ij}) \left[ \frac{CE_i^{ND} f_{1,i'} + C \tilde{E}_i^D}{E_i} \right] \left[ \frac{M_{nm,j}^{ND} f_{2,j'} + \tilde{M}_{nm,j}^D + \frac{M_{om,j} f_{3,j'}}{P_{om,j} f_{4,j'}} \left( P_{im,j} f_{5,j'} + g \left( P_{nm,j}^D - \tilde{P}_{nm,j}^D \right) \right)}{P_{nm,j}^{ND} f_{6,j'} + \tilde{P}_{nm,j}^D + P_{im,j} f_{5,j'} + g \left( P_{nm,j}^D - \tilde{P}_{nm,j}^D \right)} \right]$$

### General notation

Terms:	CE	weighted estimate of correct enumerations
	E	weighted E-Sample estimate
	M	weighted estimate of matches
	P	weighted P-Sample estimate
	$f$ 's	double sampling ratio adjustment that corrects for measurement error in the full sample using the revised coding of revision sample
	$g$	adjusts for nonmovers with duplicate links that could be inmovers
Subscripts:	$i$ ( $j$ )	full E (P) Sample Poststrata
	$i'$ ( $j'$ )	revision E (P) Sample Poststrata
	nm, om, im	indicates nonmover, outmover, and inmover
Superscripts:	C	indicates the use of version C for the treatment of movers
	ND	is not a duplicate link to a census enumeration outside the search area
	D	is a duplicate link to a census enumeration outside the search area
	$\sim$	indicates that the estimate from duplicate links includes an adjustment for the probability that the particular case is a resident given that it was found to be a duplicate



More specifics

$$f_{1,i'} = \frac{CE_{i'}^{ND*}}{CE_{i'}^{ND}} \quad f_{2,j'} = \frac{M_{nm,j'}^{ND*}}{M_{nm,j'}^{ND}} \quad f_{3,j'} = \frac{M_{om,j'}^*}{M_{om,j'}} \quad f_{4,j'} = \frac{P_{om,j'}^*}{P_{om,j'}} \quad f_{5,j'} = \frac{P_{im,j'}^*}{P_{im,j'}} \quad f_{6,j'} = \frac{P_{nm,j'}^{ND*}}{P_{nm,j'}^{ND}}$$

where \* indicates that the estimate was constructed using the revised codes.

## Appendix C – Standard Errors for Results

Corresponding to Table 6, table C1 includes the standard errors for the estimators.

**Table C1. E-sample Double-Sampling Ratios by Poststratum Group**

#	Poststratum Group Description	A.C.E. Revision II		At-Risk		Difference		Difference Confidence Interval	
		$f_1$	Standard Error	$f_1$	Standard Error	$f_1$	Standard Error	Lower Bound	Upper Bound
1	All proxy cases	0.97670	0.01132	0.97917	0.01147	-0.00246	0.00159	-0.00559	0.00066
2	American Indians on Reservations (AIR) who are Non-proxy	0.97450	0.00894	0.97360	0.00894	0.00090	0.00004	0.00082	0.00097
3	Non-Proxy/non-AIR Nuclear member 0-9 M&F	0.99682	0.00185	0.99655	0.00185	0.00027	0.00010	0.00007	0.00046
4	Non-Proxy/non-AIR Nuclear member 10-17 M&F	0.99294	0.00367	0.99266	0.00367	0.00029	0.00009	0.00011	0.00046
5	Non-Proxy/non-AIR Nuclear member 18-29 M	0.99175	0.00349	0.99188	0.00350	-0.00013	0.00042	-0.00095	0.00069
6	Non-Proxy/non-AIR Nuclear member 18-29 F	0.99331	0.00474	0.99445	0.00481	-0.00114	0.00099	-0.00309	0.00080
7	Non-Proxy/non-AIR Nuclear member 30-49 M	0.99420	0.00195	0.99384	0.00195	0.00035	0.00005	0.00025	0.00045
8	Non-Proxy/non-AIR Nuclear member 30-49 F	0.99657	0.00148	0.99631	0.00149	0.00026	0.00008	0.00011	0.00041
9	Non-Proxy/non-AIR Nuclear member 50+ M	0.99189	0.00262	0.99172	0.00263	0.00017	0.00012	-0.00007	0.00041
10	Non-Proxy/non-AIR Nuclear member 50+ F	0.99592	0.00172	0.99588	0.00172	0.00004	0.00016	-0.00028	0.00036
11	Non-Proxy/non-AIR Other member 0-9 M&F	0.98178	0.01041	0.98142	0.01041	0.00035	0.00027	-0.00017	0.00088
12	Non-Proxy/non-AIR Other member 10-17 M&F	0.98826	0.00550	0.98788	0.00551	0.00037	0.00027	-0.00017	0.00091
13	Non-Proxy/non-AIR Other member 18-29 M	0.97940	0.00540	0.97936	0.00540	0.00004	0.00021	-0.00038	0.00046
14	Non-Proxy/non-AIR Other member 18-29 F	0.97507	0.00569	0.97470	0.00570	0.00037	0.00012	0.00014	0.00061
15	Non-Proxy/non-AIR Other member 30-49 M	0.97777	0.00676	0.97818	0.00680	-0.00041	0.00063	-0.00165	0.00083
16	Non-Proxy/non-AIR Other member 30-49 F	0.98937	0.00566	0.98933	0.00567	0.00004	0.00035	-0.00064	0.00072
17	Non-Proxy/non-AIR Other member 50+ M	0.98894	0.00420	0.98856	0.00420	0.00038	0.00009	0.00020	0.00057
18	Non-Proxy/non-AIR Other member 50+ F	0.98688	0.00395	0.98666	0.00396	0.00022	0.00013	-0.00003	0.00047
19	Non-Proxy/non-AIR Nuclear member 0-17 M&F	0.99493	0.00247	0.99465	0.00247	0.00027	0.00007	0.00014	0.00041
20	Non-Proxy/non-AIR Nuclear member 18-29 M&F	0.99267	0.00335	0.99340	0.00340	-0.00073	0.00069	-0.00208	0.00061
21	Non-Proxy/non-AIR Nuclear member 30-49 M&F	0.99549	0.00150	0.99518	0.00150	0.00030	0.00005	0.00020	0.00041
22	Non-Proxy/non-AIR Nuclear member 50+ M&F	0.99390	0.00167	0.99379	0.00167	0.00011	0.00012	-0.00014	0.00035
23	Non-Proxy/non-AIR Other member 0-17 M&F	0.98451	0.00650	0.98415	0.00651	0.00036	0.00019	-0.00002	0.00074
24	Non-Proxy/non-AIR Other member 18-29 M&F	0.97743	0.00394	0.97724	0.00394	0.00019	0.00013	-0.00006	0.00045
25	Non-Proxy/non-AIR Other member 30-49 M&F	0.98282	0.00457	0.98304	0.00458	-0.00021	0.00039	-0.00098	0.00055
26	Non-Proxy/non-AIR Other member 50+ M&F	0.98754	0.00305	0.98727	0.00306	0.00027	0.00010	0.00007	0.00047

**Table C1. E-sample Double-Sampling Ratios by Poststratum Group Con't**

#	Poststratum Group Description	A.C.E. Revision II		At-Risk		Difference		Difference Confidence Interval	
		$f_1$	Standard Error	$f_1$	Standard Error	$f_1$ - At-risk	Standard Error	Lower Bound	Upper Bound
27	Non-Proxy/non-AIR Nuclear member 0-29 M&F	0.99451	0.00214	0.99442	0.00214	0.00009	0.00016	-0.00022	0.00040
28	Non-Proxy/non-AIR Nuclear member 30+ M&F	0.99477	0.00113	0.99456	0.00112	0.00021	0.00007	0.00008	0.00035
29	Non-Proxy/non-AIR Other member 0-29 M&F	0.97895	0.00345	0.97872	0.00346	0.00023	0.00012	-0.00001	0.00046
30	Non-Proxy/non-AIR Other member 30+ M&F	0.98540	0.00262	0.98535	0.00263	0.00005	0.00019	-0.00032	0.00042
31	Non-Proxy/non-AIR Nuclear member 0+ M&F	0.99467	0.00137	0.99450	0.00137	0.00016	0.00008	0.00001	0.00032
32	Non-proxy/non-AIR Other member 0+ M&F	0.98269	0.00222	0.98256	0.00222	0.00012	0.00013	-0.00012	0.00037

Table C2 corresponds to Table 7.

**Table C2. P-sample Inmovers Double-Sampling Ratios by Poststratum Group Con't**

#	Postratum Group Description	A.C.E. Revision II		At-Risk		Difference		Difference Confidence Interval	
		$f_5$	Standard Error	$f_5$	Standard Error	$f_5$ - At-risk	Standard Error	Lower Bound	Upper Bound
1	Owner	0.88696	0.03226	0.88687	0.03201	0.00009	0.00404	-0.00784	0.00802
2	Non-Owner	1.00994	0.02109	1.00993	0.02108	0.00001	0.00053	-0.00103	0.00105

Table C3 corresponds to Table 8.

**Table C3. P-sample Outmover Matches Double-Sampling Ratios by Poststratum Group**

#	Postratum Group Description	A.C.E. Revision II		At-Risk		Difference		Difference Confidence Interval	
		$f_3$	Standard Error	$f_3$	Standard Error	$f_3$ - At-risk	Standard Error	Lower Bound	Upper Bound
3	Owner	0.78901	0.05620	0.78875	0.05504	0.00025	0.01131	-0.02192	0.02243
4	Non-Owner	0.81818	0.03870	0.81788	0.03634	0.00030	0.01326	-0.02569	0.02629

Table C4 corresponds to Table 9.

**Table C4. P-sample Outmover Non-Matches Double-Sampling Ratios by Poststratum Group**

#	Poststratum	A.C.E. Revision II		At-Risk		Difference		Difference Confidence Interval	
		$f_4$	Standard Error	$f_4$	Standard Error	RevII $f_4$ - At-risk $f_4$	Standard Error	Lower Bound	Upper Bound
3	Owner	0.76287	0.04992	0.76249	0.04694	0.00038	0.01694	-0.03283	0.03359
4	Non-Owner	0.83096	0.03474	0.83050	0.02782	0.00046	0.02077	-0.04024	0.04117

Table C5 corresponds to Table 10. RevII is the A.C.E. Revision II

**Table C5. P-sample Nonmover Matches Double-Sampling Ratios by Poststratum Group**

#	Poststratum	A.C.E. Revision II		At-Risk		Difference		Difference Confidence Interval	
		$f_2$	Standard Error	$f_2$	Standard Error	RevII $f_2$ - At-risk $f_2$	Standard Error	Lower Bound	Upper Bound
5	American Indians on Reservations (AIR)	0.99567	0.03038	0.99504	0.01077	0.00063	0.02839	-0.05502	0.05629
6	Non-AIR/Owner 0-9 M&F	1.00141	0.01149	1.00116	0.00342	0.00025	0.01096	-0.02124	0.02173
7	Non-AIR/Owner 10-17 M&F	1.00114	0.01174	1.00090	0.00524	0.00023	0.01051	-0.02036	0.02083
8	Non-AIR/Owner 18-29 M	1.00344	0.01417	1.00319	0.00863	0.00025	0.01124	-0.02178	0.02228
9	Non-AIR/Owner 18-29 F	1.00064	0.01457	1.00037	0.00791	0.00027	0.01223	-0.02371	0.02425
10	Non-AIR/Owner 30-49 M	1.00424	0.01255	1.00397	0.00369	0.00027	0.01199	-0.02324	0.02378
11	Non-AIR/Owner 30-49 F	1.00325	0.00948	1.00305	0.00282	0.00020	0.00905	-0.01754	0.01795
12	Non-AIR/Owner 50+ M	1.00446	0.01038	1.00424	0.00273	0.00022	0.01002	-0.01941	0.01986
13	Non-AIR/Owner 50+ F	1.00246	0.01015	1.00224	0.00282	0.00022	0.00975	-0.01889	0.01932
14	Non-AIR/Non-Owner 0-9 M&F	1.00600	0.01587	1.00566	0.00509	0.00034	0.01503	-0.02912	0.02979
15	Non-AIR/Non-Owner 10-17 M&F	0.99989	0.01741	0.99953	0.00683	0.00036	0.01601	-0.03103	0.03174
16	Non-AIR/Non-Owner 18-29 M	1.01411	0.02461	1.01361	0.01032	0.00050	0.02233	-0.04327	0.04427
17	Non-AIR/Non-Owner 18-29 F	1.00866	0.02394	1.00814	0.00663	0.00051	0.02300	-0.04456	0.04559

**Table C5. P-sample Nonmover Matches Double-Sampling Ratios by Poststratum Group Con't**

#	Poststratum Postratum Group Description	A.C.E. Revision II		At-Risk		Difference		Difference Confidence Interval	
		$f_2$	Standard Error	$f_2$	Standard Error	RevII $f_2$ – At-risk $f_2$	Standard Error	Lower Bound	Upper Bound
18	Non-AIR/Non-Owner 30-49 M	1.00313	0.02095	1.00268	0.00690	0.00044	0.01978	-0.03833	0.03921
19	Non-AIR/Non-Owner 30-49 F	0.99992	0.01950	0.99950	0.00532	0.00042	0.01876	-0.03635	0.03719
20	Non-AIR/Non-Owner 50+ M	0.99607	0.01961	0.99568	0.00827	0.00040	0.01778	-0.03445	0.03525
21	Non-AIR/Non-Owner 50+ F	1.00025	0.01690	0.99992	0.00854	0.00033	0.01458	-0.02826	0.02891
22	Non-AIR/Owner 18+ M	1.00422	0.01130	1.00397	0.00240	0.00025	0.01104	-0.02140	0.02189
23	Non-AIR/Owner 18+ F	1.00258	0.00996	1.00236	0.00206	0.00022	0.00975	-0.01889	0.01933
24	Non-AIR/Non-Owner 18+ M	1.00493	0.02070	1.00448	0.00487	0.00045	0.02011	-0.03897	0.03987
25	Non-AIR/Non-Owner 18+ F	1.00267	0.01930	1.00224	0.00390	0.00042	0.01890	-0.03662	0.03747

Table C6 corresponds to Table 11.

**Table C6. P-sample Nonmover Matches Double-Sampling Ratios by Poststratum Group**

#	Poststratum Postratum Group Description	A.C.E. Revision II		At-Risk		Difference		Difference Confidence Interval	
		$f_6$	Standard Error	$f_6$	Standard Error	A.C.E. Revision II $f_6$ – At- risk $f_6$	Standard Error	Lower Bound	Upper Bound
5	American Indians on Reservations (AIR)	1.00058	0.01077	1.00057	0.01078	0.00001	0.00074	-0.00143	0.00146
6	Non-AIR/Owner 0-9 M&F	1.00125	0.01363	1.00095	0.00329	0.00030	0.01322	-0.02562	0.02621
7	Non-AIR/Owner 10-17 M&F	1.00073	0.01219	1.00048	0.00511	0.00025	0.01106	-0.02144	0.02193
8	Non-AIR/Owner 18-29 M	0.99695	0.04347	0.99788	0.01305	-0.00093	0.04149	-0.08224	0.08039
9	Non-AIR/Owner 18-29 F	1.00487	0.00795	1.00489	0.00788	-0.00002	0.00112	-0.00221	0.00216
10	Non-AIR/Owner 30-49 M	1.00258	0.01417	1.00227	0.00350	0.00031	0.01373	-0.02661	0.02723
11	Non-AIR/Owner 30-49 F	1.00302	0.01099	1.00278	0.00277	0.00024	0.01063	-0.02060	0.02107
12	Non-AIR/Owner 50+ M	1.00446	0.01236	1.00419	0.00300	0.00027	0.01199	-0.02322	0.02376

**Table C6. P-sample Nonmover Matches Double-Sampling Ratios by Poststratum Group  
Con't**

#	Poststratum Postratum Group Description	A.C.E. Revision II		At-Risk		Difference		Difference Confidence Interval	
		$f_6$	$f_6$ Standard Error	$f_6$	$f_6$ Standard Error	A.C.E. Revision II $f_6$ - At- risk $f_6$	Standard Error	Lower Bound	Upper Bound
13	Non-AIR/Owner 50+ F Non-AIR/Non-Owner 0-9	1.00330	0.01177	1.00305	0.00277	0.00026	0.01144	-0.02216	0.02268
14	M&F Non-AIR/Non-Owner 10-17	1.00464	0.02109	1.00419	0.00527	0.00046	0.02042	-0.03956	0.04047
15	M&F Non-AIR/Non-Owner 18-29	1.00764	0.02270	1.00715	0.00632	0.00049	0.02180	-0.04224	0.04321
16	M Non-AIR/Non-Owner 18-29	1.02601	0.01840	1.02636	0.01019	-0.00034	0.01539	-0.03051	0.02982
17	F Non-AIR/Non-Owner 30-49	1.01453	0.02616	1.01397	0.00728	0.00056	0.02512	-0.04868	0.04980
18	M Non-AIR/Non-Owner 30-49	0.99797	0.02474	0.99744	0.00662	0.00053	0.02383	-0.04618	0.04724
19	F Non-AIR/Non-Owner 50+	0.99810	0.02444	0.99756	0.00513	0.00053	0.02389	-0.04629	0.04736
20	M	0.99861	0.01871	0.99823	0.00817	0.00038	0.01682	-0.03259	0.03335
21	Non-AIR/Non-Owner 50+ F	1.00498	0.01734	1.00463	0.00676	0.00036	0.01597	-0.03095	0.03166
22	Non-AIR/Owner 18+ M	1.00253	0.00552	1.00243	0.00282	0.00011	0.00475	-0.00920	0.00941
23	Non-AIR/Owner 18+ F Non-AIR/Non-Owner 18+	1.00338	0.00970	1.00317	0.00203	0.00021	0.00949	-0.01839	0.01881
24	M	1.00736	0.01049	1.00715	0.00480	0.00021	0.00934	-0.01809	0.01851
25	Non-AIR/Non-Owner 18+ F	1.00507	0.02245	1.00457	0.00366	0.00050	0.02215	-0.04292	0.04391