



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

12/31/02

MASTER FILE

DSSD A.C.E. REVISION II MEMORANDUM SERIES # PP-9

PRED CENSUS AND SURVEY MEASUREMENT STAFF MEMORANDUM SERIES:
CSM-A.C.E. REVISION II-12R

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

From: Mary Mulry *signed 12/31/02* *MM*
Chairs, A.C.E. Revision II Quality Indicators Subgroup
Statistical Research Division

Through: David Hubble *signed 12/31/02* *RJH*
Assistant Division Chief, Evaluations
Planning, Research, and Evaluation Division

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

Subject: At-Risk Codes Evaluation Study Plan

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)

At-Risk Codes Evaluation Study Plan
Tammy Adams, Eli Krejsa, and Mary Mulry
December 31, 2002

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 Accuracy and Coverage Evaluation 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

	E-sample	P-sample
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. QUESTION TO BE ANSWERED

How much error potentially was introduced into the A.C.E. Revision II dual system estimates (DSE) by the automated coding, rather than clerical coding, of some cases?

3. METHODOLOGY

3.1 Error Factors

To assess the potential error in the DSE due to the at-risk cases, we will use 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 category¹ that were in the PFU/EFU Review. We will use the following approach to calculate the potential error in the DSE:

- Create Donor Cells – These are cases in a given combined keyed code category² in the PFU/EFU Review.
- Calculate Error Factors – We will calculate 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 will use the same correct enumeration probability³ for all cases with that why code category.

3.1.1. Creating Donor Cells

For each at-risk case, the computer code will be 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.

¹ 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”.

² 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.

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

Using these why codes, donor cells for error factors are 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 are formed to calculate the error factor that accounts for the coding of both forms and the final result. Each combination results 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.

3.1.2 Calculating Error Factors

Error factors are 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.

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

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

where:

$\text{ceprobi}_{FinalRevII}$ ⁴=the probability of correct enumeration for that case’s A.CE final code. The probability is usually 1 or 0, but for unresolved cases it can be within that range.

$\text{ceprobi}_{Bestkey}$ =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.

3.1.3 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:

⁴ Note: we use the *ceprobi*, as opposed to *ceprobf*. *ceprobf* captures the adjustment for duplicates in the non-E-sample which we do not want to capture. *ceprobi* does not capture it.

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

PFU / EFU reiew, with UE code in why code cat

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

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

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

3.1.4 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:

$$f_{1,i',ar} = \frac{CE_i^{ND*}}{CE_i^{ND}} = \frac{\sum ceprob_f * finalwgt}{CE_i^{ND}}$$

all cases in RevII

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

3.2 Sampling error assessment

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

4. DATA REQUIREMENTS

We require the following files:

- Measurement Output Files: EVC and PVC
- Missing Data Files: EMD and PMD
- Estimation Files: Revision E-sample file, Revision P-sample file

5. RESPONSIBILITIES

- Study Plan Modification – Eli and Mary
- Calculate Error Rates - Tammy and Eli
- Calculate f - Tammy
- Evaluation Results - Tammy and Eli

6. SCHEDULE

Activity	Due Date	Responsible
Write study plan	12/31/02	Eli, Tammy, and Mary
Define donor groups	12/19/02	Eli and Tammy
Calculate error factors	12/19/02	Eli and Tammy
Calculate fs	12/30/02	Tammy
At-risk draft report	12/11/02	Eli and Tammy
At-risk final report	12/31/02	Eli and Tammy

7. LIMITATIONS

We are making assumptions about the P-sample error rate based on the E-sample error rates. We will be using 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.

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

8. 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. Revised 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. Revised 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. Revised II Estimates 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

Keyed Why Code Combination (PFUKY EFUKY)	Description of Why Codes from each form	Keyed Enumeration Status	Cell Size
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	CE	843

	residence		
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
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 Review⁵. 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

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

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
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:

Table A3. Combined Group Donor and Recipient Counts

Combined Group	Donors (PFU/EFU Review cases)	Recipients (At-Risk cases)
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.

Table A4. P-sample Recipient Cells

Keyed Why Code Combination (PFUKY EFUKY)	Description of Why Codes from each form	Keyed Residence Status	Cell Size
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
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 = noninterivew, 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,	NR	142

	EFU = FUP had another residence but stayed here most of the time		
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
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 Review⁶. 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.

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

Keyed Why Code Combination (PFUKY EFUKY)	Keyed Residence Status	Donor Cell Size	Collapsed Donor Group	Collapsed Donor Group Description
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

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

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

Combined Group	Donors (PFU/EFU Review cases)	Recipients (At-Risk cases)
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. Revised II DSE

$$\text{Re DSE } C_{ij} = (\text{Cen}_{ij} - II_{ij} - LA_{ij})$$

$$\left[\frac{CE_i^{ND} f_{1,i} + C\tilde{E}_i^D}{E_i} \right] \left[\frac{M_{om,j} f_{3,j'}}{P_{om,j} f_{4,j'}} \left(P_{im,j} f_{5,j'} + g \left(P_{nm,j}^D - \tilde{P}_{nm,j}^D \right) \right) \right]$$

$$\left[\frac{M_{nm,j}^{ND} f_{2,j'} + \tilde{M}_{nm,j}^D}{P_{nm,j}^{ND} f_{6,j'} + \tilde{P}_{nm,j}^D + P_{im,j} f_{5,j'} + g \left(P_{nm,j}^D - \tilde{P}_{nm,j}^D \right)} \right]$$

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

~ 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.