

Chapter 3: Correcting Data for Measurement Error

Tamara Adams and Elizabeth Krejsa 12/16/02

3.1. Introduction

As a result of the ESCAP II decision process and our subsequent direction to revise the A.C.E. estimates, revised results from matching and coding operations were produced to correct for measurement errors in the A.C.E. Revision II Sample for the A.C.E. Revision II. This chapter covers the sources of measurement error with respect to the A.C.E. Revision II and how we implemented the correction of these errors for the A.C.E. Revision II

The quick history of this project follows: As part of the A.C.E. survey processing activities we obtained census data for persons in the A.C.E. sample to determine their residence/enumeration status. Later, the Evaluation Followup (EFU) was undertaken on a subsample of about 70,000 of these A.C.E. cases and matching and residence coding were conducted for this subsample.

Then, for the ESCAP II decision very experienced coders (Analysts) at the National Processing Center examined a Review sample of about 17,500 cases, including discrepancies between the A.C.E. production work and the EFU work.

We resolved as many of the remaining discrepancies as possible; that is, we corrected for as much measurement error as possible for the cases in the Revision sample of the A.C.E. Revision II

We first obtained the results of the A.C.E. Person Followup (PFU) and had the EFU interviews keyed. We then looked at groups of cases that had the same answers from the two interviews, compared them with similar cases from the Review sample and decided how to code all the cases in each group using computer algorithms.

Then we sent the remaining cases, problem cases that were not codable using computer algorithms—about 25,000 to the Analysts at NPC to resolve. We describe the types of problems and our solutions in the sections below.

3.2. Goals and Background

We attempted to correct as much measurement error as possible, considering resource and timing constraints¹. There were several major sources of error:

- Residence and Enumeration Status
 - Problem -- The original A.C.E. did not detect all of the erroneous enumerations (Adams and Krejsa, 2001, Fay, 2002). The Evaluation Followup (EFU) detected approximately 1.4 million additional erroneous

¹ In order to complete the A.C.E. Revision II estimates on time, we were allotted 12 weeks of coding time. We estimated that approximately 25,000 cases could be coded in that time frame by the analysts in NPC.

enumerations in the E-sample (Adams and Krejsa, 2001). Since the coding of enumeration status in the E-sample was identical to the coding of residence status in the P-sample, we expected to see similar results in P-sample residence status coding as we saw in E-sample enumeration status coding (i.e., additional nonresidents found as a result of the EFU).

- Solution -- To correct for the residence status errors, we used a recoding of the Evaluation Followup Interview in combination with the original A.C.E. to determine the best residence or enumeration status for each person within the evaluation clusters.
- Matching Error
 - Problem -- The Matching Error Study (MES) showed a net difference in match codes between the production matching results and the evaluation matching results of 0.41% in the E-sample and 0.20% in the P-sample. This net difference translated into an increase in the Dual System Estimate (DSE) of 483,938 (Bean 2001).
 - Solution -- To correct for the matching error, we used the results of the Matching Error Study in conjunction with the results of the Revised A.C.E. Revision II recoding to determine the appropriate match status for each person.
- Mover Status
 - Problem -- Raglin and Krejsa (2001) found a 2.6 percent gross difference rate in the mover status between the original A.C.E. and the Evaluation Followup. This translated into a negative bias of 465,000 in the DSE (assuming no other biases).
 - Solution – To correct for mover status errors, we used the results of the Evaluation Followup. The EFU questionnaire contained questions designed to probe for a person’s mover status. We captured this information during the clerical recoding and during the initial coding of the Evaluation Followup form.

These types of measurement errors were corrected either by computer or clerically.

We considered two sources of error to be out-of-scope or negligible with respect to the Revised A.C.E. and did not correct for those errors. These errors included:

- Geocoding Errors – Certain geocoding errors detected by various geocoding evaluations were not included in the A.C.E. Revision II². Within the P-sample, 245,926 production nonmatched residents were found to be located outside the search area³ and 195,321 production correct enumerations in the E-sample were found to be located outside the search area (Adams and Liu, 2001). Some of the

² As part of the A.C.E., we conducted several evaluations of geocoding error on various subsamples of the A.C.E., most notably Targeted Extended Search 2 (TES2) and Targeted Extended Search 3 (TES3). The results of these evaluations can be found in Adams and Liu, 2001.

³ For the 2000 A.C.E., the search area, or area in which a person can be considered a correct enumeration or match, was the cluster and any census block touching the cluster.

correct enumerations outside the search area were identified by the EFU interview and hence were reflected in the revised coding.⁴

- Duplicates outside the Search Area – Duplicates found outside the search area as a result of computer matching (see Chapter 5 of this document) were not handled by clerical coding. These duplicates were accounted for in the DSE using estimation techniques. (See Chapter 6 of this document.)

3.3. Residence Status and Enumeration Status

As stated above, the original A.C.E. did not detect enough of the erroneous enumerations. To correct for this, the best residence status code was determined based on the field followup data available. (Duplicates were corrected by a separate process). The following data were available for coding:

- Person Interview (PI) – The PI was the original A.C.E. enumeration of the P-sample. It was a Computer-Assisted Personal Interview questionnaire designed to fully enumerate the persons in the A.C.E. It was conducted by either phone or personal visit between April and September, 2000.
- Person Followup (PFU) – The PFU was the followup used to assign residence and enumeration status whenever those items were not determined after the before followup matching (Childers, 2001). It was conducted by personal visit in October and November, 2000, approximately 6-7 months after Census Day.
- Evaluation Followup (EFU) – The EFU was an evaluation of the A.C.E. designed to more readily detect unusual living situations using additional probes and additional interviewing techniques (e.g., flashcards). It was conducted by personal visit in January and February, 2001, approximately 9-10 months after Census Day.

During the missing data processes the results of the PI interview were used to assign the A.C.E. residence status by computer to all people in A.C.E. who did not need followup.

The PFU was used to assign residence status for anyone who was eligible for followup (Childers, 2001). The PFU is similar to the PI. The PFU process interviewed both P-sample people and E-sample people.

The EFU followed up a sample of people sent to PFU and a sample of those not sent to PFU. In this way, the residence/enumeration status of a representative sample of people eligible for field followup can be evaluated.

There were coding errors in both the PFU and the EFU coding (Bean, 2001 and Adams and Krejsa, 2001, respectively). The EFU was also not coded strictly according to census residence rules. To evaluate the E-sample for ESCAP II, the Census Bureau conducted the PFU/EFU Review in the summer of 2001. A subsample of the E-sample

⁴ Some of the cases within the TES2 were also evaluated using the Evaluation Followup questionnaire. For those cases, we included the results of the geocoding evaluation within the Evaluation Followup; however, if a case was in the TES2 and not in the EFU, we did not include any geocoding evaluation results.

people in the EFU was re-reviewed by expert matchers using rules consistent with census residence rules. These analysts were assumed to make negligible error; therefore, we considered the PFU/EFU Review to be free of coding error.⁵

For the A.C.E. Revision II, we needed coding with the same level of quality as the PFU/EFU Review for a large enough sample in both the P-sample and E-sample to provide accurate subgroup estimates. Twelve weeks coding time were allotted to clerically code approximately 25,000 people. However, there were over 100,000 people needing codes. To assign the highest quality codes while meeting scheduled dates, we decided to use data keyed from both the PFU form and EFU form to augment clerical coding procedures. We used an automated coding algorithm based on the questionnaires to determine the code from the keyed data. We assigned a code using both the PFU keyed data and the EFU keyed data. We then used a three-step process to assign final codes to each case and describe how each of these were carried out in the next 3 sections:

- Validation – Determine if using the keyed data produces high-quality coding for various subsets of cases using the PFU/EFU Review as a truth deck.
- Targeting – Target only those cases for clerical review that have codes produced by the computer from the keyed data that are not of high enough quality.
- Clerical Coding – Clerically code only those cases that cannot be coded using the computer.

3.3.1. Validation of Keyed Data

To validate the quality of coding produced by the keyed data algorithm, we programmed the skip patterns for both questionnaires to determine an appropriate match code and why code⁶ for each case. Then, for both the PFU and EFU forms, we examined the percentage agreement with the original coding (either production coding or the coding of the EFU form) for the respective form, the percentage agreement with the PFU/EFU Review and the residual risk.

The residual risk of disagreement (i.e. potential bias) represented the risk we would take in accepting the code based on the keyed data for categories defined by questionnaire responses and the corresponding match code.

$$risk = Agree_K - Agree_{Rev}$$

where:

$Agree_K$ = The weighted number of cases whose code from the keyed data agreed with the original production code

⁵ Throughout this document, “coding error” refers to any clerical error that was made during any of the previous coding operations. The A.C.E. Revision II does not attempt to correct for respondent error made during any of the field operations.

⁶ A why codes is assigned to reflect why each person record is assigned its residence or enumeration status.

Agree_{Rev}=Of those cases where the code from the keyed data agrees with the original production code, the weighted number of cases whose code from the keyed data agreed with the PFU/EFU Review code

We say risk, rather than an error, because some conversions may not have had a full effect on the DSE. For example, people who were in group quarters have a residual risk of 26,517 after computer coding. These represent cases that probably should have been coded as a erroneous enumeration but were not. Some of those 26,517 could be unresolved cases which have a probability less than one of being correct.

We decided to reject the automated coding results for a given why code category if the residual risk was too high or there were not enough cases to make an informed decision. The exception to this rule was the category consisting of cases without any indication of living in a group quarters or other residence – it was by far the largest category for both the PFU and EFU so we expected a higher residual risk⁷.

3.3.2. Targeting Cases for Clerical Review

After we determined whether to accept the code from the keyed data for each category, we targeted cases for clerical review. Analysts performed the clerical review; these were the highest level of clerical matcher in production and were assumed to make negligible errors in coding due to experience and additional training.

Cases went to clerical review according to the logic below. In general, cases were only exempt from clerical review if both the PFU and EFU codes from the keyed data were accepted and agreed and the mover statuses agreed.

- The case was not in the PFU/EFU Review if any of the following were true:
 - The code from the keyed data for either form was not accepted for that case
 - The code from the keyed data was accepted for both forms but at least one of the codes from the keyed data did not agree with its original code (i.e., the PFU code from the keyed data did not agree with production; the EFU code from the keyed data did not agree with the original EFU code)
 - For P-sample people, the mover status from the keyed data did not agree with mover status assigned during the EFU coding
 - There was write-in information in open-ended questions on the form that we could not code
 - The case was a possible match in before-followup matching and the production and original EFU code disagreed
 - The case was a duplicate in either the original EFU coding or production after-followup coding

⁷ We used an absolute risk, rather than a relative risk. Therefore, larger categories tended to have higher risks.

- The case was not yet flagged for clerical review and the PFU code from the keyed data did not agree with EFU code from the keyed data and one of the cases was not unresolved for certain reasons
- The case was in the PFU/EFU Review and was conflicting or had a mover status disagreement between the keyed data and the original EFU mover status

For P-sample in-movers, we had no validation data. In those cases, we sent to clerical review any case where the original EFU mover status did not match the mover status from the keyed data or the residence status from the keyed data did not match the original EFU residence status. Cases such as noninterviews or cases where mover dates could not be read were also sent to clerical review.

Certain cases were exempt from clerical review because we could code them based on information available in data files. These cases included:

- Census Usual Home Elsewhere – If the person claimed a Usual Home Elsewhere on certain types of census forms, then they were counted as a correct enumeration within the cluster and did not need clerical review.
- Geocoding Errors from Initial Housing Unit Matching – If a case should not have been sent to PFU or EFU and was only sent due to clerical error in the initial production matching, then it did not need clerical review.

3.3.3. Clerical Review

The clerical review for A.C.E. Revision II was an analyst-only operation. We collected the following types of information:

- Match Code for each form
- Why Code for each form
- Respondent for each form
- Whether the respondents are the same for the two interviews
- Best Code – A code indicating which form is the better of the two forms
- Smooshed Code – Information from both forms combined to make a code to represent the true situation
- Mover Status – Mover Status from the EFU form for P-sample people

The match codes were assigned using the census residence rules to construct coding rules for the flow of the questionnaire.

The best code could be one of four values:

- Both=The enumeration statuses were the same
- PFU=The PFU form provided better information
- EFU=The EFU form provided better information

- Conflicting=Similar caliber respondents (e.g., Husband and Wife; two neighbors) provided contradictory information for the case

To ensure replicability, we applied computer edits to the best code. If the analyst did not follow pre-specified rules, then the analyst had to re-review the case or leave a note indicating the situation.

3.4. Correction of Mover Status Assignment Errors

For each P-sample case, we determined the mover status based on EFU. We used this mover status to determine whether or not the person needed clerical review.

3.5. Correction of Matching Errors

To correct for matching error, we used the results of the Matching Error Study. We were most interested in correcting for false matches and false nonmatches. Many other matching errors were as a result of incorrect residence status coding, which were corrected as stated above in section 3.2. We used the production duplicates since most duplicates were not eligible for EFU. To determine the correct match status to use, each of the possible combinations of match status were reviewed to determine the appropriate match status for each type of case. In general, we used MES when a match is changed to a nonmatch or a nonmatch to a match; in the remainder of cases, we used the match status from the original EFU coding. This correction was assigned by computer.

3.6. Data Outputs

After the clerical operation was completed, two files were established – a P-sample file and an E-sample file. The files contained match codes and why codes (where appropriate) for production, EFU, PFU/EFU Review, Keyed Data, and A.C.E. Revision II Clerical Review. We also assigned a final code in the following hierarchy: A.C.E. Revision II Clerical Review, PFU/EFU Review, Keyed Data. This code reflected the final match, residence, and enumeration status for the A.C.E. Revision II process.

3.7. Limitations

There were several limitations on the data for the A.C.E. Revision II:

- Sample Size – The sample used to estimate measurement error is 2,259 clusters, containing about 10% of the persons in the sample used in the production A.C.E. Due to the small sample size, some subgroup estimates may not be as accurate as the production A.C.E.
- Conflicting Cases –Conflicting cases occurred when the PFU and EFU interviews had respondents of the same caliber (either both non-proxy or proxy respondents who were in the position to have similar knowledge about the household, for instance, two neighbors) and those two respondents gave contradictory

- information. Since we did not have the option of an additional field followup, these cases were coded as conflicting and were reviewed separately and imputed.
- Data Collection Error – We coded cases to the best of our ability. However, we did not attempt to correct for data collection error. Respondent and interviewer errors could not be rectified without a field followup.

3.8. References

Adams, T. and Krejsa, E. (2001) “ESCAP II: Results of the Person Followup and Evaluation Followup Forms Review,” ESCAP Report No. 24. U.S. Census Bureau.

Adams and Liu (2001) “ESCAP II: Evaluation of Lack of Balance and Geographic Errors Affecting A.C.E. Person Estimates,” ESCAP Report No. 2. U.S. Census Bureau.

Bean, S. (2001) “ESCAP II: Accuracy and Coverage Evaluation Matching Error,” ESCAP Report No. 7. U.S. Census Bureau.

Childers, Danny R. (2001). Accuracy and Coverage Evaluation: The Design Document. DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1R dated January 24, 2001.

Fay, R. F. (2002) “Evidence of Additional Erroneous Enumerations from the Person duplication Study.” ESCAP II Report No. 9. Revised U.S. Census Bureau.

Keathley, Don H. (2001) “EFU Sample Design, Stratification, Selection, and Weighting.” Planning, Research, and Evaluation Division TXE/2010 Memorandum Series: CM-GES-S-02-R2 dated July 24, 2001.

Raglin, D.A., and Krejsa, E.A. (2000) “ESCAP II: Evaluation Results for Changes in Mover and Residence Status in the A.C.E.,” ESCAP Report No. 16. U.S. Census Bureau.