

## Chapter 4: A.C.E. Revision II Missing Data Methods

Michael Beaghen, 12/31/02

Missing data arises because it is not possible to obtain interviews for all sample cases or to obtain answers to all interview questions. This was as true for the A.C.E. Revision II as it was for the A.C.E. To put the A.C.E. Revision II missing data methods in perspective we will review briefly what missing data the A.C.E. adjusted for. For the A.C.E. P sample, we applied household non-interview adjustments to compensate for non-interviewed households and imputation methods to handle missing characteristics such as age or tenure. Further, where we were unable to assign a definitive residency status or match status, we assigned probabilities of match and probabilities of residency. There was no non-interview adjustment for the A.C.E. E sample, nor was there an imputation for missing characteristics as the census imputations were used. However, for E-sample cases with unresolved enumeration status we assigned probabilities of correct enumeration. See Ikeda & McGrath (2001) for details on the A.C.E. missing data methodology.

As will be discussed in Chapter 6, the A.C.E. Revision II estimation utilized both the original A.C.E. coding results on the full A.C.E. sample and the Revision II coding results on the smaller Revision II sample (please note that the A.C.E. Revision II subsample of the A.C.E. is referred to as the Revision II sample, and the new coding operation as the Revision II coding). The missing data adjustments for the A.C.E. E sample and P sample were unchanged from those used to produce A.C.E. estimates with the exception of the imputation for missing age. It was necessary to impute age again for the A.C.E. P sample because the A.C.E. Revision II post-strata had different age groupings.

The Revision II P sample used the same imputations for missing characteristics that the A.C.E. did, including the new age imputation. However, since Revision II measurement methodology had important differences from the A.C.E. measurement methods, it was necessary to develop new missing data methods. The A.C.E. Revision II missing data confronted three general types of new missing data problems:

- new non-interviewed households: Revision II P-sample households that were considered interviews in the A.C.E. were identified as non-interviews in the Revision II coding when it was determined that none of the P-sample people there were valid Census Day residents;
- Revision II E-sample and P-sample cases with unresolved match, enumeration, or residency status because of incomplete or ambiguous interview data from the PFU or the EFU;
- Revision II E-sample or P-sample cases with conflicting enumeration or residency status because contradictory information was collected in the PFU and the EFU interviews and it could not be determined which was valid.

#### **4.1 Age Imputation for the A.C.E. P Sample**

For the original A.C.E., P-sample people with missing age were assigned to age categories as defined by the post-stratification plan. The A.C.E. Revision II P-sample post-stratification divided the original A.C.E. post-stratification group of 0-17 years old into groups of 0-9 and 10-17 years old. Those people with missing age who had been assigned to the 0-17 group were reassigned to either the 0-9 or the 10-17 group. This reassignment assumed that the age distribution of people missing age was uniform within the 0-17 age grouping. Other people with unresolved age remained in the age group they had been originally assigned to.

#### **4.2 The Household Non-Interview Adjustment for the Revision II P Sample**

The A.C.E. household non-interview adjustment generally spread the weights of P-sample non-interviewed housing units over interviewed housing units in the same block cluster with the same housing unit structure type. Housing units were determined to be non-interviews in two ways; first, an interview was not conducted during the A.C.E. person interview; second, based on the results of the A.C.E. person followup it was determined that a whole household of P-sample people should not have been listed in the first place, and that another household may have been residents at that housing unit. Separate household non-interview adjustments were implemented for Census Day and A.C.E. Interview Day.

The A.C.E. Revision II non-interview adjustment methodology for A.C.E. Interview Day was essentially unchanged from the A.C.E. There was, however, an important change from the A.C.E. methodology for the non-interview adjustment for Census Day residency. In A.C.E. Revision II we defined a cell for new non-interviews due to whole households of A.C.E. non-movers who were determined to be in-movers or non-resident out-movers by the Revision II coding. The new non-interview cell spread the weights of these non-interviewed units over housing units with at least one person who indicated he/she lived at another address or who was identified as potentially fictitious in the A.C.E. We assumed these new non-interviews would have both a low match rate and a low residency rate similar to this group. Otherwise the non-interview adjustment for Census Day used methodology similar to that of the A.C.E.

#### **4.3 Revision II E-Sample and P-Sample Assignment of Probabilities of Correct Enumeration, Census Day Residency and Match Status**

In the A.C.E., P-sample people with unresolved Census Day residency or match status occurred in one of two ways. First, the A.C.E. person interview (PI) may not have provided sufficient information for match and followup. Second, the A.C.E. person followup (PFU) may not have collected adequate information to allow us to determine a person's Census Day residency status or their match status. A.C.E. E-sample people with unresolved enumeration status likewise arose in this second manner. In the A.C.E. Revision II unresolved cases also arose because of the EFU. How we imputed for a case depended on how it came to be unresolved.

#### *4.3.1 Imputation for Revision II P-Sample People with Insufficient Information for Match and Followup*

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

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

The probability of match was assigned based on the overall match rate, divided into groups based on mover status and housing unit match status as was done in the A.C.E., and additionally on conflicting household status.

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

In contrast to P-sample people with insufficient information, the residency status for Revision II P-sample people and the correct enumeration status for Revision E-sample people often changed from the A.C.E. to the Revision coding. These statuses changed because the Revision coding processed not just the original information from the PFU, but also the new information from the EFU. Thus while the EFU information resolved many cases that were unresolved in the A.C.E. on account of the PFU, EFU cases with incomplete or ambiguous information were a new source of unresolved cases in the Revision II coding. There were about the same number of E-sample unresolved cases in the Revision II as in the A.C.E., more than six million, with about half of these representing new unresolved cases. In contrast, the Revision II coding generated substantially more P-sample unresolved cases than the A.C.E., 4.6 million versus 2.7 million. We saw this increase because all the Revision II P-sample except those with insufficient information went to EFU, including whole households of non-matched people who had not gone to PFU. These people were assumed in the A.C.E. to be resolved and could have become unresolved because of the EFU.

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

residency. In the A.C.E. Revision II we abandoned the before followup imputation cells and defined our cells based on after followup information. This change was the single most important improvement in the A.C.E. Revision II missing data methodology.

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

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

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

It is useful to make a distinction between what we call uninformative and informative unresolved cases:

- uninformative unresolved; the followup was a non-interview or an incomplete interview, though there was no evidence of an erroneous enumeration or non-resident.
- informative unresolved; a followup interview was conducted and there was evidence of an erroneous enumeration or non-resident.

Note that when one interview was uninformative unresolved, but the other interview was resolved, the Revision II coding chose (i.e., the code was based on) the resolved interview. On the other hand, when the unresolved interview was informative, the Revision II coding could choose the unresolved interview over the resolved one. See Adams & Krejsa (2002) for details of the Revision II coding.

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

At this point it may help to give an example of an after followup group. One cell of unresolved E-sample people or recipients was defined as people with evidence from the EFU interview that they had moved in since Census Day, or moved out before Census Day, though the EFU interview did not provide the address they moved to or from. We could not determine the enumeration status of these people since we did not know whether the Census Day address was

in the A.C.E. cluster. The corresponding donor pool consisted of those resolved people who indicated in the followup that they moved in after Census Day, or moved out before Census Day; these were generally people who provided the mover address in the EFU. We had an analogous after followup group for people unresolved because of they indicated they were movers on the person followup interview. We characterize these groups as informative because the followup provided evidence of an erroneous enumeration.

**Table 1** shows the nine EFU after followup groups, and **Table 2** shows the nine PFU after followup groups. People who moved in after Census Day or moved out before Census Day were the largest informative after followup group. Another important informative after followup group consisted of people who, according to the followup, had another residence such as a vacation home, though the followup did not indicate whether the other residence or the sample address was the Census Day residency. The non-interview groups and ‘didn’t answer other residence questions’ group were the larger uninformative groups. In all, there were nine PFU after followup groups and nine EFU after followup groups.

Some of the larger EFU groups were subdivided by A.C.E. operational variables such as whether or not the household went to PFU, or whether the household was conflicting. The uninformative after followup groups tended to have imputed probabilities of correct enumeration or residency

**Table 1: EFU After Followup Groups**

Informative Groups
The followed up person ‘Lived elsewhere’ or at an ‘other residence’, but the address was not given
Followed up person moved in after Census Day or out before Census Day, but Census Day address not given
Respondent indicated the followed up person ‘Never lived here’ at the sample address, but did not provide the Census Day address
The followed up person had an ‘Other residence’, but did not indicate whether sample address or the other residence was the Census Day residency
Followed up person moved in or moved out, but no move dates given
Uninformative Groups
The respondent indicated the followed up person ‘Lived here’ at the sample residence, but did not answer the other residence question
Noninterview (1); the respondent answered the current residence question, but did not answer the group quarters and other residence question
Noninterview (2); the respondent did not answer the usual residence question, nor the group quarters and other residence questions
Potentially fictitious person, no respondents knew of the followed up person

close to one, typically in the range of 0.92 to 0.99, whereas the informative after followup groups had lower probabilities, often less than 0.25. The probability of correct enumeration or residency is calculated as the weighted proportion of correct enumerations or Census Day residents in donor pool. For example,

$$\text{Prob. of correct enum.} = \frac{\text{Weighted CE's in Donor Pool}}{\text{Weighted Resolved Enumerations in Donor Pool}}$$

For the P-sample, probabilities of residency and match status were calculated analogously.

**Table 2: PFU After Followup Groups**

Informative Groups
The followed up person ‘Lived elsewhere’ or at an ‘other residence’, but the address was not given
Followed up person moved in after Census Day or out before Census Day, but Census Day address was not given
Noninterview (2); the respondent indicated the followup person ‘did not live here’ at the sample address, but did not indicate the other address and did not answer the group quarters and other residence questions
The followed up person had an ‘Other residence’, but did not indicate where the usual residence was
Uninformative Groups
The respondent indicated the followed up person ‘Lived here’ at the sample residence, but did not answer the other residence question
Noninterview (1); the respondent answered the usual residence question, but did not answer the group quarters and other residence questions
Noninterview (3); the ‘lived here’ question is DK/refused, and the group quarters and other residence questions were not answered
Noninterview (4); blank questionnaire
Potentially fictitious person, no respondents knew of the followed up person

#### 4.4 Imputation for Revision II E-Sample and P-Sample Conflicting Coding Cases

When the A.C.E. person followup (PFU) and the evaluation followup (EFU) interviews had contradictory information and we could not determine which was correct, the Revision II coding assigned the case a code of conflicting (conflicting coding is not to be confused with conflicting households, which was described in *Section 4.3.1*). All cases found to be conflicting in the Revision II automated coding were sent to analysts for clerical review. By examining the handwritten notes of interviewers, the analysts could often determine which of the interviews was the better and appropriately assign a code. There were some cases where the interviews appeared to be of equal quality, such as when both respondents were household members or both

respondents were of equal caliber proxy. For these conflicting cases, the interviews seemed equally likely to be correct based on the expertise of the analysts. Therefore, probabilities of 0.5 were assigned both for correct enumeration status of Revision II E-sample conflicting cases and for Census Day residency status of Revision II P-sample conflicting cases. It should be noted that the recoding of the Revision II samples resulted in considerably less conflicting cases than the PFU/EFU Review sample. The PFU/EFU Review sample had about 2.6 million weighted people (Adams & Krejsa, 2001) in contrast to only about 100,000 weighted people in the Revision II samples.

#### **4.5 References**

Adams, T. & Krejsa, E. (2001): '*ESCAP II: Results of the Person Followup and Evaluation Followup Forms Review*,' ESCAP Report No. 24, U.S. Census Bureau

Adams, T. & Krejsa, E. (2002): '*A.C.E. Revision II Measurement Subgroup Documentation*,' DSSD A.C.E. Revision II Memorandum Series #PP-6

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

Ikeda, M., 2001: '*Accuracy and Coverage Evaluation Survey: Some Notes Related to Accuracy and Coverage Evaluation Missing Data Procedures*,' DSSD Census 2000 Procedures and Operations Memorandum Series Chapter Q-77

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