

**CONFERENCE OF EUROPEAN STATISTICIANS**

**UN/ECE Work Session on Statistical Data Editing**  
(27 – 29 May 2002, Helsinki, Finland)

Topic (i): Planning and management of statistical data editing

**ANALYSIS OF THE COVERAGE EDIT FOLLOW-UP OPERATION FROM CENSUS 2000**

**Contributed Paper**

Submitted by the United States Bureau of the Census<sup>1</sup>

**Abstract:** For Census 2000, the United States Bureau of the Census reviewed the data from self-administered questionnaires. The review consisted of checks or edits to ensure that the respondent provided consistent data covering all household members. This operation was called Coverage Edit Follow-up. The Coverage Edit Follow-up program included two types of edit failures. The first edit was a consistency check on the number of persons in the housing unit; called “count discrepancy”. The second edit was to obtain person data for households that reported there were more persons than there was room for on the questionnaire; called “large household follow-up”. The Coverage Edit Follow-up operation occurred from May 8 to August 13, 2000. The Coverage Edit questionnaire was a computer assisted telephone interview (CATI) developed in HTML. A series of coverage probes were developed to ensure that each household roster was correct before collecting person data for any additional persons. This paper will discuss the design of coverage probes as well as results from the operation.

**I. INTRODUCTION**

1. For Census 2000, the United States Bureau of the Census collected data from both self-administered and enumerator administered questionnaires. Self administered returns were either paper questionnaires returned through the mail or questionnaires completed over the Internet. The self administered questionnaires were reviewed to ensure that the respondent provided consistent data covering all household members. The review consisted of an automated algorithm designed to improve the coverage of persons within housing units. This operation was called Coverage Edit Follow-up.
2. The Coverage Edit Follow-up operation was designed to improve within household coverage and data quality. Within household coverage improvement was accomplished through a series of nine probes during the follow-up interview. The coverage probes were developed to ensure that each household roster was correct before collecting demographic data for any additional persons. Improvement in the quality of the data resulted by reducing the number of persons requiring imputation of their demographic characteristics.
3. The Coverage Edit Follow-up program included two types of edit failures. The first edit was a consistency check on the number of persons in the housing unit; called “Count Discrepancy”. Count Discrepancy cases resulted when the respondent reported household size differed from the number of persons for whom the respondent provided data. The discrepancy could have been the result of the respondent not fully understanding who to include on their questionnaire, i.e., the residence rules. Count Discrepancy edit was performed to identify cases with potential within household coverage error.

---

<sup>1</sup> Prepared by James B. Treat (james.b.treat@census.gov).

4. The second edit was called “Large Household”. Large Household cases resulted when the respondent reported a household size larger than there were room for on the questionnaire; i.e., more than six persons. The demographic characteristics for the additional persons were collected; thus, improving the quality of the data by reducing the amount of imputation.

5. The Coverage Edit Follow-up operation occurred from May 8 to August 13, 2000. The Coverage Edit Follow-up operation consisted of a computer assisted telephone interview (CATI). The questionnaire instrument was developed in HTML.

## **II. DESIGN OF THE QUESTIONNAIRE**

6. The Coverage Edit Follow-up questionnaire was designed with several requirements. First, the questionnaire wording needed to coincide with the residence rules. Second, the questionnaire need to address the major issues related to within household coverage error. Finally, the length of the interview needed to be considered. Longer interviews result in increased respondent burden which in turn leads to a higher respondent refusal rate.

7. The Bureau of the Census conducted cognitive testing on a paper version of the CATI questionnaire. In addition, agent field trails were conducted using experienced computer assisted telephone interviewers on a CATI version of the questionnaire. From both tests comments were incorporated into the final questionnaire design. The Coverage Edit Follow-up interview contained the following four sections:

- Establishing that the correct household had been reached;
- Reviewing the roster with the respondent and providing the respondent an initial opportunity to correct the roster;
- Asking a series of nine coverage probes to establish the correct household roster;
- Collecting the demographic characteristics for any persons:
  - added to the household roster during the roster review;
  - added to the household roster during the coverage probes ;
  - originally listed on the form for whom there was no room on the form to report.

8. The nine coverage probes were designed around the residence rules. Based on the respondent’s answers to the probes, persons were either added to or deleted from the household roster. Five of the nine coverage probes were designed to identify and add persons incorrectly left off the questionnaire. The remaining four coverage probes were designed to identify and remove persons erroneously included on the questionnaire.

9. We believed the respondent would better understand the probes if given a consistent structure and use of examples. For example, the wording of the all the add probes as well as two of the delete probes referenced “around the beginning of April” and not specifically April 1, 2000 (Census Day). Most of the nine probes provided examples to the respondent. The five add probes began with “Other than those persons you included,...” and the four delete probes began with “Were any of these people...”.

10. The add probes targeted children, relatives, non-relatives, persons away temporarily or moving, and persons with no other permanent place to stay. The delete probes targeted persons living away at college, in the military, in an institution, and with multiple addresses. At the end of the paper is the exact wording of each of the nine probes.

11. There were three additional features of the Coverage Edit Follow-up questionnaire. First, at any point during the interview, if the respondent identified persons that should be added to or deleted from the household roster the interviewer had the ability to add or delete the person(s). This feature was referred to as the “Interrupt Screen”. Second, if the respondent indicated to the interviewer that a person was on the household

roster multiple times, the interviewer could identify and flag the duplicated person(s). The identification of duplicate persons occurred either during the review of the household roster or through the Interrupt Screen. Duplicated persons were removed from the final household roster. Finally, the interviewer had the ability to edit the names on the household roster based on respondent input. Note that these name edits were used only to aid the agent during the interview - these name corrections were not passed along to Census 2000 processing.

### III. RESULTS

12. For this program, 82,008,049 forms (paper and Internet) were checked for either being a Count Discrepancy or Large Household. Of the 82 million forms, 2,544,072 forms required follow-up. This represented 3.1 percent of the eligible universe. Count Discrepancy cases were 1,148,449 (45.1%) of the Coverage Edit Follow-up workload. The remaining 1,395,623 cases were Large Households; representing 54.9% percent of the follow-up workload.

13. Since this program was a telephone operation with no personal visit component, valid telephone numbers were a requirement. The Census 2000 paper and Internet questionnaires included a question asking for the respondent's telephone number. In anticipation of errors in the data capture process or the respondent not providing a valid telephone number, the Census Bureau instituted an automated telephone lookup operation. Missing or inconsistent telephone numbers were checked prior to making contact with the household. If contact was made using the respondent provided telephone number and the interviewer determined they had not contacted the correct household, the case was recycled into the lookup operation. The coverage edit follow-up program allowed for 10 call attempts to make initial contact with the household. Once contact was made, an additional 10 callback attempts were allowed to complete the interview. Once contact was made with the household, the vast majority of the time the interview did not require a callback to completed the interview.

14. Table 1 contains a distribution of the 2.3 million cases attempted and requiring follow-up by the outcome of the interview. There were four possible interview outcomes: Complete, Refusal, No valid telephone number and Non-interview.

15. The Complete interviews were largest of the four categories; representing 53.5 percent of the workload. In cases that were completed, respondents minimally answered all of the coverage probe questions, ensuring the household roster was correct. The next largest outcome category was the "No valid telephone number". For nearly a quarter of the workload, we were unable to obtain a valid telephone number even with a lookup operation. The Refusal category was 8.6 percent of the workload. Finally, the Non-interview cases were 13.8 percent of the workload. This category includes the following types of cases:

- No contact with a household member after 10 call attempts
- Contact was made with a household member, callback was requested by the respondent but no contact was made with a household member after 10 callback attempts
- Unable to contact an eligible respondent
- No contact with a household member after less than 10 call attempts
- Contact was made with a household member, callback was requested by the respondent but no contact was made with a household member after less than 10 callback attempts
- Case was never attempted

16. The last three non-interview categories are the result of ending the Coverage Edit Follow-up operation prior to completing all of the work. The operation was planned to started on April 5, 2000 and end on June 19, 2000. Due to resource issues the program started late and end late; May 8 to August 13, 2000. We needed to stop the program on August 13, 2000 in order to begin subsequent processing in order to meet the December 31, 2000 Congressionally mandated deadline for delivering the population counts to the President.

**Table 1: Coverage Edit Follow-up Interview Outcome**

<b>Interview Outcome</b>	<b>Coverage Edit Follow-up Workload</b>	
	<b>Number</b>	<b>Percent</b>
Total	2,338,420	100.0%
Completed	1,251,971	53.5%
Refusal	201,385	8.6%
No valid phone number	562,049	24.0%
Non-interviews	323,015	13.8%

17. As discussed earlier, there were five coverage probes designed to identify and add persons inadvertently left off the questionnaire. In addition, the Interrupt Screen provided the opportunity to also add persons inadvertently left off the questionnaire. Table 2 contains the distribution of added persons by the coverage probe, including the Interrupt Screen. There were 152,149 persons added to the Census from the Coverage Edit Follow-up interviews. The majority of the added persons were the results of the Interrupt Screen, 54.9 percent. More than likely, the Interrupt Screen feature was used when the interviewer reviewed the household roster with the respondent; since the review of the household roster occurred prior to the coverage probes section of the interview. The coverage probe designed to add children was the next largest category (21.9 percent of the added persons) followed by the coverage probe designed to add other relatives (aunt, uncles, grandparents, cousins, etc.). The Non-relative coverage probe represented 6.9 percent of the adds. Finally, the coverage probes designed to add persons temporarily away/moving or persons with no permanent residence each represented less than 3.0 percent of the adds. From Table 2 we see a declining improvement/ payoff starting with the Interrupt Screen and moving through the coverage probes. This could be a function of the ordering of the probes and household roster review. If the coverage probes were in a different order we might observed the same overall coverage improvement with a different distribution by probe.

**Table 2: Persons added by the coverage probe**

<b>Coverage Probes</b>	<b>Added Persons</b>	
	<b>Number</b>	<b>Percent</b>
Total number of added persons	152,149	100.0%
Children	33,246	21.9%
Relatives	17,088	11.2%
Non-relatives	10,442	6.9%
Temporarily away or moving	3,440	2.3%
No permanent residence	4,436	2.9%
Interrupt Screen	83,497	54.9%

18. As discussed earlier, there were four coverage probes designed to identify and remove persons erroneously included on the questionnaire. In addition, the Interrupt Screen provided the opportunity to also delete persons. Table 3 contains the distribution of deleted persons by the coverage probe, including the Interrupt Screen. There were 206,913 persons deleted from the Census as the result of the Coverage Edit Follow-up operation. The majority of the deleted persons were the results of the Interrupt Screen, 79.4 percent. As with the added person, more than likely the Interrupt Screen feature was used when the interviewer reviewed the household roster with the respondent. As stated above, the review of the household roster occurred prior to the coverage probes section of the interview. The coverage probe designed to deleted students who live away at college was the next largest category representing 9.2 percent of the deletes. The coverage probe designed to identify persons with multiple residence represented 7.9 percent of the deleted. Finally, the coverage probes designed to delete persons living away while in the military or persons in institutions represented 1.0 percent and 2.5 percent of the deletes, respectively. Note that the order effect observed with the add probes is not as pronounced with the delete probes.

**Table 3: Persons deleted from the coverage probe**

Coverage Probes	Deleted Persons	
	Number	Percent
Total number of deleted persons	206,913	100.0%
College students	19,103	9.2%
Military	2,022	1.0%
Institution	5,165	2.5%
Multiple residence	16,255	7.9%
Interrupt Screen	164,368	79.4%

19. Finally, the interview provided the feature to identify persons that are duplicated on the household roster. As discussed earlier, the identification of duplicate persons occurred either during the review of the household roster or through the Interrupt Screen; interview location. Table 4 contains the distribution of duplicated persons by the location in the interview where they were identified. There were 50,654 persons identified as duplicates; thus, they were deleted from the Census. The majority of the duplicated persons were through the Interrupt Screen, 77.8 percent. The remaining 22.2 percent were identified during the review of the roster.

**Table 4: Persons identified as duplicates and deleted**

Interview Location	Duplicated Persons	
	Number	Percent
Total number of duplicated persons	50,654	100.0%
During roster review	11,235	22.2%
Interrupt Screen	39,419	77.8%

20. Table 5 provides the overall coverage improvement from the Coverage Edit Follow-up operation. A total of 409,716 persons were impacted by this operation. The majority of the actions resulted in the person being removed from the Census; 50.5 percent were deletes and 12.4 percent were duplicates that were also deleted. The net coverage improvement from this operation reduced the Census 2000 population count by 105,418 persons.

**Table 5: Overall coverage improvement**

<b>Coverage Probes</b>	<b>Persons</b>	
	<b>Number</b>	<b>Percent</b>
Adds	152,149	37.1%
Deletes	206,913	50.5%
Duplicates	50,654	12.4%
Total	409,716	100.0%
Net coverage gain	(105,418)	

#### **IV. CONCLUSIONS**

21. The Coverage Edit Follow-up program improved the quality of Census 2000 by adding 152,149 persons that would have been missed from the Census. In addition, the program removed 257,567 persons that would have been erroneously included in the Census. However, only 53.5 percent of the workload was completed. For the remaining 1,086,449 cases (46.5 percent), we were unable to obtain a completed interview. For nearly a quarter of the overall workload (or 51.7 percent of the non-completed cases) we were unable to obtain a valid telephone number. Therefore, we should consider a personal visit component to this operation. The personal visit follow-up could also handle the refusal cases which represented 8.6 percent of the overall workload.

#### **V. NINE COVERAGE PROBES**

22. Below are the nine coverage probes from the Coverage Edit Follow-up questionnaire. The first five probes (sections B through F) were designed to identify and add persons incorrectly left off the questionnaire. The last four probes (sections H through K) were designed to identify and remove persons erroneously included on the questionnaire. The probes are listed in the order in which they appeared on the questionnaire. In addition, the wording which preceded the five add probes and four delete probes are also included (sections A and G).

##### **A. Introduction to the add probes**

23. We've found that sometimes people were left off the census form, especially people who stayed somewhere temporarily or were new to the household, among other reasons. I am going to ask you some questions just to make sure that no one was missed.

##### **B. Question 1 - Children**

24. Other than those persons you included, were there any children who were living or staying there around the beginning of April? Be sure to consider any newborns, foster children, step children, or children in shared custody arrangements.

##### **C. Question 2 - Relatives**

25. Other than those you included, were there any relatives, such as aunts, uncles, grandparents, cousins, or any other kinds of relatives who were living or staying there around the beginning of April?

##### **D. Question 3 - Non-relatives**

26. Other than those you included, were there any other persons not related to you who were living or staying there around the beginning of April? For example, someone who rents a room from you or a friend staying with you temporarily while looking for a place to live.

**E. Question 4 - Persons away temporarily or moving**

27. Other than those you included, were there any persons who were either away temporarily or moving around the beginning of April? For example, a household member who was visiting with friends or relatives, on vacation, on a business trip, or in the process of moving.

**F. Question 5 - Persons with no other permanent place to stay**

28. Think back to the beginning of April. Were there any people staying there who had no other permanent place to stay? Please tell me their names even if you do not consider them to be regular members of your household.

**G. Introduction to the delete probes**

29. OK. Now I have some questions about people who were included on your form and might have been counted elsewhere. The census counts people in places where groups of people stay such as colleges, nursing homes, jails and the like. For these next questions, please think about the people who were included on your form.

**H. Question 6 - Persons living away at college**

30. Were any of these people college students living or staying away around the beginning of April?

**I. Question 7 - Persons in the military**

31. Were any of these people members of the U.S. Armed Forces living or staying away around the beginning of April?

**J. Question 8 - Persons in an institution**

32. Were any of these people living away in a place such as a nursing home, prison or jail, mental hospital, or similar place on April 1st?

**K. Question 9 - Persons with multiple addresses**

33. Some people have more than one place to live. Examples include a second residence where they stay to be closer to work, a friend's or relative's home, or a vacation home. Did any of the people on the list I read you earlier have another place where they live or stay?

**VI. ACKNOWLEDGMENTS**

I would like to thank David Sheppard of the United States Bureau of the Census for his helpful suggestions.