

UNRESOLVED STATUS IN THE ACCURACY AND COVERAGE EVALUATION

Xijian Liu, John Jones, Roxanne Feldpausch, US Bureau of the Census¹
 Xijian Liu, US Bureau of the Census, 4700 Silver Hill Road, Washington, DC 20233

Keywords: Accuracy and Coverage Evaluation, Matching, Unresolved Status

1. Introduction

The 2000 Accuracy and Coverage Evaluation (A.C.E.) measured the coverage of Census 2000. Housing units within the sample clusters were listed and matched to the January 2000 version of the Decennial Master Address File. After reconciling the nonmatches, person interviews were conducted in P-sample housing units, A.C.E. housing units that were confirmed to have existed within the block clusters. P-sample persons were Census Day residents of P-sample housing units. E-sample persons were a sample of the persons enumerated in the census. The P sample and the E sample were geographically overlapped in the block clusters. During A.C.E. person matching, P-sample persons were matched against census persons, and match codes were assigned to P-sample persons and census persons. Certain P-sample and E-sample persons needing additional information to determine their match status, residence status or enumeration status were sent for person follow-up interviewing. Information obtained from follow-up interviews was used in the after follow-up coding; and a final match code was assigned. Confirmed nonresidents were removed from the P sample. If there was not enough information to determine the person's match status, residence status, or enumeration status, the person was coded unresolved status.

This paper assesses the quality of A.C.E. data by analyzing the unresolved status. We answer the questions of what percent of the A.C.E. persons had unresolved status and what were the causes of these unresolved cases. In section 2, we investigate unresolved status in the overall P-sample data and the overall E-sample data that were used for estimation. In section 3, we investigate the unresolved status among persons in follow-up interviewing. In section 4, logistic regression is performed to analyze the association of selected variables with the unresolved status. Conclusions are stated in section 5. All results throughout this paper were weighted using weights that reflect the

probability of selection in all stages of sampling. P-sample weights also reflect a noninterview adjustment. Our analysis excluded Puerto Rico. Factors not considered in this analysis include data keying errors, matching errors, etc.

2. Unresolved Status in the P-sample and in the E-sample

In the 2000 A.C.E., there was a small amount of P-sample persons having unresolved residence status or unresolved match status and E-sample persons having unresolved enumeration status. During the missing data operation, whole household insufficient information for matching and follow-up coded by computer were converted to noninterviews. Other persons with unresolved status had their probability of residence, probability of match or probability of correct enumeration imputed. In Table 1 below, unresolved (residence) status in the P sample in 2000 includes 1.2% of the P sample having both residence and match status unresolved. There were only unresolved match status in the P sample in the 1990 PES. More details on the missing data procedure for the 2000 A.C.E. can be found in Cantwell et al (2001).

Table 1. Percent Unresolved Status

	2000 A.C.E.	1990 PES
P sample	2.2	1.8
E sample	2.6	1.3

1990 figures are from Cantwell et al (2001) and Childers (2001)

More than one half of the P-sample people with unresolved residence status had insufficient information for matching and follow-up, i.e., these P-sample people did not have a complete name and at least two characteristics. Table 2 gives the percent unresolved status in the P sample by match codes for the 2000 A.C.E.. Possible matches and insufficient information for matching and follow-up had both residence status and match status unresolved.

¹ Xijian Liu, John Jones, and Roxanne Feldpausch are mathematical statisticians in the Decennial Statistical Studies Division of the U.S. Census Bureau. This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

Table 2. Unresolved Status (Percent of P sample)

Match code	Total	Imputed as		
		Residnt, match	Residnt, non- match	Non- residnt
Matched, unresolved residence status	0.164	0.149	---	0.015
Not matched, unresolved residence status	0.892	---	0.582	0.310
Possibly matched	0.018	0.015	0.002	0.001
Insuff. infor. for matching and follow-up	1.175	0.836	0.156	0.183
All unresolved codes	2.249	1.000	0.740	0.509

Almost ninety percent of the unresolved enumeration status in the E sample were nonmatches. Table 3 gives the percent unresolved status in the E sample by match codes for the 2000 A.C.E. An unresolved geography code was assigned to incomplete geography cases in the targeted extended search. Other unresolved geography cases were assigned a code of not matched with unresolved enumeration status.

Table 3. Unresolved Status (Percent of E sample)

Match code	Total	Imputed as	
		Correct enumer.	Erron. enumer.
Matched, unresolved enumeration status	0.135	0.128	0.007
Not matched, unresolved enumeration status	2.329	1.723	0.606
Possibly matched	0.015	0.013	0.002
Unresolved geography	0.126	0.116	0.010
All unresolved codes	2.605	1.980	0.625

Proxy interviews were interviews with nonhousehold members such as neighbors, apartment managers, and real estate agents. Outmovers were residents at the sample address on Census Day who were not residents at the sample address on the A.C.E. interview day. Information on whole household outmovers was collected from proxy interviews. Table 4 gives the proxy interview rate by mover status. Only 4.5% of the P-sample were collected from proxy interviews.

Table 4. Proxy Interview Rates by Mover Status

Mover status	Percent Proxy Interviews
Nonmovers	3.1
Outmovers	75.0
All P-sample	4.5

Proxy interviews were a major source of unresolved status as well as insufficient information for matching and follow-up in the P sample. Almost one half of the unresolved residence status came from proxy interviews, although people whose interview was a proxy interview consisted only 4.5% of the P sample. Outmovers were a major source of proxy interviews. Whole household outmovers were all from proxy interviews. Table 5 gives the unresolved residence status in the P sample by mover status and respondent type. Column (a) shows the percent unresolved, which is the number of unresolved cases in the category divided by the P-sample total in the same category. Column (b) shows the percent of unresolved, which is the number of unresolved cases in the category divided by the number of all unresolved cases in the P sample.

Table 5. Unresolved Status by Mover Status and Respondent type

Mover status and respondent type	P-sample percent unresolved (a)	Percent of unresolved (b)
Nonmover, non-proxy	1.2	47.9
Nonmover, proxy	17.3	23.0
Outmover, non-proxy	9.2	3.6
Outmover, proxy	22.5	25.4
All P-sample	2.2	100.0

Similar to Table 5, Table 6 gives the insufficient information for matching and follow-up in the P sample by mover status and respondent type.

Table 6. Insufficient Information for Matching and follow-up by Mover Status and Respondent type

Mover status and respondent Type	P-sample percent insufficient information (a)	Percent of insufficient information (b)
Nonmover, non-proxy	0.5	41.3
Nonmover, proxy	11.1	28.3
Outmover, non-proxy	8.9	15.7
Outmover, proxy	11.8	14.7
All P-sample	1.2	100.0

3. Unresolved Status in the Person Follow-Up.

After the before follow-up person matching, certain cases were sent to follow-up to gather additional information to accurately code the residence status of a P-sample person or the enumeration status of an E-sample person. They were:

- P-sample partial household nonmatches,
- Conflicting households (address matched, but A.C.E. and census got different households of people),
- Possible matches,
- P-sample whole household nonmatches and the A.C.E. person interview was with a proxy respondent,
- P-sample matches and nonmatches with unresolved residence status,
- E-sample nonmatches,
- Those needing housing unit geography verification, and others that would benefit from follow-up.

Table 7 shows the percent unresolved in follow-up by various follow-up categories. Follow-up category in the first column is defined by before follow-up person match status and preliminary housing unit matching status. Throughout this section, percent unresolved for P-sample people is the number of unresolved residence status divided by the total number of P-sample people in the follow-up (confirmed resident, confirmed nonresident and unresolved residence status), and the percent unresolved for E-sample people is the number of unresolved enumeration status divided by the total number of E-sample people in the follow-up (confirmed correct enumeration, confirmed erroneous enumeration, and unresolved enumeration status).

Table 7. Percent Unresolved in Follow-up

Follow-up category	P sample	E sample
Partial household non-match	12.3	13.1
Conflicting household	27.2	28.0
Matches sent to follow-up	15.8	14.6
Possible match	6.8	6.3
Whole household nonmatch, address matched	22.2	18.8
Whole household nonmatch, address did not match or no HU matching	10.9	10.7
All in follow-up	15.6	14.6

Breaking down the unresolved status by match codes, an overwhelming majority of the unresolved status in the follow-up were coded as nonmatch with unresolved status. Table 8 gives the percent of follow-up persons by unresolved match code.

Table 8. Percent of Follow-Up Persons by Unresolved Match Code

Match code (unresolved)	P sample	E sample
Matched, unresolved residence status	2.3	0.8
Not matched, unresolved residence status	12.7	13.0
Possibly matched	0.2	0.1
Insufficient information for matching and follow-up	0.4	---
Unresolved geography	---	0.7
All unresolved codes	15.6	14.6

Proxy interviews which consisted 21.1% and 27.4% of the P sample and E sample people in the follow-up were a major source of unresolved status in the follow-up. More than one half of the unresolved status in follow-up came from proxy follow-up interviews. Table 9 and Table 10 give the unresolved status by follow-up interview outcomes.

The results in Table 8 and Table 9 indicate that a majority of the unresolved status in the follow-up came from follow-up interviews which collected partial information. This information was used in the status imputation.

Table 9. P-sample Unresolved Status by Follow-up Interview Outcome

Interview outcome	P-sample percent unresolved	Percent of unresolved	P-sample percent of follow-up
Complete, nonproxy	5.7	25.1	68.0
Complete, proxy	31.1	52.3	26.2
Noninterview, or respondent type not clear	60.6	22.6	5.8
All in follow-up	15.6	100.0	100.0

Table 10. Census Unresolved Status by Follow-up Interview Outcome

Interview Outcome	E-sample Percent unresolved	Percent of unresolved	E-sample Percent of follow-up
Complete, Nonproxy	4.3	19.3	65.0
Complete, Proxy	32.7	61.5	27.4
Noninterview, or respondent type not clear	36.9	19.2	7.6
All in follow-up	14.6	100.0	100.0

Next, we analyze unresolved status using information on the follow-up form. The core Census Day residence questions on the follow-up form were:

- A Census day residence question that asked whether the person lived at the address on Census Day;
- A group quarters question that asked whether the person lived at places such as college dorms, nursing homes, etc. on Census Day;
- An other residence question that asked whether the person had another residence on Census Day.

There were four choices for each question: “Yes”, “No”, “Don’t know”, or “Refused”. The answer to a question could also be left blank. To better understand the reasons of having unresolved status in the follow-up interview, the responses to these residence questions were investigated. In Table 10, unresolved cases were classified into the following categories (see Childers and Liu (2001) for more details).

- *Lived elsewhere on Census Day, incomplete Census Day address:* The answer to the Census Day residence question was “No”. In this case, whether the person was a resident/correct enumeration was determined by

whether the address was inside the cluster (or in a surrounding block). If this address was invalid or blank, the person’s residence/enumeration status could be unresolved.

- *Group quarters or other residence unresolved.* Answer to the Census Day residence question was “Yes”. Answer to group quarters question or answer to other residence question was “Don’t Know”, or “Refused”.
- *Others.* This includes all other cases.

Table 11 shows that the top reasons for having unresolved status were the reluctance of a household member to give the Census Day address of the person and the inability of a proxy respondent to tell whether the person had another residence on Census Day.

Table 11. Reasons of Having Unresolved Status by Respondent Type in Follow-up

Category	A.C.E. person		Census person	
	Non-proxy	Proxy	Non-proxy	Proxy
Lived elsewhere	51.4	23.6	41.9	17.1
Group quarters or other residence unresolved	15.9	55.1	19.7	58.8
Others	22.7	21.3	38.4	24.1
Total	100.0	100.0	100.0	100.0

4. What variables are associated with unresolved status

Consider a population of N individuals. Let p_i be the probability of having unresolved status for the i th individual. Assume the probability of having unresolved status can be explained by a logistic model with k independent variables:

$$\log(p_i/(1 - p_i)) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}.$$

We used logistic regression to analyze the unresolved residence status in the P sample, and unresolved enumeration status in the E sample. We use the Taylor linearization method and SUDAAN to compute the Wald chi-square statistic.

For P-sample persons, we considered variables of respondent type in the person interview, mover status, tenure, race domain, age group, census region, and mode of person interview. The variable gender appeared to have a very weak effect on the model and was eliminated from the model. Proxy interview had major effects on having

unresolved status. Persons whose interview was with a proxy respondent are much more likely to have unresolved status than persons whose interview was with a household member. Age and mode of interview also showed moderate effect in the model. Other variables also showed statistical significance, although the effects were relatively weak. Table 12 shows the results of logistic regression of unresolved residence status in the P sample.

For E-sample persons, we considered variables of tenure, race domain, gender, age group, census region and census response method. Tenure had the strongest effect in the model. Response method and age group also showed moderate effects. Other variables only showed weak effects in the model. Table 13 shows the results of logistic regression of unresolved enumeration status in the E sample.

5. Conclusions

In the 2000 A.C.E., unresolved residence status consisted of 2.2% of the P-sample, unresolved match status consisted of 1.2% of the P-sample and unresolved enumeration status consisted of 2.6% of the E-sample.

Proxy respondents in person interviewing was a major source of P-sample unresolved status. Mode of interviewing and age group also had moderate effects on P-sample unresolved status. Tenure had the strongest effect on having E-sample unresolved enumeration status. Census response method and age group also had moderate effect having E-sample unresolved enumeration status.

Among the persons in follow-up, unresolved residence status consisted of 15.6% of the A.C.E. persons and 14.6% the census persons.

A majority of the unresolved status in the follow-up came from follow-up interviews that collected partial information. Follow-up interviews with proxy respondents were a major source of unresolved status in the follow-up.

The top reasons for having unresolved status in follow-up interviewing were the reluctance of a household member to give the Census Day address of the person and the inability of a proxy respondent to tell whether the person lived in group quarters or had another residence on Census Day.

Table 12. Logistic Regression for Unresolved Residence Status in the P Sample

Variable/Level	Wald Chi-square	d.f.	Odds Ratios	P-Value
Respondent type	2,548.14	1		0.0000
Proxy			12.27	0.0000
Nonproxy			1.00	---
Age group	262.70			0.0000
1-17			1.86	0.0000
18-29			1.75	0.0000
30-49			1.31	0.0000
50+			1.00	---
Mode of interview	97.65			0.0000
Telephone			0.59	0.0000
Person visit			1.00	---
Race domain	99.72			0.0000
Hispanic			1.25	0.0001
Non-Hispanic black			1.51	0.0000
Non-Hispanic Asian			1.43	0.0000
All other races			1.00	---
Mover status	66.09			0.0000
Outmover			1.54	0.0000
Nonmover			1.00	---
Region	57.70			0.0000
Northeast			0.71	0.0000
Midwest			0.72	0.0000
South			0.95	0.3392
West			1.00	---
Tenure	33.60			0.0000
Renter			1.25	0.0000
Owner			1.00	---

Table 13. Logistic Regression for Unresolved Enumeration Status in the E Sample

Variable/Level	Wald Chi-square	d.f.	Odds Ratios	P-Value
<i>Tenure</i>	380.89	1		0.0000
Renter			3.44	0.0000
Owner			1.00	---
<i>Response method</i>	200.14	2		0.0000
Enumerator, proxy			2.30	0.0000
Enumerator, non-proxy			1.64	0.0000
Mail			1.00	---
<i>Age group</i>	203.91	3		0.0000
0-17			1.11	0.0335
18-29			1.89	0.0000
30-49			1.25	0.0000
50+			1.00	---
<i>Gender</i>	67.82	1		0.0000
Male			1.15	0.0000
Female			1.00	---
<i>Region</i>	30.75	3		0.0000
Northeast			0.60	0.0000
Midwest			0.79	0.0175
South			0.99	0.8724
West			1.00	---
<i>Race Domain</i>	13.80	3		0.0032
Hispanic			1.14	0.0611
Non-Hispanic black			1.16	0.0132
Non-Hispanic Asian			1.27	0.0045
All other races			1.00	---

Childers, Danny R., and Liu, Xijian , Coding of Potential Fictitious and Lived Elsewhere on Census Day, Census 2000 Procedures and Operations (2001) Memorandum Series S-DT-07.

6. References

Cantwell, Patrick et al, Accuracy and Coverage Evaluation: Missing Data Results. Census 2000 Procedures and Operations Memorandum Series B-7 (2001).

Childers Danny R, 1990 E-sample Documentation. Census 2000 Procedures and Operations Memorandum Series