ADDRESS REPORTING ERROR IN THE 1990 POST-ENUMERATION SURVEY

Kirsten West, Mary Mulry, Randall Parmer and Joanne Petrik, U.S. Bureau of the Census*
Kirsten West, Census Bureau, Washington, D.C. 20233

1. BACKGROUND

The Census Bureau used dual system estimation to estimate the 1990 Census undercount. It is based on both the census and a sample of the population, the P-sample. Dual system estimation assumes that P-sample respondents can be linked, or matched, correctly to the census at their census day address. This study evaluates address reporting and the error in the number of people matching a census enumeration due to address misreporting.

Census Day was on April 1, 1990. The PES was conducted in July and August, 1991. Thus, some of the respondents have moved between the time the census was conducted and the PES was in the field collecting data. In spite of extensive probes on the PES Interview Questionnaire, respondents may fail to report that they moved in the interim. This type of error may cause the matching operation to search the census in an area other than where the respondent was enumerated and to assign a nonmatch status to respondents that were enumerated.

The data for this evaluation were collected in the Evaluation Followup (EFU). The geographical and demographic characteristics of the new mover cases were examined to detect if the effect was concentrated in any area. The data were analyzed by minority and nonminority evaluation post-stratum and processing office. Finally, the error in the dual system estimator (DSE) due to address misreporting was assessed.

The design of the EFU, including the sampling design, appears in Section 2. Section 3 contains the results of the study. The final section summarizes the results.

2. METHODOLOGY

A sample of PES block clusters were selected for evaluation analysis. The data were collected in the Evaluation Followup (EFU) operation. The data were then processed in a manner similar to the production matching operation and the results were subsequently weighted to represent the U.S. population. The sample selection, the weighting procedure, the data collection and the matching operation are described below.

2.1 Sample Selection

The PES Evaluation sample is a stratified systematic subsample of the PES sample block clusters. The PES sampling strata were grouped into thirteen evaluation sampling groups. To minimize the variance, the PES sampling strata were further grouped within the thirteen sampling groups, using size of weights as the criteria. Two to six subgroups were formed constituting the evaluation sampling strata. Large Indian reservations were combined to form an additional sampling stratum. Finally, the PES original sample small blocks defined

one more stratum. All together, the sample design included a total of fifty sampling strata.

The sample was allocated first to the sampling groups. A minimum sample size was determined for each sampling group, and the remainder was allocated to the largest sampling groups in an attempt to minimize the overall variance. Within each sampling group, the sample was allocated to each sampling stratum proportional to the number of people that were not computer matches within the stratum. The sample block clusters were then sorted and selected with equal probability within each of the fifty sampling strata. The resulting sample was the 920 block cluster evaluation sample.

Within each of these block clusters, all nonmatches and unresolved cases were to be sent to the field for interview. Matches from whole household matched households were subsampled within each cluster at a rate of 1-in-19. In addition, matches were included when they were in households with nonmatches.

2.2 Data Collection

The data collection for the Evaluation Followup took place in February, 1991. Approximately 11,000 households nationwide were visited by Census Bureau current survey interviewers or other experienced permanent Census Bureau employees. The interviewers were restricted to work in areas different from the areas where they worked during the PES. This was done to ensure independence in data collection between the production and the evaluation. The data were collected from thirteen regional field offices and census centers.

2.3 Matching Operation

The results from the EFU interviewing phase were used in a matching operation similar to the PES After Followup Matching operation. A team of matching experts reviewed the completed EFU questionnaires and assigned a match code to the persons selected for the EFU sample. This team of matching experts consisted of Matching Technicians (Techs) and Matching Review Specialists (MRS), the highest level and most trained of the matchers from the PES. The matching experts could not work on cases they had worked on in the production.

In addition to the team of matching experts for EFU, a team leader was installed in each of the processing offices. This team leader was a member of the permanent matching staff from the Jeffersonville Processing Office. The team leader concept was not used during the PES production matching phases.

The basic matching rules and guidelines for the EFU matching operation did not change from production PES. However, the matching experts were instructed to use more judgment than clerical matchers. The matching experts were also instructed to utilize all notes on the

EFU questionnaires. Often, the notes on the questionnaires could provide sufficient information to resolve the match status. If no new information useful in assigning a match code were collected during the EFU, the matchers were instructed to retain the production match codes.

3. RESULTS

The analysis results are presented next. First, the data are displayed by outcome of interview to evaluate the success of the data collection phase in locating knowledgeable respondents. Section 3.2 shows the number of new movers found by the Evaluation Followup. The match status at the new address is analyzed in section 3.3. The results are also displayed by minority and nonminority aggregations of evaluation post-strata (section 3.4) and by processing office (section 3.5) in order to evaluate whether the match status of the new movers were different in any given demographic or geographical area. Finally, section 3.6 discusses the impact of address misreporting on the estimate of undercount as calculated by the Census Bureau's dual system estimator (DSE).

3.1 Outcome of The Evaluation Followup

The EFU had an overall response rate of 98.65% - or a 1.35% noninterview rate. The low noninterview rate was accomplished in part because proxy respondents were allowed. However, even the number of proxy respondents was low. Almost 91% of the respondents were household members. Three percent of the respondents were neighbors. Apartment managers and landlords constituted close to 3%. The remaining group of respondents were classified as 'other' (unweighted data).

3.1.1 Type of Respondent in The Evaluation Followup (unweighted data)

Type of Respondent	%
Household Member	90.5
Neighbor	3.2
Apartment Manager	.9
Landlord	1.9
Other	3.6
Total	100.0
(N)	(15102)

3.2 Change in Mover Status

There were 334 respondents who changed from a nonmover in PES to a mover in EFU. This represents 2.2% of the P-sample in the EFU. Weighted to the PES unweighted totals, there were 2,416 new movers (s.e.= 510.41). This represents 0.69% of the sample. Similarly, weighted to the national level, there were 1,409,921 new movers (s.e. = 305,489.30). This represents 0.66% of the total population.

The 334 cases that changed from nonmover in production to mover in the evaluation are analyzed in

this evaluation. It should be noted that there were also 184 cases among the production movers that became nonmovers in the evaluation. These 184 cases represent 1,084 cases weighted to the PES P-sample unweighted totals and 474,158 cases weighted to the national level.

3.3 Match Status at New Address

The match statuses at the new address are shown in Table 3.3.1. Among the new movers, who previously matched at their sample address, 62% matched at their new address. On the other hand, among the cases that previously matched, 37% became unresolved. Among the nonmatches, 40% became matches, 12% remained nonmatches, 6% became out of scope, and 41% unresolved. Among the cases that were unresolved in production, 48% became matches, 5% out of scope, and 47% remained unresolved (Table 3.3.1).

3.4 Match Status at New Address by Minority and Nonminority Aggregation

The data were also grouped by minority and nonminority status. This was done by aggregating the evaluation post-stratum. There are 13 strata in the Evaluation Followup representing the following demographic and geographical groupings:

1. Northeast, Central City, Minority
2. Northeast, Central City, Nonminority
3. U.S., Noncentral City, Minority
4. Northeast, Noncentral City, Nonminority
5. South, Central City, Minority
6. South, Central City, Nonminority
7. South, Noncentral City, Nonminority
8. Midwest, Central City, Minority
9. Midwest, Central City, Nonminority
10. Midwest, Noncentral City Nonminority
11. West, Central City, Minority
12. West, Central City, Nonminority
13. West, Noncentral City, Nonminority + American Indian

The minority grouping includes evaluation post-strata 1, 3, 5, 8 and 11. The nonminority grouping includes evaluation post-strata 2, 4, 6, 7, 9, 10, 12 and 13.

Table 3.4.1 displays the match status of the new movers in the minority aggregation for data weighted to the national level. Overall, there were fewer matches in the evaluation than in the production and especially, there were less nonmatches in the evaluation, but many more unresolved - looking at the general distribution as shown in the margins of the table. Not shown by this grouping is the finding that the largest number of cases changing from a match to an unresolved match code was found in stratum 3 which represents noncentral minority cities in the United States. Nonmatches became unresolved in all minority strata except stratum 1 which represents the Northeast - central city area. On the other hand, among the new movers who were unresolved in the production PES, new matches were concentrated in stratum 11 - central cities in the West.

Table 3.4.2 presents the data for nonminorities and it shows a different picture. For the nonminority

Table 3.3.1 Changed From Nonmover in PES to Mover in EFU - Estimates and Coefficients of Variation. Weighted to the Total Population

Results of Evaluation	Results of Production			Total
	Match	Nonmatch	Unresolved	
Match	358,252 (68.04%)	301,649 (21.18%)	35,234 (56.01%)	695,134 (36.37%)
Nonmatch	0 (0.00%)	94,403 (37.86%)	0 (0.00%)	94,403 (37.86%)
Out of Scope	3,632 (68.03%)	48,226 (32.02%)	3,315 (91.17%)	55,173 (29.63%)
Unresolved	218,429 (59.83%)	312,530 (20.91%)	34,252 (52.94%)	565,211 (26.08%)
Total	580,313 (47.64%)	756,807 (16.77%)	72,800 (41.03%)	1,409,921 (21.67%)

Table 3.4.1 Changed From Nonmover in PES to Mover in EFU - Minority Status Estimates and Coefficient of Variation. Weighted to the Total Population

Results of Evaluation	Results of Production			Total
	Match	Nonmatch	Unresolved	
Match	13,650 (58.49%)	101,075 (44.70%)	3,238 (100.0%)	117,969 (38.99%)
Nonmatch	0 (0.00%)	27,349 (53.91%)	0 (0.00%)	27,349 (53.91%)
Out of Scope	3,632 (68.03%)	29,607 (38.94%)	308 (100.0%)	33,547 (35.16%)
Unresolved	176,984 (70.82%)	83,364 (26.64%)	5,073 (91.61%)	265,421 (47.99%)
Total	194,266 (64.55%)	241,395 (26.26%)	8,619 (65.73%)	442,286 (43.97%)

Table 3.4.2 Changed from a Nonmover in PES to Mover in EFU - Nonminority Status Estimates and Coefficient of Variation. Weighted to the Total Population

Results of Evaluation	Results of Production			Total
	Match	Nonmatch	Unresolved	
Match	344,602 (70.69%)	200,574 (22.53%)	31,997 (60.84)	577,173 (43.06%)
Nonmatch	0 (0.00%)	67,054 (48.55%)	0 (0.00%)	67,054 (43.08%)
Out of Scope	0 (0.00%)	18,618 (41.74%)	3,006 (100.0%)	21,624 (48.97%)
Unresolved	41,444 (79.51%)	229,165 (26.84%)	29,179 (60.07%)	299,788 (24.00%)
Total	386,046 (63.68%)	515,411 (16.90)	64,182 (45.69%)	965,639 (37.51%)

grouping, there were more matches in the evaluation than in the production, but also more unresolved cases and fewer nonmatches. Not shown by the aggregation is the finding that the largest number of cases changing from a match in production to a mover with a match at a new address came from strata 10 and 12 - nonminority strata in the West (central city) and the Midwest (noncentral city). It should be noted that the coefficients of variation were very high in these strata.

3.5 Match Status at New Address by Processing Office

The match status of the new movers were examined for each processing office. The results are presented in summary tables 3.5.1 through 3.5.3 for the three types of production match codes: match, nonmatch and unresolved.

Looking first at cases that were coded match in the production, it is the finding that in Jacksonville, Kansas City, San Diego and in particularly in Jeffersonville, many movers that matched in production, matched at their new evaluation followup address - but note the size of the coefficients of variation. There were no matches that changed to nonmatches. Finally, all offices except Jeffersonville, but in particular Austin, had cases changing from match to unresolved, but the coefficients of variation are high (Table 3.5.1).

Table 3.5.1 Changed From Nonmover in PES to Mover in EFU. Estimates and Coefficients of Variation. Weighted to the Total Population. By Processing Office and Match Status. Match in Production

Processing Office	Production versus Evaluation Followup		
	Match to Match	Match to Nonmatch	Match to Unresolved
Albany	0 (0.00%)	0 (0.00%)	55,632 (71.56%)
Austin	161,719 (99.43%)	0 (0.00%)	153,801 (80.85%0
Baltimore	0 (0.00%)	0 (0.00%)	607 (100.00%)
Jacksonville	5,892 (100.00%)	0 (0.00%)	7,390 (70.73%)
Jeffersonville	183,006 (100.00%)	0 (0.00%)	0 (0.00)
Kansas City	6,826 (77.74%)	0 (0.00%)	272 (100.00%)
San Diego	809 (100.00%)	0 (0.00%)	728 (100.00%)
Total	358,252 (68.04%)	0 (0.00%)	218,429 (59.83%)

Table 3.5.2 Changed From Nonmover in PES to Mover in EFU. Estimates and Coefficients of Variation. Weighted to the Total Population. By Processing Office and Match Status. Nonmatch in Production

Processing Office	Production versus Evaluation Followup		
	Nonmatch to Match	Nonmatch to Nonmatch	Nonmatch to Unresolved
Albany	9,142 (48.57%)	15,011 (100.00%)	16,113 (44.77%)
Austin	74,727 (59.00%)	14,190 (96.43%)	75,736 (57.64%)
Baltimore	7,963 (61.69%)	0 (0.00%)	28,752 (37.01%)
Jacksonville	67,930 (39.66%)	23,127 (50.70%)	56,848 (33.20%)
Jeffersonville	21,235 (63.86%)	26,051 (97.93%)	37,782 (46.86%)
Kansas City	85,660 (35.43%)	5,756 (61.41%)	35,041 (30.92%)
San Diego	34,991 (47.19%)	10,269 (78.19%)	62,257 (60.99%)
Total	301,649 (21.18%)	94,403 (37.86%)	312,530 (20.91%)

239

Table 3.5.3 Changed From Nonmover in PES to Mover in EFU. Estimates and Coefficients of Variation. Weighted to the Total Population. By Processing Office and Match Status. Unresolved in Production.

Processing Office	Production versus Evaluation Followup		
	Unresolved to Match	Unresolved to Nonmatch	Unresolved to Unresolved
Albany	0 (0.00%)	0 (0.00%)	0 (0.00%)
Austin	12,564 (78.58%)	0 (0.00%)	12,556 (97.60%)
Baltimore	0 (0.00%)	0 (0.00%)	0 (0.00%)
Jacksonville	0 (0.00%)	0 (0.00%)	11,383 (100.00%)
Jeffersonville	0 (0.00%)	0 (0.00%)	0 (0.00%)
Kansas City	22,415 (76.23%)	0 (0.00%)	9,866 (70.85%)
San Diego	255 (100.00%)	0 (0.00%)	447 (100.00%)
Total	35,234 (56.01%)	0 (0.00%)	34,252 (52.94%)

Among the cases that were nonmatches in production, all offices found matches. Baltimore did not have any nonmatches coded to nonmatches. Many cases changed from nonmatches to unresolved in all offices (Table 3.5.2).

Finally, focusing on cases that were unresolved in the production, it is the finding that three offices had no unresolved cases that changed mover status (Albany, Baltimore and Jeffersonville). Austin, Kansas City and San Diego converted many new mover cases among the unresolved to matches (high coefficient of variation). In Jacksonville, all unresolved production cases remained unresolved in the Evaluation Followup. There were no conversions to nonmatches.

3.6 Effect of Address Misreporting on The DSE

The last step in the analysis of the data involved calculating the implications of the results on the dual system estimator estimate of undercount. The Census Bureau's DSE is described below.

To estimate the number of distinct people enumerated in the census, we use $\hat{N}_{CE}$ obtained as follows. Let I_C denote the number of persons imputed into the original enumeration, let I_E denote the number of persons counted in the census for whom names are not available, let $\tilde{I}_E$ denote the weighted number of census enumerations (from the E sample) with insufficient information for matching, and let $\tilde{E}_E$ denote the weighted number of erroneous enumerations that were included in the E sample. If all those quantities were known, the estimated size of the population that could possibly be matched would be $\hat{N}_{CE} = \hat{N}_C - I_C - \tilde{I}_E - \tilde{E}_E$. As $\tilde{E}_E$ and $\tilde{I}_E$ are estimated, we substitute their estimates $\hat{I}_E$ and $\hat{E}_E$, and obtain $\hat{N}_{CE} = \hat{N}_C - I_C - \hat{I}_E - \hat{E}_E$.

Next, let N_{CP} denote the weighted number of P-sample selections who were enumerated in the census, and let $\hat{N}_{CP}$, the weighted number of matches, be an estimate of that quantity. We estimate N by

$$\hat{N} = \hat{N}_P \hat{N}_{CE} / \hat{N}_{CP}$$

The DSE is used to estimate the percent net undercount, or the net undercount rate, in the original enumeration,

$$\hat{U} = 100(\hat{N} - \hat{N}_C) / \hat{N}$$

The error in $\hat{N}_P$ caused by misreporting of Census Day address by outmovers and inmovers is n_{pa}. The error in $\hat{N}_{CP}$, is m_a, the error introduced by respondents reporting the wrong Census Day address.

Estimation of expected values of m_a and n_{pa} is based on weighted data from the Evaluation Follow-up of the P sample cases, where n_{pa} is the number of persons moving out of or into the post-stratum (note that a move is defined as an outcome resulting in a change in post-stratum) and m_a represents the difference, false matches - false nonmatches.

The estimation of the variances of m_a and n_{pa} and their covariance uses the variance estimation program VPLX (Fay, 1990). To estimate the individual effect of address reporting error on the DSE, the mean, variance and covariance of m_a and n_{pa} were estimated. Then assuming that address reporting error was the only error, a simulation was performed where the distribution of an error-free DSE and net undercount rate were calculated (Mulry and Spencer, 1991). The bias in the net undercount rate $\hat{B}(\hat{U})$ was estimated by the difference between $\hat{U}$ and the mean of the simulated distribution of the undercount rate. We also calculated the standard deviation of the estimated bias.

Table 3.6.1 Individual Effect of Address Reporting Error on the Net Undercount Rate by Evaluation Post Stratum

Evaluation Post-Stratum	$\hat{U}$	$\hat{B}(\hat{U})$	St.Dev. $\hat{B}(\hat{U})$
Minorities			
NE, Central City (1)^	6.83	0.33*	0.12
S, Central City (5)	5.68	0.29	0.21
MW, Central City (8)	3.97	0.49*	0.22
W, Central City (11)	6.14	0.11	0.06
U.S.,Noncentral City (3)	5.43	0.88*	0.43
Nonminorities			
NE, Central City (2)	-.75	0.52*	0.22
NE, Noncentral City (4)	0.01	0.18	0.13
S, Central City (6)	1.94	0.18	0.21
S, Noncentral City (7)	1.82	0.56*	0.28
MW, Central City (9)	1.28	0.61	0.42
MW, Noncentral City (10)	0.39	0.24*	0.10
W, Central City (12)	2.13	-1.36	1.57
W, Noncentral City (13)	1.84	0.57	0.31
National	2.11	0.31*	0.11

^Evaluation Post-Stratum Number

*Significant at the 0.05 level

As shown in Table 3.6.1, the bias in the dual system estimate of undercount due to address misreporting is positive in all strata except the West, central city, nonminority stratum where its value is negative (-1.36) Note, however, the magnitude of the standard deviation. The bias is significant at the 0.05 level in all the minority strata except South and West, central cities. In the nonminority strata, the bias is not significant in the Northeast, noncentral city, South, central city, Midwest, central city, in the West, central city and the West, noncentral city. At the national level, the bias in the undercount estimate due to address misreporting is 0.31 with a standard deviation of 0.11.

4. SUMMARY

In the EFU, 334 respondents changed from a nonmover in PES to a mover. This represents, 0.66% of the total population. The geographical and demographical characteristics of the movers were examined to detect if the effect was concentrated in any area. The data were analyzed by minority status and processing office. The EFU found less matches and nonmatches, but more unresolved cases than the production in the minority strata. On the other hand, in the nonminority strata, there were more matches and unresolved cases in the EFU and fewer nonmatches. The effect was not concentrated in any given processing office. It should be noted that many results had high coefficients of variation.

The bias in the DSE estimate of undercount due to address misreporting was also assessed. The effect was small, but significant at the 95% level of confidence in three out of five minority strata, and in three out of eight nonminority strata. At the national level, the bias is also small, but significant, amounting to less than a third of one percent bias in the undercount estimate.

References

Fay, R.E. (1990) VPLX: Variance Estimates for Complex Samples, *Proceedings of the Section on Survey Research, American Statistical Association*, 266-290.

Mulry, M. H. and Spencer, B. D. (1991) Total Error in the PES Estimates of Population, *Journal of the American Statistical Association*, 86. In Press.