MATCHING ERROR IN THE CENSUS 2000 ACCURACY AND COVERAGE EVALUATION*

Susanne L. Bean, U.S. Bureau of the Census
4700 Silver Hill Road, Washington, DC 20233-9200

KEY WORDS: nonsampling error, clerical matching, Accuracy and Coverage Evaluation, Dual System Estimation

* 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.

1. BACKGROUND

A potential source of error in the Accuracy and Coverage Evaluation (A.C.E.) coverage estimates is a matching operation which determined whether the respondents in the population sample (P-sample) were enumerated in the census and whether the enumerations in the enumeration sample (E-sample) were correct. In preparing for Census 2000, the A.C.E. planners put much effort into improving the person matching process from 1990. These improvements include: completing all matching in one location, utilizing a computer system in the clerical matching process, targeting the surrounding block search area, and automating the quality assurance process. To evaluate this source of nonsampling error, the Matching Error Study (MES) conducted an independent rematch in A.C.E. block clusters selected for the evaluation sample[1].

The Census Bureau conducted a MES for the Census 2000 Dress Rehearsal Integrated Coverage Measurement (ICM) and for the 1990 Post Enumeration Survey (PES). The MES for the Dress Rehearsal was unable to measure significant matching error, presumably because there was a 100 percent quality assurance (QA) during the ICM[2]. The match code discrepancy rates (which represent the magnitude of the difference between the person-level ICM and MES matching) for both the P-sample and E-sample were less than one percent in all sites.

The 1990 MES found that the PES generally tended to overestimate the P-sample nonmatches, especially when matching Central City, Minority persons. The magnitude of the biases in the population sizes due to matching error by evaluation poststratum (based on region, urbanicity, and minority status) ranged from approximately 0.7 percent to 1.3 percent. Of particular concern, nonmatches for Blacks were overestimated by about 4.5 percent (which equated to an approximately 0.7 percent positive bias in the total Black population). (Davis and Biemer, 1991a) The erroneous enumerations, on the other hand, were underestimated by about 5 percent for nonminorities (resulting in a positive bias in the overall population of about 0.25 percent). (Davis and Biemer, 1991b)

2. METHODS

2.1 A.C.E. production matching

The first phase of production person matching was computer matching. Then, there were two phases of clerical person matching: a before-followup (BFU) match and an after-followup (AFU) match following the A.C.E. Person Followup (PFU) interview. There were three levels of matchers within each clerical phase: clerks, technicians, and analysts.

Production clerical person matching used the Person Matching Review and Coding System (Per MaRCS) software. During the BFU phase of production, clerks coded all computer nonmatches and possible matches. The technicians reviewed all cases coded RV (need review) by the clerks. The technicians also conducted a QA procedure on a sample-basis of clusters done by the clerks (if a clerk was not approved for sample QA then the technicians conducted a 100 percent review). The analysts reviewed all cases coded RV by the technicians and conducted a similar QA procedure on clusters reviewed by technicians. During the AFU phase of production, the clerks used information gathered during the PFU interview to resolve the status of cases sent to followup. The technicians and analysts then reviewed cases and performed a QA as in the BFU phase. By the end of the AFU, a final code was assigned to all cases indicating P-sample match and residence status and E-sample enumeration status.

2.2 Evaluation matching

The MES rematch utilized only the two highest levels of matchers (technicians and analysts). For the MES rematch, the matchers began from scratch (i.e., did not

[1]The evaluation sample consists of 2259 clusters, which is about a fifth of the A.C.E. clusters. Block clusters with high proportions of minorities and high nonmatch rates were selected disproportionately. (Keathley, 2001)

[2]For the Dress Rehearsal, the Bureau planned to QA only a portion of the work, but logistical concerns necessitated a 100 percent QA. For 2000, QA was done on a sample basis once the matcher reached a specified level of proficiency. (Byrne, 2001)

have access to the production matching results) and used the same procedures as production matching. If the production and the rematch matchers disagreed, another matcher reconciled the difference (the reconciliation phase used only the analysts, the most highly trained matching personnel). In the reconciliation phase, the analyst looked at the production and rematch results and then decided what the true matching result should be.

The results of the study are based on the assumption that agreement of two matchers along with the reconciliation of conflicting match codes yields match results that are as close to truth as possible under the limitations of the evaluation. Another important assumption of the MES is that the production matching and evaluation rematching operations are independent. "Independent" means that the MES matchers did not work clusters they worked during A.C.E. production[3] and did not have access to the A.C.E. production match codes during the rematch phase (i.e., the match code assignments made during the MES rematch were not influenced by production matching).

3. LIMITS

As in 1990, a limitation to this study involves the assumption of independence between the production matching and the evaluation rematching operations. The matching technicians and analysts were involved in production matching, as well as being used exclusively for evaluation matching. Although different matchers must be used to rematch a given case, matchers often discuss difficult cases with others in the group. This challenges the independence assumption for an undetermined portion of the cases. The lack of independence could lead to an underestimate of the actual level of matching error. However, due to the large size of the A.C.E. and evaluation samples, memory of cases should be minimal.

4. RESULTS

4.1 Was there a reduction in matching error in the 2000 A.C.E. compared to the 1990 PES?

Yes, as discussed below, the production and MES matching results were more consistent in 2000.

To compute match probabilities for the P-sample, the A.C.E. collapses the detailed match codes into the following match status classifications: match, nonmatch, unresolved, or remove from P-sample. (Childers, 2001) "Match" means the P-sample case matched a census enumeration. If there is no match for the P-sample case, then it is a "nonmatch". "Remove from P-sample" means the person is in a housing unit that was geocoded to the cluster in error, a nonresident of the cluster on Census Day, a duplicate of another P-sample person, or discrepant[4]. A P-sample case is "unresolved" if the match status cannot be resolved or the case has insufficient information for matching. In the estimation stage, the unresolved cases receive an imputed match probability.

Table 1 presents the 2000 comparison of the P-sample production match status classifications with those from MES. The table presents data weighted to the national level. Standard errors are presented in parentheses underneath the estimates.

To summarize this data and compare to 1990 results, overall P-sample gross difference and net difference rates were calculated. The gross difference rate is the proportion of cases whose matching classifications were different for production and MES. The net difference rate is the sum of the absolute differences between the production and MES totals for each category divided by the population total.

Overall, the 1990 P-sample gross difference rate was 1.55 percent and the net difference rate was 0.93 percent. In 2000, the P-sample gross difference is approximately 0.46 percent and the net difference is approximately 0.41 percent. Therefore, the 2000 gross difference and net difference rates for the P-sample demonstrate a reduction in matching error from 1990. Further, the 2000 pattern of changes, that is more matches and fewer nonmatches in the MES, is consistent with the 1990 findings.

To calculate enumeration probabilities for the E-sample, the A.C.E. collapses the detailed match codes into the following enumeration status classifications: correct enumeration, erroneous enumeration, or unresolved. (Childers, 2001) "Correct enumeration" means the person is a resident of the block cluster on Census day. "Erroneous enumeration" means the person is in a housing unit that was geocoded to the cluster in error, a nonresident of the cluster on Census Day, a duplicate of another P-sample person, or discrepant. E-sample cases which have insufficient information for matching are also erroneous enumerations. E-sample cases are "unresolved" if their residence status or match status cannot be resolved. In the estimation stage, the unresolved cases receive an imputed enumeration probability.

Table 2 presents the 2000 comparison of E-sample production and MES enumeration status classifications. The table presents data weighted to the national level.

[3]Parts of the production AFU matching were done in batches where cases were worked as they came in from the field and not altogether as a cluster. The MES did not restrict coders from working cases they worked in the batch phase, but any memory effect would be very minimal.

[4]Discrepant results are errors that do not include honest mistakes made by the interviewers or respondents and could be falsification, but the amount is uncertain.

Overall, the 1990 E-sample gross difference rate was 2.32 percent and the net difference rate was 1.07 percent. In 2000, the E-sample gross difference is approximately 0.62 percent and the net difference is approximately 0.20 percent. Therefore, the 2000 gross difference and net difference rates for the E-sample demonstrate a reduction in matching error from 1990. Further, the 2000 pattern of changes, that is fewer correct enumerations and more erroneous enumerations in the MES, is consistent with the 1990 findings.

4.2 How does matching error affect the 2000 Dual System Estimates?

As discussed below, the national production dual system estimate (DSE) was significantly higher (by 483,938 with a standard error of 92,877) due to matching error.

The dual system estimator is

$$DSE = \frac{(DD)\left(\frac{CE}{N_E}\right)}{\frac{M}{N_P}} = (DD)\left(\frac{\text{CE Rate}}{\text{Match Rate}}\right)$$

where

DSE = the dual system estimate of the population in housing units on Census Day

DD = census data-defined persons eligible and available for A.C.E. matching

CE = the weighted estimate of correct enumerations in the E-sample

N_E = the weighted estimate of E-sample people

M = the weighted estimate of matches in the P-sample

N_P = the weighted estimate of P-sample people

DD is a census count which is not affected by matching. Therefore, the effect of matching error on the DSE will be reflected in the error in the ratio of CE rate to match rate.

Note: The calculations of match rates, correct enumerations rates, and ratios of these two rates use data which reflect imputation of match, residence, or correct enumeration probabilities for unresolved cases. Therefore, unresolved cases in Tables 1 and 2 contributed to the total number of matches and correct enumerations.

Matching error significantly decreased the national production match rate. This would falsely increase the production DSE (holding all other errors constant). Considering P-sample matching error only (i.e., matching error in the match rate), the national production DSE was overstated by 385,152 (with a standard error of 83,608).

Matching error did not significantly effect the production correct enumeration rate. Considering E-sample matching error only (i.e., matching error in the correct enumeration rate), the national production DSE was overstated by 98,925 (with a standard error of 61,388).

The ratio of CE rate to match rate demonstrates the combined effect of matching error in the match rate and correct enumeration rate. The national ratio of CE rate to match rate, and thus the production DSE, was significantly higher due to matching error. Considering the combined effect of P- and E-sample matching error, the national production DSE was 483,938 higher[5] (with a standard error of 92,877) than the MES DSE.

4.3 What types of matching errors were there?

After examining some patterns of differences, three sources of error will be discussed below: Targeted Extended Search matching, updating census cases with insufficient information for matching, and identifying discrepant cases.

The P-sample gross difference and net difference rates are less than one percent. However, when examining the differences that do exist, some interesting patterns emerge. Of particular concern are the differences in Table 1 between the match row and match column. The match to remove cell (cases that production identified as a "match" but MES said were "remove from P-sample") is about 12.5 percent of the size of its complement, the remove to match cell (26,995 versus 216,311). Further, the match to nonmatch cell (cases that production identified as a "match" but MES said were a "nonmatch") is about 23.3 percent of its complement, the nonmatch to match cell (105,281 versus 451,097).

Although the weighted difference between the match to remove and the remove to match cells was large (26,995 versus 216,311), the unweighted cell counts are close (132 versus 142). The difference between the match to remove and the remove to match cells in Table 1 is mostly due to weighting (noninterview adjustment in particular).

Table 3 highlights three types of P-sample nonmatch errors: error in searching for matches in the surrounding block (SB) search area, error resulting from errors in updating E-sample cases with insufficient information for matching, and error in identifying discrepant cases. The first two types of errors contribute to the match/nonmatch difference mentioned above. (Nonmatch to match falls under false nonmatches and match to nonmatch falls under missed nonmatches.) The table presents unweighted data.

[5] The three estimates of the increase in the national production DSE (P-sample error only, E-sample error only, and combined effect) are based on ratio estimators. Thus, the increase due to the combined effect is not exactly equal to the sum of the increases due to P-sample error only and E-sample error only.

The A.C.E selected about a fifth of the A.C.E. clusters for Targeted Extended Search (TES), where matchers looked for matches in the first ring of blocks surrounding the cluster after searching for matches within the cluster. Matchers failed to find the match in the surrounding block more often than they incorrectly identified a match in the surrounding block (91 versus 11). These errors contribute to the match/nonmatch difference and make up about 17 percent of the false nonmatches and five percent of the missed nonmatches.

Another source of error on the P-sample side stems from a problem on the E-sample side. Matchers had the ability to use information from images of the census forms to update census cases which entered the matching phase with insufficient information for matching. If the matcher failed to update an E-sample case, then the corresponding case in the P-sample could not be matched to this E-sample record (leading to false nonmatches). Conversely, if the matcher made an E-sample case eligible for matching by incorrectly updating the case, then they could incorrectly match this case to a P-sample record that should have been left a nonmatch (thereby creating a missed nonmatch). Matchers missed matches more often than they created incorrect matches due to problems in updating E-sample cases with insufficient information for matching (70 versus 27). These errors also contribute to the match/nonmatch difference and make up approximately 13.1 percent of the false nonmatches and 12.3 percent of the missed nonmatches.

Finally, approximately 9.3 percent of the false nonmatches were cases that should have been coded as discrepant (one type of "remove from P-sample") and 16.4 percent of the missed nonmatches were cases that production incorrectly identified as discrepant. The vast majority of these (48 of the 50 missed discrepant cases and all of the false discrepant cases) were cases that production coded as a nonmatch with unresolved residency status instead of discrepant or vice versa. Most of these errors are probably due to confusion about when to code cases discrepant and when to code them unresolved residency based on the Person Followup (PFU) knowledgeable respondent rules[6].

The E-sample gross difference and net difference rates are also less than one percent. Again, some interesting patterns emerge when examining the differences that do exist. Table 4 highlights three types of E-sample erroneous enumeration errors: differences due to duplicates, error in updating E-sample cases with insufficient information for matching, and error in identifying discrepant cases. The table presents unweighted data.

False duplicates made up approximately 11.7 percent of the false erroneous enumerations and missed duplicates made up approximately 15.7 percent of the missed erroneous enumerations. Differences between production and MES due to switched primaries (production found the duplicate/primary pair but picked the wrong person to be the primary according to the matching procedures) represent approximately 20.9 percent of the gross false erroneous enumerations and 18.5 percent of the gross missed erroneous enumerations. However, the switched primary differences are not an error when looking at the overall production or MES totals of any of the enumeration status categories, because in terms of the net they balance themselves out.

As mentioned earlier, matchers could use information from images of the census forms to update census cases which entered the matching phase with insufficient information for matching. About 22 percent of the false erroneous enumerations were cases that production missed the update from image (i.e., left the case insufficient in error). On the other hand, about seven percent of the missed erroneous enumerations were cases that the production matchers updated when they should not have.

Finally, almost 20 percent of the false erroneous enumerations were cases production incorrectly identified as discrepant and almost 26 percent of the missed erroneous enumerations were cases that should have been coded as discrepant. The vast majority of these (79 of the 88 false discrepant cases and 124 of the 130 the missed discrepant cases) were cases that production classified as discrepant instead of unresolved enumeration status or vice versa. As with the P-sample, most of these errors are probably due to confusion about when to code cases discrepant and when to code them unresolved based on the PFU knowledgeable respondent rules.

5. RECOMMENDATIONS

In preparing for Census 2000, the A.C.E. planners put much effort into improving the person matching process from 1990. In 2000, all the matching was done in one location (while the matching in 1990 was done in seven processing offices throughout the country) which allowed for more consistent training and supervision of the matchers. In addition, the matchers used a computer system to review and code the cases (1990 was done on paper) which made the matching process more efficient and allowed for built in checks and edits to improve data quality. Further, the searching in the surrounding block areas was targeted to clusters where matches and duplicates were likely to be found outside the cluster (in 1990 these searches were not targeted and there was anecdotal evidence that matchers did not bother to look in

[6]If the person was unknown to three "knowledgeable" respondents in the PFU, than the case was to be coded discrepant. However, if the person was unknown but there were less than three knowledgeable respondents, than the case was to be coded unresolved.

surrounding blocks because they rarely found anything). Another improvement for 2000 was in the quality assurance area through the use of automated procedures to flag cases for review.

The reductions in matching error from 1990 to 2000 provide evidence that the changes made from 1990 improved the quality of the 2000 A.C.E. matching.

Even with these improvements, matching error inflated the national production dual system estimate (by 483,938 with a standard error of 92,877) and therefore overstated the undercount estimate (holding all other errors constant). Therefore, to further reduce matching error in the future, planners should continue efforts to improve the matching process. Three specific areas which should be considered are:

- *Targeted Extended Search* - One area where matchers made errors was searching for matches in the surrounding blocks, despite attempts to improve this process by targeting the clusters selected and restricting the work to a subset of matchers approved for this type of work. Planners should explore ways to further simplify the TES procedures and improve the quality assurance for TES clusters.
- *Updating census cases with insufficient information* - Another area where errors were made even though A.C.E. planners attempted to implement improvements was in the updating of census cases with insufficient information for matching. The first step the software had the matchers work was examining these cases. Planners should attempt to identify ways to further ensure the matchers perform this step, perhaps through further emphasis in training and additional quality assurance checks.
- *Discrepant cases* - The last area which caused some problems for matchers was distinguishing between when to code cases discrepant and when to code them unresolved based on the Person Followup (PFU) knowledgeable respondent rules. Planners should make sure these rules are defined clearly in advance and enhance training in this area.

6. REFERENCES

Bean, S., Bench, K. Davis, M.C., Hill, J.M., Krejsa, E., and Raglin, D. (1999). Error Profile for the Census 2000 Dress Rehearsal. Census 2000 Dress Rehearsal Evaluation Results Memorandum Series, #C4 dated July, 1999.

Byrne, Rosemary (2001). MaRCS Specifications for Quality Assurance (QA) for Person Matching. DSSD Census 2000 Procedures and Operations Memorandum Series Chapter S-QA-13 dated February 7, 2001.

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.

Davis, M.C. and Biemer, P. (1991a). Estimates of P-sample Clerical Matching Error from a Rematching Evaluation. 1990 Coverage Studies and Evaluation Memorandum Series, #H-2 dated July 11, 1991.

Davis, M.C. and Biemer, P. (1991b). Measurement of the Census Erroneous Enumerations - Clerical Error Made in the Assignment of Enumeration Status. 1990 Coverage Studies and Evaluation Memorandum Series, #L-2 dated July 11, 1991.

Keathley, Don (2001). Evaluation Followup Sample Design, Stratification, Selection, and Weighting. PRED TXE/2010 Memorandum Series: CM-GES-S-02-R2 dated July 24, 2001.

Table 1. 2000 Comparison of Production and MES Match Status for the P-sample

Production Results	MES Results					
	Match	Nonmatch	Remove	Unresolved	**Total**	Percent
Match	240,436,019 (6,077,063)	105,281 (21,267)	26,995 (8,074)	66,496 (14,044)	240,634,791 (6,079,637)	89.96
Nonmatch	451,097 (59,911)	20,507,741 (680,409)	119,286 (30,279)	26,193 (6,680)	21,104,317 (690,802)	7.89
Remove	216,311 (38,434)	146,862 (24,403)	2,218,093 (239,223)	7,832 (3,312)	2,589,099 (257,297)	0.97
Unresolved	37,937 (12,614)	21,687 (10,414)	0 (0)	3,090,461 (164,209)	3,150,085 (166,439)	1.18
Total	241,141,364 (6,087,044)	20,781,571 (684,064)	2,364,374 (241,860)	3,190,983 (165,263)	267,478,292 (6,554,111)	100.00
Percent	90.15	7.77	0.88	1.19	100.00	

Table 2. 2000 Comparison of Production and MES Enumeration Status for the E-sample

Production Results	MES Results				
	Correct Enumeration	Erroneous Enumeration	Unresolved	**Total**	Percent
Correct Enumeration	250,509,005 (6,187,926)	363,054 (43,618)	364,858 (82,384)	251,236,917 (6,195,998)	93.49
Erroneous Enumeration	321,185 (39,124)	10,061,330 (364,291)	250,210 (32,572)	10,632,724 (374,247)	3.96
Unresolved	133,779 (20,028)	236,263 (40,041)	6,499,708 (487,748)	6,869,750 (492,644)	2.56
Total	250,963,969 (6,193,270)	10,660,647 (378,339)	7,114,776 (518,992)	268,739,391 (6,486,545)	100.00
Percent	93.39	3.97	2.65	100.00	

Table 3. Types of P-sample Nonmatch Errors

Production Results	**N**	**%**	**MES Results**	**N**	**%**
False Nonmatch	536	100.0	Missed Nonmatch	220	100.0
Missed Match to SB	91	17.0	False Match to SB	11	5.0
Missed Match due to Missed Update from Image	70	13.1	False Match due to False Update from Image	27	12.3
Missed Discrepant	50	9.3	False Discrepant	36	16.4
Other	325	60.6	Other	146	66.4

Table 4. Types of E-sample Erroneous Enumeration Errors

Production Results	**N**	**%**	**MES Results**	**N**	**%**
False Erroneous Enumeration	445	100.0	Missed Erroneous Enumerations	502	100.0
Duplicates			Duplicates		
False Duplicate	52	11.7	Missed Duplicate	79	15.7
Switched Primary - (Not an Error)	93	20.9	Switched Primary - (Not an Error)	93	18.5
Missed Update from Image	100	22.5	False Update from Image	35	7.0
False Discrepant	88	19.8	Missed Discrepant	130	25.9
Other	112	25.1	Other	165	32.9