Census 2000 Evaluation N.14
June 20, 2002

Evaluation of Matching Error

FINAL REPORT

This evaluation study reports the results of research and analysis undertaken by the U.S. Census Bureau. It is part of a broad program, the Census 2000 Testing, Experimentation, and Evaluation (TXE) Program, designed to assess Census 2000 and to inform 2010 Census planning. Findings from the Census 2000 TXE Program reports are integrated into topic reports that provide context and background for broader interpretation of results.

Susanne L. Bean
Planning, Research, and Evaluation Division

U S C E N S U S B U R E A U
Helping You Make Informed Decisions

CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

A potential source of error in the Accuracy and Coverage Evaluation 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 Accuracy and Coverage Evaluation planners put much effort into improving the person matching process from 1990. To evaluate this source of nonsampling error, the Matching Error Study conducted an independent rematch in Accuracy and Coverage Evaluation block clusters selected for the evaluation sample.

For the rematch, the matchers began from scratch (i.e., did not 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.

Conclusion

As discussed below, the reductions in matching error from 1990 to 2000 provide evidence that the changes made from 1990 improved the quality of the 2000 Accuracy and Coverage Evaluation matching process.

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.

Was there a reduction in matching error in the 2000 Accuracy and Coverage Evaluation compared to the 1990 Post Enumeration Survey?

Yes, the production and Matching Error Study matching results were more consistent in 2000. The results that support this finding are:

- Overall, the 1990 P-sample gross difference rate was 1.55 percent and the net difference rate was 0.93 percent. In 2000, the gross rate is 0.46 percent and the net rate is 0.41 percent. Therefore, the 2000 gross difference and net difference rates for the P-sample demonstrate a reduction in matching error from 1990.

- Overall, the 1990 E-sample gross difference rate was 2.32 percent and the net difference rate was 1.07 percent. In 2000, the gross rate is 0.62 percent and the net rate is 0.20 percent. Therefore, the 2000 gross difference and net difference rates for the E-sample demonstrate a reduction in matching error from 1990.

- Overall, the 2000 P-sample relative difference rate for matches is -0.21 percent and the 2000 E-sample relative difference rate for correct enumerations is 0.11 percent. The 2000 relative difference rate for matches is similar to the 1990 rate (-0.18 percent). The 2000 relative difference rate for correct enumerations shows a reduction from the 1990 rate (0.57 percent).

- In 1990, the relative difference rate for matches by the 1990 Evaluation Poststratum groups ranged from -1.38 to 0.46 percent, whereas in 2000 the rate ranged from -0.74 to 0.16 percent by the Preliminary Evaluation Poststratum groups. The relative difference rate for correct enumerations ranged from -0.54 to 1.08 percent in 1990, and in 2000 the rate ranged from -0.04 to 0.92 percent. In comparing the ranges of relative difference rates for matches and correct enumeration rates by evaluation poststratum groups, we again find a reduction in matching error from 1990 to 2000.

How does matching error affect the 2000 Dual System Estimates?

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

At the Preliminary Evaluation Poststratum level, matching error inflated the production dual system estimates in all but one of the sixteen groups. **Using a Bonferroni multiple comparison test, the production dual system estimates were significantly higher due to matching error in two of the sixteen Preliminary Evaluation Poststratum groups: 8 and 14** (Non-minority - Non-owner - Large or Medium Metropolitan Statistical Area - Mail Out / Mail Back - high Return Rate [8] and Minority - Non-owner - Large or Medium Metropolitan Statistical Area - Mail Out / Mail Back - high Return Rate [14]). These poststratum groups comprise approximately 16 percent of the population.

In addition, the components of the dual system estimate affected by matching error, the match rate and the correct enumeration rate were in agreement with these outcomes:

- Using the multiple comparison test, matching error significantly decreased the production match rates in two Preliminary Evaluation Poststratum groups: 8 and 14. Further, the national production match rate was significantly lower. This would falsely increase the production dual system estimate for these two groups and at the national level (holding all other errors constant).

- Using the multiple comparison test, matching error had no significant effect on the correct enumeration rates for any poststrata. Further, the national production correct enumeration rate was not significantly different due to matching error.

Were there clerical errors in identifying duplicates in the Accuracy and Coverage Evaluation search area?

There were only minor errors in the coding of duplicates. The results that support this finding are:

- Of the 1584 P-sample duplicates production identified, 2.4 percent were false duplicates which inaccurately increased the total number of production "remove from P-sample" cases. Of the 1601 duplicates the Matching Error Study identified, 3.3 percent were missed by production which inaccurately diminished the total number of production "remove from P-sample" cases.

- Of the 1504 E-sample duplicates production identified, 3.5 percent were false duplicates which inaccurately increased the total number of production erroneous enumerations. Of the 1526 duplicates the Matching Error Study identified, 5.2 percent were missed by production which inaccurately diminished the total number of production erroneous enumerations.

The Matching Error Study examined the clerical identification of duplicate cases only in the universe defined for production.

What other types of matching errors were there?

There were three types of errors which should be considered when attempting to improve the matching process in the future:

- *Targeted Extended Search* - One area where matchers made errors was searching for matches in the surrounding blocks. Planners should explore ways to further simplify the Targeted Extended Search procedures and improve the quality control for these clusters.

- *Updating census cases with insufficient information* - Another area where errors were made was in the updating of census cases with insufficient information for matching. Planners should attempt to identify ways to further ensure the matchers perform this step, perhaps through further emphasis in training and additional quality control 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 knowledgeable respondent rules. Planners should make sure these rules are defined clearly in advance and enhance training in this area.

1. BACKGROUND

1.1 Questions to be answered

This report answers four questions:

- Was there a reduction in matching error in the 2000 Accuracy and Coverage Evaluation (A.C.E.) compared to the 1990 Post Enumeration Survey (PES)?
- How does matching error affect the 2000 Dual System Estimates (DSEs)?
- Were there clerical errors in identifying duplicates in the A.C.E. search area?
- What other types of matching errors were there?

1.2 Purpose of the Matching Error Study

A potential source of error in the 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].

1.3 Findings of previous matching error studies

The Census Bureau conducted a Matching Error Study for the Census 2000 Dress Rehearsal Integrated Coverage Measurement (ICM) and for the 1990 Post Enumeration Survey (PES). The MES for the Census 2000 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

---

[1]The evaluation sample consists of about 2260 clusters, which is about a fifth of the A.C.E. clusters. (Keathley, 2001a)

[2]For the Dress Rehearsal ICM, the Bureau planned to QA only a portion of the work, but logistical concerns necessitated a 100 percent QA. For the 2000 A.C.E., QA was done on a sample basis once the matcher reached a specified level of proficiency (periodically, the matching software reevaluated the decision to sample). The sample QA involved a dependent rematch on 1/6 of the clerks (the lowest level of matchers) and 1/10 of the technicians (the middle level of matchers) work In addition, cases meeting special “must do” criteria were reviewed. (Byrne, 2001)

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 (the first level of matchers) coded all cases coded nonmatched or possibly matched during the computer matching operation. 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.[3] During the AFU phase of production, the clerks used information gathered during the PFU interview to resolve the match, residence, and enumeration status for cases sent to followup. The technicians and analysts then reviewed cases coded RV and performed a QA procedure as they did in the BFU phase. By the end of the AFU analyst review, a final match code was assigned to all cases.

2.2 Evaluation matching

Production person matching used three levels of matchers: clerks, technicians, and analysts. The MES rematch, on the other hand, utilized only the two highest levels of matchers (technicians and analysts). For the MES rematch, the matchers began from scratch (i.e., did not have access

[3]See Footnote 2 for more information on the QA procedures.

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[4] and did not have access to the A.C.E. production match codes during the rematch phase (i.e., the match code assignments made by the matchers during the MES rematch were not influenced by production matching).

2.3 Evaluation sample design

The sampling frame for the evaluation sample was the collection of A.C.E. sampled block clusters. A subsample of approximately 2260 A.C.E. block clusters (or 70,000 housing units) was selected for the evaluation sample. Block clusters with high proportions of minorities and high nonmatch rates were selected disproportionately. (Keathley, 2001a)

2.4 Weighting

The final person weights for the MES had three components, which were the inverses of the three selection probabilities. The first selection probability was the probability of a block cluster being selected for A.C.E. The second was the conditional probability of a block cluster being selected for the evaluation sample. The last was the probability of a person within the evaluation block cluster being selected for the study. For MES, the last probability was one, because all P-sample and E-sample persons in the evaluation sample block clusters were rematched.

3. LIMITS

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

---

[4]MES matchers did not work clusters they worked during production. However, parts of the production After Followup matching were done in a batch phase where cases were worked as they came in from the field and not altogether as a cluster. The MES did not restrict users from working cases they worked in the batch phase, but any memory effect would be very minimal.

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.

4.1.1 Differences in matching between production and MES at the national level

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[5]. 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 1a is 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.

[5]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.

Table 1a. 2000 Comparison of Production and MES Match Status for the P-sample[6]

Production Results	MES Results					
	Match	Nonmatch	Remove	Unresolved	**Total**	Percent
Match	240,436,019	105,281*	26,995*	66,496	240,634,791	89.96
	(6,077,063)	(21,267)	(8,074)	(14,044)	(6,079,637)	
Nonmatch	451,097*	20,507,741	119,286	26,193	21,104,317	7.89
	(59,911)	(680,409)	(30,279)	(6,680)	(690,802)	
Remove	216,311*	146,862	2,218,093	7,832	2,589,099	0.97
	(38,434)	(24,403)	(239,223)	(3,312)	(257,297)	
Unresolved	37,937	21,687	0	3,090,461	3,150,085	1.18
	(12,614)	(10,414)	(0)	(164,209)	(166,439)	
Total	234,965,956	20,055,492	2,086,798	3,190,983	267,478,292	100.00
	(6,087,044)	(684,064)	(241,860)	(165,263)	(6,554,111)	
Percent	90.15	7.77	0.88	1.19	100.00	

* - See Section 4.4 for more details on differences between cells in the match row and in the match column.

Table 1b is the 1990 comparison of the P-sample production match status classifications with those from MES. (Davis and Biemer, 1991a)

[6]See Appendix A for the 2000 unweighted comparison of match status for the P-sample.

Table 1b. 1990 Comparison of Production and MES Match Status for the P-sample

Production Results	MES Results					
	Match	Nonmatch	Remove	Unresolved	**Total**	Percent
Match	218,476,178	245,551	269,344	252,816	219,243,889	91.10
	(9,175,999)	(60,897)	(59,875)	(84,415)	(9,208,243)	
Nonmatch	678,189	16,016,878	159,302	549,876	17,404,245	7.23
	(125,668)	(1,249,316)	(49,639)	(93,699)	(1,289,655)	
Remove*	0	0	0	0	0	0.00
	(0)	(0)	(0)	(0)	(0)	
Unresolved	491,990	801,471	292,197	2,417,440	4,003,097	1.66
	(81,867)	(130,560)	(73,429)	(207,416)	(308,238)	
Total	210,344,690	17,063,899	611,329	3,042,018	230,153,333	100.00
	(9,203,182)	(1,290,031)	(131,698)	(269,847)	(10,059,221)	
Percent	91.27	7.09	0.30	1.34	100.00	

* - In 1990, estimates for production remove were always zero since the cases did not have weights.

Table 1c summarizes the data found in Tables 1a and 1b by giving the overall P-sample gross difference and net difference rates for 1990 and 2000. This 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.

Table 1c. Gross Difference and Net Difference Rates for the P-sample

	Gross Difference Rate	Net Difference Rate
1990	1.55%	0.93%
2000	0.46%	0.41%
	(0.04%)	(0.06%)

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[7] 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 2a is the 2000 comparison of the E-sample production and MES enumeration status classifications. The table presents data weighted to the national level.

[7]Some of the gross differences reflect differences in identifying which record was the primary in a duplicate/primary pair. That is, production found the same duplicate/primary pair as MES, but picked the wrong person to be the primary according to the matching procedures. However, some of the matching procedures are "cosmetic" rules which do not really affect the DSE process. If all the differences between production and MES in these switched primary cases were cosmetic, then the overall P-sample gross difference rate could be as low as 0.38 percent. See Section 4.3 for more details on differences in identifying duplicates.

Table 2a. 2000 Comparison of Production and MES Enumeration Status for the E-sample[8]

Production Results	MES Results				
	Correct Enumeration	Erroneous Enumeration	Unresolved	**Total**	Percent
Correct Enumeration	250,509,005	363,054	364,858	251,236,917	93.49
	(6,187,926)	(43,618)	(82,384)	(6,195,998)	
Erroneous Enumeration	321,185	10,061,330	250,210	10,632,724	3.96
	(39,124)	(364,291)	(32,572)	(374,247)	
Unresolved	133,779	236,263	6,499,708	6,869,750	2.56
	(20,028)	(40,041)	(487,748)	(492,644)	
Total	244,736,919	10,252,738	6,999,820	268,739,391	100.00
	(6,193,270)	(378,339)	(518,992)	(6,486,545)	
Percent	93.39	3.97	2.65	100.00	

Table 2b is the 1990 comparison of the E-sample production and MES enumeration status classifications. (Davis and Biemer, 1991b)

[8]See Appendix A for the 2000 unweighted comparison of enumeration status for the E-sample.

Table 2b. 1990 Comparison of Production and MES Enumeration Status for the E-sample

Production Results	MES Results				
	Correct Enumeration	Erroneous Enumeration	Unresolved	**Total**	Percent
Correct Enumeration	225,528,071	1,620,295	1,564,489	228,712,855	93.66
	(9,562,390)	(209,666)	(612,654)	(9,651,682)	
Erroneous Enumeration	1,002,013	11,244,969	363,118	12,610,101	5.16
	(163,328)	(791,646)	(69,501)	(885,229)	
Unresolved	877,458	240,630	1,759,871	2,877,959	1.18
	(261,219)	(55,369)	(248,142)	(454,430)	
Total	217,681,824	12,104,582	3,005,323	233,664,004	100.00
	(9,664,821)	(889,890)	(693,983)	(45,958,612)	
Percent	93.12	5.37	1.51	100.00	

Table 2c summarizes the data found in Tables 2a and 2b by giving the overall E-sample gross difference and net difference rates for 1990 and 2000.

Table 2c. Gross Difference and Net Difference Rates for the E-sample

	Gross Difference Rate	Net Difference Rate
1990	2.32%	1.07%
2000	0.62%	0.20%
	(0.05%)	(0.07%)

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[9] and the net difference is approximately 0.20 percent. Therefore, the 2000 gross difference and net difference

[9]Some of the gross differences reflect differences in identifying which record was the primary in a duplicate/primary pair. That is, production found the same duplicate/primary pair as MES, but picked the wrong person to be the primary according to the matching procedures. However, some of the matching procedures are "cosmetic" rules which do not really affect the DSE process. If all the differences between production and MES in these switched primary cases were cosmetic, then the overall E-sample gross difference rate could be as low as 0.52 percent. See Section 4.3 for more details on differences in identifying duplicates.

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.

4.1.2 Relative error associated with the matching at the national level

A goal of the MES is to evaluate the relative error in the number of P-sample matches (M) and in the number of E-sample correct enumerations (CE). The bias in the number of matches has an inverse relationship with the bias in the dual system estimates (DSEs). The bias in the number of correct enumerations, however, has a direct relationship with the bias in the DSEs. The MES is assumed to produce results closer to the true match results, therefore the biases in the M and CE terms are the expected values of the net difference between production and MES. In other words, the biases in M and CE, respectively, are

$$B(M) = E(M_P - M_M)$$

and

$$B(CE) = E(CE_P - CE_M)$$

where

M_P = the weighted matches from production matching

M_M = the weighted matches from the MES

CE_P = the weighted correct enumerations in the E-sample from production matching

CE_M = the weighted correct enumerations in the E-sample from the MES

The relative difference rate (RDR) is defined as RDR = (production - rematch) / rematch. Thus, the RDR(M) and RDR(CE), respectively, are

$$RDR(M) = \frac{M_P - M_M}{M_M}$$

and

$$RDR(CE) = \frac{CE_P - CE_M}{CE_M}$$

Since the numerator of the RDR is an estimator of the bias, the RDR is an estimate of the relative bias. Table 3 presents the RDR(M) and RDR(CE) for both 1990 and 2000.

Note: In 1990, the relative difference rates were calculated using the data in Tables 1b and 2b, which do <u>not</u> reflect imputation of match or enumeration probabilities for unresolved cases. To

make the 2000 rates comparable, they are also computed using data which do not reflect imputation for unresolved cases (data from Tables 1a and 2a). During the estimation stage, a portion of the P-sample unresolved cases contribute to the total number of matches and a portion of the E-sample unresolved cases contribute to the total number of correct enumerations.

Table 3. Relative Difference Rates

	RDR(M)	RDR(CE)
1990	-0.18%	0.57%
2000	-0.21%	0.11%
	(0.03%)	(0.04%)

Overall, the 2000 P-sample RDR(M) is approximately -0.21 percent and the 2000 E-sample RDR(CE) is approximately 0.11 percent. The 2000 overall RDR(M) is similar to the 1990 rate (-0.18 percent). (Davis and Biemer, 1991a) The 2000 overall RDR(CE) shows a reduction from the 1990 RDR(CE), which was 0.57 percent. (Davis and Biemer, 1991b)

4.1.3 Relative error associated with the matching at the poststratum level

Table 4 gives 16 Preliminary Evaluation Poststratum groups defined by minority status, tenure, size of Metropolitan Statistical Area (MSA), whether Mail Out / Mail Back (MO-MB) or other type of enumeration area (TEA), region, and return rate.

Table 4. 2000 Preliminary Evaluation Poststratum Group Definitions

Preliminary Evaluation Poststratum	Preliminary Evaluation Poststratum Group Definition
1	Non-minority - Owner - Large and Medium MSA - MO-MB - NE,MW - high Return Rate
2	Non-minority - Owner - Large and Medium MSA - MO-MB - S,W - high Return Rate
3	Non-minority - Owner - Large and Medium MSA - MO-MB - NE,MW - low Return Rate
4	Non-minority - Owner - Large and Medium MSA - MO-MB - S,W - low Return Rate
5	Non-minority - Owner - Small MSA and Non-MSA - MO-MB - high Return Rate
6	Non-minority - Owner - Small MSA and Non-MSA - MO-MB - low Return Rate
7	Non-minority - Owner - All other TEAs
8	Non-minority - Non-owner - Large or Medium MSA - MO-MB - high Return Rate
9	Non-minority - Non-owner - Large or Medium MSA - MO-MB - low Return Rate
10	Non-minority - Non-owner - Small MSA & Non-MSA - MO-MB - All other TEAs
11	Minority - Owner - Large and Medium MSA - MO-MB - high Return Rate
12	Minority - Owner - Large and Medium MSA - MO-MB - low Return Rate
13	Minority - Owner - All other TEAs
14	Minority - Non-owner - Large or Medium MSA - MO-MB - high Return Rate
15	Minority - Non-owner - Large or Medium MSA - MO-MB - low Return Rate
16	Minority - Non-owner - All other TEAs

Table 5a shows the RDR(M) and RDR(CE) for each of the 16 Preliminary Evaluation Poststratum groups. The table presents data weighted to the national level.

Table 5a. 2000 Relative Difference Rates by Preliminary Evaluation Poststratum Group

Preliminary Evaluation Poststratum	RDR(M) %	RDR(CE) %
1	-0.19	-0.02
	(0.10)	(0.05)
2	-0.13	0.05
	(0.06)	(0.06)
3	-0.22	0.20
	(0.17)	(0.10)
4	0.16	-0.04
	(0.10)	(0.11)
5	-0.14	0.01
	(0.10)	(0.05)
6	-0.74	-0.03
	(0.51)	(0.12)
7	0.00	0.08
	(0.04)	(0.04)
8	-0.35	0.12
	(0.14)	(0.08)
9	-0.38	0.14
	(0.21)	(0.32)
10	-0.04	0.26
	(0.08)	(0.15)
11	-0.22	0.15
	(0.09)	(0.09)
12	-0.67	0.36
	(0.32)	(0.20)
13	-0.10	0.12
	(0.18)	(0.14)
14	-0.51	0.06
	(0.13)	(0.09)
15	-0.26	0.92
	(0.12)	(0.82)
16	-0.68	0.32
	(0.27)	(0.25)

At the Preliminary Evaluation Poststratum group level, the relative difference rates for matches and correct enumerations are all less than one percent in absolute magnitude.

Although the 1990 Evaluation Poststratum groups were defined differently than the 2000 Preliminary Poststratum groups, looking at the 1990 relative difference rate by poststratum

would still give an idea of an expected range. Table 5b presents the RDR ranges for 1990 and 2000. (Davis and Biemer, 1991a and 1991b)

Table 5b. Range of Relative Difference Rates by Poststratum Group

	RDR(M)		RDR(CE)	
	Low	**High**	**Low**	**High**
1990	-1.38%	0.46%	-0.54%	1.08%
2000	-0.74%	0.16%	-0.04%	0.92%

In 1990, the RDR(M) by the 1990 Evaluation Poststratum groups ranged from -1.38 to 0.46 percent, whereas in 2000 the RDR(M) ranged from -0.74 to 0.16 percent by the Preliminary Evaluation Poststratum groups. The RDR(CE) ranged from -0.54 to 1.08 percent in 1990, and in 2000 the RDR(CE) ranged from -0.04 to 0.92 percent. In comparing the ranges of relative difference rates for matches and correct enumeration rates by evaluation poststratum groups, we again find a reduction in matching error from 1990 to 2000.

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. At the Preliminary Evaluation Poststratum level, the production DSEs were significantly higher (using a Bonferroni multiple comparison test) due to matching error in two of the sixteen groups: 8 and 14.

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.

Table 6 presents the production[10] and MES match rates for each of the Preliminary Evaluation Poststratum groups. It also gives the difference between the production and MES rates and the p-value corresponding to the null hypothesis that the difference is zero. The final column in the table indicates whether the production and MES rates are significantly different at alpha = 0.10. The table presents data weighted to the national level.

Note: The calculations below use data which reflect imputation of match probabilities for the unresolved cases. Therefore, some of the unresolved cases in Table 1a contributed to the total number of matches.

[10]In Section 4.2 "production" refers to the baseline. Baseline estimates use production data for just the evaluation sample.

Table 6. 2000 Match Rates by Preliminary Evaluation Poststratum Group[11]

Preliminary Evaluation Poststratum	Match Rate				
	Production %	MES %	Difference %	P-value	Sig?
1	95.86	96.06	-0.20	0.03	✓
	(0.59)	(0.59)	(0.09)		
2	95.38	95.40	-0.02	0.67	
	(0.37)	(0.36)	(0.05)		
3	92.75	93.04	-0.29	0.09	✓
	(1.21)	(1.20)	(0.17)		
4	91.24	91.15	0.09	0.15	
	(1.19)	(1.19)	(0.06)		
5	95.62	95.71	-0.09	0.32	
	(0.62)	(0.62)	(0.09)		
6	91.51	92.01	-0.50	0.07	✓
	(1.70)	(1.62)	(0.28)		
7	92.78	92.86	-0.07	0.31	
	(0.77)	(0.77)	(0.07)		
8	90.18	90.38	-0.20	0.00	✓
	(0.69)	(0.69)	(0.07)		
9	86.80	87.13	-0.33	0.04	✓
	(1.09)	(1.06)	(0.16)		
10	88.73	88.77	-0.04	0.46	
	(0.73)	(0.73)	(0.05)		
11	91.26	91.28	-0.02	0.82	
	(0.53)	(0.52)	(0.08)		
12	87.92	88.06	-0.14	0.44	
	(1.07)	(1.07)	(0.18)		
13	90.34	90.23	0.10	0.62	
	(1.06)	(1.07)	(0.21)		
14	86.75	87.10	-0.35	0.00	✓
	(0.68)	(0.66)	(0.12)		
15	83.24	83.36	-0.12	0.30	
	(1.06)	(1.03)	(0.11)		
16	85.43	85.77	-0.34	0.12	
	(1.03)	(1.01)	(0.22)		
National	91.87	92.00	-0.13	0.00	✓
	(0.22)	(0.22)	(0.03)		

[11]See Appendix B for the match rate components, M and N_p.

Matching error significantly decreased the production match rates in six Preliminary Evaluation Poststratum groups: 1, 3, 6, 8, 9, and 14.[12] Further, the national production match rate was significantly lower. This would falsely increase the production DSE for these six groups and at the national level (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).

Table 7 presents the production and MES correct enumeration rates for each of the Preliminary Evaluation Poststratum groups. It also gives the difference between the production and MES rates and the p-value corresponding to the null hypothesis that the difference is zero. The final column in the table indicates whether the production and MES rates are significantly different at alpha = 0.10. The table presents data weighted to the national level.

Note: The calculations below use data which reflect imputation of enumeration probabilities for the unresolved cases. Therefore, some of the unresolved cases in Table 2a contributed to the total number of correct enumerations.

[12]If a Bonferroni multiple comparison test is used with alpha* = 0.10/16 ≈ 0.006, then the only poststratum groups with significant differences in match rates are 8 and 14.

Table 7. 2000 Correct Enumeration Rates by Preliminary Evaluation Poststratum Group[13]

Preliminary Evaluation Poststratum	Correct Enumeration Rate				
	Production %	MES %	Difference %	P-value	Sig?
1	97.55	97.59	-0.04	0.41	
	(0.30)	(0.30)	(0.05)		
2	96.80	96.77	0.03	0.55	
	(0.36)	(0.37)	(0.05)		
3	95.00	94.85	0.15	0.12	
	(0.86)	(0.88)	(0.10)		
4	95.88	95.95	-0.08	0.42	
	(0.57)	(0.52)	(0.10)		
5	97.25	97.24	0.01	0.89	
	(0.37)	(0.37)	(0.06)		
6	95.43	95.45	-0.02	0.85	
	(0.94)	(0.95)	(0.10)		
7	96.16	96.07	0.09	0.07	✓
	(0.29)	(0.29)	(0.05)		
8	93.43	93.34	0.10	0.04	✓
	(0.58)	(0.58)	(0.05)		
9	92.19	92.38	-0.19	0.39	
	(0.72)	(0.69)	(0.22)		
10	93.73	93.69	0.04	0.65	
	(0.45)	(0.46)	(0.09)		
11	95.85	95.73	0.12	0.14	
	(0.37)	(0.39)	(0.08)		
12	92.66	92.45	0.21	0.22	
	(0.86)	(0.91)	(0.17)		
13	95.23	94.97	0.26	0.37	
	(0.52)	(0.61)	(0.29)		
14	92.82	92.78	0.04	0.62	
	(0.43)	(0.44)	(0.08)		
15	90.65	90.77	-0.11	0.40	
	(0.84)	(0.83)	(0.13)		
16	92.62	92.74	-0.11	0.38	
	(0.61)	(0.63)	(0.13)		
National	95.40	95.36	0.03	0.11	
	(0.14)	(0.14)	(0.02)		

[13]See Appendix B for the correct enumeration rate components, CE and N_E.

Matching error significantly increased the production correct enumeration rates in two Preliminary Evaluation Poststratum groups: 7 and 8.[14] This would falsely increase the production DSE for these two groups (holding all other errors constant). 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).

Table 8 presents the production and MES ratios of CE rate to match rate for each of the Preliminary Evaluation Poststratum groups. It also gives the difference between the production and MES ratios and the p-value corresponding to the null hypothesis that the difference is zero. The final column in the table indicates whether these production and MES ratios are significantly different for alpha = 0.10. The table presents data weighted to the national level.

[14]If a Bonferroni multiple comparison test is used with alpha* = 0.10/16 ≈ 0.006, then no poststratum groups have significant differences in correct enumeration rates.

Table 8. 2000 Ratios of Correct Enumeration Rate to Match Rate by Preliminary Evaluation Poststratum Group[15]

Preliminary Evaluation Poststratum	Ratio of CE Rate to Match Rate				
	Production	MES	Difference	P-value	Sig?
1	1.0177	1.0160	0.0017	0.06	✓
	(0.0069)	(0.0069)	(0.0009)		
2	1.0149	1.0143	0.0005	0.40	
	(0.0039)	(0.0038)	(0.0006)		
3	1.0243	1.0195	0.0048	0.03	✓
	(0.0110)	(0.0113)	(0.0022)		
4	1.0508	1.0527	-0.0019	0.21	
	(0.0134)	(0.0132)	(0.0015)		
5	1.0170	1.0160	0.0011	0.31	
	(0.0071)	(0.0071)	(0.0011)		
6	1.0428	1.0374	0.0054	0.11	
	(0.0205)	(0.0197)	(0.0034)		
7	1.0364	1.0346	0.0018	0.05	✓
	(0.0087)	(0.0086)	(0.0009)		
8	1.0361	1.0327	0.0034	0.00	✓
	(0.0085)	(0.0085)	(0.0009)		
9	1.0621	1.0603	0.0019	0.47	
	(0.0125)	(0.0119)	(0.0026)		
10	1.0563	1.0554	0.0009	0.42	
	(0.0085)	(0.0085)	(0.0011)		
11	1.0503	1.0488	0.0015	0.22	
	(0.0064)	(0.0065)	(0.0013)		
12	1.0540	1.0499	0.0041	0.13	
	(0.0150)	(0.0150)	(0.0027)		
13	1.0542	1.0525	0.0017	0.67	
	(0.0125)	(0.0127)	(0.0040)		
14	1.0700	1.0653	0.0047	0.00	✓
	(0.0086)	(0.0084)	(0.0016)		
15	1.0891	1.0889	0.0002	0.93	
	(0.0142)	(0.0139)	(0.0023)		
16	1.0841	1.0812	0.0030	0.34	
	(0.0131)	(0.0128)	(0.0031)		
National	1.0383	1.0365	0.0018	0.00	✓
	(0.0024)	(0.0024)	(0.0004)		

The ratio of CE rate to match rate demonstrates the combined effect of matching error in the match rate and correct enumeration rate. The error in this ratio reflects the effect of matching

[15]See Appendix B for the DSEs.

error on the DSE. Matching error significantly inflated the production DSEs in five Preliminary Evaluation Poststratum groups: 1, 3, 7, 8, and 14.[16] Further, the national 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[17] (with a standard error of 92,877) than the MES DSE (see Table B3).

4.3 Were there clerical errors in identifying duplicates in the A.C.E. search area?

As discussed below, there were only minor errors in the coding of duplicates.

Note: The MES examined the clerical identification of duplicate cases only in the universe defined for production.

Table 9 looks at the coding of P-sample duplicates, which are one type of "remove from P-sample" (RP) match status classification. The table presents unweighted data. The production and MES duplicates are broken down by whether or not the cases are production errors that affect the total number of RP cases. There are two types of production duplicate errors: production wrongly classified the case as a duplicate (false duplicate) or production failed to find the duplicate altogether (missed duplicate). There are three types of cases which fall into the "no production error" category: production and MES both identified the case as a duplicate (agree completely), production found the duplicate/primary pair but picked the wrong person to be the primary according to the matching procedures (agree, but switched primary), or production classified the case as a duplicate when it should have been another type of "remove from P-sample" case or vice versa (agree, but different type of RP).

[16]If a Bonferroni multiple comparison test is used with alpha* = 0.10/16 ≈ 0.006, then the only poststratum groups with significant differences in DSEs are 8 and 14.

[17]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. Therefore, 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.

Table 9. 2000 Comparison of Production and MES P-sample Duplicate Coding

Production Results	N	%	MES Results	N	%
Total Duplicates	1584	100.0	Total Duplicates	1601	100.0
Production Error - False Duplicate	38	2.4	Production Error - Missed Duplicate	52	3.3
No Production Error	1546	97.6	No Production Error	1549	96.7
Agree Completely	1419	89.6	Agree Completely	1419	88.6
Agree, but Switched Primary	127	8.0	Agree, but Switched Primary	127	7.9
Agree, but Different Type of RP	0	0.0	Agree, but Different Type of RP	3	0.2

Of the P-sample duplicates production identified, approximately 2.4 percent were false duplicates which inaccurately increased the total number of production "remove from P-sample" cases. Of the duplicates MES identified, approximately 3.3 percent were missed by production which inaccurately diminished the total number of production RP cases.

Note: Cases which fall into the switched primary category appear in Table 1a as a gross error, because production is a remove and MES is a match, nonmatch, or unresolved match status (or vice versa). However, these differences are <u>not</u> an error when looking at the overall production or MES totals of any of the match status categories, because in terms of the net they balance themselves out. There could also be concern that switching of the primary could cause changes at the poststratum-level if the duplicate and primary are in different poststratum groups. However, there were few cases which could cause a change in poststratum groups and these cases do not appear to change any of the conclusions regarding the significance of differences in match rates or DSEs in Section 4.2.

Table 10 looks at the coding of E-sample duplicates, which are one type of "erroneous enumeration" (EE) enumeration status classification. The table presents unweighted data. The production and MES duplicates are broken down by whether or not the cases are production errors that affect the total number of EE cases. As with the P-sample, there are two types of production duplicate errors: false duplicate or missed duplicate. Again, there are three types of cases which fall into the "no production error" category: agree completely; agree, but switched primary; or agree, but different type of EE.

Table 10. 2000 Comparison of Production and MES E-sample Duplicate Coding

Production Results	N	%	MES Results	N	%
Total Duplicates	1504	100.0	Total Duplicates	1526	100.0
Production Error - False Duplicate	52	3.5	Production Error - Missed Duplicate	79	5.2
No Production Error	1452	96.5	No Production Error	1447	94.8
Agree Completely	1334	88.7	Agree Completely	1334	87.4
Agree, but Switched Primary	93	6.2	Agree, but Switched Primary	93	6.1
Agree, but Different Type of EE	25	1.6	Agree, but Different Type of EE	20	1.3

Of the E-sample duplicates production identified, approximately 3.5 percent were false duplicates which inaccurately increased the total number of production "erroneous enumeration" cases. Of the duplicates MES identified, approximately 5.2 percent were missed by production which inaccurately diminished the total number of production EE cases.

Note: Cases which fall into the switched primary category appear in Table 2a as a gross error, because production is an erroneous enumeration and MES is a correct enumeration or unresolved enumeration status (or vice versa). However, these 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. There could also be concern that switching of the primary could cause changes at the poststratum-level if the duplicate and primary are in different poststratum groups. However, there were few cases which could cause a change in poststratum groups and these cases do not appear to change any of the conclusions regarding the significance of differences in correct enumeration rates or DSEs in Section 4.2.

4.4 What other types of matching errors were there?

There were three other types of errors which 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, net difference, and relative 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 noted in Table 1a 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 the size of its complement, the nonmatch to match cell (105,281 versus 451,097).

Table 11 presents unweighted data to examine the match/remove difference and how it is affected by whole household (HH) removes (cases in households where all the data defined people were identified as "remove from P-sample", RP).

Table 11. 2000 Unweighted Match to Remove Versus Remove to Match

Production Results	N	%	MES Results	N	%
Match to Remove	132	100.0	Remove to Match	142	100.0
MES Whole HH RP	109	82.6			
MES Partial HH RP	23	17.4			

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 MES noninterview adjustment is one factor in the MES weight. For cases identified as whole household removes, the noninterview adjustment is zero and thus the weight is zero. Approximately 82.6 percent of the match to remove cases are whole household removes according to MES and therefore would have a zero MES noninterview adjustment and thus a zero MES weight. By contrast, all of the remove to match cases have nonzero MES noninterview adjustments (because they are matches according to MES). Therefore, the difference between the match to remove and the remove to match cells in Table 1a is mostly due to weighting.

Table 12 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.

Table 12. 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

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.[18] 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

[18]There were 4338 (unweighted) E-sample cases which had insufficient information for matching (which is approximately 2.6 percent of the E-sample). Of these insufficient cases, MES matchers were able to update and make sufficient for matching 854 cases (which is approximately 19.7 percent of the cases that were originally insufficient for matching).

and when to code them unresolved residency based on the Person Followup (PFU) knowledgeable respondent rules[19].

The E-sample gross difference, net difference, and relative difference rates are also less than one percent. Again, some interesting patterns emerge when examining the differences that do exist. Table 13 highlights three types of E-sample erroneous enumeration errors: differences due to duplicates (data on duplicates is from Table 10), error in updating E-sample cases with insufficient information for matching, and error in identifying discrepant cases. (Duplicates, cases with insufficient information for matching, and discrepant cases are three of the types of erroneous enumerations.) The table presents unweighted data.

Table 13. 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

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, these differences are <u>not</u> 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. (See Section 4.3 for more information on duplicates.)

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

As mentioned earlier, 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. Over 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 Person Followup (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 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 process.

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 Targeted Extended Search (TES) procedures and improve the quality control for these 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 made 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 control 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, Susanne (2001a). Study Plan for Evaluation N14: Evaluation of Matching Error. Internal Memorandum dated January 30,2001.

Bean, Susanne (2001b). Matching Error Study (MES) Clerical Matching Guide. PRED TXE/2010 Memorandum Series: CM-MES-S-02-R01 dated February 6, 2001.

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.

Ikeda, Michael (2000). Accuracy and Coverage Evaluation Survey: Specifications for Creating A.C.E. Final Outcome Codes and A.C.E. Interview Status Codes. DSSD Census 2000 Procedures and Operations Memorandum Series, Q-16 dated December 12, 2000.

Ikeda, Michael and McGrath, David (2001). Accuracy and Coverage Evaluation Survey: Specifications for the Missing Data Procedures. DSSD Census 2000 Procedures and Operations Memorandum Series, Q-62 dated July 9, 2001.

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

Keathley, Don (2001b). Post-Stratum Codes. PRED TXE/2010 Memorandum Series: CM-EFU-G-01-R01 dated June 27, 2001.

APPENDIX A

Unweighted Comparisons of Matching Results

Table A1. 2000 Unweighted Comparison of Production and MES Match Status for the P-sample

Production Results	MES Results				
	Match	Nonmatch	Remove	Unresolved	**Total**
Match	129,786	80	132	56	130,054
Nonmatch	319	18,333	190	27	18,869
Remove	142	125	3,150	6	3,423
Unresolved	33	15	1	2,340	2,389
Total	130,280	18,553	3,473	2,429	154,735

Table A2. 2000 Unweighted Comparison of Production and MES Enumeration Status for the E-sample

Production Results	MES Results			
	Correct Enumeration	Erroneous Enumeration	Unresolved	**Total**
Correct Enumeration	149,463	286	278	150,027
Erroneous Enumeration	241	9,924	204	10,369
Unresolved	185	216	6,321	6,722
Total	149,889	10,426	6,803	167,118

APPENDIX B

Components of the Match Rate and CE Rate and DSEs

Table B1. 2000 Match Rate Components by Preliminary Evaluation Poststratum Group

Preliminary Evaluation Poststratum	Baseline Results		MES Results	
	M	N_P	M	N_P
1	37,350,299	38,964,943	37,418,501	38,954,738
2	32,745,847	34,331,757	32,758,767	34,337,418
3	3,287,511	3,544,427	3,305,306	3,552,670
4	6,617,166	7,252,617	6,592,003	7,231,914
5	25,431,803	26,596,722	25,457,995	26,598,444
6	6,525,475	7,130,530	6,561,043	7,130,583
7	30,734,242	33,124,827	30,784,882	33,153,004
8	16,908,795	18,750,125	16,960,664	18,766,599
9	5,223,062	6,017,620	5,244,331	6,019,276
10	17,688,664	19,934,496	17,690,702	19,928,421
11	19,739,808	21,629,162	19,767,550	21,655,066
12	4,127,675	4,694,843	4,148,126	4,710,488
13	8,282,940	9,169,133	8,286,601	9,183,815
14	18,363,622	21,168,901	18,462,926	21,198,564
15	6,214,629	7,466,074	6,232,958	7,477,472
16	7,665,084	8,971,990	7,721,869	9,002,570

Table B2. 2000 Correct Enumeration Rate Components by Preliminary Evaluation Poststratum Group

Preliminary Evaluation Poststratum	Baseline Results		MES Results	
	CE	N_E	CE	N_E
1	37,205,708	38,139,126	37,220,109	38,139,126
2	33,637,605	34,749,855	33,627,642	34,749,855
3	3,255,879	3,427,076	3,250,704	3,427,076
4	6,380,273	6,654,664	6,386,009	6,655,230
5	26,404,055	27,151,244	26,407,117	27,156,961
6	7,198,170	7,542,513	7,199,532	7,542,513
7	32,248,508	33,537,935	32,219,002	33,537,935
8	16,531,590	17,693,220	16,514,083	17,693,220
9	5,499,810	5,965,837	5,512,237	5,967,158
10	17,938,142	19,138,181	17,928,956	19,136,878
11	21,096,433	22,008,718	21,069,850	22,008,718
12	4,953,749	5,345,943	4,944,326	5,347,959
13	8,858,422	9,301,933	8,833,834	9,301,933
14	19,580,572	21,095,511	19,572,886	21,096,013
15	7,310,354	8,063,939	7,319,353	8,063,939
16	8,257,137	8,914,878	8,267,215	8,914,878

Table B3. 2000 Dual System Estimates by Preliminary Evaluation Poststratum Group

Preliminary Evaluation Poststratum	Baseline DSEs	MES DSEs
1	35,746,742	35,686,049
	(243,836)	(244,062)
2	31,274,891	31,258,450
	(118,921)	(117,041)
3	5,287,759	5,263,141
	(56,986)	(58,525)
4	8,486,858	8,501,847
	(108,137)	(106,322)
5	25,819,449	25,792,116
	(181,188)	(179,661)
6	6,331,024	6,297,941
	(124,529)	(119,500)
7	34,702,416	34,643,076
	(291,442)	(288,417)
8	20,124,301	20,059,120
	(165,831)	(165,273)
9	6,978,065	6,965,841
	(81,870)	(77,872)
10	19,411,699	19,394,933
	(156,775)	(156,194)
11	24,958,488	24,921,867
	(151,234)	(153,795)
12	5,253,452	5,233,021
	(74,525)	(75,003)
13	9,830,531	9,814,605
	(116,518)	(118,091)
14	24,746,922	24,638,047
	(198,077)	(194,767)
15	7,611,245	7,609,805
	(99,085)	(97,009)
16	9,645,390	9,618,817
	(116,723)	(114,207)
National[20]	275,762,677	275,278,739
	(636,435)	(629,351)

[20]The difference between the national baseline and MES DSEs is 483,938 with a standard error of 92,877.