ESTIMATES OF UNDERCOUNT IN THE 1980 CENSUS

Charles D. Cowan and Robert E. Fay, U.S. Bureau of the Census

1. PEP Methodology

The Post Enumeration Program was conducted in two parts. The first, designated as the P-Sample, was designed to measure gross undercoverage in the 1980 census. As will be explained later, because of methodological problems the P-Sample actually overestimates gross undercoverage and an adjustment must be made in the estimation process. The second part, designated as the E-Sample, was designed to estimate gross overcoverage in the census; this would include duplicate and erroneous enumerations.

The P-Sample was collected as two supplements to the Bureau's Current Population Survey (CPS) in April and August of 1980. In each supplement a complete household roster was collected, including responses on demographic items like age, race, sex, ethnicity, and relationship in household. At the time of the interview or prior to it, a sketch map was obtained, locating the CPS housing unit relative to cross streets and major intersections. Using the information from the sketch maps and interviews, CPS housing units were located in the census, and persons within housing units were matched to the census roster for that housing unit. When a person or family could not be located, the focus of search was expanded to the whole enumeration district (ED), the area covered by one census enumerator. If a person or family still could not be found, a follow-up interview was conducted with the CPS household to determine if there was a reason why the person could not be found (e.g., lived at a different address April 1).

Using the information from the initial interview and the follow-up interview, a second attempt was made at matching. At this point in the matching process, a decision was made for each case as to whether it matched to the census, did not match to the census, or that no decision could be made because of insufficient information. Insufficient information cases included refusals in the CPS, cases which could not be located geographically in the census (the census ED could not be determined with certainty), and cases for which the follow-up information was incomplete (interviewer error or follow-up refusal). Treatment of these cases will be more completely described in the next section.

One methodological problem which plagued the matching occurred if the census household which was to be matched was incorrectly located in the census. When the CPS sketch map was obtained, the CPS household would be located using the sketch map in a census ED. If the corresponding census interview was allocated to the wrong location geographically, i.e., placed in the wrong census ED, then the search for the census household would fail. This implies that the household was correctly counted in the census, correctly enumerated in the PES, but could not be matched between the two sources because the information would be located in two different EDs and so stored separately. This was called a geocoding error. Geocoding errors would lead to overestimates of the undercount in the census.

In addition, there were other problems in the census that would lead to matching problems. Census households that were "closed-out" at the end of the census enumeration period or other census households that had not had enough information collected in the field were imputed in the census, but the imputation was done on the computer. This means a household or persons would be counted in the census but could not be matched because the census form did not have sufficient information to determine whether the persons were the same. Again, although the people in census imputed households were correctly enumerated in both sources, they could not be matched between sources, leading to overestimates of the undercount.

To measure overcoverage in the census, a second sample called the E-Sample was drawn from the census. The 110,000 households in this sample were drawn in the same PSUs used for the CPS. Both the P-Sample and the E-Sample contained sufficient samples to make estimates of under and over coverage for the 50 states and 16 large SMSAs. The E-Sample was used to measure three quantities: erroneous enumerations in the census, duplicates, and geocoding errors. After the sample was drawn for the E-Sample, an interview was conducted at the sample household in November, 1980. Questions were asked to determine if the census respondent had actually lived at that address at the time of the census, or should have been located elsewhere. The questionnaire was also designed to determine the number of persons listed in the census who were born after Census Day or died before Census Day. Finally, the questionnaire was used to uncover fabricated persons or households.

In addition, at the time of the interview a sketch map of the location of the E-Sample unit was made. The map was used to re-geocode the address of the unit, this geographic code was compared to the one assigned to the housing unit by the census. If, after reconcilation, the address in the census was found to be incorrect, this information was also recorded on the E-Sample questionnaire to document the number of geocoding errors in the census.

The third use of the E-Sample was to determine the number of duplicate enumerations in the census. This was done by comparing all E-Sample questionnaires in an ED (10 sample units) to all other census questionnaires in the same ED. A person by person match was conducted to determine the number duplicated in the census.

The information from the two samples was combined to make estimates of the total population by age, race/ethnicity, and sex groups within each state. These estimates were summed to the state level to yield estimates for states, and then further summed across states to give a national estimate of the undercount. A similar method was used to obtain estimates for 16 SMSAs and central cities.

Estimates of the total population are in

the same form as what is referred to as the dual system or capture-recapture estimator.

$$N_T^k = \frac{N_P^k (N_C^k - E^k - G^k - D^k - I^k)}{M^k}$$

where

N_T^k is an estimate of the total population in demographic/geographic subgroup k

N_P^k is the weighted sample total of the number of persons in the P-Sample in subgroup k

M^k is the weighted number of persons who match between the census and the P-Sample in subgroup k

N_C^k is the census count of persons in subgroup k

E^k is the weighted number of persons who were census erroneous enumerations from the E-Sample in subgroup k

G^k is the weighted number of persons in incorrectly geocoded housing units in the census from the E-Sample in subgroup k

D^k is the weighted number of duplicate counts in the census from the E-Sample in subgroup k

I^k is the count from the census of field-related imputations in subgroup k

The value N_T^k can be compared to the census count N_C to determine if the census value is low or high relative to the estimate of the population size.

2. Effect of Missing Data

2.1 The E-Sample

Since the net error of the census was generally only a few percent, the effect of missing data in PEP presents serious issues of interpretation, even though the rates of missing data are modest by typical survey standards. The P- and E-samples both encountered appreciable problems of missing data relative to the objective of estimating a small net census error; of the two, the issue of missing data for the E-sample is simpler and will be described first.

The E-sample provided measures of geocoding error, definitional error, and duplication in the census. Measurement of definitional error was the only component of the three requiring household interviews; collection of data on geocoding and duplication was performed by interviewing and clerical staff and consequently had low rates of incomplete data.

The survey questionnaire for the E-sample divided measurement of definitional error into three distinct phases: one section contained questions to be asked of a respondent from the original census housing unit; a second section was to be asked of neighbors if no respondent from the original housing unit could be contacted or would cooperate; and a third section was to be asked of the Post Office if no neighbor could be reached who knew of the proper address and composition of the census household on census day. Because information from "proxy" respondents - neighbors or the Post Office - was explicitly permitted by this design, the rate of complete non-response was much lower than typical household surveys requiring a household respondent. Unambiguous determination of whether the sampled census enumerations were definitionally correct required completion of a number of questions within the appropriate section of the questionnaire; approximately 1.5 percent of the sample had partial information recorded on the questionnaire that was insufficient to classify the sample case unambiguously.

After editing for logical relationships, the missing data corresponded to a "nested" or "monotone" pattern (described, for example, by Little (1982)). Essentially, the missing observations were imputed using an assumption of ignorable nonresponse (again described by Little (1982) and other references in missing data), taking advantage of the nested pattern of nonresponse to implement an imputation based on statistical matching.

In the application of statistical matching, cases with incomplete response were matched to completed cases with the same pattern of responses for the observed data; additional characteristics such as race/ethnicity, state/division/ region, household composition, and geographic location were also considered in attempting to obtain a similar match. To be consistent with the ignorable response model, the imputation was performed in waves, considering only specific patterns of response at a time. The first wave of imputation involved a statistical match of incomplete cases with observed responses complete up to the last applicable question to similar complete cases, leaving all other incomplete cases to the side. The second wave of imputation matched incomplete cases with observed responses, incomplete by at most two applicable questions, to complete cases and cases imputed in the first wave; etc.) (A more complete description is given by Fay (1984a).)

Out of approximately 3.5 million persons (a weighted estimate) requiring imputation approximately 1.0 million persons were imputed to the classification of definitionally incorrect enumeration under these procedures, but most of these cases corresponded to a relatively small number of specific situations. When a member of the original census household was interviewed, approximately 0.3 million were imputed as erroneous enumerations among the instances in which the person did stay at the address on April 1, 1980 but had another address on that date as well (but the correct address was not determined) or actually did not stay at the sample address on April 1, 1980 (but the appropriate census address was not determined). Where a neighbor was interviewed instead, approximately 0.2 million were imputed as erroneous enumerations from among the instances when the neighbor responded that the sample person stayed at the address but not on April 1, 1980 (but did not know the correct

address where the sample person should have been enumerated), and 0.1 million were imputed as erroneous when the sample person was unknown to the neighbor but no information was obtained from the Post Office. Another 0.3 million were imputed as erroneous enumerations for cases in which the Post Office did not know with certainty whether the person ever stayed at the address.

2.2. The P-Sample

The implications of missing data in the P-sample were even larger than those for the E-sample. The P-sample procedures consisted in principle of an initial attempt to match the census based on information obtained from the CPS interview; a follow-up interview, generally during the winter and spring of 1981, to obtain more complete information for cases not initially matched; and a final match to determine a final status of matched or not matched. There were three major sources of missing data in executing this design: inability to complete the follow-up interview; inability to determine the correct census ED from the follow-up information; and the existence of specific classes of cases not matched to census after an initial effort, for which no follow-up interviews were attempted.

Approximately 23 percent of the cases assigned to follow-up interviews from the April sample were incomplete, and the corresponding figure for August is approximately 19 percent. Furthermore, the classes of cases with the highest non-match rates to the census among the completed follow-up cases typically had higher than average rates of incompleteness. Adjustment for these incomplete cases depends upon the assumed mechanisms for nonresponse: the imputation model used for the estimates presented here assumed that nonresponse was independent of census match status conditional upon other observed covariates, such as race/ethnicity and the specific status of the initial match attempt before being sent to follow-up. In effect, the nonrespondents were assumed (conditional upon covariates) a random sample of the assigned follow-up cases.

Even if the follow-up interview was completed, some respondents could not provide adequate information on their April 1, 1980 address to identify a specific ED and allow a valid determination of matched or non-matched status. Since essentially all CPS sample housing units could eventually be coded to census geography, this issue is almost entirely restricted to "movers" - persons with different April 1, 1980 and CPS addresses. Since the correct EDs had never been searched for these persons, the assumption made by the imputation procedure was that such persons were missed at the same rate as movers with complete data.

The preceding two assumptions: the independence (conditional upon covariates) of census match status and nonresponse in follow-up, and that movers with incomplete geographic information would be missed from the census at the same rate as movers with complete geographic information, interact in a complex manner. A fuller discussion of this model was given earlier (Fay and Cowan 1983).

Additional classes of cases failed initially to match the census on the basis of available information, but no follow-up interview was attempted. The largest subset of these cases arose from attempting to match April 1980 CPS noninterview households with known composition to the census. Because this operation was conducted considerably after the original effort, no follow-up interviews were assigned for the nonmatching cases. In a general sense, nonmatches from this initial matching operation represented instances of missing data, since the effect of a follow-up interview to collect information that may have enabled a match is unknown. Such cases were treated in imputation in the same manner as follow-up noninterviews.

Several coverage evaluation efforts in conjunction with earlier census had set cases with incomplete information aside for purposes of estimation, in effect assuming that such cases were missing from the census at the same rate as all complete cases. The imputation procedures employed for PEP treated nonrespondents in follow-up as a random sample (controlling for specific covariates) from the population assigned follow-up interviews; this assumption effectively implies a much higher rate of unmatchable persons among the incomplete cases than the earlier assumptions. The issue here may be stated in terms of causal models for nonresponse (Fay and Cowan 1983), Fay 1984b).

No definitive validation of the imputation assumptions has been accomplished, but a study of incomplete cases from the April CPS for Washington, D.C. provides some evidence guardedly in support of the assumptions. Table 1 presents results from this study (Keeley 1984). Table 1 shows the total of 73 study cases divided into 56 cases interviewed in the April 1980 CPS and 17 cases that were noninterviews at that time. For each of these two groups separately, the table compares the treatment in estimation in PEP using the imputation procedures with the study outcome. (Four persons were included in the PEP household noninterview adjustment, and Table 1 shows the equivalent treatment in fractional persons.) Even after intense investigation, 32 percent (= 18/56) of the cases incomplete in the original follow-up remained unresolved, as did 24 percent (= 4/17) of CPS noninterviews that initially could not be matched to the census.

Among completed study cases, the imputation procedures appeared to have only modest success at predicting which cases were true nonmatches, but they were far more successful in estimating average level. For CPS-interviewed households, the imputation assigned a non-matched rate of 48 percent (= 18.4/38) compared to the observed rate of 63 percent (= 24/38). Similarly, for CPS noninterviews, imputation assigned a nonmatch rate of 85 percent (= 11/13) compared to an observed 92 percent (= 12/13). It is quite clear from these data that treating incomplete PEP cases as similar to the population as a whole (with a non-match rate less than 15 percent) would have given a gross underestimate of the true nonmatches among the incomplete PEP cases. Certainly, this study is too small to be taken as an absolute validation of the specific imputation procedures used for PEP, but it does show that modeling of mechanisms of

nonresponse in coverage studies can produce more appropriate estimates than simply setting aside incomplete cases.

3. Results

The previous section of this paper discussed the problems with missing data encountered in this survey. Because of the problems with missing data and interpretation of some of the portions of the E-Sample questionnaire, a variety of estimates were made that allowed different treatments of the data. This section will present the different estimates made, gross measures of undercount and overcount, and differences in estimates due to alternative methods of stratifying the estimates.

Different Estimates

In working with the E-Sample questionnaires as they returned from the field, it became apparent that responses to the last section of the questionnaire might be questionable. The initial sections of the questionnaire asked about persons listed in the census, whether these persons had ever lived at the census address sampled and, if so, had they lived at that address on Census Day. If the person listed on the census form could not be spoken to directly, the interviewers were instructed to obtain information from other persons living at the census address at the time of the census interview, next door neighbors, and apartment managers, in that order of preference for contracts. As a last resort, the interviewer was instructed to contact the postal carrier whose route covered the census address to ascertain if the person listed on the census form actually lived at the address. This was designed as a last ditch effort to differentiate between fabrications in the census and persons who had actually lived at the census address and moved out before the E-Sample contact.

However, in reviewing the responses of the postal carriers, it had become obvious that there were times that the postal carriers could not recall whether the census person had lived at the address, or was confused about the time the person had lived at the address. Consequently, all the estimates of the undercount were run two ways, once with the postal carrier responses as given, and once with all postal carrier responses imputed. The proportion of cases in the E-Sample requiring contact with the postal carriers was small, so the total number of cases imputed for this reason was small, but use or nonuse of the carrier responses makes a substantial difference in the estimates as presented.

For the missing data problems in the P-Sample, the predominant problem was how to deal with CPS noninterviews, including refusal cases and households where no one was home to be interviewed. For the April CPS supplement, for a large proportion of the households we also had interviews with the same households in March and May of 1980. For those households where the household roster was the same in both months, that roster was used for searching the census to cut back on the nonresponse for April. Consequently, for half of the April estimates we used the March/May rosters to search the census to determine if the household was in the census. The other half of the April estimates delete the households that were CPS noninterviews from the estimation process, and a weighting adjustment is made for these households. No corresponding inclusion of noninterview households was incorporated in the August estimates.

Furthermore, the August interviews had an additional problem because of the large proportion of persons who had moved between April and August. There was a concern that movers may suffer a significant recall bias in remembering and reporting their April address. To study this problem, some August estimates were run with all the mover information collected used, and other August estimates drop the mover information from the estimates for all movers, and a noninterview weighting adjustment is made for these cases.

Table 2 presents the results at the national level for different race/ethnic groups. The emphasis in this table is the difference between estimates due to missing data problems, methodological problems, and different responses in the two sample months.

In Table 2, one can see that there is a substantial range in the net undercount rate, a range of 0.2 percent to 2.0 percent for the total population, and broader ranges for minority groups of 2.7 percent to 6.7 percent for Blacks, and 3.6 percent to 7.6 percent for Nonblack-Hispanics. For the other category, which is predominantly Whites, the estimates range from an overcount of 0.4 percent to a net undercount of 1.1 percent. The biggest differences in the estimate come from the deletion of movers in the August sample, and in the use of April vs. August. Note especially that from April to August the estimates of the undercount decline for Blacks, but they increase for Hispanics. Other changes have less signficant effects.

Gross Estimates of the Undercount and Overcount

Table 3 presents gross undercount and gross overcount estimates. As explained in the first first section of this paper, the net undercount rate is determined as a function of the two gross gross rates. The net undercount rate can be written as the ratio of the coverage rate in the census to the correct enumeration rate in the census. Algebraically, this can be written as

$$\text{Net Undercount Rate} = 1.0 - \frac{1.0 - \text{Gross Undercount Rate}}{1.0 - \text{Gross Overcount Rate}}$$

As can be seen from Table 3, the gross undercount rate is more variable than the gross overcount rate. It is also interesting to note that the two rates do not move in concert. For example, comparing the Northeast to the West, although the gross undercount rate is higher in the West than in the Northeast, the gross overcount rate is lower. Both of these changes from West to Northeast tend to cause the net undercount rate to increase.

4. The E-Sample as an Evaluation of the 1980 Census

.... then we must speak
Of those counted not wisely but too well.

Although the design of the E-Sample was principally dictated by the intent to balance gross omissions from the P-Sample by an estimate of erroneous enumerations defined in such a way as

to estimate net census error, the E-Sample results by themselves, particularly when supplemented by additional data from the Housing Unit Coverage Study (HUCS), provide an interesting perspective on characteristics of the 1980 census.

The three components of overenumeration measured by the E-Sample were initially treated as a hierarchy: geocoding error, definitionally incorrect enumeration, and duplication. Cases classified as geocoding errors were enumerated in enumeration districts (EDs) in the census outside the set of EDs constituting the probable area of search of the P-Sample. Cases not classified as geocoding errors were classified as definitionally incorrect enumerations if the sample persons should not have been enumerated at that address by census definitions. Lastly, duplicates were measured only among persons not classified as erroneous enumerations in the preceding two senses, and only within the probable area of search for the P-Sample.

This initial estimation strategy served the primary purpose of avoiding double counting of persons as erroneous enumerations for more than one reason, but obscured the interaction between these sources of error. In particular, one potentially important interaction of this sort was between geocoding and duplication, since a housing unit enumerated in the wrong ED may also have been enumerated in the correct ED because of coverage improvement efforts, or vice versa.

The Housing Unit Duplication Study (HUCS) took cases from the E-Sample that had been geocoded to more than one ED in the handling of the case and searched for between ED enumerations; by proper weighting of these data it is possible to estimate instances of duplicate enumerations between EDs (with the exception of the presumably small proportion of cases enumerated in two incorrect EDs outside of the P-Sample area of search, with no enumeration in the correct ED).

Table 4 presents the combined information from the original PEP E-Sample set 8 which includes the data obtained from the Post Office, and the results on between ED enumeration from HUCS. For purposes of discussion, the first two columns of the table, showing national totals, are first considered. The estimate of total erroneous enumerations, 3.4 percent, is the sum of 1.0 percent classified as geocoding errors, 1.6 percent classified in PEP as definitionally incorrect enumerations, and .8 percent as duplicates.

Approximately 30 percent of those classified as geocoding errors were duplicated in another ED, making the estimate of total duplicates from this study approximately 1.1 percent overall. (This estimate does not include, however, any allowance for persons counted in two entirely different locations, such as duplicate enumeration of college students at a school and a home or the enumeration in census follow-up of persons moving into an address after April 1 who were also enumerated at their April 1 address.) Total duplicate enumerations measured by the study are divided between about 17 percent within the same ED and .4 percent between EDs. The estimate of 1.6 percent for definitional erroneous enumerations was composed of .4 percent classified on the basis of either complete or partial responses from the Post Office, and 1.2 percent from other sources.

Table 4 divides the national total into three components: the relatively small portion of the country enumerated by conventional census procedures, persons in households enumerated on mail returns (according to the classification 1980 census computer file) and those enumerated on non-mail returns. Some differences are striking. Both conventional areas, 1.7 percent, and mail returns in mail areas, 2.5 percent, show estimates below the overall average. Non-mail returns in mail areas, that is, households enumerated by census follow-up procedures, show higher estimates of percent overenumeration in all three major categories, with the most dramatic dramatic differences with respect to definitional errors and duplications.

The higher definitional errors on non-mail returns have a number of explanations. Mail returns essentially represent information obtained over a relatively short period of time, and the conventional census was generally substantially completed more quickly than the census in mail areas. By contrast, non-mail returns in mail areas were obtained through the protracted effort of census follow-up, which required several months. This temporal difference creates a situation in which errors of residence are far more likely, such as enumeration of persons moving into housing units vacant on April 1. Another factor is the effect of "curbstoning" enumeration of fictitious persons by census enumerators in place of proper data collection. To the extent that the "curbstoned" cases may have been entirely fabricated, including names, the E-Sample procedures would have required referral of such cases to the Post Office for final determination. Since there is little or no incentive for "curbstoning" of mail returns, it is not suprising to find over 60 percent of all erroneous enumerations classified as the basis of Post Office information to have been on non-mail returns, even though non-mail returns accounted for less than 20 percent of the overall count.

Estimates for duplicate enumerations tell another interesting story. Although the relative proportion of duplication within EDs is quite different, the absolute estimated numbers of duplicates is quite close between mail and non-mail returns in mail areas (761.0 v. 773.7 thousand). Furthermore, this relationship consistently appears in various partitions of the data that have been attempted (urban, rural, central cities of SMSAs balance of SMSAs, non-SMSA urban, non-SMSA rural). A mechanism to account for this consistent pattern is that most duplicates may occur as a household enumerated on a mail return was re-enumerated in follow-up. Households enumerated in the same EDs on mail/non-mail pairs of questionnaires have equal probability of contributing to the estimated duplicates for mail and non-mail returns.

References

Fay, R. (1984a) "Results from the 1980 PEP E-Sample Meaurement of Erroneous Enumeration," unpublished memorandum dated July 30, 1984.

(1984b) "Causal Models for Patterns of Nonresponse," manuscript submitted to the Journal of the American Statistical Association.

and Cowan, C. (1983) "Missing Data Problems in Coverage Evaluation Studies," Proceedings of the Section on Survey Research Methods, American Statistical Association, Washington, D.C.

Keeley, Catherine (1984) "Preliminary Evaluation Results Memorandum No. 74, the Post Enumeration Program Unresolved Cases Study Pretest," unpublished U.S. Census Bureau memorandum.

Little, R.S.A. (1982) "Models for Nonresponse in Sample Surveys," Journal of the American Statistical Association, 77, 237-250.

Table 1. Summary of Results from the Unresolved Cases Study Pretest

Study Results	PEP Imputation/Weighting: Matched	Non-Matched	Total
	Cases Originally Assigned to Follow-Up		
Matched	9.6	4.4	14
Non-Matches	10	14	24
Unresolved	4	14	18
Total	23.6	32.4	56
	Cases from CPS Noninterviews		
Matched	1	0	1
Non-Matched	1	11	12
Unresolved	2	2	4
Total	4	13	17

Table 2: Estimates of the Net Undercount Using Different Methods for Dealing with Data Problems

	April: CPS Nonint Retained		CPS Nonint Deleted	
	Post Off Retained	Post Off Deleted	Post Off Retained	Post Off Deleted
Total	1.0%	1.4%	0.8%	1.3%
Black	5.6	6.7	5.2	6.3
Nonblack Hispanic	4.4	5.6	4.1	5.3
Other	0.0	0.3	-0.1	0.2

	August -CPS Noninterviews Deleted: Movers Retained		Movers Deleted
	Post Off Retained	Post Off Deleted	Post Off Retained
Total	1.6%	2.0%	0.2%
Black	4.3	5.4	2.7
Nonblack Hispanic	6.5	7.6	3.6
Other	0.8	1.1	-0.4

Table 3: Estimates of Gross Undercount and Gross Overcount for Regions (Gross Undercount/ Gross Overcount)

	April: CPS Nonint Retained		CPS Nonint Deleted	
	Post Off Retained	Post Off Deleted	Post Off Retained	Post Off Deleted
U.S.	5.6/3.3	5.6/2.9	5.4/3.1	5.4/2.9
NE	5.1/3.6	5.1/3.3	5.0/3.6	5.0/3.3
NC	4.3/2.4	4.3/2.1	4.2/2.4	4.2/2.1
S	6.5/4.1	6.5/3.6	6.4/4.1	6.4/3.6
W	6.1/2.6	6.1/2.0	5.9/2.6	5.9/2.0

	August - CPS Noninterviews Deleted: Movers Retained		Movers Deleted
	Post Off Retained	Post Off Deleted	Post Off Retained
U.S.	5.1/3.3	6.1/2.9	4.7/3.3
NE	5.3/3.6	5.3/3.3	4.2/3.6
NC	4.5/2.4	4.5/2.1	3.4/2.4
S	7.6/4.1	7.6/3.6	5.9/4.1
W	6.7/2.6	6.7/2.0	5.0/2.6

Table 4 Components of Estimated Census Overenumerations and Duplications from the 1980 PEP E-Sample (Weighted Estimates in Thousands)

	Total U.S. Est	Total U.S. Pct	Conventional Est	Conventional Pct	Mail Return in Mail Area Est	Mail Return in Mail Area Pct	Non-Mail Return in Mail Area Est	Non-Mail Return in Mail Area Pct
Estimated complete enumerations (excluding census imputations for noninterviews)	220802.7	100.0	8791.0	100.0	173881.1	100.0	37992.7	100.0
Estimated erroneous enumerations								
Total	7581.8	3.4	145.2	1.7	4299.5	2.5	3121.3	8.2
Geocoding errors in census								
Total	2104.1	1.0	27.7	.3	1484.7	.9	583.0	1.5
Matching Duplicate in correct ED	618.0	.3	.0	.0	406.3	.2	211.7	.6
Other	1486.1	.7	27.7	.3	1078.4	.6	371.3	1.0
Definitionally incorrect								
Total	3602.7	1.6	98.6	1.1	1841.5	1.1	1658.1	4.4
Based on Post Office response	982.0	.4	21.3	.2	346.9	.2	611.3	1.6
Other	2620.7	1.2	77.3	.9	1494.7	.9	1046.9	2.8
Duplication								
Total	1874.9	.8	18.9	.2	973.3	.6	880.2	2.3
Within the same ED	1551.1	.7	16.4	.2	761.0	.4	773.7	2.0
Between ED's in the same area of search	323.8	.1	2.6	.0	212.2	.1	106.4	.3