# ANALYSIS OF CENSUS OMISSIONS: PRELIMINARY RESULTS

**Christopher L. Moriarity, National Center for Health Statistics**
**Danny R. Childers, U.S. Bureau of the Census**
**Christopher L. Moriarity, NCHS, 6525 Belcrest Road, Room 915, Hyattsville, MD 20782**

KEY WORDS: undercount, Post-Enumeration Survey, Housing Unit Coverage Study

## I. Introduction

This paper reports preliminary results of an analysis of data from the 1990 Post-Enumeration Survey (PES). The purpose of the analysis was to search for factors that may have been related to the omission of persons from the 1990 Census.

The PES was designed to measure the net coverage of the population in the 1990 Census. As such, the PES does not provide a direct estimate of census omissions, i.e., persons not included in the census count who should have been included. Hogan (1993) provides a complete background discussion of the 1990 PES.

The PES consists of two samples: the P-sample and the E-sample. The P-sample consists of all persons enumerated via PES personal visit interviews in the PES sample areas and is designed to furnish information about omissions. The E-sample consists of census enumerations in the PES sample areas and is designed to furnish information about erroneous enumerations. The persons in the two samples are matched after the PES interviews. For a variety of reasons, some P-sample people are nonmatches. One reason is that there is no corresponding census enumeration, i.e., the person was omitted from the census. Other reasons are:

1. The corresponding census person had insufficient information for matching.
2. The case was unresolved, even after a followup operation.
3. The P-sample person is compared to census enumerations (both E-sample persons and non E-sample persons) only in a specified search area.
4. Errors occurred during the matching process, causing a P-sample person who should have been matched to be a nonmatch.

The dual system estimator used to estimate net census coverage does not require the PES interview process to compile a complete list of persons in the PES sample areas. Although attempts are made to make the P-sample coverage as wide as possible, some persons omitted from the census are likely to also be omitted from the P-sample.

Any analysis with a goal of learning more about census omissions must recognize that although PES P-sample nonmatches provide valuable information about census omissions, not all P-sample nonmatches represent census omissions. Conversely, the P-sample does not necessarily include everyone in PES sample areas that were omitted from the census.

## II. Limitations

We will not address the following issues in this paper:

1. Persons omitted from the census who were also omitted from the P-sample ("correlation bias").
2. Errors during the matching process.
3. P-sample persons not matched because the search area was limited. This includes cases where the matching census person was assigned to incorrect census geography outside the search area ("geocoding error").

## III. Methodology - General

We used the imputed match probabilities from the PES estimation process for P-sample persons whose match status was unresolved. We used the imputed classification data (e.g., tenure, race/ethnicity) contained in the P-sample. Note that the imputation rates were low.

We used the weights from the PES estimation process for all of our estimates.

We classified census enumerations on Advance Census Reports in "List/Enumerate" areas in the same group as census enumerations returned by mail in other areas. The List/Enumerate procedure includes an enumerator visit to all households and was done in sparsely populated rural areas in the 1990 Census. Advance Census Reports were delivered by the U.S. Postal Service in List/Enumerate areas prior to the enumerator visit. For mail returns and Advance Census Reports, the method of enumeration is "self-enumeration". Minimal involvement by enumerators occurs in both cases, even though the Advance Census Reports are picked up by a census enumerator.

All estimates of uncertainty, such as standard errors and covariances, were computed using VPLX, a general-purpose variance estimation software package developed by Robert E. Fay, Senior Mathematical Statistician at the Census Bureau (Fay, 1990). VPLX generated design-based stratified jackknife estimates.

We performed all hypothesis tests at a confidence level of 10 percent. We did not employ a multiple comparison methodology for our hypothesis tests.

## IV. Overall Results for P-sample Nonmatches

The P-sample nonmatches can be grouped into several categories, using match codes assigned during the matching process. These match codes permit some interpretation of nonmatches, although some errors occurred during the assignment of the codes (Davis, 1991). We do not use the match codes to group our results in this paper.

Overall, the weighted estimates from the PES P-sample are approximately 18.8 million nonmatches out of approximately 241.2 million persons (7.8 percent nonmatch rate).

In addition to the possibility of classifying nonmatches using the match codes, another meaningful subdivision of the P-sample is possible. Since the PES interviews were conducted after the census, some persons in the P-sample had moved between Census Day and the PES interviews. Such persons were called "inmovers". For a small number of P-sample persons, it could not be determined if the person was an inmover or not. This leads to the three classifications (P-sample final status) "nonmover", "inmover", and "no status".

For nonmovers in the PES P-sample, the weighted estimates are approximately 13.8 million nonmatches out of approximately 221.4 million persons (6.2 percent nonmatch rate).

For inmovers, the weighted estimates are approximately 4.7 million nonmatches out of approximately 18.9 million persons (24.8 percent nonmatch rate).

For no status persons, the weighted estimates are approximately 0.3 million nonmatches out of approximately 1.0 million persons (29.7 percent nonmatch rate).

Note that 73.5 percent of the nonmatches were nonmovers, 25.0 percent of the nonmatches were inmovers, and 1.5 percent of the nonmatches were "no status" cases. This contrasts with 91.8 percent, 7.8 percent, and 0.4 percent of total P-sample persons being nonmovers, inmovers, and "no status" cases, respectively.

Comparison of these results makes it clear that the distribution of nonmatches is very different by P-sample final status classification. Inmovers and no status cases have much higher nonmatch rates than nonmovers. This could be an indication of higher omission rates for these two groups and/or problems with PES attempts to match these persons to the census.

Tables 1 through 4 summarize the weighted distributions of P-sample total persons and nonmatches cross-classified by tenure, age, race/ethnicity, or sex, and by P-sample final status classification. Here, the tenure, age, race/ethnicity, and sex classifications are based on the P-sample.

**Table 1: Tenure by P-sample final status**

| Total Persons: | Overall | Nonmovers | Inmovers | No Status |
|---|---|---|---|---|
| Owner | 68.2% | 70.7% | 40.2% | 41.8% |
| Renter | 31.8% | 29.3% | 59.8% | 58.2% |
| **Nonmatches:** | **Overall** | **Nonmovers** | **Inmovers** | **No Status** |
| Owner | 44.5% | 50.2% | 28.1% | 36.9% |
| Renter | 55.5% | 49.8% | 71.9% | 63.1% |

As one might expect, renters are a majority of the inmover and no status persons.

Note that although renters make up less than 1/3 of the population, 55.5 percent of the nonmatches were renters.

**Table 2: Age by P-sample final status**

| Total Persons: | Overall | Nonmovers | Inmovers | No Status |
|---|---|---|---|---|
| 0-17 | 26.3% | 26.4% | 25.3% | 21.5% |
| 18-29 | 18.9% | 17.1% | 39.9% | 25.5% |
| 30-49 | 30.1% | 30.6% | 24.1% | 29.0% |
| 50+ | 24.7% | 25.9% | 10.8% | 24.0% |
| **Nonmatches:** | **Overall** | **Nonmovers** | **Inmovers** | **No Status** |
| 0-17 | 28.6% | 29.5% | 26.4% | 21.7% |
| 18-29 | 30.3% | 26.5% | 41.0% | 33.4% |
| 30-49 | 27.1% | 28.0% | 24.2% | 28.8% |
| 50+ | 14.1% | 15.9% | 8.4% | 16.1% |

Table 2 shows that persons aged 18-29 make up a larger share of the inmovers and no status cases, relative to their overall share of the population.

Although persons aged 18-29 are a smaller proportion of the population than any of the other three groupings used here, 30.3 percent of the nonmatches came from this group. This was the highest proportion from the four age groups.

**Table 3: Race/Ethnicity by P-sample final status**

| Total Persons: | Overall | Nonmovers | Inmovers | No Status |
|---|---|---|---|---|
| Black | 11.0% | 11.0% | 10.3% | 12.0% |
| Nonblack Hispanic | 9.2% | 9.1% | 10.1% | 8.2% |
| All Other | 79.8% | 79.8% | 79.6% | 79.8% |
| **Nonmatches:** | **Overall** | **Nonmovers** | **Inmovers** | **No Status** |
| Black | 20.0% | 21.8% | 15.2% | 12.2% |
| Nonblack Hispanic | 14.9% | 15.5% | 13.6% | 8.4% |
| All Other | 65.1% | 62.7% | 71.3% | 79.5% |

Table 3 shows very little change in race/ethnicity across the P-sample final status. Note that "all other" includes all persons who were not Black and not Hispanic.

The proportions of nonmatches for Blacks and Nonblack Hispanics exceed their proportions in the population - 20.0 percent and 14.9 percent, respectively. The nonmatch proportions are in much closer agreement with population proportions for inmovers and no status cases than for nonmovers.

**Table 4: Sex by P-sample final status**

| Total Persons: | Overall | Nonmovers | Inmovers | No Status |
|---|---|---|---|---|
| Male | 48.6% | 48.5% | 49.7% | 53.4% |
| Female | 51.4% | 51.5% | 50.3% | 46.6% |
| **Nonmatches:** | **Overall** | **Nonmovers** | **Inmovers** | **No Status** |
| Male | 52.4% | 52.4% | 52.1% | 56.2% |
| Female | 47.6% | 47.6% | 47.9% | 43.8% |

Males appear to make up a slightly larger proportion of inmovers and no status cases, relative to their proportion in the general population.

Males were consistently a slight majority of the nonmatches in each P-sample final status grouping.

We focus below on additional analysis of P-sample nonmovers in housing units where a link could be made to a census unit (housing unit or group quarters). These results are of great value, but do not by themselves tell the full story of omission information contained in the P-sample. The inmovers and no status cases contain valuable information that merit study at some future time.

The analysis described below was inspired in part by similar work by one of us (Childers) on Housing Unit Coverage Study (HUCS) data. Results have been released as internal Census Bureau memoranda (Childers, 1993). The HUCS results are not directly comparable to the results presented here, so no comparison is made.

## V. Methodology For Linking P-sample Nonmovers to Census Data

We used information from the PES matching process. Where available, the information linked P-sample persons in households to the census unit (housing unit or group quarters). Some P-sample persons could not be linked to a census unit and dropped out of subsequent processing steps. For example, P-sample persons in housing units missed by the census dropped out at this stage.

For all P-sample households where some persons were linked to a census unit and some were not, we assigned the same link to all persons not linked during the PES matching process. However, we took no action if there were links to more than one census housing unit (this was not a common occurrence, but it did occur).

Note that P-sample nonmatches can be viewed as "whole household" nonmatches versus "partial household" nonmatches. (The classification of a P-sample household into one of these two groups is complicated, however, by the presence in some P-sample households of persons whose final match status was unresolved. These persons had a match probability imputed.) The information from the PES matching process permitted a linkage to occur for some "partial household" nonmatches and some "whole household" nonmatches.

After linking as many P-sample persons to census enumerations as possible, we then merged the linked persons to reference files containing census data. These data were data for housing units in the PES sample blocks from automated census files, and data

(enumeration date and respondent) obtained from a clerical review operation of census questionnaires. P-sample persons linked to census group quarters or census housing units outside the PES sample blocks dropped out at this stage. (Note that reference files containing census data outside the PES sample blocks were not readily available.)

Note that the nonmatch rate for P-sample nonmover cases who had dropped out up to this point was approximately 40 percent. These cases merit study in future research.

In addition to the processing steps described above, we took steps to identify some P-sample nonmatches that may have matched to census enumerations if sufficient information had been available. As indicated in the introduction, some P-sample nonmatches occurred only because the corresponding E-sample person had insufficient information. That is, these persons were not omitted in the census.

We looked at each linked household where E-sample persons with insufficient information occurred. If the P-sample person count was less than or equal to the census person count, and the E-sample count of persons with insufficient information was greater than or equal to the P-sample nonmatch count, we considered the P-sample nonmatches "matched" to the E-sample cases with insufficient information. We omitted all P-sample persons in these households from subsequent analyses (1687 unweighted P-sample persons, approximately 1.0 million weighted P-sample persons).

This procedure was not rigorous. For example, we did not carefully compare any characteristics that may have been present such as age, sex, or race/ethnicity. We do not claim that our effort is the best possible "match" between P-sample nonmatches and E-sample cases with insufficient information. We do feel that, if anything, we have been conservative. It is likely that a more rigorous "matching" process would have identified some additional P-sample households where the P-sample nonmatches are "balanced" by the E-sample persons with insufficient information for matching (these are treated like erroneous enumerations in the dual system estimator). However, we feel that we have obtained a good first approximation and can refer to the remaining P-sample nonmatches that we analyzed as "omissions".

Overall, we "matched" approximately 0.8 million (weighted) nonmatches to E-sample cases with insufficient information. This is approximately 4.1 percent of total nonmatches, and approximately 5.6 percent of nonmover nonmatches.

## VI. Results for P-sample Nonmovers in Linked Households, Not "Matched" to E-sample Cases With Insufficient Information

The overall estimates for P-sample nonmovers in linked households, not "matched" to E-sample cases with insufficient information, are approximately 6.1 million omissions (nonmatches) out of approximately 203.5 million persons (3.0 percent). Note that even though approximately 5/6 of the weighted P-sample total is present, only about 1/3 of the weighted nonmatches are present. Comparing to the universe of P-sample nonmovers, about 92 percent of the weighted total and about 44 percent of the weighted nonmatches are present.

All omission results given below are for P-sample nonmovers in linked households, not "matched" to E-sample cases with insufficient information. For the sake of brevity, this will not be repeated.

We have chosen below to present results separately for P-sample persons linked to census vacant units versus P-sample persons linked to census occupied units. Note that most or all of the rates for persons linked to households enumerated by a personal visit that are quoted in Section VI.B below would rise if they were recomputed using P-sample persons linked to both occupied and vacant census units. The rates that we present below for persons linked to households enumerated by mail would not change, or change by a very small amount.

### A. Results for Census Vacant Units

One result that can be obtained from this linkage process is an estimate of the number of P-sample persons omitted from the census because the associated census housing unit was erroneously enumerated as vacant. We estimate that approximately 1.6 million persons ($\sigma$=0.1 million) were omitted from the census for this reason.

This is a surprising result. The 1990 Census included a special program where most census units enumerated as vacant were revisited to confirm the vacancy status. Although this operation did convert a substantial number of housing units from vacant to occupied (Sledge, 1992), many persons were still missed in housing units considered to be "vacant" by the census.

### B. Results for Census Occupied Units

Analyses of PES E-sample data that examined erroneous enumeration (EE) rates by census operation variables and other variables were reported in Griffin and Moriarity (1992) and Moriarity (1993). We report analogous P-sample missed person rates for these variables. We have rounded all miss rate estimates and standard error estimates to the nearest tenth of a percent.

#### 1. Miss Rates by Method of Enumeration

Griffin and Moriarity (1992) reported large differences in EE rates by enumeration method. Persons enumerated by mail had a significantly lower EE rate than persons enumerated by a personal visit.

We found an analogous result for omissions. The miss rate was 1.3 percent ($\sigma$=0.1 percent) for P-sample persons in housing units that were enumerated in the census by mail, and 5.0 percent ($\sigma$=0.2 percent) for P-sample persons in housing units that were enumerated by a personal visit. These rates are significantly different.

Some census units had persons enumerated both by mail and by personal visit. All P-sample persons linked to such households were assumed to be enumerated by mail, since it would have been difficult or impossible to determine which P-sample persons were enumerated by which method.

For the sake of brevity below, we use less lengthy and precise phrases such as "the miss rate for mail returns" instead of "the miss rate for P-sample persons in housing units that were enumerated in the census by mail". ALL rates given below are PERSON miss rates.

#### 2. Miss Rates by Tenure

Griffin and Moriarity (1992) reported that EE rates for renters were significantly higher than the EE rates for owners, across enumeration methods.

We found that the miss rates behaved in a similar fashion. Overall, the miss rate for renters was 3.8 percent ($\sigma$=0.2 percent), 1.6 percent ($\sigma$=0.1 percent) for owners. For households enumerated by mail, the miss rate for renters was 2.1 percent ($\sigma$=0.1 percent), 1.1 percent ($\sigma$=0.1 percent) for owners. For households enumerated by a personal visit, the miss rate for renters was 6.5 percent ($\sigma$=0.4 percent), 3.8 percent ($\sigma$=0.2 percent) for owners. All of the miss rates by tenure are significantly different.

Note that the tenure classification we use in this section came from the census.

#### 3. Miss Rates by Form Length

Griffin and Moriarity (1992) reported that EE rates were not significantly different by form length ("short form" versus "long form") for persons enumerated by mail. The EE rate for persons enumerated by personal visit was higher when the persons were enumerated on short forms than when the persons were enumerated on long forms.

Overall, the miss rates were 2.3 percent ($\sigma$=0.1 percent) for P-sample persons linked to census units that were enumerated on short

forms, 2.0 percent ($\sigma$=0.1 percent) for P-sample persons linked to census units that were enumerated on long forms. These rates are significantly different. For persons enumerated by mail, the miss rate was 1.3 percent ($\sigma$=0.1 percent) when a short form was used and 1.3 percent ($\sigma$=0.1 percent) when a long form was used. These rates are not significantly different. For persons enumerated by personal visit, the miss rate was 5.4 percent ($\sigma$=0.3 percent) when a short form was used and 3.5 percent ($\sigma$=0.3 percent) when a long form was used. These rates are significantly different.

These results suggest that the use of long forms did not cause higher miss rates.

## 4. Miss Rates in Households Enumerated Via Last Resort/Closeout Procedures

As discussed in Griffin and Moriarity (1992) and Moriarity (1993), last resort/closeout procedures that occurred at the end of the period of enumeration by personal visit had the effect of increasing EE rates. We found a similar effect here. The miss rate was 4.1 percent ($\sigma$=0.2 percent) for P-sample persons linked to census units where regular personal visit procedures were used, while the rate was 18.5 percent ($\sigma$=2.2 percent) when last resort/closeout procedures were used. These rates are significantly different.

To provide some perspective, approximately 3.1 million P-sample persons were linked to households enumerated via last resort/closeout procedures, and approximately 0.6 million of these persons were missed.

## 5. Miss Rates by Enumeration Date - Households Enumerated by Mail

Griffin and Moriarity (1992) reported that EE rates for persons enumerated by mail generally rose over time, using the enumeration date recorded on the back of the census form as the time reference.

We noted a similar trend for miss rates. Table 5 presents miss rates for four weekly time periods around Census Day, April 1, 1990. The total number of P-sample persons linked to census households enumerated in the time period is also given. Standard errors of the miss rates are given in parentheses.

**Table 5**

| Time period | Miss Rate(%) | Total Persons |
|---|---|---|
| Week of 3/18 | 1.0 (0.1) | 36.7 million |
| Week of 3/25 | 1.3 (0.1) | 68.3 million |
| Week of 4/1 | 1.6 (0.2) | 16.4 million |
| Week of 4/8 | 2.1 (1.0) | 3.3 million |

The miss rate for the week of April 8 is not significantly different from the other rates. The other rates are all significantly different from each other. These data suggest that miss rates increased with time; those who filled out their census form early gave the most complete and accurate household roster.

## 6. Miss Rates by Enumeration Date - Households Enumerated by Personal Visit

Griffin and Moriarity (1992) reported that the EE rates for persons enumerated by personal visit rose over time. Table 6 uses the same biweekly periods given in Griffin and Moriarity (1992) to present miss rates by enumeration date. The total number of P-sample persons linked to census households enumerated in the time period is also given. Standard errors of the miss rates are given in parentheses.

**Table 6**

| Time period | Miss Rate(%) | Total Persons |
|---|---|---|
| 4/15-4/28 | 1.6 (0.3) | 2.0 million |
| 4/29-5/12 | 3.0 (0.2) | 10.4 million |
| 5/13-5/26 | 4.1 (0.3) | 13.5 million |
| 5/27-6/9 | 7.4 (0.7) | 10.1 million |
| 6/10-6/23 | 9.1 (1.6) | 3.6 million |
| 6/24-7/7 | 9.5 (1.5) | 1.6 million |

The miss rates for the first three time periods are significantly different from all other rates. The rates for the last three time periods are not significantly different. These data suggest that the proportion of missed persons increased over time.

## 7. Miss Rates by Census Respondent

Griffin and Moriarity (1992) reported that the EE rates for households where the respondent was a household member were lower than for households where the respondent was a "proxy" (not a household member). This difference was consistent across method of enumeration.

For mail returns, the miss rate was 1.3 percent ($\sigma$=0.1 percent) with a household respondent, 6.9 percent ($\sigma$=2.1 percent) with a proxy respondent, and 1.8 percent ($\sigma$=0.2 percent) when the respondent was missing or could not be classified. All of these rates are significantly different.

For enumerator returns, the miss rate was 3.7 percent ($\sigma$=0.2 percent) when a household member was the respondent, 19.2 percent ($\sigma$=2.4 percent) when the respondent was a proxy, and 9.3 percent ($\sigma$=1.1 percent) when the respondent was missing/unidentifiable. Again, all of these rates are significantly different.

These results clearly suggest that a household member is the preferred respondent.

## 8. Miss Rates by the Presence/Absence of Nonrelatives in the P-sample Household Roster

Moriarity (1993) reported that the presence of nonrelatives on the census roster appeared to be a factor of higher EE rates on mail returns, but not on enumerator returns.

We find that the miss rate for mail returns is 1.1 percent ($\sigma$=0.1 percent) when the P-sample household consisted of related persons, 4.7 percent ($\sigma$=0.3 percent) when nonrelatives were present. For enumerator returns, the miss rate is 4.4 percent ($\sigma$=0.2 percent) when everyone was related, 9.2 percent ($\sigma$=0.7 percent) when nonrelatives were present. These rates are significantly different for both methods of enumeration.

We conclude that the presence of nonrelatives is associated with increasing miss rates for both methods of enumeration.

## 9. Miss Rates by Type of Enumeration Area

Census procedures varied by type of enumeration area (TEA). The PES matching process also varied by type of enumeration area. Table 7 presents estimated miss rates and total P-sample persons by TEA by method of enumeration. Standard errors of the miss rates are given in parentheses. Recall that the presence of "mail returns" in the List/Enumerate TEA was explained in the methodology section.

**Table 7**

| Mail Returns: TEA | Miss Rate(%) | Total Persons |
|---|---|---|
| Tape Address Register | 1.5 (0.1) | 89.1 million |
| Prelist | 1.2 (0.1) | 45.5 million |
| Update/Leave | 1.2 (0.1) | 14.1 million |
| List/Enumerate | 0.7 (0.1) | 3.7 million |
| **Enumerator Returns: TEA** | **Miss Rate(%)** | **Total Persons** |
| Tape Address Register | 6.8 (0.4) | 26.9 million |
| Prelis | 3.2 (0.3) | 12.7 million |
| Update/Leave | 3.1 (0.3) | 4.1 million |
| List/Enumerate | 1.7 (0.2) | 5.7 million |

For both mail returns and enumerator returns, the Prelist and Update/Leave miss rates are not significantly different. All other rates are significantly different.

The PES matching process searched the PES sample block for a census enumeration that corresponded to the P-sample person. In Tape Address Register areas, the TEA that most urban areas fell into, the "search area" also included one ring of adjacent census blocks. In Prelist and Update/Leave areas, the TEAs where the remaining urban areas, suburban areas, and many rural areas fell into, the search area included two rings of adjacent census blocks. In List/Enumerate areas, the most sparsely populated rural areas, the search area was usually the largest of all, an entire census enumeration district, or "Address Register Area".

It is interesting to note that there is an inverse relationship between miss rates and search areas. Also, TEAs with the same search areas have similar miss rates.

### 10. Miss Rates by District Office Procedure

Census district offices in the central portions of large urban centers were classified as "Type I". For a Type I district office, census questionnaires returned by mail went directly to the associated census processing office where questionnaire edits were performed. In the other district offices, census questionnaires returned by mail went to the district office where questionnaire edits were performed.

The miss rate for mail returns in Type I district offices was 2.0 percent ($\sigma$=0.2 percent), 1.2 percent ($\sigma$=0.1 percent) in the other types of district offices. These rates are significantly different.

### 11. Miss Rates by P-sample Household Size - Mail Returns

Griffin and Moriarity (1992) reported that the EE rate for persons enumerated by mail was higher for households of size 8 or more, compared to all smaller households. For enumerator returns, the EE rate was higher for one person households, compared to all larger households.

We found a similar result for miss rates for mail return households. The miss rate was 8.4 percent ($\sigma$=0.9 percent) for P-sample households with a nonmover count of 8 or more, compared to 1.3 percent ($\sigma$=0.1 percent) where the nonmover count was 7 or less. These rates are significantly different. For additional perspective, approximately 1.2 million P-sample persons were in the "8+" category; approximately 0.1 million of them were missed by the census.

It was not possible for us to directly compare to the EE results for enumerator returns, since we separated our analysis by links to a vacant census unit versus links to an occupied census unit.

### 12. Miss Rates by Type of Structure

The census questionnaire included a question (Question H2) that asked the respondent to indicate what type of building the respondent lived in. Choices included a one family detached home, an apartment building with 3-4 units, etc.

Griffin and Moriarity (1992) and Moriarity (1993) reported EE rates by type of building. Moriarity (1993) grouped structures into four types that appeared to have similar EE rates: One Family, Multi-Unit, Trailer and Other, and Missing. "Missing" indicates that no response was given to Question H2. Persons living in one family structures consistently had the lowest EE rates across enumeration methods.

Table 8 presents estimated miss rates and total P-sample persons by type of structure, grouped in the same way as Moriarity (1993). Standard errors are given in parentheses.

**Table 8**

| Mail Returns: Structure Type | Miss Rate(%) | Total Persons |
|---|---|---|
| One Family | 1.1 (0.1) | 117.5 million |
| Multi-Unit | 2.2 (0.2) | 23.8 million |
| Trailer and Other | 1.4 (0.2) | 8.3 million |
| Missing | 3.3 (0.6) | 2.9 million |
| **Enumerator Returns: Structure Type** | **Miss Rate(%)** | **Total Persons** |
| One Family | 4.3 (0.2) | 29.4 million |
| Multi-Unit | 7.2 (0.7) | 12.7 million |
| Trailer and Other | 3.7 (0.5) | 3.8 million |
| Missing | 4.8 (0.7) | 3.4 million |

For mail returns, the miss rate is higher for P-sample persons linked to multi-units and to missing values of Question H2 than for persons linked to one family structures. However, the miss rates for multi-units and missing values are not significantly different. The one family miss rate is not significantly different from the trailer and other miss rate.

For enumerator returns, the miss rates are not significantly different for one family, trailer and other, and missing value. The miss rate is higher for P-sample persons linked to multi-units.

The data clearly indicate that miss rates are higher in multi-units, across method of enumeration.

## VII. Conclusions/Recommendations

We have described important differences in miss rates by census procedures such as method of enumeration (mail versus personal visit) and regular personal visit enumeration versus last/resort closeout enumeration. The miss rates by enumeration date suggest that the most complete household data are obtained early.

For the year 2000 Census, we suggest that consideration be given to a well-organized outreach campaign, starting well before Census Day, to encourage prompt self-enumeration by a household member. These factors appear to be well-correlated with the highest quality data. The outreach should try to target groups identified with higher miss rates, e.g., minorities, and persons in younger age groups.

We encourage the continuation of current research designed to improve the census questionnaire. Although the miss rate for P-sample persons linked to a mail return was only 1.3 percent, this represents approximately 2.0 million omissions in the census.

We suggest that consideration be given to shortening the enumeration period and using demographic and statistical methods as an integral part of the 2000 Census. The miss rates from "eleventh hour" procedures such as the last resort/closeout period of personal visit enumeration are very high. We feel that the errors that would inevitably occur by using statistical models, administrative records, etc., are likely to be less than the errors that occur during the late stages of an actual enumeration.

The results for type of enumeration area raise some reasonable doubt that variations in PES search areas could have contributed somewhat to the higher observed nonmatch rate in urban areas. If an operation similar to the PES occurs as part of the 2000 Census, we suggest that uniform search area definitions be used, to the extent that resources permit this.

## VIII. Suggestions for Future Research

This paper has presented descriptive statistics from one portion of the P-sample. We suggest that analytic methods be employed to determine which variables best explain the missed persons. One possible approach is log-linear modeling using the CPLX software developed by Robert E. Fay (1989). CPLX is designed for analysis of categorical data from complex surveys such as the 1990 PES, which was a stratified cluster sample. In addition to the variables documented here, additional variables such as block-level variables described in Alho, Mulry, Wurdeman, and Kim (1993) and/or geographic variables may help to explain variations in missed person rates. Indicator variables for missing data may also be well-related to higher miss rates. Without such additional work, it is not possible to decide, for example, if the observed higher miss rate for P-sample persons linked to mail returns in Type I district offices is related to the different questionnaire processing procedures or to other characteristics of Type I district office areas (e.g., higher minority proportion).

Additional analysis is merited for the portions of the P-sample not analyzed in depth here, particularly the inmovers. This research will not be easy to perform. It would be possible to link some inmovers to census data in a fashion similar to what we have described here, if census data are obtained for the census units the inmovers are linked to.

## REFERENCES


Alho, J., Mulry, M., Wurdeman, K., Kim, J. (1993). "Estimating Heterogeneity in the Probabilities of Enumeration for Dual System Estimation", Journal of the American Statistical Association, 88, 1130-1136.

Childers, D. (1993). Housing Unit Coverage Study Results Memoranda, internal memoranda, U.S. Bureau of the Census.

Davis, M. (1991). "Preliminary Final Report for 1990 PES Evaluation Project P7: Estimates of P-sample Clerical Matching Error from a Rematching Operation", internal memorandum, U.S. Bureau of the Census.

Fay, R. (1989). "CPLX - Contingency Table Analysis for Complex Sample Designs, Program Documentation", unpublished report, U.S. Bureau of the Census.

Fay, R. (1990). "VPLX: Variance Estimates for Complex Samples", Proceedings of the Section on Survey Research Methods, American Statistical Association, 266-271.

Griffin, D., and Moriarity, C. (1992). "Characteristics of Census Errors", Proceedings of the Section on Survey Research Methods, American Statistical Association, 512-517.

Hogan, H. (1993). "The 1990 Post-Enumeration Survey: Operations and Results", Journal of the American Statistical Association, 88, 1047-1060.

Moriarity, C. (1993). "Characteristics of Census Error - Additional Results", internal memorandum, U.S. Bureau of the Census.

Sledge, G. (1992). "Documentation of the Vacant/Delete Movers Operation", internal memorandum, U.S. Bureau of the Census.



This paper reports results of research undertaken by staff of the National Center for Health Statistics and the Census Bureau. The first author is a former employee of the Census Bureau, and this research was performed as a Census Bureau project. The views expressed are attributable to the authors and do not necessarily reflect those of the National Center for Health Statistics nor the Census Bureau.