RESULTS FROM THE 1990 SEARCH/MATCH OPERATION

Susan C. Wajer
Bureau of the Census, Washington, D.C. 20233

KEY WORDS: coverage improvement, erroneous enumerations

INTRODUCTION

The Search/Match (S/M) operation was conducted during the 1990 Decennial Census to help ensure that all persons were enumerated at their usual residence. All persons must be counted at their usual residence for apportionment purposes. A usual residence is "the place where the person lives and sleeps most of the time". S/M was designed to improve both within household and whole household coverage.

There were six different search forms processed during S/M. Many persons listed on a search form were not at their usual residence on census day; for example, they may have been at a hotel on census day. We wanted to ensure that they were counted at their usual home, therefore we searched the census questionnaire at their reported usual residence to determine if they were counted there. If they were not counted at their reported usual residence, we added them to the census at that address.

This paper describes the S/M operation, one of the Census Bureau's largest coverage improvement programs, and presents final results from the operation. The results detail estimates of the number of persons added to the census by each search form type, and the errors associated with these enumerations.

BACKGROUND

The concept of the S/M operation is really quite simple - to verify that persons reported on any of the search forms (the search forms are described later in this section) were enumerated at their census day address. If any of the persons were not found at their reported census day address, they were added to the census at this address. Although the concept seems simple, the S/M operation was long and complex. S/M took place from July 1990 through December 1990. The following is a brief description of the S/M operation.

All search forms were sent to the census processing offices, of which there were seven nationwide. All search forms were sorted by form type (the form types being the six different search forms listed below), and from there on in the processing were kept separated by form type. Each search form had a S/M status label affixed, the purpose of which was to record the results of the S/M processing steps on each form. This processing information would tell the disposition of each search form, and the data recorded on the label would later be used for various analyses.

After the S/M status labels were affixed to the search forms, the form was reviewed to determine if it was searchable. A searchable form had to contain both of the following:

1. **Complete data**: A name and at least two of the population questions - sex, age, race, hispanic origin, marital status or relationship - for at least one person.
2. **Searchable address**: The search address reported on the search form had to have either a) House number, street name, city, state, ZIP Code, or b) Rural route (or comparable route), box number, city, state, ZIP Code.

If the search form did not contain these necessary items, it was not processed further. If the search form contained the required items, it went to the next step of processing - what we called geocoding, a term coined from **geographic coding**. The geocoding step was usually performed simultaneously with the Address Control File (ACF) address match, or ACF browse.

The geocoding and the ACF browse processing steps involved searching, or browsing the ACF to see if: a) The search address could be geocoded. Every piece of geography in the U.S. is associated with a set of numerical codes stored in our TIGER (Topologically Integrated Geographic Encoding and Referencing) files. The TIGER files are the files which generate our detailed maps, down to the block level; b) The **exact** search address or the basic street address (if the search address was a multi-unit) was on the ACF.

If the address could not be geocoded, no further processing was done on the case. If the address was geocoded, it fell into one of two categories - geocoded but not found on the ACF, or found on the ACF. If the exact address was on the ACF, a copy of the census questionnaire for that address was printed. From there, the search form and the copy of

the census questionnaire were sent to the next step of S/M - matching/transcription (M/T).

If the exact address was geocoded but not found on the ACF, we sent the address to the United States Postal Service (USPS) to check if the address was correct and deliverable. Once it was verified by the USPS as deliverable, the process of searching the ACF was again repeated, in case there were changes made to either the address by the USPS or to the ACF since it was last checked. If the address was still not found on the ACF, the address was added and the search form was sent to the next step of S/M processing - M/T. If the address was found on the ACF, a copy of the census questionnaire was printed, and the search form and the census questionnaire copy were sent to M/T. Finally, if the address returned from the USPS was undeliverable, no further processing of the case was done.

The M/T portion of the S/M operation involved reviewing a copy of the census questionnaire for the search address to determine if the persons reported on the search forms had been enumerated on the actual census questionnaire for their reported census day address. Any search persons not found to be counted at their reported census day address were added to the census at that address.

There were six different search forms processed during the S/M operation, each designed to enumerate either persons staying temporarily at a special place (such as a hotel), but who had a usual home elsewhere (UHE), or to ensure accurate coverage of certain subpopulations. The different search forms are described below.

Individual Census Reports (ICRs)

Enumeration of places such as hotels, and the nonresponse followup and field followup operations, generated ICRs. ICRs were completed for individuals found at a special place, or for visitors or nonfamily residents found at housing units during the nonresponse followup and field followup operations who felt they may not have been counted. An ICR listed only one person. If the respondent indicated that they were at the special place temporarily and usually lived somewhere else, the ICR was processed during S/M.

Military Census Reports (MCRs)

Group quarters (GQ) enumeration generated MCRs. Military GQs are a large subset of all GQs. All military personnel completed an MCR. The MCR listed only one person. If the respondent listed an off-base UHE address **and** they indicated that the address was family-type housing, the form was processed during S/M.

Shipboard Census Reports (SCRs)

GQ enumeration also generated SCRs. All shipboard personnel, both military and maritime, completed an SCR form. The SCR listed only one person. If a respondent listed a UHE address, the SCR was processed in S/M.

Parolee/Probationer Information Records (PPIRs)

The Census Bureau conducted the Parolee/Probationer Coverage Improvement Program (PPCIP) and a subsequent PPCIP followup (FU) program. Certain subpopulations traditionally suffer from differential undercount. Differential undercount occurs when a subset of the total population is underrepresented in the census count at a disproportionately higher rate than other subsets of the total population. The PPCIP (and the subsequent FU program) was initiated to address the differential undercount of Black males.

During the PPCIP, parolees/probationers completed PPIRs and returned them to their parole/probation officer. The parole/probation officer mailed the PPIRs in bulk to the Census Bureau's processing offices. The PPIR listed only one person. If the respondent listed their census day address on the PPIR, and indicated that they did not stay elsewhere (such as a detention center) around census day, then the PPIR was processed during the S/M operation.

The response rate to the PPCIP was very low - only about 25 percent of all participating states' parolees/probationers (Reference #3); thus, the PPCIP FU was developed. The PPCIP FU program also generated PPIRs. For the FU program, state and local governments' Department of Corrections supplied administrative lists for selected areas designated by the Census Bureau. The Department of Corrections validated the addresses on the administrative list as the parolee/probationer's census day address. A PPIR was completed for each parolee/probationer that had a validated census day address.

The results from these forms are presented **separately** by operation, i.e., the PPCIP versus the PPCIP FU program. The PPIRs are from the PPCIP, and PPIRFUs are from the FU program.

Were You Counted? (WYC)

The print and electronic media generated WYC forms. Respondents who believed their household, or persons within their household, were missed in the 1990 Decennial Census either completed a WYC form that was displayed in local print media or called the Census Bureau's processing offices or district offices. In some cases, the processing office or district office staff completed WYC forms for the respondents who called to report that they had not been counted. The WYC form could list more than one person. All searchable WYC forms were processed during the S/M operation.

The D-190 Search Record

A D-190 search record was generated for either whole households that usually lived elsewhere, or for recent movers that lived elsewhere on census day. All searchable D-190 search records were processed during the S/M operation.

Normal census procedures generated whole household usual home elsewhere (WHUHE) cases. If a respondent indicated on his/her census questionnaire that the usual residence of the entire household was somewhere other than the address where they received their census questionnaire, the district office or processing office staff completed a D-190 search record for the household. After verification, the household was removed from the census questionnaire where they reported that they do not usually reside. The persons were listed on the D-190 search record and the D-190 was sent to S/M to determine whether they were counted at their usual residence, and if not, to add them there.

The vacant/delete operation generated mover-UHE cases. This operation revisited vacant and deleted housing units. If an enumerator located a respondent who indicated that he/she moved into the unit sometime after census day and did not complete a questionnaire at his/her census day address, the enumerator completed a census questionnaire for the household, indicating that this household recently moved. District office or processing office staff then completed a D-190 search record for the household.

METHODOLOGY

After the S/M operation, a sampling plan was developed for this evaluation. The sampling plan involved a two-stage sort of all search forms and systematic sampling. The two-stage sort was 1) by form type (ICR, MCR, etc.) and 2) by processing outcome (unsearchable, matched, etc.). Dependent upon the total number of forms in each category, a random start and a take every interval was assigned. The sample sizes were determined such that we would be able to calculate reliable estimates of person adds by form type. Data from the sampled search forms were keyed. The estimates of the number of persons added to the census are from these keyed data. The estimates are rounded to the nearest hundred.

The estimates of the number of persons added to the census from each form type are weighted estimates (based on the take every interval described above). There is no standard error due to sampling associated with the estimates of persons added from the form types that had only one person listed on them (ICRs, MCRs, SCR, PPIRs and PPIRFUs). This is because there is no variation in the data, i.e., all of these search forms could have only one person added to the census. However, there is nonsampling error associated with these estimates, which cannot be measured for this analysis. For example, sorting errors in the sorting and sampling operation described above are examples of nonsampling errors. Note that the standard errors that are measurable are **not** rounded.

The Post Enumeration Survey (PES) was a national survey that was conducted after the census to measure census undercounts and overcounts. Data from the PES were used for this analysis.

The search forms that identified persons to be added to the census were sorted into PES versus non-PES blocks. Data from these search forms that were in PES sample blocks were keyed. These sample data were used to generate the estimated erroneous enumeration (EE) rates. An EE is defined as an enumeration that was considered incorrect because the person should not have been counted at the specified address on census day. For example, they may have been born after, or died before, census day, or they may have had a usual residence elsewhere. EEs are also duplicate enumerations, fabricated enumerations, and enumerations that were assigned to the wrong census geography due to a geocoding error. The PES final enumeration status for these person adds was deemed to be the "truth".

Note that this paper **does not** include estimates of the EE rates for the persons added to the census from MCRs or SCRs. This is due to the PES sample being used for the EE rate estimates. By design, the PES sample did not include barracks on military bases or ships.

90 percent confidence intervals on the EE rates are presented. Rather than the traditional confidence interval formula, the Bonferroni Method for multiple confidence statements (Johnson and Wichern, 1988) was used. With this method, confidence statements about all of our intervals can be made simultaneously with 90 percent confidence. Note that this method makes the intervals more conservative (i.e., larger). For p simultaneous confidence statements, the Bonferroni method uses $z_{(\frac{\alpha}{2p})}$ instead of $z_{(\frac{\alpha}{2})}$, yielding longer confidence intervals.

LIMITATIONS

There are several limitations to these results that should be taken into consideration when using these data.

First, the S/M operation, the sorting and sampling of the search forms, and the keying of the sampled forms were all clerical operations. Thus, the data obtained from these operations are subject to clerical errors.

Second, only one search form was accepted per household. Thus, if more than one search form was processed for a household, only the persons on the last processed form were added to the census.

Third, the PES was **not** designed to measure S/M errors. Any S/M person adds that were in the PES sample fell in sample by chance, not design. Therefore, this is not the best possible measurement of EE rates. However, this is the only available measurement of EE rates.

Lastly, for the purposes of this analysis the PES final enumeration status is deemed to be the correct determination. It must be recognized that there were errors in the PES that lead to limitations in these results. PES followup activities occurred in the fall and winter of 1990. It is possible that the time between April 1, 1990 and PES followup resulted in recall and other errors. In some instances response error could lead to an incorrect categorization of an enumeration as erroneous. These factors should be taken into consideration before drawing conclusions from these results.

RESULTS

ICRs

There were about 203,000 ICRs received for processing during S/M. We estimate that approximately 36,100 persons were added to the census from ICRs. About 17.78 percent of the ICRs received for processing during S/M resulted in a person add.

As previously discussed, the estimate of the number of persons added to the census from ICRs does not have a standard error due to sampling since only one person could be reported on a form.

The estimated EE rate for persons added to the census on ICRs is about 15.61 percent. This form type had the second smallest representation in the sample, which contributed to a high standard error (10 percent). As a result, the 90 percent confidence interval includes zero; the 90 percent confidence interval is between zero and 42.51 percent. The EE rate is different from zero since enumeration errors were found and measured.

MCRs

During S/M, we received about 697,400 MCRs. We estimate that we added approximately 56,000 persons to the census from MCRs, which represents about an 8.0 percent person add rate. Again there is no associated sampling standard error for this estimate, only unmeasurable nonsampling error as described above. As mentioned earlier in this paper, there are no EE rates for persons added to the census from MCRs.

SCRs

This form type had the smallest number of processed forms during S/M - we received only about 79,600 SCRs during S/M. From these, we estimate that about 14,000 persons were added to the census. Thus, about 17.59 percent of all the SCRs received resulted in a person add. As was the case for the ICRs and the MCRs, there is no standard error for this estimate. Similar to the military bases, there is no estimate of the EE rate for persons added to the census from SCRs.

PPIRs and PPIRFUs

The Census Bureau received about 484,000 PPIRs from the initial PPCIP. We estimate that approximately 127,400 persons were added to the census from the initial program. Thus, about 26.32 percent of all PPIRs received resulted in a person add. As was true for the other form types that reported only one person, there is no sampling error associated with this estimate of added persons.

Here is where we begin to see the EE rates rising. The estimated EE rate of persons added to the census

from PPIRs is 45.15 percent. A 90 percent confidence interval for the true EE rate is between 32.88 percent and 57.42 percent.

Although this EE rate seems high, the converse implies that almost 55 percent of all the persons added to the census from the PPCIP were confirmed to be correctly enumerated. Given that these persons are believed to be a traditionally hard to enumerate population, it is likely that many of these persons would have remained missed in the census if we had not added them during this program. It is also reasonable to assume that some of the erroneously enumerated persons would have remained uncounted if we had not added them to the census, even though we may have ended up adding them in the wrong block. It must also be reiterated that these EE estimates obtained from the PES are not the best measurement of EE rates since the PES sample was not designed to measure errors in S/M enumerations.

We received approximately 1,000,000 PPIRFUs, and we estimated that about 354,800 persons were added to the census from the PPCIP FU program. This represents about a 35.48 percent person add rate of all PPIRFUs received. An estimate of the EE rate for the persons added to the census from PPIRFUs is about 62.29 percent, and a 90 percent confidence interval for the true EE rate is between 51.10 percent and 73.48 percent. One reason for this high EE rate may be the source of the addresses that were reported on PPIRFUs - state parole/probation offices' administrative records. No verification of the accuracy of the administrative lists was conducted by the Census Bureau prior to its usage.

WYC?

The Census Bureau received about 352,800 WYC forms. From these forms, about 260,000 persons were added to the census. The standard error of this estimate is 2,511 persons. Approximately 34.55 percent of all WYC forms resulted in at least one person being added to the census.

The estimated EE rate for the persons added to the census from WYC forms is 35.20 percent. A 90 percent confidence interval for the EE rate is between 20.36 percent and 50.05 percent.

D-190 Search Records - WHUHEs and Mover-UHEs

We received about 375,300 D-190 search records that were WHUHE cases to be processed during S/M. From these, about 162,800 persons were added to the census. This estimate has a standard error of 2,645 persons. The estimated EE rate of the persons added to the census that were WHUHE cases is 40.48 percent. This estimate also had a high standard error (11.12 percent), contributing to a very wide 90 percent confidence interval. A 90 percent confidence interval for the true EE rate is between 10.58 percent and 70.39 percent.

Approximately 85,300 of the D-190 search records that were WHUHE cases resulted in a person add. Thus, about 22.73 percent of all WHUHE cases resulted in at least one person being added to the census during S/M.

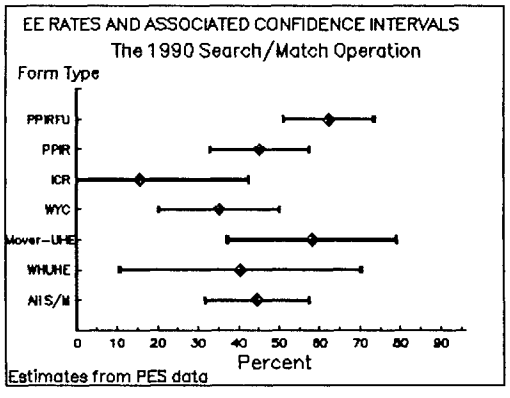
The Census Bureau received about 95,600 D-190 search forms that were mover-UHE cases. From these cases, we estimate that approximately 73,100 persons were added to the census. The standard error of this estimate is 1,282 persons.

Approximately 34,900 D-190 search records that were mover-UHE cases resulted in a person add. Thus, about 36.51 percent of all mover-UHE cases resulted in at least one person add.

The estimated EE rate for the persons identified as movers that had a UHE is high - about 58.21 percent. A 90 percent confidence interval for the true EE rate is between 37.28 percent and 79.14 percent. This high estimated EE rate for movers suggests that there may have been recall bias present during PES followup operations. That is, persons may have had difficulty recalling exactly when they moved. This mobile group of movers may also be difficult to enumerate in the PES too.

Graph #1 depicts the various EE rates and their associated confidence intervals for each of the form types just discussed. The diamonds depict the point estimate of the EE rates and the lines shows the length of the confidence intervals.

Graph #1



Combined: PPIRs/PPIRFUs

The PPIRs and the PPIRFUs were combined to produce one estimate of EEs for all parolees/probationers added to the census. When these forms were combined, the estimated EE rate for all parolee/probationer person adds is approximately 57.24 percent. A 90 percent confidence interval for the true EE rate is between 48.87 percent and 65.61 percent.

All Other Search Forms Combined

This group includes ICRs, mover-UHEs, WHUHEs and WYC forms. The estimated EE rate for persons added to the census from these form types is about 35.74 percent. A 90 percent confidence interval for the true EE rate is between 20.01 percent and 51.47 percent.

A hypothesis test for the difference between two proportions was conducted to see if the EE rates for the PPIR/PPIRFUs versus all other measurable search forms were significantly different. As a result, we can state with 90 percent confidence that the EE rate of about 57.24 percent for the PPIRs/PPIRFUs is higher than the EE rate of 35.74 percent for all other measurable search forms.

CONCLUSIONS

The individual estimates of EE rates by form type are not very reliable; they have relatively high standard errors caused by small sample sizes. In addition, almost all of the intervals for the individual EE rates are overlapping, thus suggesting that the rates may not be significantly different. But when the form types were collapsed into the groups PPIR/PPIRFUs versus all other search forms, it was shown that these rates were significantly different, and that the PPIRs/PPIRFUs had the higher EE rate. This confirms that the PPCIP and the FU programs added a greater rate of persons erroneously than the group of other search forms. Most errors seemed to result from a misunderstanding by the parolee/probationer about what was their "usual residence" on census day. The concept of "usual residence" my be difficult to apply for persons with tenuous or multiple attachments to residences such as parolees/probationers. We recommend that this program be carefully scrutinized before any future implementation, in order to capitalize on adding a large number of persons to the census from this subgroup, but to also improve on the quality of the enumerations.

Although evidence exists that confirms errors were introduced from the S/M operation, data also show that in most cases, and for most form types, persons added to the census from these search forms were correctly enumerated. This point must not be overlooked when examining EE rates.

It is clear from these results that we need to examine the methodology for adding persons to the census from the S/M operation to improve census coverage. Search forms that collect this type of data must clarify the importance of collecting the address that corresponds to the individual's usual residence on census day. Clearly defining where persons should be counted and developing the tools and procedures to collect sufficient information to ensure their correct enumeration is critical to the future success of the S/M operation.

REFERENCES

1. Johnson, Richard A. and Wichern, Dean W. Applied Multivariate Statistical Analysis. 2nd ed., 1988. Prentice Hall, Englewood Cliffs, NJ 07632.
2. Tillman, Amy. 1990 Decennial Census Preliminary Research and Evaluation Memorandum Series No. 105, "Preliminary Results: 1990 Search/Match Workloads by Result from the Sample Selection Sort Operation Conducted in the Processing Offices". U.S. Department of Commerce, Bureau of the Census. January 10, 1992.
3. Beverage, Susan C., Amy L. Tillman, and Fay F. Nash. 1990 Decennial Census Preliminary Research and Evaluation Memorandum Series No. 132, "Preliminary Results From the Parolee/Probationer Coverage Improvement Program and the Parolee/Probationer Coverage Improvement Followup Program". U.S. Department of Commerce, Bureau of the Census. February 4, 1992.

This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.