PRELIMINARY COVERAGE IMPROVEMENT RESULTS FROM TESTS FOR THE 1990 CENSUS

Robert G. Edson and James B. Treat, Bureau of the Census[1]
Robert G. Edson Statistical Support Division Washington, DC 20233

A. INTRODUCTION

Over the past two decades, the U.S. Bureau of the Census has transformed the decennial census from a list and enumeration procedure applied in all areas of the country into an operation that relies almost exclusively on a mailout/mailback procedure with associated coverage improvement programs. Approximately 60 percent of the population in 1970 was mailed a questionnaire with instructions for its completion and return by mail. Households for which a questionnaire was not returned by mail were visited by enumerators during a nonresponse follow-up operation.

Due to its effectiveness in the 1970 census, use of the mailout/mailback procedure was expanded in the 1980 census to enumerate about 95 percent of the population. These persons were located in two types of areas: Tape Address Register (TAR) and Prelist.

TAR areas consisted of the highly urbanized portions of the United States which met three conditions:

1. A commercial mailing list existed for the areas;
2. The areas were covered by city mail delivery; and
3. Addresses in these areas could be assigned geographic codes from the geographic computer files maintained by the Bureau of the Census.

For these areas, the address list was compiled by purchasing a commercial mailing list and then updating this list through several coverage improvement operations.

In areas where a mail census was desired but which the Bureau of the Census could not classify as TAR, an address list was compiled by a prelist operation followed by several coverage improvement programs. The Census Bureau conducted the prelist operation by having enumerators prepare an address list about a year and a half before the census. The enumerators were given a map of an area and a listing register. The enumerators canvassed the area in a systematic fashion and listed every housing unit they found.

Concurrent with the inception of the mail census was the development of extensive coverage improvement programs. The basis of the mail census was the preparation of an accurate address list, the subsequent self-response of persons who received a questionnaire, and enumerator contact at addresses which did not return a questionnaire by mail. Early in the development of the mail census, it was realized that a single procedure for address list compilation and enumeration would never result in the best possible census. As a result, a number of overlapping operations were developed with potentially diverse strengths and weaknesses. The system of programs increased the cost of the census, but they also produced a more complete census. The coverage of the 1970 census was assessed to have been better than that in any previous census, and the 1980 census was reported to be the most accurate ever.

The full paper presents a discussion of research and development toward more accurate and cost effective coverage improvement programs for TAR areas in the 1990 census. Included in this discussion is a description of the programs utilized in the 1980 census. The U.S. Bureau of the Census has completed three major tests (the 1985 Test Censuses of Jersey City, New Jersey and Tampa, Florida and the 1986 Test Census of Central Los Angeles County, California) and is in the process of conducting a dress rehearsal for the 1990 census. A description of these programs and their results in the test censuses are given in the full paper. Due to time constraints, the presentation was limited to a thorough discussion of one of the programs. This restriction is reflected in this proceedings summary.

B. 1980 CENSUS PROGRAMS FOR TAR AREAS

The coverage improvement programs conducted in TAR areas for the 1980 census can be classified into two categories:

- o Those designed to improve the coverage of the address list prior to data collection, and
- o Those aimed at improving coverage during data collection.

These programs were conducted in the TAR areas as described below. Results from some of these programs are documented in [1] - [3].

1. Programs to Improve List Coverage

The programs performed before Census Day to compile the address list consisted of three basic operations: Advance Post Office Check (APOC), Precanvass operation, and Post Office Casing and Time-of-Delivery Checks.

The APOC was a two-stage operation conducted at post offices by the United States Postal Service (USPS) mail carriers during the summer of 1979. In the first stage, called APOC I, the postal carriers were given address cards on which the addresses contained on the commercial mailing list were printed. Carriers identified the addresses as either residential, nonresidential, or undeliverable (addresses to which mail could not be delivered). The residential addresses were further classified as either deliverable with corrections, deliverable without corrections, or

deliverable addresses which were not on the commercial list (added addresses). The second stage, called APOC II, was conducted several weeks after APOC I. In APOC II, the carriers were given the addresses which initially had been classified as undeliverable along with a "cover sample" of deliverable addresses and asked to classify these addresses as deliverable or undeliverable. Addresses classified as undeliverable in both checks were considered to be nonexistent.

After the APOC, the updated commercial list was processed by the Bureau of the Census first in a computer operation and then in a clerical procedure to assign geographic classification codes to the addresses. This was called geocoding. Addresses geocoded to a given area were structured into address registers which contained approximately 300 to 600 addresses. These were referred to as Master Address Registers and served as the basic control list for the remaining census operations. The addresses which were not assigned geographic codes underwent a field geocoding operation.

The Precanvass operation took place about two months before Census Day. For this operation, a separate set of registers were prepared which contained a listing of the address and the corresponding number of housing units for each structure in a given area. Census enumerators were given these registers along with maps of the address register area and instructed to travel through the entire area to verify the list. This procedure was called canvassing the area. The enumerators compared the structures they found in the area to the listings in the register to add missed residential addresses, delete nonexistent addresses, make addresses corrections, and to verify for each structure that the number of units listed was correct. When an enumerator found that a structure contained more units than were listed, the enumerator listed the apartment designation of each unit in the structure. Subsequently, an office operation was carried out in which the apartment designations obtained during Precanvass were matched to the Master Address Registers to determine which units had been missed. Additional field reconciliation was required when apartment designations could not be matched. Finally, the units added to the Precanvass registers were transcribed by hand to the Master Address Registers and a mailing piece (an envelope containing a census questionnaire, instructions, and a return envelope) was addressed by hand for each added unit.

Mailing pieces were then addressed by computer for the units that were originally listed in the Master Address Registers. These mailing pieces were combined with the hand addressed mailing pieces and given to the USPS so that the carriers could conduct the Casing and Time-of-Delivery Checks. The Casing Check took place two to three weeks before Census Day and was very similar to the APOC operation except that the carriers reviewed the addresses on the mailing pieces instead of those on the address cards. However, the carriers were not asked to correct deliverable addresses. The Time-of-Delivery Check occurred as postal carriers delivered the mailing pieces. The carriers were instructed to prepare "add" cards for each residential address on the route for which they had not received a mailing piece. The rationale for conducting these two separate checks so close together was based on several considerations. First, the Census Bureau planned to prepare mailing pieces for Casing Check adds in time for delivery. Second, the Casing Check was conducted in a controlled environment where the major objective was to update the address list, not to deliver mailing pieces.

2. Programs to Improve Coverage During Data Collection

There were several programs conducted in TAR areas after Census Day to improve the coverage of housing units and people in the 1980 census. Many of these are discussed in the paper. A description of the Nonhousehold Sources Program (NHHS) is presented below.

The 1980 NHHS was an administrative records check process aimed at reducing differential undercount of minorities by improving the within household coverage of these persons in highly urbanized areas. Lists containing the names, addresses, and various demographic characteristics of people were obtained from the Motor Vehicles Department for each state, the U.S. Immigration and Naturalization Service, and a public assistance file provided by the city of New York. The lists were screened to identify persons in areas with large concentrations of minority populations. These persons were matched by address and person data to the census. Nonmatches to the census were followed-up and those persons who had been missed were added to the census at the given address.

C. TEST OF THE NONHOUSEHOLD SOURCES PROGRAM IN THE 1986 TEST CENSUS

1. Background

The NHHS was tested in each of the urban sites involved in pretests for the 1980 census [4]. In 1980, the NHHS was conducted in selected urban areas which had been identified as having a high proportion of minorities. The lists used in 1980 were the driver's license files, a file obtained from the Immigration and Naturalization Service, and a public assistance file provided by the city of New York. The results of the 1980 program were not encouraging (see [3]) due to three major problems:

a. The add rate of persons in core urban areas considered to be hard to

enumerate (11.6 percent of the follow-up workload) was much higher than that in the other urban areas which were thought to be easier to enumerate (4.2 percent of the follow-up workload. There were a total of about 127,000 persons added.

b. Operational problems prevented about 58,000 potential person adds from being made to the census. These represented about 31.4 percent of the person adds that should have been made.

c. Included on the NHHS lists were a large number of persons (20.2 percent of the list in core urban areas) whose supplied addresses were at large multiunit structures but which did not include apartment designations. These cases were identified during the clerical address matching operation. To avoid the possibility of examining a prohibitive number of questionnaires for each such person to determine the person's enumeration status, these cases were excluded from further processing.

The cost of the NHHS in 1980 was $9.8 million or about $77 per added person.

The 1986 Test Census included the only test of the NHHS in the 1990 planning cycle. The independent lists used in 1986 were the driver's license file and smaller listings for draft registration and persons receiving aid under the Job Training and Partnership Act. To address the major problems experienced in 1980, the test was limited to a core urban area and procedures were automated as much as possible. Automated matching systems were developed to make the program more cost effective and less dependent on clerical operations. Automation would also allow the processing of a large number of records and would facilitate the identification of addresses which did not have apartment designations.

Results from the 1986 NHHS are documented below. Corresponding results from the 1980 pretests and census are supplied when meaningful comparisons can be made. To this end, the figures cited for the 1980 census reflect the results only for the core urban areas involved in the 1980 NHHS.

2. List Preparation and Matching

There were a total of 197,442 person records on the NHHS lists and about 90 percent of these were from the driver's license file. The addresses on the lists were processed through a computer geocoding operation and those which were not geocoded were deleted from the lists. The lists were unduplicated based on address and person data, and records lacking age or sex codes were deleted from the lists. Finally, since 1980 census results showed that the differential undercount was higher for persons age 16-55, records for persons 15 and younger or 56 and older were deleted. These procedures produced an effective list of 148,548 person records.

A two phase operation to match the NHHS records by address and person data (name, sex, age) to the census was used. In the first phase, a direct computer match was made to the census files. This operation partitioned the cases into three groups: exact matches, those with basic addresses missing from the census, and the residual cases. The residual cases were sent to the second phase, a computer assisted clerical matching operation performed by ten clerks using computer terminals. The clerks used all available data from the census and the NHHS files to determine if the address for a given NHHS person could be linked with one and only one census address. This was required since, as for previous tests of the NHHS, follow-up was limited to one address for a specific case. Address nonmatches of this type were ineligible for follow-up. For address matches, the clerks examined the data from the questionnaire for the associated census address to decide if the NHHS person matched to a person on the questionnaire. Table 1 shows the number of matches as a percentage of the total number of cases processed by match phase for the four tests of the NHHS.

The address and person match rates from 1986 appear to be relatively close to those for the 1980 census and pretests. However, the out-of-scope name match rate for 1986 was much lower than that for the Camden pretest and the 1980 census. For a large proportion of the Camden out-of-scope cases, the address matched to the census but the unit was classified as vacant or nonexistent because the unit did not return a questionnaire by mail and was not contacted by enumerators until after the match to the census. The 1980 census out-of-scope cases were primarily addresses in multiunit structures which lacked apartment designations. Neither of these two were frequent occurrences in 1986, so the out-of-scope rate was much lower.

It should be noted that the direct computer and the computer-assisted clerical matching was possible in 1986 since census names and person data were captured for all questionnaires processed through the completion of Nonresponse Follow-up. Since the capture of names in the 1990 census has been determined to be too costly, this type of matching operation will not be employed in 1990. Computer matching for 1990 is applicable only for address matching and for screening the NHHS lists to identify addresses for multiunit structures without apartment designations.

3. Follow-up and Coverage Yield

The NHHS follow-up operations in 1986 were conducted for 39,102 of the 71,571 cases eligible for follow-up after matching. Records for addresses linked with four or more NHHS persons were deleted from follow-up for two reasons; to limit respondent burden in supplying in-

formation, and because there was concern that a large number of records may have indicated that the records were out of date. Records geocoded to blocks involved in the Post Enumeration Survey (PES) were not sent to follow-up so that the NHHS would not bias the estimates of undercoverage produced from the PES. Due to time constraints, each address was either processed in telephone follow-up or personal visit follow-up but not both. In all prior applications of the NHHS, all follow-up cases had first undergone a telephone follow-up. Those not contacted by telephone then went to personal visit follow-up.

To assign the follow-up universe to either telephone or personal visit follow-up, the PES stratification was considered. For the PES sample design, the central Los Angeles County test site had been divided into a number of areas. Data from the 1980 census were used to determine the ethnic make-up of the persons and the proportion of multiunit structures in each area. The areas were then assigned to a group of Asian strata, a group of Spanish strata, and a group of Other strata. This procedure corresponded to the way in which minority areas were identified in 1980 for the NHHS. The cases were randomly divided between telephone and personal visit follow-up with the goal of selecting about 10,000 cases for personal visit. The results of follow-up are given in Table 2.

Note that there were two types of adds from the NHHS - primary and secondary. Primary adds were persons included on the NHHS list who were added as a result of follow-up. Secondary adds were persons not included on the NHHS list who were identified as census misses during the follow-up interview. Unless specifically noted, all adds discussed below include both primary and secondary adds.

The figures in Table 2 include 2,310 adds which were correctly made to the 1986 Test Census and 110 adds which should have been made but were not due to processing errors. The 590 erroneous adds to the census which should not have been made based on the follow-up interview are not reflected in the tables.

The percent of persons added per follow-up attempt was measured in two ways; as a percent of the total number of persons in follow-up, and as a percent of the total number of housing units in follow-up. Respectively, the two percents are lower and upper bounds of the coverage gain that would have resulted if the 1980 version of the follow-up procedures were used in 1986. A precensus screening operation which eliminated all but one record per housing unit was used in 1980. In the 1986 Test Census, follow-up was conducted for up to three NHHS persons at a given housing unit to determine if this procedure produced a higher add rate then follow-up of one person per unit. To limit respondent burden and to avoid the use of outdated person data, follow-up was not conducted at units associated with four or more NHHS persons. If the 1980 follow-up procedure was used in 1986, some of the primary adds may have been secondary adds, but the coverage increase probably would not have been as high since in 1980 secondary adds were made at a lower rate per follow-up visit (3.6 percent) than primary adds (8.0 percent).

To make meaningful comparisons between the four NHHS tests, the results from 1986 were adjusted to account for the differences in the follow-up procedures used in 1986. The adjustment was made as follows:

a. Lower and upper bounds for the estimated number of person adds from telephone follow-up of <u>all</u> follow-up cases were determined by multiplying the number of housing units in follow-up (25,466 for the entire site; 7,581 for Spanish areas) by, respectively, the lower and upper bounds on the true rate at which persons were added in telephone follow-up.

b. Similarly, lower and upper bounds for the estimated number of person adds from personal visit follow-up of cases unresolved in telephone follow-up were calculated as the product of the number of noncontacted units in telephone follow-up (6,606 for the entire site; 1,960 in Spanish areas) and, respectively, the lower and upper bounds on the true add rate for personal visit follow-up.

c. The estimated lower (upper) bound on the total number of adds by the NHHS was the sum of the two lower (upper) bounds calculated in steps a and b.

This procedure produced the add rates shown in Table 3 for the 1986 test site and for Spanish areas in the 1986 Test Census. Rates are also provided for the 1980 pretests and census.

The data in Table 3 indicate several interesting points. First, the 1986 follow-up add rates are essentially the same as those for core urban areas in the 1980 census and these rates are substantially lower than the 1980 census pretest findings for Travis and Camden. This is an indication that the automated screening used in 1986 did little to increase the effectiveness of follow-up. Second, the adjusted add rates for 1986 are essentially the same for the entire site as they are for the Spanish areas only. The rate in Spanish areas was expected to be higher since the census coverage was expected to be worse in Spanish areas than in other areas.

4. <u>Operational Problems</u>

There were several operational problems associated with the 1986 NHHS.

a. Telephone and personal visit follow-up enumerators had some trouble following the flow of the interview as

described on the NHHS form.

b. About 590 persons were erroneously added to the census due to enumerators who did not following skip patterns on the NHHS Interview Records and, thus, erroneously listed persons as household members. The problem was recognized early in the NHHS operation and clerical procedures were developed to identify NHHS forms for these cases and delete the erroneous adds.

c. About 110 persons who should have been added to the census were not added during NHHS processing. These represented about 0.3 - 0.4 percent of the follow-up workload in 1986. The corresponding figure in 1980 was 6.4 percent.

d. The follow-up operations for the 1986 Nonhousehold Sources Program started approximately 18 weeks after Census Day. The length of time between Census Day and the start of the NHHS follow-up in 1980 varied from one district office to the next. For the Travis County and Camden pretests this period was about 12 and 7 weeks, respectively. The late start of follow-up in 1986 could have contributed to the low add per follow-up attempt rate.

The problems discussed above occurred in the same phase of the NHHS operation as did the major 1980 census processing errors. The follow-up and subsequent operations to include the NHHS adds in the census counts continue to present problems. Improvement of the instructions for conducting the NHHS follow-up may be possible. However, by its nature the interview is complicated. Certain probing questions are needed to determine the existence of persons who may have been missed in the census and the address of their residence on Census Day.

5. Cost

The only cost data available from the NHHS in 1986 are those related to personal visit follow-up. This operation cost $24,700 or about $2.47 per follow-up visit. This is similar to the $2.50 per follow-up visit for the NHHS in the 1980 census.

When available, the cost of the NHHS in 1986 will not be comparable to that from the 1980 census due to the procedural differences between the two operations. The 1986 cost also will not be useful in planning for the 1990 census. The costly clerical operation for matching names was not needed in 1986 because all of the names were keyed to support the automated procedures. As previously mentioned, this is too costly to be done for the 1990 census.

6. Implications for the 1990 Census

The 1986 NHHS was designed to determine if the major problems associated with the 1980 NHHS could be solved through the use of automated procedures. The results indicate that, at best, the only problems that can be solved by automation are those related to addresses that are unsuitable for matching and follow-up. The operational problems associated with determining who should be added to the census have not been solved. It is apparent from 1986 that the NHHS interview is difficult for census enumerators to conduct. While particular problems noted with the 1986 interview may be solved, there is no evidence that the follow-up operation will ever be conducted better than it was in 1980. The importance of having clerks review the NHHS results was demonstrated in 1980 and 1986 to be a necessary part of the program. Unfortunately, clerical operations are also subject to error.

Automation will not significantly reduce the cost of the NHHS for the 1990 census. Automated matching will only be possible for the address matching phase. If anything, this was the most accurate and cost effective part of the 1980 procedure. Names will not be available for automated person matching, so matching will have to be done clerically. Given the nature of data processing planned for 1990, the control and implementation of a matching operation may be more difficult and costly than in 1980 when all available questionnaires for an enumeration area were in a single box and could be easily examined by a matching clerk. In 1990 the census questionnaires will have to be cycled on a flow basis from the census edit operation through a special matching group which will identify the NHHS follow-up cases and then cycle the questionnaires back into the census flow so that the units can be processed in NHHS follow-up.

Automation does not appear to offer any promise of improving the effectiveness of the NHHS follow-up. It was thought that automated procedures would produce a follow-up workload with a high proportion of person adds. Automation was used to a much higher degree in 1986 than would be possible in 1990, yet the coverage gain was equivalent to that realized in 1980 and much lower than the gain for the 1980 pretests. While the 1986 follow-up was subject to some operational problems, it is doubtful that they alone are responsible for the large difference between the add rates from 1986 and the 1980 pretests.

To summarize, the test of the NHHS in 1986 showed that automation could not solve many of the problems that made the 1980 NHHS an ineffective and costly program. In addition, there were no results to suggest that the program could be modified to make it more effective in justifying its purpose which is to decrease the differential undercount of minority persons. This undercount exists but the NHHS does not seem to be very useful in locating census misses.

D. CONCLUSIONS

The development of coverage improvement programs in conjunction with the

mail census has resulted in an increasingly more accurate census. Yet, the Bureau of the Census continues to perform research and conduct evaluations of new or improved coverage improvement programs in three sites (St. Louis, Missouri; East Central Missouri; and Eastern Washington) in which the 1988 Dress Rehearsal for the 1990 census will be conducted. The goal of much of this work is to make the programs more cost effective, efficient, and compatible with the increased automation of the census process. The final coverage improvement system for the 1990 census has not yet been established. However, based on the findings described here, the decision has been made to exclude the NHHS from the 1990 census procedures.

REFERENCES

All references are proceedings from papers presented at previous annual meetings of the American Statistical Association.

[1] K. Thomas and D. Whitford. "Post Office Effectiveness", 1984 Section on Survey Research Methods.

[2] M. Fan, M. Sutt, and J. Thompson. "Evaluation of the 1980 Census Precanvass Coverage Improvement Operations", 1984 Section on Survey Research Methods.

[3] C. Keeley and J. Thompson. "The 1980 Census Nonhousehold Sources Program", 1984 Section on Survey Research Methods.

[4] J. Thompson. "The Nonhousehold Sources Coverage Improvement Program", 1978 Social Statistics Section.

FOOTNOTES

[1] 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.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the following persons for their aid in preparing this paper: Dennis Schwanz, Ed Robison, and Rose Hernandez for their thoughtful review; Don Dalzell and his staff for their programming support; Dave Franklin and Chris Moriarity for their analysis of data; and Jeanne Washington for her excellent typing.

Table 1 Match Rates for the Nonhousehold Sources Program

Test	Matching Operation: Address Match	Name Match	Out-of-Scope*
Travis County	96.1	44.2	Not available
Camden	98.6	45.0	19.6
1980 Census Core Urban	96.9	46.3	30.9
1986 Test Census	94.6	42.2	6.9

*The out-of-scope name match results include records which matched to vacant/delete units, matched to a questionnaire for which names had not been captured, or were for a unit with no apartment designation at a structure with 16 or more units.

Table 3 Nonhousehold Sources Follow-up Add Rates

Test	Adds as a Percent of Housing Units in Follow-up
Travis County	21.0
Camden	23.4
1980 Census Core Urban	11.6
1986 Total	6.6 - 10.2
1986 Spanish	6.5 - 9.9

Table 2 1986 Nonhousehold Sources Program: Follow-up Workload and Persons Added by Follow-up Type and by PES Stratum

Follow-up Results	TELEPHONE: TOTAL	PES Area: Span.	Asian	Other	PERSONAL VISIT: TOTAL	PES Area: Span.	Asian	Other
Persons in FU	23,778	5,487	5,826	12,465	15,324	5,949	6,261	3,114
HUs in FU	15,470	3,644	3,788	8,038	9,996	3,937	4,058	2,001
Adds	1,011	208	270	533	1,409	646	568	195
o As % of Persons in FU	4.3	3.8	4.6	4.3	9.2	10.9	9.1	6.3
o As % of HUs in FU	6.5	5.7	7.1	6.6	14.1	16.4	14.0	9.7