

## THE 1980 CENSUS NONHOUSEHOLD SOURCES PROGRAM

Catherine Keeley and John Thompson, U.S. Bureau of the Census

### A. Background

The objective of the 1980 Census Non-household Sources (NHHS) Program was the reduction of the differential undercoverage of minority populations. The program was conducted only in certain areas with large minority populations and was designed to enumerate persons who had been missed in households for which a census questionnaire had been received. This program was one of few census operations directed at improving within household coverage.

Names and addresses of potential non-enumerations were obtained by a match between the census records and lists of names and addresses from outside sources. The Bureau had acquired lists of names and addresses from the Department of Motor Vehicles (DMV) for 43 states, and the District of Columbia, the U.S. Immigration and Naturalization Service (INS), and the 1979 New York City Public Assistance file.

The lists were processed by computer in three phases to eliminate out-of-scope and duplicate records. Out-of-scope records were those from tracts with low precensus estimates of minority populations. This operation resulted in a total list size of approximately 6.8 million records. A label with name and address was printed for the in-scope records. The labels were then sent to the appropriate district office (DO). Only DO's with in-scope tracts received labels. In the DO's each preprinted label was attached to a separate NHHS Form.

In the district offices the lists were matched to the census records after the first follow-up of non-mail return questionnaires. The matching operation was performed by regular census office clerks. In order to keep the program cost effective, the matching phase was not designed to be an exhaustive match reconciliation procedure.

There were four stages of processing the NHHS Form in the district offices. First, the addresses were matched to the Master Address Register (MAR). A NHHS Form was eliminated from further processing if:

1. The address was not found in the MAR.
2. The address was indicated as a vacant or deleted unit.
3. The apartment had no address designation and was found to be at a multi-unit structure with 10 or more units.
4. The address had no apartment designation and was found to be at a multi-unit structure with nine or fewer units and the surname could not be matched to a resident of the building.

When a unit was found in the MAR, the census serial number was added to the NHHS Form.

During the first stage, approximately 1.4 million addresses (20 percent of the lists) were eliminated from further processing leaving approximately 5.4 million addresses available for the second stage of processing.

The second stage of processing was an attempt to match each case to a census questionnaire. At this stage a case was eliminated from further processing if:

1. The census questionnaire was found and the NHHS person was enumerated.
2. The census questionnaire was missing or was missing or was not filled out. Units corresponding to these types of questionnaires were visited by enumerators.

If the census questionnaire was found and the NHHS person was not enumerated, the census roster and relationship to head, sex, race, date of birth, and Spanish origin for each person on the roster were entered on the NHHS Form.

The third stage was a follow-up interview conducted to determine whether or not the NHHS person should have been listed on the census questionnaire. The number of cases going to follow-up was approximately 1.8 million (27 percent of the lists).

The follow-up interviews were conducted by telephone whenever possible. If a household could not be contacted by telephone, a personal visit interview was conducted. If the respondent was the NHHS person or knew the NHHS person, the interviewer asked if the NHHS person lived there on census day. If the NHHS person lived there on census day, the NHHS person and his/her characteristics were added to the NHHS Form and the interviewer checked for other missed persons in the household. If any others were found, their names and characteristics were also added to the NHHS Form. If the respondent did not know the NHHS person and the respondent was on the census roster, the interviewer checked for other missed persons.

The fourth and final stage was adding all missed persons enumerated during the follow-up to the census questionnaires. The address register was updated to reflect the number of persons on the census household roster plus the number of persons added to the household roster through this operation.

### B. Results

Preliminary analysis of the data indicate that a total of 126,848 persons were added to the census as a result of the NHHS program. Of this total 81,520 were persons whose names were on the NHHS lists, and 45,328 were persons whose names did not appear on the lists but who resided at an address on the lists and were enumerated as part of the NHHS follow-up.

These persons represent approximately 1.2

and 0.7 percent of the 6.8 million persons who were included on the NHHS lists. This results in a total yield of about 1.9 percent from processing the NHHS lists in the 1980 Census. This result is substantially lower than our pretest experiences had led us to expect. We had anticipated that the proportion of persons added during the 1980 Census would be about 10 percent of the lists processed. This discrepancy between the results of the 1980 census program and the pretest expectations will be discussed in a later section (Section C).

The ensuing presentation of the data will distinguish between the two types of added persons. The two types of added persons are: 1) those whose names appeared on the NHHS lists called "NHHS Adds"; and 2) those whose names were not on the list but who were enumerated as part of the NHHS Program follow-up called "Other Adds."

In examining the characteristics of persons added to the census by the NHHS Program, it will soon become apparent that while the NHHS Program did not add as many persons as was expected, it was successful in adding minority populations, specifically Black and Spanish. For comparability with reported census data the race and Spanish origin results will be presented first separately. However, it is difficult to determine the program's success in adding minorities when race and Spanish origin are reported separately because of the overlap of race and Spanish origin. Therefore, the persons added by the NHHS Program were also examined by three race/origin groups: 1) Persons who are White and "Other" races that are non-Spanish, that is, those non-Black and non-Spanish; 2) Persons who are White and "Other" races that are Spanish; and 3) all Black persons, those Spanish and non-Spanish.

The first step in the analysis was to examine race and Spanish origin separately for the persons added. Approximately the same number of Blacks and Whites were added. For NHHS Adds, 33 percent (26,568) were White and 31 percent (25,010) were Black. For Other Adds, the proportion of Black and White added persons was the same as the NHHS Adds. There is some evidence that 34 percent (27,614) of the NHHS Adds were Spanish and that approximately the same proportion of Other Adds were Spanish. Additionally, 67 percent (54,632) of the NHHS Adds were males and 33 percent (26,888) were females. For Other Adds, approximately the same number of males and females were added.

A portion of the analysis of the NHHS Program involved comparing "centralized" census areas to "decentralized" census areas. All the district offices in which the NHHS Program was conducted were located in either a centralized or decentralized census area. "Centralized" and "decentralized" are descriptions of two operating procedures used in the 1980 Census within the areas enumerated by the mail-out/mail-back technique. A centralized procedure was followed in the central cities of large metropolitan areas, and a decentralized

procedure was used elsewhere. In the decentralized procedure, all the filled-in questionnaires which were mailed back were given to enumerators who, working from their homes, checked in the questionnaires, edited them, and contacted those housing units for which incomplete questionnaires were returned, or in which householders failed to return them at all.

The centralized procedure, used in the largest cities, was characterized by having as much of the operation as possible take place under close supervision in one office. Enumerators were assigned only those housing units for which a questionnaire had not been returned or which had returned an incomplete questionnaire that could not be completed by telephone.

Blacks comprise 13 percent of the population covered by the NHHS Program. However, 31 percent (25,010) of the NHHS Adds and 31 percent (13,953) of Other Adds were Black. Spanish are 8 percent of the total population. Of the NHHS Adds, there is some evidence that 34 percent (27,614) were Spanish and 37 percent (16,805) of the Other Adds were Spanish.

In the centralized area, Blacks were added at the same rate as the population percentage. In the decentralized area, Blacks comprise 9 percent of the population while 25 percent (6,939) of the NHHS Adds and 27 percent (5,687) of the Other Adds were Black. Spanish are 7 percent of the decentralized population. There is some evidence that 25 percent (7,078) of the NHHS Adds were Spanish.

The next step in the analysis was to examine the data using the combined race and Spanish origin variable. That is, the three race/origin groups of: 1) non-Black and non-Spanish; 2) Spanish; and 3) Black. The Spanish and non-Spanish Blacks were combined because so few Black persons were of Spanish origin. Only 2 percent of the Black NHHS Adds and 1 percent of the Black Other Adds were of Spanish origin.

Table 1 shows the breakdown of NHHS Adds and Other Adds for the three race/origin groups. The largest group of added persons were Black and Spanish. The coverage of young Black and Spanish males was a special concern of the NHHS Program. For NHHS Adds, about 48 percent (39,524) were Black and Spanish in the age group 16-44 years. About 33 percent (26,798) of the NHHS Adds were Black and Spanish males age 16-44 years.

#### C. Evaluation of the NHHS Operations

To examine the substantial difference between the expected and actual 1980 census NHHS yield rates, the first step will be to review the results of two major phases of internal processing of the NHHS Form: 1) the address search, and 2) the census questionnaire search to determine if the person was enumerated. At each of these stages the NHHS Forms are dropped from further processing if certain criteria are met. These first two stages are crucial. If conditions are present that result in a significant number of records being dropped at these stages, the overall yield of the program is jeopardized.

The second step will be to consider the results of the follow-up phase and the phase in which non-enumerated persons were added to the census. Finally, the three lists used in the program will be examined to determine their relative effectiveness in producing census adds.

When examining the results of the address match phase of the program it became apparent that a substantial portion of the addresses in the centralized area (20 percent) were matched to an ineligible multi-unit structure and were not processed any further. The cases that continued onto the second stage of processing were those classified as a match or possible match to the MAR. These were about 5.4 million names and addresses.

The next stage of processing was the census questionnaire search to determine if the person had been enumerated. Cases were assigned a follow-up interview if the NHHS person's name was not found on a census questionnaire. In the centralized area 22 percent of the total lists were sent to follow-up and 31 percent in the decentralized area.

Of the 6.8 million NHHS Records, 1.8 million were followed-up to find out if the person had been missed in the census. When the follow-up was completed, the missed persons were added to census questionnaires.

At the conclusion of the follow-up and census adding phase, cases were classified as:

1. NHHS Adds
2. Other Adds
3. NHHS Should Have Been Added -

Persons whose names appeared on the NHHS lists and should have been added to the census but were not due to error.

4. Other Should Have Been Added - Persons whose names were not on the lists but who were enumerated as part of the NHHS program follow-up and should have been added to the census but were not due to error.

5. Should Not Have Been Added - Persons erroneously enumerated.

6. Don't know if Case Was Added - The person was found on a census questionnaire other than the one that had been initially checked, therefore, it is not certain whether the person was added because of the program or had always been enumerated.

7. Non-Contacts - The person was never contacted.

8. Knew NHHS Person - The respondent knew the NHHS persons but the NHHS person did not live there and no one was added.

9. Did Not Know NHHS Person - The respondent did not know the NHHS person and no one was added.

10. Person on Census Questionnaire - The person was on a census questionnaire during the initial check and the case had erroneously been sent to follow-up.

As has been noted before, 126,848 persons were added to the Census. The yield from the follow-up cases in the centralized area (8 percent for NHHS Adds and 3.6 percent for Other Adds) was above the yield in the decentralized area (2.4 percent for NHHS Adds and 1.8 percent

for Other Adds). Another 58,103 persons (49,324 NHHS Adds and 8,779 Other Adds) should have been added but were not due to procedural errors. It is unknown specifically why these persons were not added, but some reasonable speculation can be made. The NHHS Program was conducted late in the census processing during the time when there was great pressure to finish the work and close the district offices. The adding of persons to the census was done manually by clerks and it was a slow and cumbersome operation which may have become more error prone given the pressure to hurry. In addition, other major operations were going on simultaneously which made it difficult to locate census questionnaires and make the adds. Also, this phase of the program had never been pretested. Therefore, problems associated with adding persons had not been anticipated and, consequently, safeguards against missing adds had not been established. On the positive side, it should be noted that very few persons were added to the census who should not have been added (0.2 percent).

To summarize, based on pretest experience, the yield rate of added persons due to the NHHS Program was expected to be approximately 10 percent of the lists processed. The yield rate for the 1980 census program was about 2 percent.

It is evident that there were three problem areas in the 1980 census program that contributed to the low yield rate. First, the decentralized area yield of adds was well below the centralized area. Second, in the centralized area a large proportion of addresses were dropped from processing because they were classified as ineligible multi-unit structures. Third, only 70 percent of the persons who should have been added to the census actually were. If these conditions had not existed, that is, 1) if the decentralized rate of added persons from the follow-up cases was as high as the centralized rate; 2) if 20 percent of the list in the centralized area was not dropped from processing because the addresses were in ineligible multi-unit structures; and 3) if all persons had been added correctly to the census, it is predicted that the national rate of added persons would have been increased to approximately 5 percent.

These findings suggest areas for improvement of the yield rate for the program. First, consideration should be given to limiting future use of the program to highly urban areas. The program was most efficient in producing census adds in centralized areas. In the decentralized area the yield rates were low for the work done, consequently, the cost to add a person in the decentralized area was twice the cost to add a person in the centralized area (Section D).

Second, consideration should be given to automating various phases of the program. All the work in the district offices had been done manually. This necessitated setting limits on the search of the address register because of time restrictions. It was not possible to search on a name basis for persons whose addresses did not have apartment designations

in large multi-unit structures. An automated matching system would provide the means for checking all addresses quickly and efficiently.

Another area in which automation could improve the program is in the phase in which persons were added to the census. An automated control system would eliminate the need for physically searching for census questionnaires in order to add persons, thereby reducing the risk of missing added persons due to not being able to locate questionnaires. Overall, an automated control system should speed the entire process and provide immediate feedback on the status of cases during all phases of the program.

As stated previously, the lists of names and addresses used in the NHHS Program had been obtained from three sources; the Department of Motor Vehicles (DMV), the U.S. Immigration and Naturalization Service (INS) and the 1979 New York City Public Assistance File. To evaluate the relative effectiveness of the three lists in producing added persons to the 1980 census, the lists were examined by:

1. The proportion of potential nonenumerations (i.e., the proportion of cases requiring a follow-up to determine enumeration status).

2. The proportion of persons added to the census as a result of the follow-up.

3. The proportion of Black and Spanish persons added to the census as a result of the follow-up.

Overall, 27 percent (1.8 million) of the original 6.8 million cases required a follow-up interview. In comparing the centralized area to the decentralized area, there is some evidence that the rate of follow-up cases (i.e., potential non-enumerations) in the decentralized area was 31 percent (1.15 million) as opposed to 22 percent (673,000) in the centralized area. However, the rate of NHHS Adds from the follow-up cases was only 2.4 percent in the decentralized area compared to 8 percent in the centralized area (Table 2). There is some evidence that the yield of NHHS Adds from the total list size was almost 2 percent in the centralized area, but less than 1 percent in the decentralized area. A similar observation can be made for the Other Adds. That is, in the decentralized area, all yield rates appear to be lower than the centralized rates. The list yielding the largest percentage of follow-up cases was the DMV list in the decentralized area, (31 percent, 1.0 million) while the Public Assistance list yielded the lowest percentage of follow-up cases (9 percent, 18,000).

The rate of all adds for the Public Assistance and INS lists was higher than the DMV list for both NHHS Adds and Other Adds.

In the centralized area, the rate of NHHS Adds for the Public Assistance and INS lists was 15 percent and 13 percent, respectively, while the DMV rate was 6.5 percent. In the decentralized area, the rate of NHHS Adds for the INS and the DMV list was approximately the same. The same pattern can be observed among the Other Adds.

In examining the proportion of Black and Spanish persons added to the census by list it appears that all lists provided Black and Spanish adds. There is some evidence that the highest rate of Black adds were from the Public Assistance list. Also, there is some evidence that the Public Assistance list yielded more Black and Spanish adds than the DMV list.

#### D. Costs

The disparity in the total number of cases processed and the yield of added persons for the centralized and decentralized areas is reflected in the cost of the NHHS Program. The NHHS Program is estimated to have cost approximately \$6.3 million, of which about \$2.8 and \$3.5 million was spent in centralized and decentralized areas, respectively. This amounts to about \$35.00 per person added in centralized areas and \$73.00 per person added in decentralized areas. Clearly the NHHS Program was less expensive to operate in centralized areas than in decentralized.

#### E. Summary

The major goal of the NHHS Program was to increase the coverage of minority populations. The program succeeded in adding 126,848 persons to the 1980 Census and the largest group added were Black and Spanish persons. Additionally, about one-third of all NHHS Adds were Black and Spanish males age 16-44 years.

Some conclusions can be tentatively stated based on this preliminary analysis. First, the program was expensive and did not add persons at the anticipated rate. However, the program has potential if operational problems can be corrected. Had these problems not existed, the yield rate for 1980 would have been 5 percent, thereby making the program more cost effective. Second, while the add rate was below expectations, the program was successful in adding minority populations. Finally, each of the three lists used in the program provided adds for various minority populations, therefore, consideration should be given to identifying other sources of administrative records for use in this program.

#### F. Methodology

The NHHS Program was conducted nationally during the 1980 Census in 303 district offices covering a total population of approximately 172.5 million persons. The preliminary analysis presented here is based on a national sample of NHHS Forms. The sample may be viewed essentially as a two-stage sample design. The first stage sample units (PSU's) consist of the sample of district offices chosen to test coverage improvement techniques. The second stage is a sample of NHHS Forms within the selected district offices.

The first stage sample of district offices was instituted as part of the overall Coverage Improvement Evaluation Program. The NHHS Program sample of 48 district offices was a stratified simple random sample drawn from the universe of 303 district offices. The sample was selected from five strata representing types of census areas. For purposes of analysis, these five strata were collapsed into two types of census areas: centralized and

decentralized census areas. The second stage selection was a stratified systematic sample of NHHS Forms within each of the 48 first stage district offices.

The sampled forms were reprocessed through all stages they had originally undergone in the district offices (with the exception of the follow-up interview) and classified into 12 major categories, each of which contained numerous subcategories. The categories were designed to provide the data necessary to, first, produce national estimates of persons added to the census as a result of this program, and second, to evaluate the various office and field operations of the program.

Once the forms had been categorized, various categories and subcategories were furthered sampled. The resulting sample was approximately 147,000 NHHS Forms. The NHHS Forms that were classified as "noncontacts" and "miscellaneous" were not sampled within the DO. Therefore, the total number of Forms (33,369 and 7,188, respectively) for those categories are included in the sample size. The computation and use of weights and standard errors are discussed in a forthcoming Preliminary Evaluation Results Memorandum.

Acknowledgments

The authors are especially grateful for the contributions to this study from Nick Alberti, Robert Edson, Richard Griffin, and Kathy Thomas for their thoughtful review; to Karen O'Connor for providing us with the data; to Sue Heskamp for unraveling many clerical mysteries; and to Jeanne Washington for her excellent typing and good humor through numerous revisions.

References

1. For a full discussion of the pretest results and methodology see "The Nonhousehold Sources Coverage Improvement Program," John Thompson, 1978.
2. For a full discussion of the computer processing see "1980 Census Nonhousehold Sources (NHHS) - Summary of Computer Processing - Memorandum 2-12."
3. For more information on the processing in the district offices see "Evaluation Plan for the Nonhousehold Sources (NHHS) Program" by Karen O'Connor, Survey Design Branch, Statistical Methods Division.
4. For more information on telephone and personal visit procedures and on census questionnaire search and correction see:
  - a. Telephone Follow-up Manual (Centralized) - D-513J.
  - b. Follow-up 2 Enumerator's Manual (Centralized) - D-557.
  - c. Follow-up 2 Enumerator's Manual (Decentralized) - D-558.
  - d. Coverage Improvement Searches Manual (Centralized/Decentralized) - D-513M.
  - e. Coverage Improvement Corrections Manual (Centralized/Decentralized) - D-513 "O"

Table 1: Persons Added for Race/Origin Groups

Race/Origin	NHHS Adds	Percent	Other Adds	Percent	Total Adds	Percent
Non-Spanish and Non-Black	18,052 (4,332)	22.1 (5.3)	12,290 (2,950)	27.1 (6.5)	30,342 (7,282)	23.9 (5.7)
Spanish	26,551 (14,072)	32.6 (17.3)	16,225 (8,599)	35.8 (19.0)	42,776 (22,671)	33.7 (17.9)
Black	23,480 (4,461)	28.8 (5.5)	12,186 (2,315)	26.9 (5.1)	35,666 (6,777)	28.1 (5.3)
Race and Spanish Origin Unknown	13,437 (3,091)	16.5 (3.8)	4,627 (1,064)	10.2 (2.3)	18,064 (4,155)	14.3 (3.3)
Total Adds	81,520 (18,750)	100.0	45,328 (10,425)	100.0	126,848 (29,175)	100.0

Standard errors in parentheses

Table 2: Added Persons for Follow-up Cases and List Size for All Lists

List Type	N H H S A D D S			O T H E R A D D S		
	CT	DT	Total	CT	DT	Total
<b>DMV</b>						
Number of Adds	34,815	23,801	58,616	14,794	18,241	33,035
Percent of Follow-up Cases	(11,141) 6.5 (2.1)	(6,188) 2.2 (0.6)	(12,896) 3.7 (0.8)	(4,734) 2.8 (0.9)	(4,734) 1.7 (0.4)	(7,268) 2.1 (0.5)
Follow-up Cases	531,727	1,063,293	1,595,020	531,727	1,063,293	1,595,020
Percent Adds Per List	1.5 (0.5)	0.7 (0.2)	1.0 (0.2)	0.6 (0.2)	0.5 (0.1)	0.6 (0.1)
List Size	2,277,902	3,421,074	5,698,976	2,277,902	3,421,074	5,698,976
<b>INS</b>						
Number of Adds	13,541	3,141	16,682	7,276	2,003	9,279
Percent of Follow-up Cases	(4,333) 12.8 (4.1)	(817) 4.7 (1.2)	(3,670) 9.7 (2.1)	(2,328) 6.9 (2.2)	(521) 3.0 (0.8)	(2,041) 5.4 (1.2)
Follow-up Cases	105,750	67,044	172,794	105,750	67,044	172,794
Percent Adds Per List	2.6 (0.8)	1.1 (0.3)	2.1 (0.5)	1.4 (0.4)	0.7 (0.2)	1.2 (0.3)
List Size	520,009	285,893	805,902	520,009	285,893	805,902
<b>PUBLIC ASSISTANCE</b>						
Number of Adds	2,707	----	2,707	884	----	884
Percent of Follow-up Cases	(866) 14.9 (4.8)	---- ---- ----	(866) 14.9 (4.8)	(283) 4.9 (1.6)	---- ---- ----	(283) 4.9 (1.6)
Follow-up Cases	18,150	----	18,150	18,150	----	18,150
Percent Adds Per List	1.4 (0.4)	---- ----	1.4 (0.3)	0.4 (0.1)	---- ----	0.4 (0.1)
List Size	196,164	----	196,164	196,164	----	196,164
<b>LIST UNKNOWN</b>						
Number of Adds	2,508	1,007	3,515	1,486	644	2,130
Percent of Follow-up Cases	(803) 14.5 (4.6)	(262) 4.9 (1.3)	(773) 9.3 (2.0)	(476) 8.6 (2.8)	(167) 3.1 (0.8)	(469) 5.6 (1.2)
Follow-up Cases	17,278	20,538	37,816	17,278	20,538	37,816
Percent Adds Per List	5.0 (1.6)	1.9 (0.5)	3.4 (0.7)	3.0 (1.0)	1.2 (0.3)	2.1 (0.5)
List Size	50,158	51,770	101,928	50,158	51,770	101,928
<b>ALL LISTS</b>						
Number of Adds	53,571	27,949	81,520	24,440	20,888	45,328
Percent of Follow-up Cases	(17,143) 8.0 (2.6)	(7,267) 2.4 (0.6)	(17,934) 4.5 (1.0)	(7,821) 3.6 (1.2)	(5,431) 1.8 (0.5)	(9,972) 2.5 (0.5)
Follow-up Cases	672,905	1,150,875	1,823,780	672,905	1,150,875	1,823,780
Percent Adds Per List	1.8 (0.6)	0.7 (0.2)	1.2 (0.3)	0.8 (0.3)	0.6 (0.2)	0.7 (0.2)
List Size	3,044,233	3,758,737	6,802,970	3,044,233	3,758,737	6,802,970

CT = Centralized Census Area      DT = Decentralized Census Area  
Standard errors in parentheses