

George Sledge, Thomas Harahush and Robert O'Brien, U.S. Bureau of the Census

I. Background

In planning for the 1980 census it was felt that a potential source of coverage improvement might be found by examining the occupancy status of housing units classified as vacant or nonexistent. In pretests prior to the 1970 census, it was discovered that a significant factor in the population undercount was the misclassification of occupied units as vacant. In 1970, the National Vacancy Check was conducted. It was a large-scale post-census survey in which a sample of the units classified as vacant were revisited to determine whether or not any had been misclassified. The National Vacancy Check detected a misclassification rate of 11.4 percent. As a result, over one million persons, or about 0.5 percent of the total population, were added to the 1970 population counts. These persons were added as a result of an imputation process which was based on housing unit and population miss rates estimated from the survey. These rates were also used to randomly convert vacant housing units to occupied housing units.

In planning for the 1980 census, it was decided very early to attempt to eliminate imputation procedures such as those used in 1970 for vacant units. Therefore, all pretests prior to 1980 included attempts to determine efficient methods of detecting misclassification of occupied housing units based on actual enumeration. Extensive follow-ups of vacant or non-existent units were conducted in the Travis County, Texas; Camden, New Jersey; and Oakland, California Pretests for the 1980 census. They proved to be effective in identifying occupied units that had been misclassified.

It should be noted that several means of reducing the costly follow-up of all vacant and nonexistent units were also pretested. One alternative compared the census lists of vacant and nonexistent units to the corresponding Post Office (P.O.) designation. This operation was simulated for Travis County and Camden. The Travis County and Camden Pretests also included other variations, such as telephone Quality Control (QC) and a Call Back Record, to test means of reducing the workload. None of these procedures were successes. The procedures requiring a match with post office returns reduced field reinterview workload but failed to detect about 20 percent of the errors. The telephone QC reduced the field workload significantly, but retained a high residual error of the original misclassified cases and also added additional telephone costs. The joint QC and P.O. check procedures uncovered a high percentage of errors yet reduced the field workload very little. It also added additional work and cost by having to undertake two additional office operations.

In the Oakland Pretest the Post Office (P.O.) classification was tested. However, even though the reduction of the follow-up workload was substantial, the reduction was not large

enough to justify the additional cost and error incurred by the matching, and the significant amount of residual misclassification.

From the results of the pretests, the 1980 Census Procedures included a follow-up of units classified as vacant or nonexistent with the exclusion of those units which had been determined to be vacant-UHE's (Usual Household Elsewhere), duplicates, or units deleted due to the geographic transfer operations.

II. The 1980 Census Procedure

In 1980, addresses for which no questionnaire was returned were followed-up during Follow-up I, the nonresponse follow-up operation. During Follow-up II, an enumerator completed a "Unit Status Review" Form D-160, for those addresses in mail areas designated in Follow-up I as vacant or nonexistent, (with the exceptions noted above). The Form D-160 indicated the Follow-up I status of the unit and was used to determine the Follow-up II, (FU II) status for each vacant or nonexistent unit.

After Follow-up II, the Form D-160 Unit Status Review, was checked back into the District Office (DO). If the Follow-up II enumerator's classification of the unit matched that of the Follow-up I enumerator's, no further processing was needed. The remaining Form D-160's were reviewed, edited, and sorted into three groups, occupied, vacant, or nonexistent as per FU II designation. All the units for which the classification was different than the one that had been found during Follow-up I were matched to the Master Address Registers (MARs) and a change in unit status was made in the MAR to show the new classification. As a result, all units and persons added were processed into the census counts.

III. Summary

Of the 8.4 million nonexistent and vacant units that were followed-up during Follow-up II by the Misclassified Occupied Operation, 159,468 were misclassified nonexistent when they were actually occupied units and 532,688 misclassified vacant units that were also found to be occupied units. A total of 692,156 (8%) occupied units and 1,724,085 persons were added to the census. An additional 175,242 (2%) misclassified nonexistent units were found to be vacant. These units were also added back to the census. There were 507,140 (6%) that were misclassified vacants when these units should have been deleted. However, the census procedures did not require changing vacant housing units back to nonexistent ones. A detailed discussion of the added and non-added units will be presented in a forthcoming census Permanent Memorandum series.

IV. Evaluation of Yield

A. By matching the completed Forms D-160, Unit Status Review, to the Master Address Registers and later to the census detail file, a determination was made of the following:

1. The number of vacant units converted to occupied.
2. The number of deleted units converted to occupied.
3. The number of deleted units converted to vacant.
4. The number of persons enumerated as a result of the unit status review operation.

5. The total Follow-up II workload.

6. Characteristics of added persons and housing units.

B. The basis for obtaining these data was the classification status of the housing units and whether it was determined to be misclassified occupied, misclassified vacant or correctly classified. The housing units were categorized by the data indicated on the Forms D-160 as follows:

1. Misclassified Occupied - Housing units which were classified as vacant or nonexistent in Follow-up I, but were determined to be occupied in Follow-up II, with the occupants stating that they had moved to that address prior to Census Day.

2. Misclassified Vacant - Housing units that were declared nonexistent in Follow-up I and hence deleted from the address register, but were found to exist and determined to be vacant on Census Day in Follow-up II. Misclassified vacants also included housing units that were classified as nonexistent in Follow-up I that were truly vacant, but during Follow-up II it was determined that the unit was currently occupied by a household that had moved to the address after Census Day and had not been counted in the Census.

3. Correctly classified as a vacant or nonexistent unit - Housing units which were classified as vacant or nonexistent in Follow-up I and retained that classification in Follow-up II.

C. The total yield is estimated by determining from the sample of Forms D-160 the number of persons added and units reinstated to the census counts. Below are the types of actions that were defined as misclassifications. Refer to Appendix B for a detailed explanation.

1. Enumerator Errors - If the data from the Form D-160's matched to the Census Detail File, and

a) the household lived at the address on April 1, 1980, the unit was tallied as misclassified occupied and persons in the household as being added to the census.

b) the household moved there after April 1, 1980, the unit and the persons in the household were tallied as procedural misclassified occupied adds.

c) the unit had been changed from nonexistent to vacant the unit was tallied as misclassified vacant and added to the census.

2. District Office Errors - Errors reflecting processing of information obtained from misclassified/occupied enumerators.

a) Follow-up status not indicated on the Forms D-160.

b) Address or serial number on Forms D-160 was wrong

c) Failure to change the Master Address Register (MAR) classification for the housing unit after Follow-up II

A determination of how well procedures were implemented and the effect of enumerator and procedural errors was made. A tally of detected errors by type, and an estimate of the number of persons who were not counted in the

census as a result of these errors was derived.

V. Sample Design and Methodology

A. For the evaluation of the Misclassified Occupied Study, a Coverage Improvement Evaluation Sample was used. Generally, all the District Offices (DO's) in the country were stratified by size and census enumeration procedures.

Within the areas enumerated by mail-out/mail-back techniques, two operating procedures were used. A "centralized" procedure was used in central cities of large metropolitan areas and a "decentralized" procedure was used elsewhere. For centralized (CT) areas, the population was enumerated by the mail-out/mail-back technique and all checking and editing was done in the District Offices. Enumerators only followed-up missing questionnaires or incomplete information that could not be obtained over the telephone. For decentralized (DT) areas all questionnaires mailed back were given to the enumerators at their homes for checking in and editing. Any missing information for a household was followed-up by personal visits. These procedures are stratified into the following six strata:

- Stratum I - Centralized DO's in a city with 1,000,000 or more population
- Stratum II - Remaining centralized DO's
- Stratum III - Decentralized DO's without Prelist Recanvass
- Stratum IV - Decentralized DO's with Prelist Recanvass-Urban
- Stratum V - Decentralized DO's with Prelist Recanvass-Rural
- Stratum VI - Conventional Plus Two-procedure DO's

B. For each sample coverage improvement DO, the Forms D-160's were sorted into 27 different categories which described results of the various Census and follow-up activities. Categories could be combined to obtain counts of misclassified units, correctly enumerated units, and added persons. Next, a 10 percent sample of the Forms D-160, Unit Status Review was selected from each category. The occupancy status from the Forms D-160 was compared to that in the 1980 Master Address Register and all differences tallied. The number of occupied units that were misclassified during FU I as vacant or nonexistent for those DO's was determined. Similarly, the number of vacants that were misclassified as nonexistent during FU I was ascertained. This data was used to produce an estimate of the number of units whose occupancy status had been misclassified. The number of persons added to the census by the misclassified/occupied operation for the various categories mentioned above was obtained by matching the Forms D-160 to the census detail file.

The last categorical breakdown is of Undeterminable Units. (Refer to Appendix (F)). These were errors made by the enumerators when the Form D-160 was not completed. For example, there may have been no explanation given or information was incomplete on the Form D-160, or there was a conflict of information on the Form D-160. Of the undeterminable Forms D-160,

97 percent (265,832) of these were incomplete and 2 percent (6,341) had conflicting data. Altogether these Forms D-160 represented a total of 272,173 units that were potential adds to the census. These D-160's were not processed for the evaluation, thus they did not contribute to the results listed in this report.

VI. Demographic Characteristics of Persons in Added Housing Units

The following tables reflect various demographic characteristics of persons in housing units added by the Misclassified Occupied operation. The data were obtained by matching the housing units added by the operation to the census detail file to obtain the characteristics. Total figures will not agree exactly because of rounding.

A. Race and Spanish Origin of Added Persons

	Misclassified Occupied		Census	
	Units	%	Units	%
Total	1,724,087	100	220,695,834	100
White	1,258,617	73	182,521,648	83
Black	325,549	19	26,495,028	12
Other	139,922	8	11,679,158	5
Total	1,724,087	100	220,695,834	100
Hispanic	179,664	10	14,608,671	7
Nonhispanic	1,544,427	90	206,087,163	93

Thus, the misclassified occupied operation helped with minority coverage of certain sub-populations that have historically been undercovered by the census.

B. Added Persons by Census Procedure

	Misclassified Occupied		Census	
	Units	%	Units	%
Total	1,724,087	100	220,695,834	100
Centralized	265,766	15	27,702,735	12
Decentralized	1,387,440	81	189,097,591	86
Conventional	70,880	4	3,895,508	2

C. Race by Origin by Sex

The misclassified occupied operation added the following numbers of persons in race by origin by sex categories.

Race	Origin	Sex	Total
White	Nonhispanic	Male	590,165
		Female	576,767
	Hispanic	Male	45,361
		Female	46,321
Black	Nonhispanic	Male	153,084
		Female	166,872
	Hispanic	Male	2,822
		Female	2,761
Other	Nonhispanic	Male	29,577
		Female	27,951
	Hispanic	Male	42,457
		Female	39,936

D. Allocation by Race by Household

The Misclassified Occupied data indicated that there were census cases for which data had to be allocated. This was done for persons and

households in which the enumerator failed to collect, or a household refused to supply the required census data.

Race	Vacant to Occupied		P/N	Non-existent to Occupied		P/N
	Units	Units		Units	Units	
Non-White HUs	370,524	130,988	2.82	120,329	37,859	3.18
White HUs	914,283	406,953	2.25	318,929	121,612	2.62
Total	1,284,807	537,941	2.38	439,258	159,471	2.75

For the vacant units converted to occupied units there was no significant difference between the units that had all allocated data, no allocated data, and/or partial allocated data. The same characteristics existed for the nonexistent units converted to occupied. The unexplained differences seem to occur between the vacants converted to occupied and the nonexistents converted to occupied. There was no apparent reason for the differences because both kinds of data was obtained by using the same procedure. The procedure did not require doing anything different for a vacant unit found to be occupied or a nonexistent unit found to be occupied.

VII. Cost

In estimating and reporting cost of the Misclassified Occupied Operation, several things should be considered. This operation was conducted in conjunction with many other Coverage Improvement (CI) Programs. This operation was implemented by Field Division and cost factors were often combined for similar activities for several coverage improvement operations. Therefore, estimated cost for the Misclassified Occupied Operation (M/O) must be proportioned out from the total coverage improvement operational costs.

The cost differed between the centralized and decentralized area of the country. In part, the cost varied due to the type of collection areas and the volume within each area. Also the cost is distributed among training, crewleader (Supervisory/Administrative), and enumerator functions involved for follow-up activities. The M/O follow-up cost resulted from revisiting structures classified as nonexistent and vacant in the Master Address Registers. The follow-up visit cost can be estimated directly from the data since this information was recorded. The follow-up visit rate was \$2.25 per housing unit in centralized areas and \$2.00 per housing unit in decentralized areas. The crewleader cost also contained office supervisory costs. A ratio was applied to the total crewleader cost for all coverage improvement operations (i.e., Misclassified Occupied FU II Visits divided by the Total Coverage Improvement cost times the Total Crewleader cost). The training cost included training for all phases of the CI operation. This estimate was derived in the same manner as crewleader cost. The total cost of the M/O operation broken out by centralized and decentralized areas is shown in Table A. Of

the total cost of 36.3 million dollars, 6.3 million dollars was spent in centralized areas and 29.9 million dollars was spent in decentralized areas. Of the 1,724,087 persons added, 317,699 were in centralized areas and 1,406,388 were in decentralized areas. For the total number of persons added, the cost per person was about \$21. In centralized areas, the cost was roughly \$20 and for decentralized areas, approximately \$21 per person.

Table A: M/O Costs by Area of the Country and Per Person Added

Table A: M/O Costs by Area of the Country and Per Person Added

	Areas		Total
	Centralized	Decentralized	
FU 2 Field and Off. Operations	4,757,312	21,677,236	26,434,548
Crewleader	1,514,561	7,400,848	8,915,409
Training	111,705	849,891	961,596
Total	6,383,578	29,927,975	36,311,553
Cost Per Person Added	\$20.09	\$21.28	\$21.06

VIII. Recommendations

The Misclassified Occupied operation was an effective census operation which found and added missed housing units and persons back into the census.

In both the centralized and decentralized areas, the additions of minorities were significant. The proportion of Black and Hispanic persons among the census adds from the Misclassified Occupied operation was higher than that found in the general census population. This is a definite plus for continuing this type of operation for the 1990 census. It appears that the follow-up of vacant/deleted units can be beneficial in improving differential census coverage.

However, emphasis should be placed on minimizing different types of field and office errors. There should be a better built-in system for monitoring and correcting the incoming and outgoing quality of the data being obtained for the Misclassified Occupied operation.

Overall, the Misclassified operation added a substantial number of persons to the 1980 census, however, at great expense. Clearly, in 1970 and 1980 there was a coverage problem with units that are classified as vacant or non-existent on the census address lists. These housing units probably contain more Black and Spanish persons than found in the general population. It is highly recommended that a review of vacant and deleted units be part of the 1980 census.

This operation can be controlled more effectively through an automated census.

Acknowledgments

I wish to acknowledge John Thompson for his input and expertise. I wish to also acknowledge Sue Heskamp and her staff for all their efforts in providing me with all the detailed information that was required to extract these results; and Jeanne Washington for her patience and steadfastness for continuously updating and making timely corrections over and over again.

References

1. Effect of special procedures in 1970 census Chapter VI National Vacancy Check.
2. Census Coverage Loss Due to the Misclassification of Occupied Units. ASA paper by Richard A. Griffin and Richard W. Lavalley.
3. Census Coverage Loss Due to the Misclassification of Occupied Units. 1978 ASA paper by Richard A. Griffin and Richard W. Lavalley.

H4-EDIT: COVERAGE STUDY

I. Introduction

One of the steps taken in the 1980 census to improve housing unit coverage was the edit of question H4 on the census questionnaires. Basically, question H4 asked respondents the number of units in the structure in which they resided. By comparing the respondents' answer to the number of units in the census records, it was hoped that small multi-unit structures with missing housing units could be detected. To evaluate the effectiveness of the H4-Edit, this study estimated the number of housing units added by the H4-Edit in both centralized and decentralized offices using the Census Allocation Program Evaluation Sample. The results show that an estimated 93,000 units were added to the census due to H4 procedures, which improved the housing unit count by 0.1 percent.

II. Background of H4-Edit Procedures in the Census

The H4-Edit took place in all district offices during the census; however, because of the sample used in this study, we estimated housing unit adds only in centralized and decentralized offices. The H4-Edit procedures used in centralized and decentralized offices involved several operations, while in conventional offices the procedures were less complex.

In centralized and decentralized offices, item A2 on the cover of the census questionnaire showed the number of questionnaires mailed to units with the same basic street address. During questionnaire check-in, the answer to question H4, "How many living quarters are at this address?", was compared to the A2 entry. If the answer to H4 was greater than the A2 entry and the A2 entry was less than 10, then the clerk checked the Master Address Register to see how many units were shown at the same basic street address. The Master Address Register had units which were added by Precanvass, the Yellow Card Operation, and Postal Corrections. These adds were

not included in the A2 entry, but were included in the number of units at the basic address for comparison to the answer to H4.

If the answer to Question H4 was greater than the number of units in the Master Address Register at the same basic street address, the census questionnaire failed edit. Failed edit questionnaires were sent to telephone follow-up first. If there was no response during telephone follow-up or the respondent confirmed that the number of units was greater than the number of units listed in the Master Address Register, the questionnaire was sent to personal visit follow-up. See the Appendix for a more detailed account of the H4-Edit procedures.

III. Sample Plan and Estimates

To get an estimate of the number of housing units added by the H4-Edit, it was necessary to have both census questionnaires and Master Address Registers for an ED; therefore, the Census Allocation Program Evaluation Sample (CAPE) was chosen. The CAPE sample was a systematic sample from the universe of all non-zero ED's in the Data Acceptance Census File after 100% data capture. Two hundred and eighty four ED's were selected in the sample with equal probability (1 in a 1,000). Of the 284 ED's selected, 260 ED's were in centralized or decentralized offices, and in each of these ED's all the census questionnaires were saved. The remaining 24 ED's were conventional ED's and in these ED's the questionnaires were not saved. That these 24 conventional ED's did not have their questionnaires saved is the reason this study only estimates housing unit adds from centralized and decentralized offices. However, the estimates that were produced can be considered fairly representative of the total United States, if we consider the H4-Edit procedures used in conventional offices and the fact that less than 10 percent of the district offices were conventional.

To estimate the number of housing units added by the H4 edit, we used clerks in Jeffersonville, Indiana. The clerks went through all the census questionnaires in the 260 ED's we had for this study and pulled all the census questionnaires that were marked as having failed the H4 edit during the census. Then they checked the MAR's to tally the number of adds that an address produced in Follow-up 2 due to H4. The tallies were computed for each ED in sample. Two hundred thirty of the sample ED's had no adds due to H4 edit.

Estimates for this study were produced by weighting the ED's tallied by the inverse of their probability of selection (1 in 1,000) and summing over all ED's. Sampling errors were produced for all estimates by assuming simple random sampling.

IV. Analysis

An estimated 93,000 units were added to the census by the H4 procedure, representing an improvement in the housing unit count of about 0.1 of 1 percent. Among the households mailing back their census questionnaires, about 2,120,000 addresses were followed-up as a result of having failed H4 edit. Of these 2,120,000 addresses, 1,232,000 units were

known in the census as single units and 704,000 were known as multiunit structures. The results of the H4 Follow-up could not be determined for the additional estimated 184,000 of these addresses because of a lack of sufficient address information. Of the 1,232,000 single units that failed H4 edit, units were added at an estimated 50,000 of these addresses, and of the 704,000 multi-unit structures, addresses were added at 22,000 of these addresses. At some of the addresses more than one unit was added resulting in a total of 93,000 units. If we take into account the 184,000 non-processable cases and assume these could have been processed for this study, we probably would have an estimated 100,000 units added by H4. However, this number is not significantly different from the 93,000 we estimated.

V. Costs of H4

There are not exact costs for the H4 operation in the census; however, it can be estimated that the entire H4 operation cost a total of roughly \$7.5 million to carry out. This estimate was derived by breaking the H4 operation into two phases, the editing and follow-up phases, estimating their costs and summing. For the editing phase, we assumed that the H4 edit of all census questionnaires cost about 30 percent of the total check in and editing costs after adjusting for serialization and sorting. Using cost estimates provided by the Field Division, the H-4 edit cost roughly 30 percent of \$13 million or about \$4 million. For the second phase, the H4 follow-up, it was assumed that of 2,140,000 addresses that failed H-4 Edit, half went to personal visit follow-up and the other half to telephone follow-up only. For telephone follow-up this would mean 1,070,000 calls or roughly \$500,000. For personal visit follow-up, if we assume that the average hourly enumerator pay rate and workload applies to the 1,070,000 addresses in this phase, and add in proportionately estimated crew leader, check-in and training costs, this part of the follow-up is roughly estimated at \$3 million.

VI. Conclusions

In 1970, using a similar H4 edit operation 126,000 units were added to the census for a housing count improvement of 0.2 percent. This estimate of 0.2 percent is not statistically different from the 0.1 percent we added in 1980. As evidence points out in 1970 and 1980, the amount of coverage improvement gained from H4 has been minimal. There are a number of potential reasons for this, some of which are as follows:

- 1) The H4 edit may be difficult to carry out.
- 2) The H4 question may not be a good way to find missed housing units.
- 3) It may be that by the time the H4 follow-up is done there may be very few housing units which are missed. The H4 follow-up occurs late in the census process, and well after the major address list update procedures have been undertaken. If housing unit undercoverage is low, then the amount of housing units that can be added will also be low.

From an operational view, there is evidence that respondents had difficulty answering the

H4 question and/or there were many editing problems. In our estimates it can be seen that of the estimated 1,232,000 single units that failed H4 edit, 1,182,000 remained single units and likewise of the 704,000 multi-unit structures that failed H4 edit, 682,000 received no added units from H4. Collectively, 96 percent of the units that failed H4 edit had no added units. How much error is due to respondents is difficult to ascertain, since

our processing for this study, there is also evidence of H4 editing errors were census clerks failed units in the H4 edit for no observable reason outlined in the procedures.

In terms of cost effectiveness, the H4 edit operation cost roughly \$7.5 million. If we consider the fact that we added an estimated 93,000 housing units in 1980, this suggests that we spent \$81 per unit added by H4 edit