Census 2000 Evaluation L.2
June 30, 2003

Operational Analysis of the Decennial Response File Linking and Setting of Housing Unit Status and Expected Household Size

FINAL REPORT

This evaluation reports the results of research and analysis undertaken by the U.S. Census Bureau. It is part of a broad program, the Census 2000 Testing, Experimentation, and Evaluation (TXE) Program, designed to assess Census 2000 and to inform 2010 Census planning. Findings from the Census 2000 TXE Program reports are integrated into topic reports that provide context and background for broader interpretation of results.

Miriam Rosenthal
Decennial Statistical Studies Division

U S C E N S U S B U R E A U
Helping You Make Informed Decisions

CONTENTS

EXECUTIVE SUMMARY iv

1. BACKGROUND 1

1.1 The 1990 Census 1
1.2 Census 2000 Dress Rehearsal 1
1.3 Census 2000 2

2. METHODS 3

2.1 Files used in this evaluation 3
2.2 Enumerator-replaced returns 4
2.3 Blank returns 4
2.4 Definition of valid person records 4
2.5 Quality assurance procedures 5

3. LIMITS 5

4. RESULTS 5

4.1 The linking process 5
4.2 The setting of housing unit status 7
4.3 The setting of expected household size 16

5. CONCLUSIONS 19

References 21

Appendix A: Self-response, enumerator, and other returns 22
Appendix B: Definition of a data-defined person 23

Appendix C: Classifying type of enumeration areas as Mail, Update/Leave, Update/Enumerate, or List/Enumerate 24
Appendix D: Interview Summary section of the enumerator questionnaire 25

LIST OF TABLES

Table 1. Status resolution: enumerator-replaced returns 4

Table 2. Type of enumeration area for linked returns 6

Table 3. Type of forms for linked returns 6

Table 4. Number of forms comprising a return 7

Table 5. Status resolution: overall 7

Table 6. Status resolution: self-response returns 8

Table 7. Status resolution: enumerator returns 9

Table 8. Enumerator returns: Nonresponse Followup and Coverage Improvement Followup 10

Table 9. Enumerator returns: Occupied with Unresolved Population Count 12

Table 10. Enumerator returns: Unresolved Occupied/Vacant 14

Table 11. Enumerator returns: Unresolved Occupied/Vacant/Delete 15

Table 12. Setting expected household size: Occupied paper mail returns 17

Table 13. Setting expected household size: Occupied enumerator returns, excluding orphans 19

EXECUTIVE SUMMARY

This evaluation looks at three Decennial Response File processes: linking and the setting of housing unit status and expected household size. The Decennial Response File is the first in a series of files, which ultimately produces final census population counts. Errors at this step, including omissions, deletions, and misclassifications, may impact subsequent files.

A return represented a single household enumeration. A return consisted of one or more Decennial Response File household forms. For example, we linked an enumerator continuation form to its parent enumerator form to create one return. Similarly, we linked a Be Counted Form for a partial household to a mail return form if we identified the Be Counted Form as the mail return's "continuation," containing information on the additional household members for which there was no room on the mail return questionnaire.

In mailback areas, the use of two forms to enumerate large households—requiring linking—most commonly occurred in the Nonresponse Followup and Coverage Improvement Followup operations. Typically the two forms were an enumerator first form and an enumerator continuation form.

Large households on mail returns generally did not involve linking. We produced a composite record for these large households from the originating mail return and a Coverage Edit Followup telephone interview. The Coverage Edit Followup interview collected census data for members of mail return households for which there was no room on the mailback questionnaire.

We assigned to each return a housing unit status and if we determined the status to be occupied an expected household size. Then, we applied the Primary Selection Algorithm, a computer program run on the Decennial Response File to select one return to represent each housing unit in the census, if multiple returns were present for a housing unit. The setting of housing unit status and expected household size occurred as follows:

- If there was sufficient information to determine the housing unit status, we set the housing unit status to occupied, vacant, or delete. If a housing unit was occupied, we set expected household size based on all available information, such as the following: total number of valid person records, number of names on the census questionnaire roster, the respondent-reported household size, and the enumerator-reported household size (also referred to as the Interview Summary Population). If there was insufficient information on household size, we set expected household size to unresolved.

- When there was insufficient or conflicting information on the housing unit status, we set the housing unit status and expected household size to unresolved.

This evaluation presents the results from the linking of census forms and the setting of housing unit status and expected household size. First, we linked household forms—usually no forms were linked—to identify the combination that constituted a single return. After the linking

process, we assigned to each return a housing unit status and if we determined the status to be occupied an expected household size.

Results

Of 129,389,529 returns, 1,387,085 returns, or 1.07 percent, were linked; that is, they were returns comprised of two or more forms. Of these, 39,108 returns, or 2.82 percent, had three or more forms.

The type of enumeration area with the highest rate of linked returns was in Update/Enumerate: 41,559 of 1,052,591 returns, or 3.95 percent. Large households probably caused this result. Most linked returns, 1,384,233 returns or 99.79 percent, were comprised of an enumerator first and an enumerator continuation form. Enumerators used this combination of forms to enumerate large households in the List/Enumerate, Update/Enumerate, Nonresponse Followup, and Coverage Improvement Followup operations. Linkage rates comparable to Update/Enumerate did not occur in List/Enumerate probably because of a processing error. Enumerator continuation forms in List/Enumerate—along with Update/Leave adds and Update/Enumerate adds—were erroneously omitted from the Decennial Response File.

Of 129,389,529 returns, 1,318,350 returns, or 1.02 percent, had either an unresolved expected household size or an unresolved housing unit status. The three unresolved categories were the following: Occupied with Unresolved Population Count, Unresolved Occupied/Vacant, and Unresolved Occupied/Vacant/Delete.

We sent mail returns to Coverage Edit Followup if there was an inconsistency in household size, leaving few self-response returns unresolved.

A programming error affected the status resolution for some Vacant enumerator returns. Mistakenly, we recoded any Interview Summary Population of 0 to blank. As a result, we may have classified up to 133,438 Vacant returns as Deletes and up to 258,963 Vacant returns as Unresolved Occupied/Vacant. As many as 145,367 housing units of the 191,826 housing units in the census that had their occupancy status imputed, or 75.78 percent of housing units that had their occupancy status imputed, may have been affected by this latter error.

We had 712,858 unresolved enumerator returns, or 1.51 percent of all enumerator returns (329,895 returns were Occupied with Unresolved Population Count; 329,266 returns were Unresolved Occupied/Vacant; and 53,697 returns were Unresolved Occupied/Vacant/Delete). Prominent results concerning these unresolved enumerator returns were as follows:

- The biggest reason for Occupied with Unresolved Population Count returns was an Interview Summary Housing Unit Status of Occupied and an Interview Summary Population of "POP unknown."

- Most Unresolved Occupied/Vacant returns were returns with one or more valid

person records, an Interview Summary Housing Unit Status of Vacant, and a blank Interview Summary Population.

- Most of the Unresolved Occupied/Vacant/Delete returns had no valid person records. All returns with no valid person records had an Interview Summary Housing Unit Status of Delete and an Interview Summary Population of greater than 0 or "POP unknown."

These results highlight confusion in filling the Interview Summary boxes. The challenges in filling out the Interview Summary boxes could be manifold: insufficient training in how to complete this section, limited information from proxies, conflicting data, seasonal vacants or other confusing situations, unclear dependency between the Interview Summary boxes, or omissions. Also not capturing data or misinterpreting scanned images could have contributed to unresolved enumerator returns.

For Occupied self-response (restricted to paper mailback questionnaires) and enumerator returns, setting the expected household size was usually straightforward. For 74,725,437 self-response returns, or 93.71 percent, the number of valid person records and roster names corresponded to the respondent-reported household size. For enumerator returns, most household size measures also were consistent. This can be seen in the following results for enumerator returns:

- For linked returns or unlinked returns that had the "continuation form(s) attached" box checked and the expected household size equal to the Interview Summary Population, 1,475,382 returns, or 99.11 percent, had the same Interview Summary Population and respondent-reported household size.

- For unlinked returns with the Interview Summary Population less than or equal to five 26,897,133 returns, or 99.52 percent, had the same Interview Summary Population, number of valid person records, and respondent-reported household size.

- For unlinked returns with the Interview Summary Population greater than five and the respondent-reported household size greater than five, 68,599 returns, or 99.58 percent, had the same Interview Summary Population and respondent-reported household size.

Recommendations

We recommend attempting to link only enumerator first and enumerator continuation forms, if such forms exist in the future. Doing so would simplify the linking process, cause very little loss of data, and would have almost no effect on the population counts.

We recommend a redesigned Interview Summary Section, if this section exists in the future, to improve the consistency of responses.

We recommend using Mobile Computing Devices to help ensure data capture and consistency of responses.

1.BACKGROUND

1.1 The 1990 Census

The 1990 Census household questionnaire had space to report demographic characteristics for one to seven people. The Census Bureau followed up by telephone or through a personal visit all households that returned their census form by mail with evidence of seven or more household members. We collected demographic information on a continuation form for any additional people. Large households encountered during Nonresponse Followup also had more than one form completed for them. The Search/Match process could also generate questionnaires for housing units. Search/Match attempted to place people who were not counted during the regular data collection where they said they belonged, working from Were You Counted forms and various other inputs.

The Primary Selection Algorithm, a computer program run on the data capture files, selected the form or forms representing each housing unit in the census. A part of that process was to determine whether the selected main form (containing the householder) had related continuation forms that should be selected also. The decision to select related continuation forms relied on the housing unit population count from the selected main form. The results of the Primary Selection Algorithm determined the final expected household size. We used field-keyed Interview Summary information to settle inconsistencies in the household status and expected household size.

1.2 Census 2000 Dress Rehearsal

Identifying the main, or parent, records and their associated continuation records became the first step in the process that created the Census 2000 Dress Rehearsal Decennial Response File. The linking process started with individual census forms and created housing unit returns by combining the individual forms as necessary. In most instances, the housing unit returns created by the linking process were comprised of a single census form.

For the first time, the linking process was a separate step from the Primary Selection Algorithm. This was done because of the large number of different forms and because of the increased possibility that forms not designed as continuation forms would be used as continuation forms by respondents attempting to complete the enumeration of their households.

The design of the dress rehearsal forms had significantly increased the number of households considered to be "large" (i.e., households with more members than could be accommodated by a single census form). Census dress rehearsal questionnaires contained only enough room to accommodate households of one to five people. Large mail return households were sent special questionnaires called Large Household forms on which they were asked to report the information of those household members for whom there was no room on the original form. It was necessary to link these special forms with the initial mail return questionnaire.

In the dress rehearsal, there were examples of forms used incorrectly as continuation forms. Respondents used replacement forms mailed to every address in the mail-out universe as continuation forms. Respondents used Be Counted Forms (BCFs) made available to people in public places as continuation forms. Enumerators used first forms as continuation forms instead of enumerator continuation forms. In some cases enumerators accepted mail return questionnaires during Nonresponse Followup and used the enumerator first form or an enumerator continuation form as a continuation form for the mail return.

To accomplish the linking, we established a form-based expected household size, based solely on the information available for the individual form. We used these form-based expected household sizes to decide which forms should be linked together to represent a single enumeration of the household, known as the housing unit return. Next, we determined the expected household size for each housing unit return. We used field-keyed Interview Summary information to settle inconsistencies in the household status and expected household size.

1.3 Census 2000

The Census 2000 enumerator questionnaire allowed the enumerator to report demographic characteristics for one to five people, and for more than five people on enumerator continuation forms. We used a process very similar to dress rehearsal in Census 2000 to link forms and establish an expected household size. The purpose of linking the Decennial Response File household forms was to identify the combination of household forms that constituted a single return. The linking process started by identifying one form at a housing unit as the parent form and then determined if other forms (child forms) should be linked to the parent form to form a return. Enumerator continuation forms and partial BCFs were not eligible to be parent forms. We did not link forms resulting from Computer-Assisted Telephone Interview (CATI) operations (Telephone Questionnaire Assistance (TQA) reverse-CATI, Coverage Edit Followup) because of the ability of these operations to entirely enumerate large households.

A return represented a single household enumeration. A return consisted of one or more Decennial Response File household forms. For example, we linked one or more enumerator continuation forms to their parent enumerator form to create one return. Similarly, we linked a BCF for a partial household to a mail return form if we identified the BCF as the mail return's "continuation," containing information on the additional household members for which there was no room on the mail return questionnaire.

Unlike the dress rehearsal linking process, the census mail return linking process did not involve Large Household forms. The Coverage Edit Followup interview collected census data for members of mail return households for which there was no room on the mailback questionnaire (which collected demographic characteristics for one to six people). This process produced a composite record—not considered a linked return—that combined the information from the originating mail return with the telephone interview information, eliminating the need to link forms for households with more than six people.

Next we assigned to each return a housing unit status and if the status was determined to be occupied an expected household size prior to the application of the Primary Selection Algorithm. The setting of housing unit status and expected household size occurred as follows:

- If there was sufficient information, we set the housing unit status to occupied, vacant, or delete. If a housing unit was occupied, we attempted to set expected household size.
  - We set expected household size using all available information, such as total number of valid person records, number of names on the census questionnaire roster, the respondent-reported household size, and the enumerator-reported household size (also referred to as the Interview Summary Population).
  - If there was insufficient information on household size, we set expected household size to unresolved.
- When there was insufficient or conflicting information on the housing unit status, we set the housing unit status and expected household size to unresolved.

Unlike the 1990 Census and Census 2000 Dress Rehearsal, we did not use field-keyed Interview Summary information to settle inconsistencies in the household status and expected household size.

2.METHODS

2.1 Files used in this evaluation

The Decennial Response File provided data at the census form, return, and person levels.

The variable RRAS on the Decennial Response File identified if a return was the result of linking two or more forms. We created categories of the combinations of forms comprising a single return (i.e., a set of linked forms). We tallied the counts in each of the categories in order to examine the results of the linking process.

For the setting of housing unit status and expected household size, we looked at the variables RSTATUS, status of return; and REXPOP, expected population for this return. We placed returns into one of the following three categories:

1. those with consistent data on housing unit status;
2. those with inconsistent data on housing unit status; and,
3. occupied housing units with missing data on household size.

We tallied the three categories by self-response, enumerator, and other returns. Appendix A lists which returns were self-response, enumerator, and other. We examined these data to study the

sufficiency and the characteristics of the data used to determine housing unit status and expected household size.

The final Hundred Percent Census Unedited File provided a file of housing units with imputed occupancy status—housing units assigned a housing unit status of either Occupied or Vacant. We matched this file to housing units on the Decennial Response File that had returns possibly misclassified as Unresolved Occupied/Vacant instead of Vacant. The intent was to identify the possible number of housing units for which we imputed occupancy status unnecessarily.

2.2 Enumerator-replaced returns

This evaluation excludes enumerator-replaced returns, the variable RPELIG = 3. Table 1 presents the status of the 696,691 enumerator-replaced returns. There were 15,655 blank enumerator-replaced returns and 681,036 non-blank enumerator-replaced returns.

Table 1. Status resolution: enumerator-replaced returns

Status	Number	Percent
Total enumerator-replaced returns	696,691	100.00
Occupied	94,299	13.54
Vacant	38,753	5.56
Delete	21,584	3.10
Occupied with Unresolved Population Count	517,450	74.27
Unresolved Occupied/Vacant	21,191	3.04
Unresolved Occupied/Vacant/Delete	3,414	0.49

Data source: Decennial Response File

2.3 Blank returns

This evaluation excludes blank returns, the variable RBLANK = 1. There were 181,436 blank returns in addition to the blank enumerator-replaced returns mentioned in Section 2.2. Linking and setting housing unit status and expected household size only occurred for non-blank returns.

2.4 Definition of valid person records

A person record is valid if all of the following conditions hold:

- the person record is ***not*** Data Capture Audit Resolution non-valid; and
- the person-level record is data-defined; and
- the person-level record is not canceled by Coverage Edit Followup; and
- the person-level record is not canceled by an enumerator.

See Appendix B for the definition of a data-defined person.

2.5 Quality assurance procedures

We applied quality assurance procedures throughout the creation of this report. They encompassed how we determined evaluation methods, created specifications for project procedures and software, designed and reviewed computer systems, developed computer procedures, analyzed data, and prepared this report.

3.LIMITS

A possible limit in answering the question regarding the combination of forms that can make up a single return is that the Decennial Response File does not retain return-level data from the child form. The Decennial Response File retains the parent form and all person-level records from the child form. The variable PFT (form type) on person-level records can be used to determine the type of the child form. It may be difficult to determine the child records if the parent form and a child form are of the same form type or if there is more than one child return of the same form type.

4.RESULTS

This section presents the results of the linking process and the setting of housing unit status and expected household size.

4.1 The linking process

Table 2 shows the number of linked returns for Mail, Update/Leave, Update/Enumerate, and List/Enumerate. (See Appendix C to see how we classified the nine type of enumeration areas into these four categories.)

The lower rate of linked forms in the Mail and Update/Leave areas—or mailback areas—compared to Update/Enumerate areas, can be explained by the existence of the Coverage Edit Followup operation for large households in these areas. The Coverage Edit Followup interview resulted in a single return for large households on mail returns. In the mailback areas, two forms (an enumerator first form and an enumerator continuation form) were routinely used to enumerate large households only in the Nonresponse Followup and Coverage Improvement Followup operations.

In Update/Enumerate areas, all large households were routinely enumerated on two forms, requiring them to be linked in the Decennial Response File linking process.

The low linkage rate in List/Enumerate illustrates another result. Enumerator continuation forms in List/Enumerate—along with Update/Leave adds and Update/Enumerate adds—were not included on the Decennial Response File because of a processing error. This low linkage rate and the fact that few unlinked enumerator continuation forms, or orphan returns (9 of the 33,472 orphan returns), were in List/Enumerate is evidence of this error. This mistake could have affected the expected household size and data completeness in these areas. The coverage impact is possibly small because the Interview Summary Population, a piece of information not affected by this mistake, also informs the setting of expected household size.

Table 2. Type of enumeration area for linked returns

Type of enumeration area	Linked returns	All returns	Percent linked
Total returns	1,387,085	129,389,529	1.07
Mail	1,106,072	101,421,457	1.09
Update/Leave	239,337	26,464,251	0.90
Update/Enumerate	41,559	1,052,591	3.95
List/Enumerate	117	451,230	0.03

Data source: Decennial Response File

Table 3 shows the combination of forms comprising linked returns. We linked few forms, other than enumerator first and enumerator continuation forms. We recommend attempting to link only enumerator first and enumerator continuation forms, if such forms exist in the future. Doing so would simplify the linking process causing very little loss of data and having almost no effect on the population counts.

Table 3. Type of forms for linked returns

Form 1/form 2	Number	Percent
Total linked returns	1,387,085	100.00
Enumerator first/Enumerator continuation	1,384,233	99.79
Mail/Enumerator continuation	328	0.02
Mail/Enumerator first	2,039	0.15
Mail/Mail	332	0.02
Be Counted (whole)/Be Counted (partial)	28	0.00
Mail/Be Counted (partial)	109	0.01
More than two form types	16	0.00

Data source: Decennial Response File

Table 4 shows the number of forms that comprised the Decennial Response File returns after completing the linking process. Of 129,389,529 returns, 1,387,085 returns, or 1.07 percent, were linked; that is, they were returns comprised of two or more forms. Of these, 39,108 returns, or 2.82 percent, had three or more forms.

Table 4. Number of forms comprising a return

Forms per return	Number	Percent
Total returns	129,389,529	100.00
1	128,002,444	98.93
2	1,347,977	1.04
3+	39,108	0.03

Data source: Decennial Response File

4.2 The setting of housing unit status

In this section, we examine how we set the housing unit status; in particular, how we set the housing unit status for Nonresponse Followup and Coverage Improvement Followup; and how we determined the housing unit status of unresolved enumerator returns.

3.1.1 The setting of housing unit status: overall, self-response returns, enumerator returns, and other returns

Overall

Table 5 shows the housing unit status resolution. The housing unit status for most returns was resolved (Occupied, Vacant, or Delete). The cases with a resolved status were returns with consistent data on household size. The Unresolved Population Count cases were occupied returns with inconsistent data on household size. The Unresolved Occupied/Vacant and Unresolved Occupied/Vacant/Delete were returns with inconsistent data on housing unit status.

Table 5. Status resolution: overall

Status	Number	Percent
Total returns	129,389,529	100.00
Occupied/Vacant/Delete	128,071,179	98.98
Occupied with Unresolved Population Count	934,849	0.72
Unresolved Occupied/Vacant	329,804	0.25
Unresolved Occupied/Vacant/Delete	53,697	0.04

Data source: Decennial Response File

Self-response returns

Table 6 shows the status resolution for self-response returns. We classified self-response returns as Occupied, Vacant, or Unresolved Occupied/Vacant using the following information: the number of valid person records, the number of names on the roster, and the respondent-reported household size. Two exceptions were that Internet Data Capture returns were not classified as Vacant and TQA reverse-CATI and Coverage Edit Followup returns were not classified as Unresolved Occupied/Vacant.

For all other self-response returns (i.e., census questionnaires returned via the U.S. Postal Service), the housing unit status was Vacant or Unresolved Occupied/Vacant under the following conditions:

- Vacant: if the self-response return had no valid person records or names on the roster, and the respondent-reported household size was 0.
- Unresolved Occupied/Vacant: if the self-response return had no valid person records or names on the roster, and the respondent-reported household size was blank.

Of the 1,270,385 Coverage Edit Followup returns included in the self-response returns, 3,146, or 0.25 percent, were Vacant. Most of the returns with an Unresolved Occupied/Vacant status were Internet Data Capture returns, 388 of the 538 returns, or 72.12 percent.

Table 6. Status resolution: self-response returns

Status	Number	Percent
Total self-response returns	81,099,704	100.00
Occupied	81,080,662	99.98
Vacant	18,504	0.02
Unresolved Occupied/Vacant	538	0.00

Data source: Decennial Response File

Enumerator returns

Table 7 shows the status resolution for enumerator returns. The largest unresolved category for enumerator returns was the category of Occupied with Unresolved Population Count.

Overall 20,082,071 enumerator returns, or 42.66 percent were proxy. Of the resolved enumerator returns, 19,687,419 returns, or 42.46 percent were proxy. Of the unresolved enumerator returns, 394,652 returns, or 55.36 percent were proxy.

A programming error affected the status resolution for some Vacant enumerator returns. Mistakenly, we recoded any Interview Summary Population, the variable RISPOP, of 0 to blank. This meant we may have classified up to 133,438 Vacant returns as Deletes and up to 258,963 Vacant returns as Unresolved Occupied/Vacant.

Table 7. Status resolution: enumerator returns

Status	Number (percent)		Percent proxy
Total enumerator returns	47,080,158	(100.00)	42.66
Occupied	30,465,137	(64.71)	17.41
Vacant	14,123,339	(30.00)	93.60
Delete	1,778,824	(3.78)	65.52
Occupied with Unresolved Population Count	329,895	(0.70)	30.77
Unresolved Occupied/Vacant	329,266	(0.70)	82.48
Unresolved Occupied/Vacant/Delete	53,697	(0.11)	40.18

Data source: Decennial Response File

Other returns

Not considered to be self-response or enumerator returns were Usual Home Elsewhere returns for individuals (the variable RSOURCE = 26-29), which consisted of 604,954 Occupied with Unresolved Population Count returns, and BCF returns (the variable RSOURCE = 11, 12, 32, 33, 35), which consisted of 604,713 Occupied returns. The 604,954 Occupied with Unresolved Population Count returns were 64.71 percent of the overall 934,849 returns that were Occupied with Unresolved Population Count.

3.1.2 The setting of housing unit status: Nonresponse Followup and Coverage Improvement Followup

Table 8 examines two Census 2000 operations that used enumerator returns: Nonresponse Followup and Coverage Improvement Followup. Coverage Improvement Followup had about a 1 percent higher unresolved rate than Nonresponse Followup, 2.28 percent versus 1.34 percent. Coverage Improvement Followup had about twice the percent of Occupied with Unresolved Population Count than Nonresponse Followup, 1.25 percent versus 0.58 percent. Nonresponse Followup targeted housing units in mailback areas where a census questionnaire had not been checked-in by April 22, 2000. The Coverage Improvement Followup operation mostly checked the housing unit status of Vacants and Deletes.

The overlap between unresolved housing units in Nonresponse Followup and Coverage Improvement Followup was 4,223 housing units. So there were not many housing units

classified as unresolved as a result of both Nonresponse Followup and Coverage Improvement Followup.

Table 8. Enumerator returns: Nonresponse Followup and Coverage Improvement Followup

	Nonresponse Followup		Coverage Improvement Followup	
Status	**Number**	**Percent**	**Number**	**Percent**
Total enumerator returns	38,796,478	100.0	6,760,744	100.0
Total resolved enumerator returns	38,275,218	98.66	6,606,658	97.72
Total unresolved enumerator returns	521,260	1.34	154,086	2.28
Occupied with Unresolved Population Count	223,109	0.58	84,461	1.25
Unresolved Occupied/Vacant	251,830	0.65	63,616	0.94
Unresolved Occupied/Vacant/Delete	46,321	0.12	6,009	0.09

Data source: Decennial Response File

3.1.3 Determining housing unit status for unresolved enumerator returns: Occupied with Unresolved Population Count, Unresolved Occupied/Vacant, Unresolved Occupied/Vacant/Delete

We categorized the unresolved enumerator returns in Tables 9-11. We used the following information to classify the unresolved enumerator returns: the number of valid person records, the respondent-reported household size, Interview Summary Housing Unit Status (Item A), Interview Summary Population (Item B), Interview Summary Vacancy Type (Item C). Appendix D has an image of the Interview Summary section of the enumerator questionnaire.

The unresolved enumerator returns were a result of contradictory and missing responses on the questionnaire. Some of the contradictions may have been introduced because the Interview

Summary Section was not intuitive for unusual situations. Perhaps the enumerators needed more comprehensive instructions for complicated cases. Crew leaders not performing adequately their review process of the completed questionnaires may have also contributed to unresolved enumerator returns. Also not capturing data or misinterpreting scanned images could have contributed to unresolved enumerator returns. A redesigned Interview Summary Section would probably improve the consistency of responses. Mobile Computing Devices would help ensure data capture and consistency of responses.

Occupied with Unresolved Population Count

Table 9 shows the reasons we assigned occupied enumerator returns an Unresolved Population Count. Most cases had no valid person records: 283,252 returns, or 85.86 percent of the Occupied with Unresolved Population Count category. All returns with no valid person records had an Interview Summary Housing Unit Status of Occupied, and 190,499 of these returns, or 67.25 percent, had an Interview Summary Population of "POP unknown."

Occupied enumerator returns could have an Unresolved Population Count if the enumerator knew the house to be occupied but was unable to complete an interview or if a proxy did not have information about all of the household members.

Table 9. Enumerator returns: Occupied with Unresolved Population Count

Interview Summary Housing Unit Status (Item A)	Interview Summary Population (Item B)	Interview Summary Vacancy Type (Item C)	Respondent-reported population	Number (percent)	Percent proxy
Enumerator returns: Occupied with Unresolved Population Count				329,895 (100.00)	30.77
Return with no valid person records				283,252 (85.86)	28.82
Occupied	1-97	Any	0 or blank	91,187 (27.64)	25.44
Occupied	POP unknown (99)	Any	Any	190,499 (57.75)	30.31
Occupied	Blank	Any	> 0	1,566 (0.47)	43.87
Return with one or more valid person records				46,643 (14.14)	42.63
Occupied	POP unknown (99)	Any	Any	46,117 (13.98)	42.79
Blank	POP unknown (99)	Any	Any	526 (0.16)	28.90

Data source: Decennial Response File

Unresolved Occupied/Vacant

Table 10 shows the combinations of responses that generated an Unresolved Occupied/Vacant status. In these cases, the responses within a return contradicted each other or were missing, making it unclear whether the housing unit was Occupied or Vacant. Most of the cases, 82.48 percent of the Unresolved Occupied/Vacant, were proxy. It seems that the interviewer had difficulty reconciling the proxy information and the Interview Summary section of the census questionnaire.

A programming error affected the status resolution for some Vacant enumerator returns. Mistakenly, we recoded any Interview Summary Population of 0 to blank. We classified up to 258,963 Vacant returns as Unresolved Occupied/Vacant.

Corroborating evidence supporting assigning a Vacant housing unit status for the returns affected by the programming error was a filled Interview Summary Vacancy Type: for rent; for sale only; rented or sold, not occupied; for seasonal, recreational, or occasional use; for migrant workers; or other vacant. Of the 258,963 returns possibly misclassified as Unresolved Occupied/Vacant, 244,070 returns, or 94.25 percent, had the Interview Summary Vacancy Type filled. By combination of responses, the results were as follows:

- 39,885 returns out of 46,846 returns, or 85.14 percent, had no valid person records, an Interview Summary Housing Unit Status of Occupied, and a respondent-reported population of 0 or blank;

- 203,325 returns out of 211,257 returns, or 96.25 percent, had one or more valid person records and an Interview Summary Housing Unit Status of Vacant;

- and 860 returns out of 860 returns had one or more valid person records and a blank Interview Summary Housing Unit Status.

By looking at the final Hundred Percent Census Unedited File, we can further measure the possible impact of the programming error. Of the 258,963 possibly misclassified returns, 145,367 returns were selected by the Primary Selection Algorithm (the variable RPRSTAT = 1) and were imputed on the final Hundred Percent Census Unedited File. Of the 191,826 housing units in the census that had their occupancy status imputed, these 145,367 housing units, or 75.78 percent, were possibly Vacant housing units that should not have had their occupancy status imputed.

Enumerator confusion about completing the Interview Summary Section for seasonal vacants also probably contributed to some of the unresolved Occupied/Vacant returns. For instance, the enumerator may have determined the house was a vacation or seasonal home and skipped to the Interview Summary Section as instructed, filled in the Interview Summary Housing Unit Status (Item A) as “vacant-usual home elsewhere,” and the Interview Summary Population (Item B) because someone was living there April 1, 2000. For returns with an Interview Summary

Housing Unit Status of Vacant and not impacted by the programming error, 12,002 of the 54,106 returns, or 22.18 percent, were Usual Home Elsewhere vacants.

Table 10. Enumerator returns: Unresolved Occupied/Vacant

Interview Summary Housing Unit Status (Item A)	**Interview Summary Population (Item B)**	Interview Summary Vacancy Type (Item C)	Respondent-reported population	Number (percent)	Percent proxy
Enumerator returns: Unresolved Occupied/Vacant				329,266(100.00)	82.48
Return with no valid person records				100,277 (30.45)	84.39
Occupied	Blank	Any	0 or blank	46,846 (14.23)	83.57
Vacant	1-97	Any	Any	49,900 (15.15)	87.39
Vacant	POP unknown (99)	Blank	Any	3,216 (0.98)	53.73
Blank	1-97	Any	0 or blank	241 (0.07)	41.08
Blank	POP unknown (99)	Any	Any	74 (0.02)	58.11
Blank	Blank	Any	Any	0 (0.00)	0.00
Return with one or more valid person records				228,989 (69.55)	81.64
Vacant	POP unknown (99)	Any	Any	990 (0.30)	75.66
Vacant	Blank	Any	Any	211,257 (64.16)	86.89
Blank	Blank	Any	Any	16,742 (5.08)	15.69

Data source: Decennial Response File

Unresolved Occupied/Vacant/Delete

Table 11 shows the combinations of responses that generated an Unresolved Occupied/Vacant/Delete status. These returns have contradictions between the Interview Summary Housing Unit Status and the Interview Summary Population. Since none of the

Interview Summary Housing Unit Status boxes were titled "Delete," the enumerator may not have understood when to pick Delete when completing the Interview Summary Population.

Most of the Unresolved Occupied/Vacant/Delete returns had no valid person records: 48,615 returns, or 90.54 percent. All of these returns with no valid person records had an Interview Summary Housing Unit Status of Delete and an Interview Summary Population of greater than 0 or "POP unknown."

Another contradiction was that the return contained valid person records but had an Interview Summary Housing Unit Status or Interview Summary Population that indicated Vacant or Delete. One possible explanation is that the enumerator went down the wrong path and then tried to remedy the situation by entering the correct information in the Interview Summary section.

Table 11. Enumerator returns: Unresolved Occupied/Vacant/Delete

Interview Summary Housing Unit Status (Item A)	Interview Summary Population (Item B)	Interview Summary Vacancy Type (Item C)	Respondent-reported population	Number (percent)	Percent proxy
Enumerator returns: Unresolved Occupied/Vacant/Delete				53,697 (100.00)	40.18
Return with no valid person records				48,615 (90.54)	38.51
Delete	1-97	Any	Any	21,923 (40.83)	53.21
Delete	POP unknown (99)	Any	Any	26,692 (49.71)	26.44
Return with one or more valid person records				5,082 (9.46)	56.16
Vacant	Delete (98)	Any	Any	917 (1.71)	63.90
Delete	POP unknown (99)	Any	Any	821 (1.53)	34.23
Delete	Blank	Any	Any	3,162 (5.89)	61.07
Blank	Delete (98)	Any	Any	182 (0.34)	30.77

Data source: Decennial Response File

4.3 The setting of expected household size

In this section, we examine how we set expected household size for Occupied self-response returns and Occupied enumerator returns. We set expected household size only for Occupied returns.

4.3.1 The setting of expected household size: Occupied self-response returns

Here are definitions of variables used to set expected household size for Occupied self-response returns:

- RNPOP is the respondent-reported household size, Question 1 on the self-response questionnaire: "How many people were living or staying in this house, apartment, or mobile home on April 1, 2000?"

- DPPOP is the number of valid person records for the census questionnaire.

- ROSPOP is the number of the valid roster names for the census questionnaire (persons 7-12 on short-form returns and persons 1-12 on long-form returns).

- For short-form returns: TOTPOP = (DPPOP + ROSPOP) from the parent form.

- For long-form returns: TOTPOP = Maximum (DPPOP, ROSPOP) from the parent form.

- ALLPOP is the DPPOP for the census return (the sum of the DPPOP for the parent form and all child forms linked to the parent form).

The occupied self-response returns described in Table 12 are paper mailback questionnaires (RSOURCE = 1-10). The following outcomes determined the expected household size for self-response returns:

- Whether or not the return included linked forms,
- Whether or not a short form had a continuation roster,
- Whether or not RNPOP > 6, and
- Whether or not RNPOP = TOTPOP.

In Table 12, few self-response returns were comprised of linked forms, 2,800 returns of the 79,739,116 returns; and few short-form returns had a continuation roster, 533,299 returns of the 67,827,664 short-form returns, or 0.79 percent. For the other self-response returns (79,203,017 returns), the following occurred:

- RNPOP = TOTPOP for 74,725,437 returns, or 94.35 percent of the 79,203,017 returns (63,970,275 short-form returns, 95.06 percent of the 67,293,626 short-form returns; 10,755,162 long-form returns, 90.31 percent of the 11,909,391 long-form returns).
- RNPOP $\neq$ TOTPOP and RNPOP $\leq$ 6 for 4,429,189 returns, or 5.59 percent of the 79,203,017 returns (3,290,936 short-form returns, 4.89 percent of the 67,293,626 short-form returns; 1,138,253 long-form returns, 9.56 percent of the 11,909,391 long-form returns).

- For RNPOP ≠ TOTPOP and RNPOP ≤ 6, the expected household size was set to maximum (TOTPOP, RNPOP), which was TOTPOP for 4,182,140 returns, or 94.42 percent of these returns (3,067,426 short-form returns, or 93.21 percent of the 3,290,936 short-form returns; and 1,114,714 long-form returns, or 97.93 percent of the 1,138,253 long-form returns).

Table 12. Setting expected household size: Occupied paper mail returns

Population	Expected household size	Number	Percent
Total Occupied paper mail returns		79,739,116	100.00
Linked return and short-form return with a continuation roster			
Linked return	Maximum (TOTPOP, ALLPOP, Minimum (RNPOP, Maximum (TOTPOP+6, ALLPOP+6)))	2,800	0.00
Short-form return with a continuation roster	Maximum (TOTPOP, Minimum (RNPOP, TOTPOP+6))	533,299	0.67
Neither linked return nor short-form return with a continuation roster			
RNPOP = TOTPOP	TOTPOP	74,725,437	93.71
RNPOP > 6, RNPOP < TOTPOP	TOTPOP	47,735	0.06
RNPOP > 6, RNPOP > TOTPOP	TOTPOP	656	0.00
RNPOP ≤ 6, RNPOP ≠ TOTPOP	Maximum (TOTPOP, RNPOP)	4,429,189	5.55

Data source: Decennial Response File

4.3.2 The setting of expected household size: Occupied enumerator returns

Here are definitions of variables used to set expected household size for Occupied enumerator returns:

- RNPOP is the respondent-reported household size, Question S5 on the enumerator questionnaire: "How many people were living or staying in this (house/apartment/mobile home) on April 1, 2000?"

- DPPOP is the number of valid person records for the census questionnaire.
- ALLPOP is the DPPOP for the census return (the sum of the parent form DPPOP and all child forms linked to the parent form).

- RISPOP is the Interview Summary Population.

The occupied enumerator returns described in Table 13 exclude the 33,472 orphan returns. The following outcomes determined the expected household size for enumerator returns:

- Whether or not the return included linked forms,
- Whether or not RNPOP > 5,
- Whether or not RISPOP > 5,
- Whether or not RNPOP = DPPOP, and
- Whether or not RISPOP = DPPOP.

In Table 13, 1,498,855 returns, or 4.93 percent of all enumerator returns were linked or had the "continuation form(s) attached" box checked. For most of these returns, the expected household size was RISPOP (1,488,608 returns, or 99.32 percent). And of these, RISPOP = RNPOP for 1,475,382 returns or 99.11 percent.

For the 28,794,014 unlinked returns with RISPOP $\leq$ 5, the expected household size was the larger of either the enumerator-reported household size (RISPOP) or the number of valid person records (DPPOP) with the following results:

- Most had RISPOP = DPPOP: 27,026,490 returns or 93.86 percent.
    - Of these, 26,897,133 returns or 99.52 percent also had the same respondent-reported household size (RNPOP), RISPOP = DPPOP = RNPOP.
- For RISPOP > DPPOP, 1,491,562 returns, or 96.49 percent had the same enumerator-reported household size and respondent-reported household size, RISPOP = RNPOP.
- For DPPOP > RISPOP, 59,796 returns, or 26.97 percent had the same number of valid person records as the respondent-reported household size, DPPOP = RNPOP.

There were 68,886 returns, or 0.23 percent, that were unlinked and had RISPOP > 5 and RNPOP > 5. Most, 68,599 returns, or 99.58 percent, had the same enumerator-reported household size and respondent-reported household size, RISPOP = RNPOP. Of these, RISPOP was set as the expected household size for 47,367 returns, or 69.05 percent. Expected household size for RISPOP > 5 and RNPOP > 5 was set to the minimum (RISPOP, DPPOP+6).

Table 13. Setting expected household size: Occupied enumerator returns, excluding orphans

Population	Expected household size	Number	Percent
Total Occupied enumerator returns, excluding orphans		30,431,665	100.00
Linked return or return with "continuation form(s) attached" box checked			
	Maximum ((ALLPOP, Minimum (RISPOP, ALLPOP+6))	1,498,855	4.93
Neither linked return nor return with "continuation form(s) attached box" checked			
RISPOP ≤ 5, RISPOP > DPPOP	RISPOP	1,545,811	5.08
RISPOP ≤ 5, RISPOP = DPPOP	DPPOP	27,026,490	88.81
RISPOP ≤ 5, DPPOP > RISPOP	DPPOP	221,713	0.73
RISPOP > 5, RNPOP > 5	Minimum (RISPOP, DPPOP+6)	68,886	0.23
RISPOP > 5, RNPOP ≤ 5, RNPOP = DPPOP	DPPOP	1,096	0.00
RISPOP > 5, RNPOP ≤ 5, RNPOP ≠ DPPOP	Maximum (RNPOP, DPPOP)	68,814	0.23

Data source: Decennial Response File

5.CONCLUSIONS

Of 129,389,529 returns, 1,387,085 returns, or 1.07 percent, were linked; that is, they were returns comprised of two or more forms. Of these, 39,108 returns, or 2.82 percent, had three or more forms.

The type of enumeration area with the highest rate of linked returns was in Update/Enumerate: 41,559 of 1,052,591 returns, or 3.95 percent. Large households probably caused this result. Most linked returns, 1,384,233 returns, or 99.79 percent, were comprised of an enumerator first and an enumerator continuation form. Enumerators used this combination of forms to enumerate large households in the List/Enumerate, Update/Enumerate, Nonresponse Followup, and Coverage Improvement Followup operations. Linkage rates comparable to Update/Enumerate did not occur in List/Enumerate probably because of a processing error. Enumerator continuation forms in List/Enumerate—along with Update/Leave adds and Update/Enumerate adds—were erroneously omitted from the Decennial Response File.

Of 129,389,529 returns, 1,318,350 returns, or 1.02 percent, had either an unresolved expected household size or an unresolved housing unit status. The three unresolved categories were the following: Occupied with Unresolved Population Count, Unresolved Occupied/Vacant, and Unresolved Occupied/Vacant/Delete.

We sent mail returns to Coverage Edit Followup if there was an inconsistency in household size, leaving few (538) of the 81,099,704 self-response returns unresolved.

The overlap between unresolved housing units in Nonresponse Followup and Coverage Improvement Followup was 4,223 housing units. So there were not many housing units classified as unresolved as a result of both Nonresponse Followup and Coverage Improvement Followup.

A programming error affected the status resolution for some Vacant enumerator returns. Mistakenly, we recoded any Interview Summary Population of 0 to blank. As a result, we may have classified up to 133,438 Vacant returns as Deletes and up to 258,963 Vacant returns as Unresolved Occupied/Vacant. As many as 145,367 housing units of the 191,826 housing units in the census, or 75.78 percent, which had its occupancy status imputed may have been affected by this latter error.

We had 712,858 unresolved enumerator returns, or 1.51 percent of the 47,080,158 enumerator returns (329,895 returns were Occupied with Unresolved Population Count; 329,266 returns were Unresolved Occupied/Vacant; and 53,697 returns were Unresolved Occupied/Vacant/Delete).

The unresolved enumerator returns were a result of contradictory and missing responses on the questionnaire. Also not capturing data or misinterpreting scanned images could have contributed to unresolved enumerator returns. A redesigned Interview Summary Section would probably improve the consistency of responses. Mobile Computing Devices would help ensure data capture and consistency of responses.

Of the 1,209,667 other returns, 604,954 were Usual Home Elsewhere returns for individuals. All of these returns were classified as Occupied with Unresolved Population Count—64.71 percent of the 934,849 overall returns classified as Occupied with Unresolved Population Count.

For Occupied self-response (restricted to paper mailback questionnaires) and enumerator returns, setting the expected household size was usually straightforward. For 74,725,437 self-response returns, or 93.71 percent, the number of valid person records and roster names corresponded to the respondent-reported household size. For enumerator returns, most household size measures were consistent. This can be seen in the following results for enumerator returns:

- For linked returns or unlinked returns that had the “continuation form(s) attached” box checked and the expected household size equal to the Interview Summary Population, 1,475,382 returns, or 99.11 percent, had the same Interview Summary Population and respondent-reported household size.

- For unlinked returns with the Interview Summary Population less than or equal to five 26,897,133 returns, or 99.52 percent, had the same Interview Summary Population, number of valid person records, and respondent-reported household size.

- For unlinked returns with the Interview Summary Population greater than five and the respondent-reported household size greater than five, 68,599 returns, or 99.58 percent, had the same Interview Summary Population and respondent-reported household size.

References

U.S. Census Bureau, 1990, "Description of the Decennial Census Algorithm for the Selection of the Primary and Supplemental Records from the 1990 FOSDIC Data Capture Files," Decennial Operations Division, November 27, 1990.

U.S. Census Bureau, 2000, "Specifications for Identifying Census Continuation Forms and Setting the Expected Return Population Count on the Decennial Response File," Census 2000 Procedures and Operations Memorandum #C-5, October 25, 2000.

U.S. Census Bureau, 2001, "ESCAP II: Analysis of Census Imputations," Executive Steering Committee for A.C.E. Policy II (ESCAP II) Report #21, September 24, 2001.

Appendix A: Self-response, enumerator, and other returns

Self-response returns
x
Enumerator returns
x
Other returns
x

RSOURCE SOURCE OF RETURN (RECODE) (From Decennial Response File Processing)

blank= Not computed
x01= Paper mailback questionnaire from mailout
02= (not used)
x03= Paper mailback questionnaire from TQA mailout with NO ID
x04= Paper mailback questionnaire from Update/Leave
x05= Paper mailback questionnaire from Update/Leave ADD
x06= Paper mailback questionnaire from Update/Leave SUBSTITUTE
x07= Paper mailback questionnaire from Urban Update/Leave
x08= Paper mailback questionnaire from Urban Update/Leave ADD
x09= Paper mailback questionnaire from Urban Update/Leave SUBSTITUTE
x10= Paper mailback questionnaire from Request for Foreign Language
x11= Paper mailback questionnaire from BCF marked as whole household
x12= Paper mailback questionnaire from BCF partial household (i.e., NOT marked as whole household)
x13= Paper enumerator questionnaire from List/Enumerate
x14 = Paper enumerator questionnaire from Update/Enumerate
x15= Paper enumerator questionnaire from Update/Enumerate ADD
x16= Paper enumerator questionnaire from Update/Enumerate SUBSTITUTE
x17= Paper enumerator questionnaire from Nonresponse Follow-up (NRFU)
x18= Paper enumerator questionnaire from NRFU ADD
x19= Paper enumerator questionnaire from NRFU SUBSTITUTE
x20= Paper enumerator questionnaire from NRFU Whole Household Usual Home Elsewhere
x21= Paper enumerator questionnaire from NRFU In-mover
x22= Paper enumerator questionnaire from Coverage Improvement Follow-up (CIFU)
x23= Paper enumerator questionnaire from CIFU ADD
x24= Paper enumerator questionnaire from CIFU SUBSTITUTE
x25= Paper enumerator questionnaire from T-Night
x26= Paper questionnaire for Usual Home Elsewhere (UHE) from Service-based Enumeration (Individual Census Questionnaire)
x27= Paper questionnaire for UHE from Group Quarters (GQ) enumeration (Individual Census Report)
x28= Paper questionnaire for UHE from Military GQ enumeration (Military Census Report)
x29= Paper questionnaire for UHE from Shipboard GQ enumeration (Shipboard Census Report)
x30= Electronic short form from Internet Data Collection
x31= Electronic TQA reverse-CATI short form
x32= Electronic TQA reverse-CATI BCF for whole household
x33= Electronic TQA reverse-CATI BCF for partial household
x34= Electronic Coverage Edit Follow-up (CEFU) from long or short form
x35= Electronic CEFU from BCF for whole household
x36= Electronic CEFU from Internet Data Collection
x37= Paper enumerator continuation form – “orphan”

Appendix B: Definition of a data-defined person

A person is data defined if at least two of the data items are complete as outlined below. The completion criteria depend on variable outcomes from the Decennial Response File.

Data item	Completion criteria
Name	The combination of first and last name (PFNAME and PLNAME) contain 3 or more characters; and neither the first nor last name was blanked in key from image (neither PDKFIBF nor PDKFIBL = 1).
Relationship	The relationship is husband/wife; natural-born son/daughter; adopted son/daughter, stepson/stepdaughter; brother/sister; father/mother; grandchild; parent-in-law; son-in-law/daughter-in-law; other relative; roomer, boarder; unmarried partner; foster child; or other nonrelative (PREL = 2-16) or the relationship write-in variable contains 1 or more characters (PRELWI).
Sex	Sex is male or female (PSEX = 1-2)
Age/date of birth	The age, year of birth, month of birth, and day of birth meet the following conditions: PAGE = 0-999; or PYOB = 0-9999; or PMOB = 1-99 and PDOB = 1-99.
Hispanic origin	For any of the Hispanic origin variables, a box is marked (PSPAN*n* = 1 for any n, n = 1-5) or the Hispanic write-in variable contains 1 or more characters (PSPANWI).
Race	For any of the race variables, a box is marked (PRACE*n* = 1 for any n, n = 1-15) or any of the race write-in variables contain 1 or more characters (PRACEWI1, PRACEWI2, PRACEWI3, PRACEWIGEN)

Appendix C: Classifying type of enumeration areas as Mail, Update/Leave, Update/Enumerate, or List/Enumerate

Mail Update/LeaveUpdate/EnumerateList/Enumerate

Type of enumeration area (TEA)

x1 = Mailout/Mailback
x6 = Military in Update/Leave area

x2 = Update/Leave
x7 = Urban Update/Leave
x9 = Update/Leave (originally TEA 1)

x5 = "Rural" Update/Enumerate
x8 = "Urban" Update/Enumerate

x3 = List/Enumerate
x4 = Remote List/Enumerate (Alaska)

Appendix D: Interview Summary section of the enumerator questionnaire