

Census Unedited File Creation

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.

Kim Jonas

Decennial Statistical
Studies Division

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

Intentionally Blank

CONTENTS

LIST OF TABLES	ii
EXECUTIVE SUMMARY	iii
1. BACKGROUND	1
1.1 The purpose of Census Unedited File creation	1
1.2 Creation of the HCUF in past Censuses and tests	1
1.3 HCUF creation in Census 2000	2
2. METHODS	4
2.1 Files used in this evaluation	4
2.2 Geography included in this evaluation	4
3. LIMITATIONS	5
4. RESULTS	6
4.1 How many DMAF addresses were resolved as housing units in Census 2000? How many DMAF addresses were not included in Census 2000?	6
4.2 What was the source of the housing unit status of the addresses on the DMAF in Census 2000?	7
4.3 What was the source of the housing unit status of the addresses included on the HCUF in Census 2000?	9
4.4 What was the source of the housing unit status of the addresses deleted as duplicates in Census 2000?	11
4.5 What was the breakdown of addresses resolved as deletes in Census 2000? ...	13
5. RECOMMENDATIONS	14
5.1 Enhance the Census software quality assurance process to ensure adherence to specifications	14
5.2 Reexamine timing of late Census operations	14
5.3 Conduct further research on Duplicate Delete cases	14
References	16
Appendix 1: DMAF Addresses Resolved As Housing Units in the Census	17
Appendix 2: DMAF Addresses Not Included in the Census	19

LIST OF TABLES

Table 1: DMAF Housing Unit Addresses and the Census	7
Table 2: Source of Housing Unit Status for DMAF Addresses	8
Table 2A: DMAF Deliveries of Addresses with Unknown Status	9
Table 3: Housing Unit Status for Addresses in the Final Census	10
Table 4: Housing Unit Status for Addresses Deleted as Duplicates	11
Table 4A: Comparison Between HCUF Addresses and Addresses Deleted as Duplicates . .	12
Table 5: Housing Unit Addresses Resolved as Deletes in the Final Census	13

EXECUTIVE SUMMARY

The Census 2000 Hundred Percent Census Unedited File contains all the household and person records included in Census 2000. It has all the attributes of the final Census file, excepting imputation of person characteristics where needed. The purpose of Census Unedited File creation is to determine which addresses are in the Census, and to determine the count of persons at each such address.

Nearly 128 million addresses were either on the Decennial Master Address File as Census 2000 began, or were added to it in the course of Census 2000 operations. Approximately 117.3 million were ultimately resolved as housing unit addresses. Just over nine million addresses were determined to not be valid addresses, and roughly 1.4 million addresses were determined to be nonvalid duplicates of valid addresses on the Decennial Master Address File.

Of the 117.3 million addresses resolved as housing unit addresses, 106.7 million were determined or imputed to be occupied, and the remaining 10.6 million were determined or imputed to be vacant.

Roughly half a million addresses had their status resolved by imputation. There were 195,245 addresses determined to be valid Census addresses whose occupancy status could not be determined, and had to be imputed as a result. There were 296,617 addresses whose validity as Census addresses could not be determined. As a result, their validity and their occupancy status were both imputed. There was no enumeration data on the Decennial Response File or the Decennial Master Address File for 251,477, or 84.8 percent, of the addresses whose validity as Census addresses could not be determined.

We recommend some changes affecting Census Unedited File creation in the next Census, including to:

- Use stronger software quality assurance processes to ensure more complete adherence to specifications; and
- Refine the timing of late Census followup operations to ensure that addresses added by those operations are placed on the Decennial Master Address File in time for the questionnaires from those addresses to be included in the Census.

1. BACKGROUND

1.1 The purpose of Census Unedited File creation

The Census 2000 Hundred Percent Census Unedited File (HCUF) contains all the household and person records included in Census 2000. It has all the attributes of the final Census file, excepting imputation of person characteristics where needed.

The purpose of HCUF creation is to determine which addresses are Census housing units, and to determine the count of persons at each such address. This evaluation is solely concerned with housing unit addresses and the persons counted there.

1.2 Creation of the HCUF in past Censuses and tests

1.2.1 The 1990 Census

The 1990 Census equivalent of the Census 2000 HCUF was created from the final 1990 Census Data Capture File, the 1990 Address Control File, and the 1990 Capture Control File. The resulting file reflected the results of the census response records selected by the Primary Selection Algorithm (PSA) applied to the Data Capture File, and the final version of the Address Control File after all maintenance operations on the file ceased.

The Data Capture File contained all questionnaire response data for each housing unit. The Address Control File data included the status of each census address based on address maintenance operations and data from the field check-in operations which were recorded on the Collection Control File. These data included housing unit status data as recorded on enumerator questions in the field by census enumerators. The Data Capture File data for the housing unit records selected by the PSA were matched to the Address Control File address records to define the final census housing unit universe. The housing unit status and/or a household's size were imputed for address records that had an unknown housing unit status and for occupied housing units that had an unknown population count. A hot deck imputation method was used to impute housing unit status and population count. The resultant file was the basis for the 1990 Census final housing unit and population counts.

1.2.2 The Census 2000 Dress Rehearsal

The HCUF for the Census 2000 Dress Rehearsal was created from the Data Response File (DRF) and the Decennial Master Address File (DMAF). The DRF included the response data for Census 2000 questionnaires, and reflected the results of the PSA. The DMAF contained the housing unit addresses included in the last pre-Nonresponse Followup DMAF extract from the Master Address File (MAF) provided by Geography Division.

The HCUF record for a DMAF address was created from its DMAF record and the DRF return associated with that address, if one existed. Housing units whose occupancy status or population count was not defined from either a selected DRF return or from Nonresponse Followup check-in information prior to creation of the HCUF were assigned an occupancy status and/or a population count by hot decking when the HCUF was created.

1.3 HCUF creation in Census 2000

The Census 2000 HCUF was created by merging data from two sources: the PSA response data on the Census 2000 DRF, and the DMAF.

The HCUF was constructed in three stages. In the first stage, the data from these two files were combined to determine which housing units potentially existed in the census. Each unique DMAF address was determined either to be a potential census housing unit or to not exist as a housing unit in the Census.

Addresses determined not to be a housing unit fell into two groups: Kills and Resolved Deletes. *Kills* were identified primarily on the basis of address list development data. *Resolved Deletes* were identified primarily on the basis of housing unit response data.

A DMAF address became a Kill if the Census could not find any recent evidence of its existence. The primary means by which a DMAF address would be classified as a Kill were if no mail return was received from that address, and:

- it was a “Double Delete”, that is, it had been classified as a delete by both the Block Canvassing and Local Update of Census Addresses Field Verification operations, which were pre-Census 2000 address list building operations;
- it was an “Old Delivery Sequence File Address”, that is, it had been placed on the DMAF by virtue of being a residential address on one of the U.S. Postal Service’s (USPS) Delivery Sequence Files from 1997 or 1998, but was no longer a residential address on any of the USPS’ Delivery Sequence Files in 1999 and 2000; or
- if the address was identified as a delete by a Census 2000 enumerator in a Census 2000 operation such as Update/Leave, Nonresponse Followup, Urban Update/Leave, Update/Enumerate, or Coverage Improvement Followup, and no evidence was received from any Census 2000 operation indicating that the address was an existing residential address.

A complete listing of all means of classifying a DMAF address as a Kill is in the Kill Specification.¹

¹Treat, James B., “Specification of the Kill Universe on the Decennial Master Address File for Census 2000”, DSSD Census 2000 Procedures and Operations Memorandum Series #D-13, December 21, 2000. (Referred to throughout as the *Kill Specification*.)

Also in the first stage of HCUF creation, response data from both the DMAF and the DRF were used to assign status and population count to the remaining potential housing units. The possible statuses were: (1) Occupied, (2) Vacant, (3) Resolved as Occupied (Unknown Pop), (4) Occupancy Status Unknown, and (5) Status Unknown. *Resolved as Occupied (Unknown Pop)* meant that the housing unit was occupied but the population count was unknown. *Occupancy Status Unknown* was the status assigned when the housing unit existed but could have been either occupied or vacant. *Status Unknown* was the status assigned when the address might have been an occupied housing unit, a vacant housing unit, or not a Census housing unit at all.²

In the second stage of HCUF creation, the housing unit status and/or population count were imputed to those potential housing units with a housing unit status of Unknown Population Count, Occupancy Status Unknown and Unknown Status. During this stage, addresses given an imputed status of delete were eliminated from the HCUF. These are referred to as Imputed Deletes.³

The third and final stage of HCUF creation was to unduplicate the remaining housing units using address information and response data. The duplicate housing unit records were flagged on the HCUF, but not actually deleted from Census processing until creation of the Hundred Percent Census Edited File (HCEF).

Excepting the housing unit records flagged for deletion as duplicates, the HCUF contained data only for addresses in the final Census 2000 housing unit inventory at the end of these three stages. Every housing unit had a population count and an occupancy status of ‘occupied’ or ‘vacant’ at the completion of the HCUF processing.

² See Alberti, Nick, “Specifications for Assigning the Housing Unit Status and Population Count of the Hundred-Percent Unedited File Prior to the Imputation of Unclassified Units,” DSSD Census 2000 Procedures and Operations Memorandum Series #D-14, January 19, 2001.

³ For more information on Census 2000 count imputation procedures, see Griffin, Richard, “Census 2000: Overview of Count Imputation – Reissue of Q-2,” DSSD Census 2000 Procedures and Operations Memorandum Series #Q-78, March 18, 2002.

2. METHODS

2.1 Files used in this evaluation

Two files were used in this evaluation: the final HCUF file, and the Decennial Statistical Studies Division (DSSD) Combo file, where key variables from the MAF, DMAF, HCUF, HCEF, and some DSSD control files were brought together for each address record. The principal DSSD control files were:

- The Kill file, which was created from the current DMAF by applying the Kill identification quality control testing software; and
- The Count Imputation Output File, which was a copy of the preliminary HCUF, updated with the results of status imputation for those addresses requiring occupancy or status imputation, as described in detail in the Unclassified Estimation specification.⁴

The DSSD Combo file also includes some DSSD-generated variables. This evaluation used one such variable called IDSTAT, which identified DMAF housing unit addresses as occupied, vacant, a Kill (see Sec. 1.3), or one of several different categories of deletes.

For the most part, IDSTAT was set directly from HCUF and DMAF variables. But there were exceptions. For instance, Kills (see Sec. 1.3) were not identified by a single identifying variable on the HCUF or DMAF, but rather by linking control files for imputation and output from the Kills identification quality control testing programs applied to the current DMAF. (A more detailed discussion of the identification of Kills is in the Limitations section of this evaluation.)

2.2 Geography included in this evaluation

All statistics in this evaluation include Puerto Rico, in addition to the 50 States and the District of Columbia. Counts for “Deleted in Status Imputation” and “Deleted as Duplicates” have been provided in other Census publications, but the counts differ because the counts in those publications excluded Puerto Rico. The same is true of the counts for “Occupancy Status Unknown” and “Status Unknown” in Table 3.

⁴ Griffin, Richard, “Census 2000 Count Imputation - Results,” DSSD Census 2000 Procedures and Operations Memorandum Series #Q-81, November 5, 2002.

3. LIMITATIONS

Limitations surrounding Kills:

Contrary to the Kill Specification the DMAF did not distinguish Kills from other deleted addresses. In order to provide information on deleted addresses for this study, the DSSD divided the deleted addresses into the three categories of Resolved Deletes, Imputed Deletes, and Kills by linking control files for imputation and output from the Kills identification quality control testing programs applied to the current DMAF. Any deleted address on the current DMAF not independently identified as an Imputed Delete or a Kill from these should be a Resolved Delete.

The quality control testing program for the identification of Kills represents an independent identification of Kills. Output from these programs based on the DMAF as of December 2000 was verified at that time to be consistent with the Kills identified by the census production processing. In 2002 the DSSD applied the same quality control testing programs to the current DMAF in order to identify Kills on the DMAF for this and other studies. We discovered 13,783 addresses that were not identified as Kills when the quality control testing program for the identification of Kills was applied to the DMAF in 2000. These addresses must surely be Kills that were not identified as such by the testing program, because they were included in a mailout or update/leave enumeration area, but had no form data captured and were not included in any followup operation. These discrepancies are unexplainable if one assumes that the 2000 and 2002 versions of the DMAF are identical.

We believe that the only explanation for the status of these addresses is that they were among those addresses identified as Kills in the census production processes. Accordingly, they have been classified as Kills for purposes of this evaluation.

4. RESULTS

4.1 How many DMAF addresses were resolved as housing units in Census 2000? How many DMAF addresses were not included in Census 2000?

There were nearly 128 million housing unit addresses on the DMAF. Table 1 provides a breakdown of the addresses included and not included in the final Census.

Kills are primarily (a) addresses that had been identified by two pre-Census 2000 operations as not being valid residential housing units; (b) addresses that had been on the USPS' Delivery Sequence Files prior to 1999 but not in 1999 or 2000; or (c) identified as deletes by one or more Census 2000 operations, with no countervailing evidence of their legitimacy as a residential address from any Census 2000 operation. Kills are discussed in more detail in Section 1.3 and in the Kill Specification.

Addresses "Resolved as Deletes" are addresses identified as not being Census housing units on the basis of housing unit response data. The addresses "Resolved as Deletes" include 5,469 addresses that appear to have been deleted in error. These were addresses that were not on the DRF (i.e. no questionnaire record was placed on the DRF for these addresses), but were identified as a valid residential address by one of three Census 2000 followup operations: Nonresponse Followup, Coverage Improvement Followup, or Field Verification. For such an address to have properly been deleted, it should have been identified as not being a valid residential address by one of the three aforementioned Census 2000 followup operations. A further breakdown of the resolved deletes is provided in Section 4.5.

The Census designed and implemented a duplicate operation in the summer and fall of 2000, using a combination of address matching and person matching to correct a potential overcount of housing units. Some DMAF addresses were identified as matching other DMAF addresses, and were accordingly "Deleted as Duplicates" from the Census.

There were addresses on the DMAF for which the response data were not sufficient to determine if a valid Census housing unit existed there at the time of enumeration. Status of Occupied, Vacant, or Delete was imputed for each of these addresses before the HCUF was finalized. Addresses with Delete status imputed are included in the "Deleted in Status Imputation" category.

Table 1: DMAF Housing Unit Addresses and the Census

Resolution	Number	Percent
DMAF Addresses Resolved as Housing Units in the Census	117,323,117	91.78
DMAF Addresses Not Included in the Census:	10,505,661	8.22
Kills ⁵	9,057,195	7.09
Resolved as Deletes	8,654	0.01
Deleted as Duplicates	1,392,686	1.09
Deleted in Status Imputation	47,126	0.04
Total	127,828,778	100.00

Totals by state are provided in the Appendix.

Sources: HCUF, DMAF

- Approximately 91.8 percent of DMAF housing unit addresses were included in the Census.
- Of the 8.2 percent that weren't, the vast majority (7.1 percent) were kills, and 1.1 percent were deleted as duplicates.
- Resolved Deletes constituted a negligibly small percentage of DMAF addresses. The same was true of addresses deleted in status imputation.

A philosophy of erring on the side of inclusion was present in Census 2000. The counts of addresses killed and deleted as duplicates may have been a consequence of this approach.

4.2 What was the source of the housing unit status of the addresses on the DMAF in Census 2000?

Table 2 shows the source of information for all housing unit addresses on the DMAF, except for the kills.

Data sources: the data source for an address could be either the return type of the response record chosen by the PSA⁶ to represent the address, or the occupancy status information recorded on the DMAF from the Nonresponse Followup, Coverage Improvement Followup, or Field Verification field operations.

If the information on the DMAF from the aforementioned field operations was the data source, the data source is listed here as "Enumerator Response."

⁵ Includes 13,783 addresses deleted for unknown reasons but believed to be Kills. This is discussed in more detail in Section 3, Limitations.

⁶ For more information regarding the PSA, see Baumgardner, Stephanie, "Analysis of the Primary Selection Algorithm," Census 2000 Evaluation L.3.a, November 12, 2002.

Otherwise, if the PSA-selected return was a mail return, then the data source is listed here as “Self Response;” if the PSA-selected return was an enumerator return, then the data source is listed here as “Enumerator Response;” or if the PSA-selected return was a Be Counted Form or any type of Group Quarters questionnaire, then the data source is listed here as “Respondent Provided Address.”

If no questionnaire was placed on the DRF for an address, and the address was not assigned to any regular or followup Census 2000 operation involving an enumerator visit, then the source is given as “No Data.” Addresses with this status are primarily addresses which were added to the DMAF too late in Census 2000 for any return to be included in the DRF.

Housing unit status: addresses in the Resolved Occupancy Status categories had response data that sufficed to determine the housing unit status (Occupied, Vacant, or Delete). Those “Resolved as Valid Housing Units” had response data that sufficed to determine that the unit was occupied or vacant, and sufficed to determine the household size if occupied. Addresses “Resolved as Deletes” had response data that sufficed to determine that the address was not a Census housing unit. Addresses “Resolved as Occupied (Unknown Pop)” had response data that sufficed to determine that the unit was occupied, but were not sufficient to determine the household size.

Addresses with “Occupancy Status Unknown” had response data that sufficed to determine that the housing unit existed, but were not sufficient to determine whether it was occupied or vacant. Addresses with “Status Unknown” had response data that were not sufficient to determine whether there was a Census housing unit at the address. These addresses were eligible to be given an imputed status of Occupied, Vacant, or Delete.

Table 2: Source of Housing Unit Status for DMAF Addresses

	Data Sources			No Data	Totals	%
	Self Response	Enumerator Response	Respondent Provided Address			
Housing Unit Status:						
<i>Resolved Occupancy Status:</i>						
Resolved as Deletes	0	8,653	0	1	8,654	0.01
Resolved as Occupied (Unknown Pop) (Household Size Imputation)	0	169,902	30,232	0	200,134	0.17
Resolved as Valid Housing Unit	80,797,403	37,075,752	197,778	0	118,070,933	99.41
<i>Unresolved Occupancy Status:</i>						
Occupancy Status Unknown (Occupancy Imputation)	506	194,739	0	0	195,245	0.16
Status Unknown (Status Imputation)	0	45,113	27	251,477	296,617	0.25
Total	80,797,909	37,494,159	228,037	251,478	118,771,583	100.0
Percentages:	68.02	31.57	0.19	0.21	100.0	

Sources: HCUF, DMAF. Excludes Kills from Table 1.

- There were 491,862 addresses on the DMAF that required either occupancy or status imputation.
- Of these 491,862 addresses, 39.7 percent, or 195,245 addresses, were identified as having Occupancy Status Unknown.
 - the data source for 99.7 percent of these 195,245 addresses was enumerator responses;
 - the remaining 0.3 percent were from self responses.
- Of the 491,862 addresses, 60.3 percent, or 296,617 addresses, were identified as Status Unknown.
 - There were no data for 85 percent of these 296,617 addresses.
 - The data source for the remaining 15 percent of these was almost entirely enumerator responses.
- Units with Unresolved Occupancy Status made up only 0.4 percent of all DMAF addresses. But they were 0.6 percent of the addresses whose source was an enumerator response.

The addresses requiring status imputation (the Status Unknown addresses in Table 2) were delivered to the DMAF very late in Census 2000 processing:

Table 2A: DMAF Deliveries of Addresses with Unknown Status

	Delivery				Totals	%
	June	July	August	After August		
Deleted as Duplicates	1,559	0	149	5,322	7,030	2.37
Imputed as Deletes	11,283	51	3,562	32,230	47,126	15.89
Imputed as Housing Units	76,415	24	36,743	129,279	242,461	81.74
Total	89,257	75	40,454	166,831	296,617	100.0
Percentages:	30.09	0.03	13.64	56.24	100.0	

Sources: HCUF, DMAF.

- Nearly 70 percent of addresses with unknown status were delivered to the DMAF in the August delivery or later.
- Per Table 2, there was “No Data” for 85 percent of the addresses in Table 2A. Though not reflected in any table in this evaluation, the 85 percent “No Data” figure applies to the addresses in the August and later deliveries.

4.3 What was the source of the housing unit status of the addresses included on the HCUF in Census 2000?

Table 3 shows the final occupancy status of housing units included in the Census. The housing units are grouped by whether or not they had a resolved occupancy status, and by type of unresolved status.

Table 3: Housing Unit Status for Housing Units in the Final Census

Response	Occupied	Vacant	Totals	Vacant %
Housing Unit Status:				
Resolved Occupancy Status:	106,469,702	10,417,315	116,887,017	8.91
Self Response	80,187,952	16,228	80,204,180	0.02
Enumerator Response	26,071,164	10,401,087	36,472,251	28.52
Other	210,856	0	210,856	0.00
Unresolved Occupancy Status:	271,724	164,376	436,100	37.89
Occupancy Status Unknown (Occupancy Imputation)	107,887	85,752	193,639	44.28
Self Response	289	214	503	42.54
Enumerator Response	107,598	85,538	193,136	44.29
Other	0	0	0	
Status Unknown (Status Imputation)	163,837	78,624	242,461	32.43
Self Response	0	0	0	
Enumerator Response	24,862	17,818	42,680	41.75
Other	138,975	60,806	199,781	30.44
Total	106,741,426	10,581,691	117,323,117	9.02

Sources: HCUF, DMAF. Includes addresses resolved as housing units from table 1.

Only 8.9 percent of housing units with resolved occupancy status were found to be vacant. However, 37.9 percent of units that underwent either occupancy or status imputation were imputed as vacant. This was the result of the design of the count imputation process. The addresses having occupancy or status imputed were assigned the status and population count of a donor unit. The donor units were nearby units that were occupied (with a population count) or vacant units either included in Nonresponse Followup or Coverage Improvement Followup operations, or from field enumeration areas. For addresses having status imputed, nearby deleted and killed addresses were also in the donor pool. Donors were assigned through the nearest-neighbor hot deck method. Of units with occupancy status unknown, 44.3 percent had vacancy status imputed; of units with unknown status, 32.4 percent had vacancy status imputed. This was higher than the vacancy rates of 28.52 percent and 24.79 percent in the respective donor pools. According to Griffin⁷, this is because “[u]nclassified units – units for which we are unaware of residency status – are more likely to be neighbors to vacant and delete/kill housing units.”

⁷ Griffin, Richard, “Census 2000 Count Imputation - Results,” DSSD Census 2000 Procedures and Operations Memorandum Series #Q-81, November 5, 2002, p.13.

4.4 What was the source of the housing unit status of the housing unit records deleted as duplicates in Census 2000?

4.4.1 Status of housing unit records deleted as duplicates

Table 4 shows the housing unit status for all housing unit records deleted from Census 2000 as duplicates.

Table 4: Housing Unit Status for Housing Unit Records Deleted as Duplicates

Response	Occupied	Vacant	Totals	%
Housing Unit Status:				
Resolved Occupancy Status:	1,346,215	37,835	1,384,050	99.38
<i>Unresolved Occupancy Status:</i>				
Occupancy Status Unknown (Occupancy Imputation)	971	635	1,606	0.12
Status Unknown (Status Imputation)	5,007	2,023	7,030	0.50
Total	1,352,193	40,493	1,392,686	100.00
Percentages	97.09	2.91	100.00	

Sources: HCUF, DMAF. Includes addresses deleted as duplicates from Table 1.

4.4.2 Comparison with housing units included in the HCUF

The following table compares the percentages of vacant housing units across three categories: housing units *On the HCUF*, which are HCUF housing units that were neither permanently deleted as duplicates, nor provisionally deleted as duplicates, then reinstated; housing units *Reinstated to (the) HCUF*, which were provisionally deleted from the HCUF as duplicates, then reinstated; and housing unit records permanently *Deleted as Duplicates* from the HCUF.⁸

⁸Fay, Robert, "The 2000 Housing Unit Duplication Operations and Their Effect on the Accuracy of the Population Count," Proceedings of the Annual Meeting of the American Statistical Association, August 5-9, 2001, pp.3-4, describes a two-phase operation to identify duplicate addresses: a first phase which cast a broad net for potential duplicates which were "provisionally deleted," followed by a closer review that resulted in the reinstatement of many of the provisional deletes from the first phase.

Table 4A: Comparison Between HCUF HUs and HU Records Deleted as Duplicates

Response	Percentage of Vacant Units		
	On the HCUF ⁹	Reinstated to	Deleted as Duplicates
Totals:	9.01 %	10.46 %	2.91 %
Resolved Occupancy Status:	8.90 %	10.07 %	2.73 %
Occupancy Status Unknown:	44.26 %	45.07 %	39.54 %
Status Unknown:	32.24 %	38.48 %	28.78 %

Vacancy percentages are for the categories indicated. For example, in the third column of the second row, 2.73% of addresses with resolved occupancy status that were deleted as duplicates had been listed as vacant.

Sources: HCUF, DMAF

- Housing unit records deleted as duplicates had only one-third the vacancy rate of housing units on the HCUF.
 - 62.6 percent of housing unit pairs that were included in the Census 2000 duplicate operation were identified via person matching, which excluded vacant units.
 - This appears to explain most of the reduction in vacancy rate for housing unit records deleted as duplicates.
- Housing unit records initially included in the Census 2000 duplicate operation, but ultimately reinstated to the HCUF, had slightly higher vacancy rates than housing units on the HCUF.
 - This seemingly anomalous result may warrant further investigation.¹⁰
- The vacancy rates imputed to addresses with unknown occupancy or status were basically consistent between HCUF housing units, housing unit records deleted as duplicates, and housing units reinstated to the HCUF.

⁹ Excludes housing units reinstated to the HCUF. Percentages are essentially the same with or without the reinstated units.

¹⁰ The reinstatement rules (Howard Hogan, “Specification for Reinstating Addresses Flagged as Deletes on the Hundred Percent Census Unedited File,” DSSD Census 2000 Procedures and Operations Memorandum Series #D-11, November 7, 2000) can be interpreted to lean, in minor ways, towards reinstating duplicates that were matched by address, but this writer is not convinced that the reinstatement rules can explain more than a fraction of the difference. The reinstatement rules are not subject to easy summarization, and any reader desiring to know more about those rules is urged to read the reinstatement specification.

4.5 What was the breakdown of addresses resolved as deletes in Census 2000?

Correctly deleted addresses were addresses deleted in HCUF creation that were found to be not a valid housing unit by either Nonresponse Followup, Coverage Improvement Followup, or Field Verification. Addresses deleted in error were addresses deleted in HCUF creation despite having been identified as a valid occupied or vacant housing unit by either Nonresponse Followup, Coverage Improvement Followup, or Field Verification.

Table 5: Housing Unit Addresses Resolved as Deletes in the Final Census

Status	Count	Percent
Correctly Deleted	3,185	36.8
Deleted in Error	5,469	63.2
Total	8,654	100.0

Sources: HCUF, DMAF. Includes addresses resolved as deletes in Table 1.

Approximately 63.2 percent of addresses resolved as deletes were deleted in error.

5. RECOMMENDATIONS

5.1 Enhance the Census software quality assurance process to ensure adherence to specifications

Problems with adherence to specifications cropped up in two areas of HCUF creation.

As indicated in Table 5, 5,469 addresses that were deleted from the Census were apparently deleted in error. These were addresses that were identified as occupied or vacant by the Nonresponse Followup or Coverage Improvement Followup operations, or verified by the Field Verification operation. As the Housing Unit Status specification¹¹ indicates, such units were to be kept on the HCUF.

As discussed in the Limitations section, the DMAF did not distinguish Kills from other deleted addresses, contrary to the Kill Specification. The DSSD had to rely on control files from an independent quality control testing program to divide the deleted addresses into the three categories of Resolved Deletes, Imputed Deletes, and Kills for purposes of this evaluation.

5.2 Reexamine timing of late Census operations

If no questionnaire was placed on the DRF for an address, and the address was not assigned to any regular or followup Census 2000 operation involving an enumerator visit, then the source is given as “No Data.” As Table 2 shows, there were 251,478 such addresses that were either included in the Census, or deleted in status imputation. Most of these addresses were added to the DMAF too late in Census 2000 for any return to be included in the DRF. We need to examine whether there is a way to work out the timing of the late Census operations so that any late adds from these operations can be placed on the DMAF in time for questionnaires captured from those addresses to have a DMAF address to be linked to.

5.3 Conduct further research on Duplicate Delete cases

Table 4A shows the differences in vacancy rates between housing units on the HCUF, housing unit records deleted as duplicates, and housing units provisionally deleted as duplicates but ultimately reinstated. The extremely low vacancy rates for housing unit records deleted as duplicates can largely be explained by the heavy use of person records in identifying possible duplicate pairs. However, the vacancy rates for the reinstated housing units are slightly higher than those never considered for deletion from the HCUF. The process by which duplicates were identified and reinstated or deleted suggests that the factors causing a low vacancy rate for

¹¹ Alberti, Nick, “Specifications for Assigning the Housing Unit Status and Population Count of the Hundred-Percent Unedited File Prior to the Imputation of Unclassified Units,” DSSD Census 2000 Procedures and Operations Memorandum Series #D-14, January 19, 2001.

deleted housing unit records should, for the most part, apply to reinstated housing units. We feel this merits further research.

References

Alberti, Nick, "Specifications for Assigning the Housing Unit Status and Population Count of the Hundred-Percent Unedited File Prior to the Imputation of Unclassified Units," DSSD Census 2000 Procedures and Operations Memorandum Series #D-14, January 19, 2001.

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

Baumgardner, Stephanie, "Analysis of the Primary Selection Algorithm," Census 2000 Evaluation L.3.a, November 12, 2002.

Fay, Robert, "The 2000 Housing Unit Duplication Operations and Their Effect on the Accuracy of the Population Count," Proceedings of the Annual Meeting of the American Statistical Association, August 5-9, 2001.

Griffin, Richard, "Census 2000: Overview of Count Imputation – Reissue of Q-2," DSSD Census 2000 Procedures and Operations Memorandum Series #Q-78, March 18, 2002.

Griffin, Richard, "Census 2000 Count Imputation - Results," DSSD Census 2000 Procedures and Operations Memorandum Series #Q-81, November 5, 2002.

Nash, Fay, "ESCAP II: Analysis of Census Imputations," Executive Steering Committee for A.C.E. Policy II Report No. 21, September 24, 2001.

Treat, James (2000), "Specification of the Kill Universe on the Decennial Master Address File for Census 2000," Internal Census Bureau memorandum, DSSD Census 2000 Procedures and Operations Memorandum Series, #D-13.

Appendix 1: DMAF Addresses Resolved As Housing Units in the Census

State	Housing Units
Alabama	1,963,711
Alaska	260,978
Arizona	2,189,189
Arkansas	1,173,043
California	12,214,549
Colorado	1,808,037
Connecticut	1,385,975
Delaware	343,072
District of Columbia	274,845
Florida	7,302,947
Georgia	3,281,737
Hawaii	460,542
Idaho	527,824
Illinois	4,885,615
Indiana	2,532,319
Iowa	1,232,511
Kansas	1,131,200
Kentucky	1,750,927
Louisiana	1,847,181
Maine	651,901
Maryland	2,145,283
Massachusetts	2,621,989
Michigan	4,234,279
Minnesota	2,065,946
Mississippi	1,161,953
Missouri	2,442,017
Montana	412,633
Nebraska	722,668
Nevada	827,457
New Hampshire	547,024
New Jersey	3,310,275
New Mexico	780,579
New York	7,679,307
North Carolina	3,523,944
North Dakota	289,677
Ohio	4,783,051
Oklahoma	1,514,400
Oregon	1,452,709

Pennsylvania	5,249,750
Puerto Rico	1,418,476
Rhode Island	439,837
South Carolina	1,753,670
South Dakota	323,208
Tennessee	2,439,443
Texas	8,157,575
Utah	768,594
Vermont	294,382
Virginia	2,904,192
Washington	2,451,075
West Virginia	844,623
Wisconsin	2,321,144
Wyoming	223,854
Total	117,323,117

Appendix 2: DMAF Addresses Not Included in the Census

State	Kills	Resolved as Deletes	Deleted as Duplicates	Deleted in Status Imputation	Totals
Alabama	176,557	119	38,548	722	215,946
Alaska	23,894	12	3,136	245	27,287
Arizona	196,174	94	29,214	2,643	228,125
Arkansas	81,250	59	19,984	314	101,607
California	798,048	558	89,411	4,976	892,993
Colorado	139,968	73	17,118	572	157,731
Connecticut	103,721	69	14,305	351	118,446
Delaware	21,214	37	3,990	96	25,337
District of Columbia	19,835	13	436	53	20,337
Florida	559,955	497	70,632	3,540	634,624
Georgia	370,164	215	55,480	1,154	427,013
Hawaii	79,299	124	8,733	262	88,418
Idaho	50,374	25	7,131	448	57,978
Illinois	609,885	629	53,659	3,066	667,239
Indiana	228,943	181	31,692	1,602	262,418
Iowa	64,055	45	12,245	178	76,523
Kansas	67,391	24	11,211	199	78,825
Kentucky	123,944	57	29,273	969	154,243
Louisiana	188,355	203	32,850	378	221,786
Maine	35,268	21	7,797	110	43,196
Maryland	125,511	83	21,289	527	147,410
Massachusetts	181,188	139	28,106	761	210,194
Michigan	277,440	190	38,147	1,147	316,924
Minnesota	127,625	47	18,017	277	145,966
Mississippi	95,367	78	26,880	662	122,987
Missouri	174,276	83	26,751	524	201,634
Montana	26,554	11	3,722	188	30,475
Nebraska	36,495	39	4,562	85	41,181
Nevada	42,965	28	5,485	862	49,340
New Hampshire	28,791	18	7,449	192	36,450
New Jersey	228,920	261	39,809	630	269,620

New Mexico	74,873	32	12,424	697	88,026
New York	743,860	735	103,320	2,384	850,299
North Carolina	269,625	490	61,381	1,950	333,446
North Dakota	19,016	2	2,886	50	21,954
Ohio	286,364	399	41,775	1,061	329,599
Oklahoma	89,798	88	16,875	365	107,126
Oregon	126,775	127	16,323	1,172	144,397
Pennsylvania	414,390	258	66,531	1,651	482,830
Puerto Rico	65,454	85	21,366	928	87,833
Rhode Island	33,142	28	5,029	143	38,342
South Carolina	207,758	371	40,477	1,048	249,654
South Dakota	18,115	13	2,788	92	21,008
Tennessee	208,647	492	41,058	1,149	251,346
Texas	538,176	809	93,822	2,614	635,421
Utah	76,415	63	8,845	281	85,604
Vermont	25,050	18	5,057	73	30,198
Virginia	138,631	260	28,293	602	167,786
Washington	217,156	234	28,777	1,470	247,637
West Virginia	46,867	17	11,988	341	59,213
Wisconsin	156,781	97	24,573	1,247	182,698
Wyoming	16,875	4	2,036	75	18,990
Total	9,057,195	8,654	1,392,686	47,126	10,505,661