Census 2000 Evaluation K.1.a
October 24, 2003

DCS 2000 Data Capture Audit Resolution Process

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 Applying quality procedures 4

3. LIMITS 4

4. RESULTS 5

5. CONCLUSIONS 18

6. RECOMMENDATION 20

References 21

LIST OF TABLES AND FIGURES

Table 1. Form type by DCAR edit results 6

Table 2. Data Capture Center by DCAR edit results 7

Table 3. Household size for mail returns by DCAR edit results 8

Table 4. Household size for enumerator returns by DCAR edit results 8

Figure 1. Count Check percent per week 9

Figure 2. Status Review percent per week 9

Table 5. Duplicate person panels, mail returns 10

Table 6. Duplicate roster entries, mail returns 11

Table 7. Validation changes for person panels by Data Capture Center, mail returns 12

Table 8. Validation changes for roster entries by Data Capture Center, mail short forms 13

Table 9. Validation changes for roster entries by Data Capture Center, mail long forms 14

Table 10. Validation changes for person panels by household size, mail returns 15

Table 11. Validation changes for roster entries by household size, mail short forms 16

Table 12. Validation changes for roster entries by household size, mail long forms 17

Table 13. Whether post-Data Capture Audit Resolution mail and enumerator returns now meet the criteria to pass the edit 18

EXECUTIVE SUMMARY

This evaluation looked at results from Data Capture Audit Resolution. Data Capture Audit Resolution identified errors in interpreting scanned questionnaires.

The Data Capture Audit Resolution consisted of the following three phases:

1. an automated review of data used to set person panel and roster entry statuses;

2. an edit to compare respondent or enumerator responses on household size to a household population count derived from a tally of person panels and roster entries; and,

3. a clerical review of images and an update of data for questionnaires whose response records had conflicting household size information.

In the first phase, the Data Capture Audit Resolution identified person panels with sufficient data (data-defined persons) and roster entries with three or more legal characters. We considered these person panels and roster entries to be valid persons and counted them in the household population. It also identified duplicate person panels based on name and age data and duplicate roster entries. We counted only non-duplicate persons in the household population.

We compared the household population based on the count of valid persons to the respondent or enumerator household size responses. Cases with conflicting household size information failed the edit in the second phase of the Data Capture Audit Resolution.

In the third phase, clerks reviewed computer images of questionnaires that failed the edit. There were two types of review: the Audit Count Check and the Audit Status Review.

The Audit Count Check required that clerks review and correct the Optical Character Recognition interpretation of respondent or enumerator responses on household size only. They did not make corrections to the Optical Character Recognition fields based on a review of person panels or roster entries.

The Audit Status Review also required that clerks review and correct the Optical Character Recognition fields. In addition, they required the review and correction of the status of person panels and roster entries. (We did not alter response data in the process of correcting the status of person panels or roster entries.) The clerical staff could only correct the statuses set in the first phase of the Data Capture Audit Resolution. The Data Capture Audit Resolution process did not require that a questionnaire meet the criteria to pass the edit applied in the second phase in order for the third phase to be complete.

Conclusions -

- The DCS 2000 successfully captured the response data that was input to the determination of household size. It successfully captured numeric responses and accurately identified the presence of responses in check boxes.

- Of the 126,866,759 returns that were sent to DCAR, 124,194,637 returns, or 97.89 percent, passed the edit. Of the 2,672,122 failed edits, the Count Check process included 882,555 returns, or 33.03 percent, and the Status Review process included 1,789,567 returns, or 66.97 percent.

- The rate of edit failures varied only slightly across Data Capture Center within form type.

- The rate at which mail returns passed the DCAR edit varied greatly by household size. Vacant mail returns passed the DCAR edit at rate only about 8 out of 100. It is possible that many of the vacant mail returns represent occupied housing units. About 98 percent of mail returns with a household size between 1 and 9 passed the DCAR edit but only about 61 percent of the mail returns with a household size of 10 or more passed the edit. This may be due in part to the limit of 12 names that could be reported on a mail return.

  The rate at which enumerator returns passed the DCAR edit varied only slightly by household size. The rate decreased slightly as household size increased. It is curious that the rate for households with 10 or more persons is so much larger for enumerator returns compared to mail returns, 96 percent versus 61 percent.

- As the check-in date of the return became further removed from Census Day, the percent sent to Count Check and Status Review increased for mail returns faster than for enumerator returns, indicating more consistent quality for enumerator returns over time.

- The status of pre-audit duplicates among person panels and among roster entries on mail returns were compared to their post-Status Review status. There were 52,406 pre-audit duplicate person panels and 41,562 pre-audit duplicate roster entries. Only 507, or 0.97 percent of the person panels were determined to not be a duplicate and only 1,233, or 2.97 percent of the roster entries were determined to not be a duplicate by the Status Review process. The lower rate of change for person panels may indicate that without associated demographic characteristics, which roster entries lack, it is more difficult accurately to identify duplicates.

- The Status Review changed only a small percentage of pre-audit statues.

  - The Status Review process changed about 12 percent of the statuses for person panels with a pre-audit status of valid. The Status Review process changed about 13 percent of the statuses for person panels with a pre-audit status of invalid.

- The Status Review process changed about 29 percent of the statuses for short form mail return roster entries with a pre-audit status of valid. The Status Review process changed less than 0.5 percent of the statuses for short form mail return roster entries with a pre-audit status of invalid.

- The Status Review process changed about 10 percent of the statuses for long form mail return roster entries with a pre-audit status of valid. The Status Review process changed about 4 percent of the statuses for long form mail return roster entries with a pre-audit status of invalid.

- When the DCAR edit is applied to the post-DCAR data (i.e. the data after Status Review and Count Check edits), about one-third of the mail returns that originally failed the DCAR edit, meet the criteria to pass the DCAR edit. This is about 35 percent of those included in the Status Review process and about 32 percent of those included in the Count Check process. When the DCAR edit is applied to the post-DCAR data about 63.97 percent of the enumerator returns that originally included in the Count Check process, meet the criteria to pass the DCAR edit.

  The Count Check process only made changes only to the respondent filled or enumerator filled population counts. These results imply that the DCS2000 had much more success interpreting the numeric characters written by mail respondents than those written by Census enumerator

Recommendation -

A process similar to the DCAR should be incorporated into the 2010 Census. The DCAR corrected the data on a large number of cases that would have been included in the Coverage Edit Followup (CEFU) without the corrections made by the DCAR process. Without the DCAR process, CEFU in the Census 2000 would have included as many as 369,000 additional cases.

1. BACKGROUND

1.1 The 1990 Census

The data capture system for the 1990 Census used FOSDIC (film optical sensing device for input to computers) technology to capture and process data from census questionnaires. We filmed questionnaires and then created electronic data files from the microfilm images as follows:

- We converted filled answer circles directly to electronic data by scanning the microfilm images.

- We captured character data as electronic data using microfilm images and by keying data in a data entry process.

Prior to filming, clerks reviewed each mail return and recorded the maximum possible household population count on the questionnaire, based on the number of person panels with at least one item filled and the number of names on the household roster. A return represented a single household enumeration.

After producing the electronic data, we performed an edit on the questionnaires comparing the number of data-defined person panels to the maximum population count recorded on the questionnaire in the clerical review.

We then performed another clerical review and update on all questionnaires that failed the edit (i.e., questionnaires for which the counts differed). Clerks reviewed the person panels and roster entries and recorded the correct household size. They filled cancellation circles of questionnaire person panels if marks on the questionnaire showed that the respondent or the enumerator meant to invalidate the response. They also filled in pre-coded answer circles for answers that only had been circled or underlined.

We recycled the updated questionnaires through the data capture process and created a new response record.

1.2 Census 2000 Dress Rehearsal

In the Census 2000 Dress Rehearsal, we captured the questionnaire data using the Data Capture System (DCS) 2000. This system consisted of scanning the questionnaires and producing electronic images of the questionnaires and electronic data files. Optical Character Recognition (OCR) and Optical Mark Recognition (OMR) interpreted the questionnaire entries, converting the filled answer circles and character data into an electronic data file.

The OMR and OCR systems do not capture data perfectly. Errors can occur for many reasons, including the following:

- answer circles not completely filled,
- stray marks on the questionnaire,
- character data not clearly written, and
- the respondent not following instructions for recording the answers.

We implemented the DCS 2000 Data Capture Audit Resolution (DCAR) to identify these errors when they affected the count of persons for a household. We conducted DCAR only on mail return records for the Census 2000 Dress Rehearsal.

The DCAR assigned a capture status to person panels and roster entries based on the content of data captured. Using these capture statuses, we derived a household population count. We compared this derived population count to the household size response provided by the respondent. When the two counts differed, clerks reviewed the questionnaire images and made corrections to any data misinterpreted by the OCR and OMR systems.

1.3 Census 2000

The Census 2000 DCAR process was similar to the dress rehearsal DCAR process but was restructured to include all mail returns and enumerator returns.

The DCAR process consisted of the following three phases:

1. an automated review of data used to set person panel and roster entry statuses;

2. an edit to compare respondent or enumerator responses on household size to a household population count derived from a tally of person panels and roster entries; and,

3. a clerical review of images and an update of data for questionnaires whose response records had conflicting household size information.

In the first phase, the DCAR identified person panels with sufficient data (data-defined persons) and roster entries with three or more legal characters. We considered these person panels and roster entries to be valid persons and counted them in the household population. It also identified duplicate person panels based on name and age data and duplicate roster entries. We counted only non-duplicate persons in the household population.

We compared the household population based on the count of valid persons to the respondent or enumerator household size responses. Cases with conflicting household size information failed the edit in the second phase of the DCAR.

In the third phase, clerks reviewed computer images of questionnaires that failed the edit. There were two types of review: the Audit Count Check and the Audit Status Review.

The Audit Count Check required that clerks review and correct the OCR interpretation of respondent or enumerator responses on household size only. They did not make corrections to

the OCR fields based on a review of person panels or roster entries. The Audit Status Review also required that clerks review and correct the OCR fields. In addition, they required the review and correction of the status of person panels and roster entries. (We did not alter response data in the process of correcting the status of person panels or roster entries.) The clerical staff could only correct the statuses set in the first phase of the DCAR. The DCAR process did not require that a questionnaire meet the criteria to pass the edit applied in the second phase in order for the third phase to be complete.

The Census 2000 requirements are described in U.S. Census Bureau, 1999.

2. METHODS

2.1 Files used in this evaluation

The source of data was the Decennial Response File (DRF). The DRF includes data on each questionnaire successfully captured in the census including data on the results of the DCAR process. These data show the pre-DCAR status of both person panels and roster entries.

This study is concerned only with stateside census questionnaires to which the DCAR process is applied. These include mail questionnaires (D-1, D-2, D-1(UL), D-2(UL)) and enumerator questionnaires (D-1(E), D-2(E), D-1(E)SUPP, D-2(E)SUPP, D-1(E)(converted to continuation form or ccf), D-2(E)(ccf)).

Study Variables:

Listed below are the descriptions of the important variables that enabled us to categorize DCAR results.

RNPOP - the respondent response to inquiry “How many people were living or staying in this house, apartment, or mobile home on April 1, 2000?”

RISPOP - The household size count recorded by the enumerator in Item B of enumerator returns at the completion of the interview.

PCANCEL - The check item used by enumerators to cancel a person panel on an enumerator questionnaire.

RFT - Type of questionnaire.

Variables derived during the DCAR process

PDSTAT - The pre-audit resolution person panel status.

PDVSTAT - The post-audit resolution person panel status.

RRSTATn - The pre-audit resolution status of roster position n

RRVSTATn - The post-audit resolution status of roster position n

[PDSTAT, PDVSTAT, RRSTATn, RRVSTATn can take on the values Blank, Valid, Invalid and Canceled.]

PDKFIBL & PDKFIBF - The Key From Image (KFI) name blanking status of the person panel last and first names fields, respectively.

RRSTATBLn & RRSTATBFn - The KFI name blanking status of the last and first name fields for roster position n, respectively.

RIDPPOP - The pre-audit resolution count of Valid person panels (See PDSTAT).

RVDPPOP - The post-audit resolution count of Valid person panels (See PDVSTAT).

RROSPOP - The pre-audit resolution count of Valid roster entries (see RRSTATn).

RVROSPOP - The post-audit resolution count of Valid roster entries (see RRVSTATn).

RTOTPOP - The pre-audit resolution household POP count.

RVTOTPOP - The post-audit resolution household POP count.

RDCAREDIT - The DCAR edit results for the household. (Passed, eligible for Audit Count Check, or eligible for Audit Status Review.)

2.2 Applying 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

Information on changes made by the DCAR clerical review and updates to respondent and enumerator household size responses are not available. We will not be able to report on the number or types of changes that clerical staff made to the OCR interpreted responses (the Audit Count Check outcomes).

The changes made by the clerical review and update to the statuses of person panels from enumerator questionnaires (Forms D-1(E) and D-2(E)) were not recorded on the DRF. We know that the variables reflecting the result of these changes were not included in the data output from the DCS 2000. As a result, the action taken by the DCAR clerical review and update were not used to validate or invalidate person panels on the enumerator forms. We do not know how often updates to these person panel statuses were made and ignored. As a result, for enumerator returns we will not be able to answer questions on person panels (the Audit Status Review outcomes).

4. RESULTS

This section presents the results of the DCAR process.

Table 1 shows that 2.11 percent of the 126,866,759 DRF returns included in the DCAR failed the DCAR edit. Of the 2,672,122 failed edits, clerks performed Audit Count Checks on 33.03 percent, and they performed Audit Status Reviews on 66.97 percent.

The 1,448,534 mail short-form returns that failed the DCAR edit had 37.16 percent sent to Count Check and 62.82 percent sent to Status Review. Similarly, the 196,071 failed mail long-form returns had 40.63 percent sent to Count Check and 59.37 percent sent to Status Review. For enumerator returns, a lower percent of the failures were sent to Count Check, 25.35 percent of enumerator short forms and 26.94 percent of enumerator long forms.

The mail and enumerator returns failed the DCAR edit based on household size comparisons that were specific to each form type. For mail returns, DCAR compared the respondent-reported household size and the total persons on the form as determined by the person-panels filled and roster entries completed. For enumerator returns, DCAR compared the respondent-reported household size, the number of person-panels filled, and the Interview Summary Population.

Table 1. Form type by DCAR edit results

	Total		Mail short form		Enumerator short form		Mail long form		Enumerator long form	
Status	Number	%	Number	%	Number	%	Number	%	Number	%
Total DCAR returns	126,866,759	100.00	68,380,897	100.00	36,180,534	100.00	11,926,315	100.00	10,379,013	100.00
Pass	124,194,637	97.89	66,932,363	97.88	35,395,272	97.83	11,730,244	98.36	10,136,758	97.67
Count Check	882,555	0.70	538,604	0.79	199,040	0.55	79,657	0.67	65,254	0.63
Status Review	1,789,567	1.41	909,930	1.33	586,222	1.62	116,414	0.98	177,001	1.71

Data source: Decennial Response File

Table 2 shows that the DCAR edit results were consistent for each form type across Data Capture Centers. Each form type passed the DCAR edit for 97.53 percent to 98.64 percent of the total DCAR returns.

For each Data Capture Center by form type, failed edits were distributed consistently. The results were the same as the overall results discussed in Table 1: a lower percent of enumerator returns than mail returns failed Count Check relative to Status Review.

Table 2. Data Capture Center by DCAR edit results

Data Capture Center	Mail short form		Enumerator short form		Mail long form		Enumerator long form	
Status	Number	%	Number	%	Number	%	Number	%
Total DCAR returns	68,380,897	100.00	36,180,534	100.00	11,926,315	100.00	10,379,013	100.00
Pass	66,932,363	97.88	35,395,272	97.83	11,730,244	98.36	10,136,758	97.67
Count Check	538,604	0.79	199,040	0.55	79,657	0.67	65,254	0.63
Status Review	909,930	1.33	586,222	1.62	116,414	0.98	177,001	1.71
Baltimore	20,371,811	100.00	8,934,756	100.00	3,573,341	100.00	2,703,590	100.00
Pass	19,973,603	98.04	8,726,996	97.67	3,518,962	98.48	2,636,712	97.53
Count Check	146,451	0.72	52,117	0.58	21,133	0.59	17,633	0.65
Status Review	251,757	1.24	155,643	1.74	33,246	0.93	49,245	1.82
Jeffersonville	6,413,328	100.00	3,542,523	100.00	1,230,357	100.00	1,155,178	100.00
Pass	6,294,983	98.15	3,459,270	97.65	1,213,594	98.64	1,127,538	97.61
Count Check	44,558	0.69	21,700	0.61	6,789	0.55	7,546	0.65
Status Review	73,787	1.15	61,553	1.74	9,974	0.81	20,094	1.74
Phoenix	20,984,719	100.00	12,670,924	100.00	3,450,410	100.00	3,245,247	100.00
Pass	20,486,556	97.63	12,403,130	97.89	3,387,102	98.17	3,169,999	97.68
Count Check	181,479	0.86	68,061	0.54	25,868	0.75	20,117	0.62
Status Review	316,684	1.51	199,733	1.58	37,440	1.09	55,131	1.70
Pomona	20,611,039	100.00	11,032,331	100.00	3,672,207	100.00	3,274,998	100.00
Pass	20,177,221	97.90	10,805,876	97.95	3,610,586	98.32	3,202,509	97.79
Count Check	166,116	.81	57,162	0.52	25,867	0.70	19,958	0.61
Status Review	267,702	1.30	169,293	1.53	35,754	0.97	52,531	1.60

Data source: Decennial Response File

Tables 3 and 4 show the effect of household size on the DCAR results. The household size is the respondent-reported household size, if reported. For mail returns, vacant returns had low pass rates, 8.12 percent. This indicates that completing the forms was confusing for these cases.

For non-vacant mail returns, the DCAR pass rate decreased as the household size increased. For the largest households, the respondent may not have been able to list all of the persons in the household on the 12-person roster, causing Count Check failures.

For enumerator returns, the DCAR pass rate decreased as the household size increased, but not as precipitously as for mail returns. The enumerators were familiar with how to fill out the forms. This is also evidenced by the high pass rate for vacant returns. Confusion in using multiple continuation forms may contribute to Status Review failures for the largest households.

Table 3. Household size for mail returns by DCAR edit results

	No response		Vacant		1-4 persons		5-9 persons		10+ persons	
Status	Number	%	Number	%	Number	%	Number	%	Number	%
Total DCAR returns	4,017,565	100.00	127,501	100.00	68,270,852	100.00	7,702,197	100.00	189,097	100.00
Pass	4,007,427	99.75	10,349	8.12	67,080,778	98.26	7,449,018	96.71	115,035	60.83
Count Check	9,190	0.23	3,078	2.41	412,122	0.60	128,482	1.67	65,389	34.58
Status Review	948	0.02	114,074	89.47	777,952	1.14	124,697	1.62	8,673	4.59

Data source: Decennial Response File

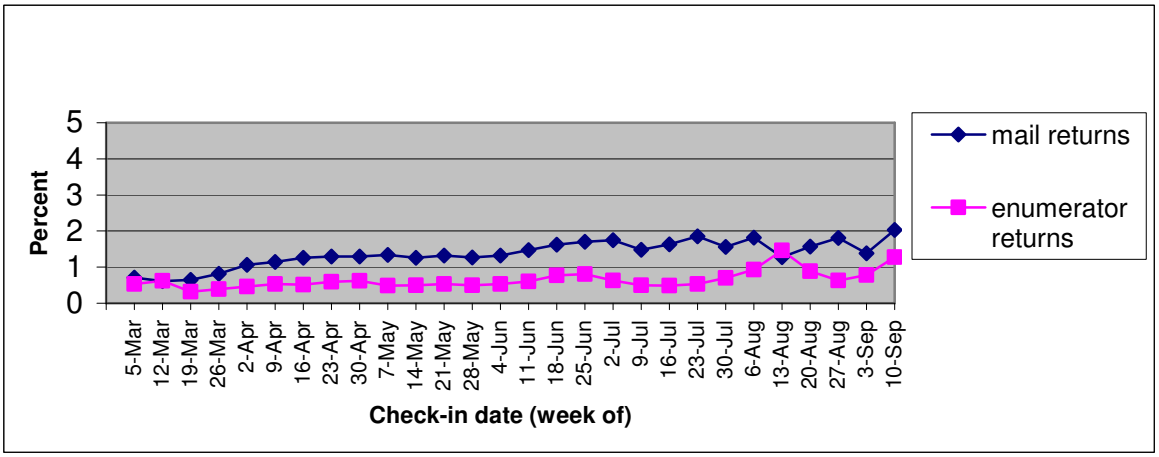
Table 4. Household size for enumerator returns by DCAR edit results

	No response		Vacant		1-5 persons		6-9 persons		10+ persons	
Status	Number	%	Number	%	Number	%	Number	%	Number	%
Total DCAR returns	8,033,804	100.00	6,331,511	100.00	28,204,132	100.00	1,453,446	100.00	2,536,654	100.00
Pass	7,930,978	98.72	6,191,283	97.79	27,569,581	97.75	1,415,079	97.36	2,425,109	95.60
Count Check	12,981	0.16	38,778	0.61	155,239	0.55	21,368	1.47	35,928	1.42
Status Review	89,845	1.12	101,450	1.60	479,312	1.70	16,999	1.17	75,617	2.98

Data source: Decennial Response File

Figure 1 shows that mail returns had a greater percent of Count Check failures than enumerator returns. The percents for mail returns gradually increased as the time from Census Day (April 1, 2000) increased, indicating a decrease in quality for later returns. The rates for enumerator returns were relatively constant with some degradation of quality for later returns.

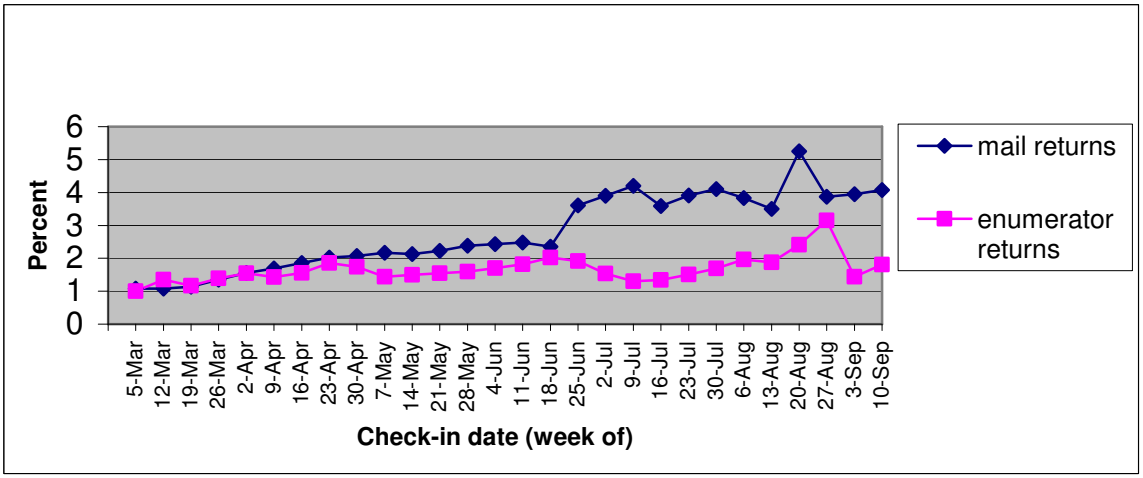
Figure 1. Count Check percent per week



Data source: Decennial Response File

Figure 2 shows that mail and enumerator returns initially had similar percents of Status Review failures. As the check-in date became further from Census Day, the mail returns had a greater percent of Status Review failures. The enumerator returns held relatively steady becoming more erratic when there were few returns at the end of the collection of data.

Figure 2. Status Review percent per week



Data source: Decennial Response File

Table 5 compares the pre-audit and post-Status Review statuses of person panels identified as duplicates in either pre-audit or post-Status Review.

For the 52,406 duplicate person panels in the pre-audit, clerks identified 16,527 person panels, or 31.54 percent, as duplicates. Clerks identified few of the pre-audit duplicate person panels, 507 or 0.97 percent, as valid. Most of the person panels became blank (or not computed) post-Status Review, 33,549, or 64.02 percent. Since few person panels turned valid, the designation of duplicate was consistent with the post-Status Review outcome. The clerks may have performed no action—the blank result—rather than confirming the duplicate status.

Of the 40,733 person panels that clerks identified as duplicates in Status Review, 16,527 person panels, or 40.57 percent, were pre-audit duplicates. Most of the post-Status Review duplicate person panels, 23,697, or 58.18 percent had been valid in pre-audit. Clerks were able to resolve some audit cases by identifying valid persons as duplicates. Relying only on the automated review would have missed some duplicates.

Table 5. Duplicate person panels, mail returns

Post-Status Review	Duplicate persons		Invalid persons		Cancel persons		Valid persons		Blank	
Pre-audit	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Duplicate persons	16,527	31.54	1,820	3.47	3	0.01	507	0.97	33,549	64.02
Short form	14,351	30.42	1,693	3.59	2	0.00	450	0.96	30,673	65.03
Long form	2,176	41.55	127	2.43	1	0.02	57	1.09	2,876	54.92
Invalid persons	180	0.11								
Short form	144	0.09								
Long form	36	0.97								
Cancel persons	326	0.13								
Short form	296	0.13								
Long form	30	0.15								
Valid persons	23,697	0.73								
Short form	20,932	0.72								
Long form	2,765	0.83								
Blank	3	0.01								
Short form	3	0.01								
Long form	0	0.00								

Data source: Decennial Response File

Table 6 compares the pre-audit and post-Status Review statuses of roster entries identified as duplicates in either pre-audit or post-Status Review.

For the 41,562 duplicate roster entries in the pre-audit, clerks identified 10,164 roster entries, or 24.46 percent, as duplicates. Most of the pre-audit duplicate roster entries became invalid post-Status Review, 29,527, or 71.04 percent. Clerks identified 1,233, or 2.97 percent of the pre-audit duplicate roster entries as valid. A higher percent of the roster entries were identified as valid post-audit than the person panels 2.97 percent vs. 0.97 percent, respectively. This may indicate that it is more difficult to create an algorithm to identify duplicate roster entries than person panels.

Of the 28,196 roster entries that clerks identified as duplicates in Status Review, 10,164 roster entries, or 36.05 percent, were pre-audit duplicates. Most of the post-Status Review duplicate roster entries, 17,621, or 62.49 percent had been valid in pre-audit. Clerks were able to resolve some audit cases by identifying valid roster entries as duplicates. Relying only on the automated review would have missed some duplicates.

Table 6. Duplicate roster entries, mail returns

Post-Status Review	Duplicate persons		Invalid persons		Cancel persons		Valid persons		Blank	
Pre-audit	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Duplicate persons	10,164	24.46	29,527	71.04	102	0.25	1,233	2.97	536	1.29
Short form	7,638	22.80	24,961	74.51	84	0.25	364	1.09	451	1.35
Long form	2,526	31.32	4,566	56.62	18	0.22	869	10.78	85	1.05
Invalid persons	15	0.04								
Short form	10	0.03								
Long form	5	0.16								
Cancel persons	282	0.07								
Short form	163	0.05								
Long form	119	0.28								
Valid persons	17,621	0.03								
Short form	6,001	0.02								
Long form	11,620	0.04								
Blank	114	0.21								
Short form	72	0.15								
Long form	42	0.59								

Data source: Decennial Response File

Table 7 compares the pre-audit and post-Status Review statuses of mail return person panels. Invalid persons consists of Duplicate, Invalid, and Cancel statuses.

Of the 464,135 pre-audit invalid person panels, 251,962 person panels, or 54.29 percent were post-Status Review invalid. Besides remaining invalid, many of the invalid person panels became blank in post-Status Review, 150,695 person panels, or 32.47 percent.

Of the 3,233,829 pre-audit valid person panels, 2,843,022 person panels, or 87.92 percent, were post-Status Review valid.

Of the 25,451 pre-audit blank person panels, 4,317 person panels, or 16.96 percent, were post-Status Review blank.

Table 7. Validation changes for person panels by Data Capture Center, mail returns

Post-Status Review	Invalid persons		Valid persons		Blank	
Pre-audit	Number	Percent	Number	Percent	Number	Percent
Invalid persons	251,962	54.29	61,478	13.25	150,695	32.47
Baltimore	60,664	49.22	17,908	14.53	44,670	36.25
Jeffersonville	14,121	42.86	5,934	18.01	12,890	39.13
Phoenix	107,665	60.32	15,199	8.52	55,613	31.16
Pomona	69,512	53.69	22,437	17.33	37,522	28.98
Valid persons	270,726	8.37	2,843,022	87.92	120,081	3.71
Baltimore	68,717	7.90	773,503	88.92	27,627	3.18
Jeffersonville	21,409	8.15	230,083	87.54	11,351	4.32
Phoenix	96,825	8.57	988,959	87.57	43,552	3.86
Pomona	83,775	8.62	850,477	87.52	37,551	3.86
Blank	18,745	73.65	2,389	9.39	4,317	16.96
Baltimore	5,217	75.86	753	10.95	907	13.19
Jeffersonville	1,173	62.16	152	8.06	562	29.78
Phoenix	6,527	72.83	875	9.76	1,560	17.41
Pomona	5,828	75.44	609	7.88	1,288	16.67

Data source: Decennial Response File

Table 8 compares the pre-audit and post-Status Review statuses of mail short-form roster entries.

Of the 406,267 pre-audit invalid roster entries, 383,760 roster entries, or 94.46 percent were post-Status Review invalid.

Of the 276,845 pre-audit valid roster entries, 194,719 roster entries, or 70.34 percent, were post-Status Review valid.

Table 8. Validation changes for roster entries by Data Capture Center, mail short forms

Post-Status Review	Invalid persons		Valid persons		Blank	
Pre-audit	Number	Percent	Number	Percent	Number	Percent
Invalid persons	383,760	94.46	1,786	0.44	20,721	5.10
Baltimore	96,333	94.60	531	0.52	4,972	4.88
Jeffersonville	26,333	88.43	105	0.35	3,341	11.22
Phoenix	139,467	94.17	585	0.40	8,048	5.43
Pomona	121,627	96.11	565	0.45	4,360	3.45
Valid persons	80,769	29.17	194,719	70.34	1,357	0.49
Baltimore	17,604	29.57	41,713	70.06	225	0.38
Jeffersonville	4,254	22.35	14,447	75.89	335	1.76
Phoenix	39,136	34.48	73,956	65.16	413	0.36
Pomona	19,775	23.33	64,603	76.22	384	0.45
Blank	47,294	98.75	599	1.25		
Baltimore	13,248	98.66	180	1.34		
Jeffersonville	3,007	98.82	36	1.18		
Phoenix	15,430	98.41	250	1.59		
Pomona	15,609	99.16	133	0.84		

Data source: Decennial Response File

Table 9 compares the pre-audit and post-Status Review statuses of mail long-form roster entries.

Of the 53,566 pre-audit invalid roster entries, 49,067 roster entries, or 91.60 percent were post-Status Review invalid.

Of the 330,638 pre-audit valid roster entries, 297,405 roster entries, or 89.95 percent, were post-Status Review valid.

Table 9. Validation changes for roster entries by Data Capture Center, mail long forms

Post-Status Review	Invalid persons		Valid persons		Blank	
Pre-audit	Number	Percent	Number	Percent	Number	Percent
Invalid persons	49,067	91.60	2,070	3.86	2,429	4.53
Baltimore	12,733	90.65	680	4.84	634	4.51
Jeffersonville	3,323	87.86	207	5.47	252	6.66
Phoenix	17,016	92.38	450	2.44	953	5.17
Pomona	15,995	92.36	733	4.23	590	3.41
Valid persons	32,952	9.97	297,405	89.95	281	0.08
Baltimore	7,884	8.63	83,405	91.31	54	0.06
Jeffersonville	5,425	19.17	22,814	80.63	57	0.20
Phoenix	11,126	10.23	97,531	89.69	83	0.08
Pomona	8,517	8.33	93,655	91.59	87	0.09
Blank	6,855	96.92	218	3.08		
Baltimore	1,910	97.40	51	2.60		
Jeffersonville	447	96.34	17	3.66		
Phoenix	2,161	95.92	92	4.08		
Pomona	2,337	97.58	58	2.42		

Data source: Decennial Response File

Table 10 compares the pre-audit and post-Status Review statuses by household size of mail return person panels.

The lowest percent of returns that switched from pre-audit invalid to post-Status Review valid were for vacant returns, 3.13 percent. The highest percent of returns that switched from pre-audit invalid to post-Status Review valid were household returns with 5-9 persons, 27.80 percent. The switch means that the data capture system assigned an invalid status that clerks reinterpreted as valid.

The highest percent of returns that switched from pre-audit valid to post-Status Review invalid were household returns with 1-4 persons, 11.52 percent. The lowest percent of returns that switched from pre-audit valid to post-Status Review invalid were for household returns with

10 plus persons, 0.83 percent. This could be that clerks were able to review more thoroughly and identify more easily invalid persons when there were fewer person panels.

Table 10. Validation changes for person panels by household size, mail returns

Post-Status Review	Invalid persons		Valid persons		Blank	
Pre-audit	Number	Percent	Number	Percent	Number	Percent
Invalid persons	251,962	54.29	61,478	13.25	150,695	32.47
Vacant	4,799	40.72	369	3.13	6,618	56.15
1-4 persons	236,609	58.23	48,650	11.97	121,078	29.80
5-9 persons	9,019	21.57	11,626	27.80	21,174	50.63
10+ persons	1,279	33.48	760	19.90	1,781	46.62
No response	256	68.63	73	19.57	44	11.80
Valid persons	270,726	8.37	2,843,022	87.92	120,081	3.71
Vacant	4,905	2.38	194,231	94.43	6,547	3.18
1-4 persons	257,525	11.52	1,932,401	86.43	45,885	2.05
5-9 persons	7,724	1.05	674,636	91.68	53,461	7.27
10+ persons	450	0.83	39,708	73.14	14,129	26.03
No response	122	5.48	2,046	91.87	59	2.65
Blank	18,745	73.65	2,389	9.39	4,317	16.96
Vacant	280	54.69	128	25.00	104	20.31
1-4 persons	18,218	74.66	2,108	8.64	4,075	16.70
5-9 persons	167	42.60	124	31.63	101	25.77
10+ persons	70	56.91	21	17.07	32	26.02
No response	10	43.48	8	34.78	5	21.74

Data source: Decennial Response File

Table 11 compares the pre-audit and post-Status Review statuses by household size of mail short form roster entries.

On mail short forms, the respondent completes a roster for persons 7-12. While many roster entries' statuses remain unchanged between pre-audit and post-Status Review, changes favor a greater percent of invalid becoming valid for larger households, 2.88 percent and 5.52 percent for 5-9 person returns and 10 plus person returns respectively vs. 0.33 percent and

0.27 percent for vacant returns and 1-4 person returns, respectively. Changes also favor a greater percent of valid becoming invalid for smaller households, 15.31 percent and 54.87 percent for vacant returns and 1-4 person returns respectively vs. 8.34 percent and 1.63 percent for 5-9 person returns and 1-4 person returns, respectively.

Table 11. Validation changes for roster entries by household size, mail short forms

Post-Status Review	Invalid persons		Valid persons		Blank	
Pre-audit	Number	Percent	Number	Percent	Number	Percent
Invalid persons	383,760	94.46	1,786	0.44	20,721	5.10
Vacant	7,508	91.87	27	0.33	637	7.79
1-4 persons	355,960	94.70	1,033	0.27	18,908	5.03
5-9 persons	16,806	91.69	527	2.88	996	5.43
10+ persons	3,185	89.69	196	5.52	170	4.79
No response	301	95.86	3	0.96	10	3.18
Valid persons	80,769	29.17	194,719	70.34	1,357	0.49
Vacant	508	15.31	2,795	84.21	16	0.48
1-4 persons	69,754	54.87	56,249	44.25	1,107	0.87
5-9 persons	10,050	8.34	110,278	91.48	223	0.18
10+ persons	419	1.63	25,260	98.33	11	0.04
No response	38	21.71	137	78.29	0	0.00
Blank	47,294	98.75	599	1.25		
Vacant	1,284	98.09	25	1.91		
1-4 persons	43,786	98.89	490	1.11		
5-9 persons	1,722	96.20	68	3.80		
10+ persons	447	96.75	15	3.25		
No response	55	98.21	1	1.79		

Data source: Decennial Response File

Table 12 compares the pre-audit and post-Status Review statuses by household size of mail long form roster entries.

On mail long forms, the respondent completes a roster of the entire household. While many roster entries' statuses remain unchanged between pre-audit and post-Status Review, changes favor a greater percent of invalid becoming valid for larger households, 9.08 percent and 6.89 percent for 5-9 person returns and 10+ person returns respectively vs. 1.94 percent and 2.87 percent for vacant returns and 1-4 person returns, respectively. Changes also favor a greater percent of valid becoming invalid for smaller households, 8.31 percent and 11.61 percent for vacant returns and 1-4 person returns respectively vs. 5.81 percent and 4.07 percent for 5-9 person returns and 1-4 person returns, respectively.

Table 12. Validation changes for roster entries by household size, mail long forms

Post-Status Review	Invalid persons		Valid persons		Blank	
Pre-audit	Number	Percent	Number	Percent	Number	Percent
Invalid persons	49,067	91.60	2,070	3.86	2,429	4.53
Vacant	1,713	90.02	37	1.94	153	8.04
1-4 persons	39,245	92.70	1,217	2.87	1,874	4.43
5-9 persons	6,966	86.65	730	9.08	343	4.27
10+ persons	1,039	88.35	81	6.89	56	4.76
No response	104	92.86	5	4.46	3	2.68
Valid persons	32,952	9.97	297,405	89.95	281	0.08
Vacant	1,115	8.31	12,301	91.64	7	0.05
1-4 persons	27,046	11.61	205,577	88.27	260	0.11
5-9 persons	4,519	5.81	73,224	94.17	14	0.02
10+ persons	257	4.07	6,059	95.93	0	0.00
No response	15	5.79	244	94.21	0	0.00
Blank	6,855	96.92	218	3.08		
Vacant	324	97.59	8	2.41		
1-4 persons	5,899	96.93	187	3.07		
5-9 persons	415	95.62	19	4.38		
10+ persons	183	98.92	2	1.08		
No response	34	94.44	2	5.56		

Data source: Decennial Response File

Table 13 shows that about one-third of post-DCAR mail returns meet the criteria to pass the edit. A slightly higher percent, 35.29 percent, pass the Status Review than the Count Check, 32.36 percent.

The enumerator returns had 169,073 returns, or 63.97 percent, post-DCAR that meet the criteria to pass the edit.

The higher percent of enumerator returns passing the edit post-DCAR might be because of different reasons for failing the edit. Enumerator returns might have had a simple handwriting correction. Mail returns might also have had more complex problems such as skipping questions or partially completing questions, which could not be remedied in the edit phase.

That some of the forms were now able to pass the DCAR edit indicates that the clerical review was beneficial.

Table 13. Whether post-Data Capture Audit Resolution mail and enumerator returns now meet the criteria to pass the edit

		Pass	
	Total	Number	Percent
Mail returns	1,644,605	562,201	34.18
Count Check	618,261	200,043	32.36
Status Review	1,026,344	362,158	35.29
Enumerator returns	264,294	169,073	63.97
Count Check	264,294	169,073	63.97

Data source: Decennial Response File

5. CONCLUSIONS

- The DCS 2000 successfully captured the response data that was input to the determination of household size. It successfully captured numeric responses and accurately identified the presence of responses in check boxes.

- Of the 126,866,759 returns that were sent to DCAR, 124,194,637 returns, or 97.89 percent, passed the edit. Of the 2,672,122 failed edits, the Count Check process included 882,555 returns, or 33.03 percent, and the Status Review process included 1,789,567 returns, or 66.97 percent.

- The rate of edit failures varied only slightly across Data Capture Center within form type.

- The rate at which mail returns passed the DCAR edit varied greatly by household size. Vacant mail returns passed the DCAR edit at rate only about 8 out of 100. It is possible that many of the vacant mail returns represent occupied housing units. About 98 percent of mail returns with a household size between 1 and 9 passed the DCAR edit but only about 61 percent of the mail returns with a household size of 10 or more passed the edit. This may be due in part to the limit of 12 names that could be reported on a mail return.

  The rate at which enumerator returns passed the DCAR edit varied only slightly by household size. The rate decreased slightly as household size increased. It is curious that the rate for households with 10 or more persons is so much larger for enumerator returns compared to mail returns, 96 percent versus 61 percent.

- As the check-in date of the return became further removed from Census Day, the percent sent to Count Check and Status Review increased for mail returns faster than for enumerator returns, indicating more consistent quality for enumerator returns over time.

- The status of pre-audit duplicates among person panels and among roster entries on mail returns were compared to their post-Status Review status. There were 52,406 pre-audit duplicate person panels and 41,562 pre-audit duplicate roster entries. Only 507, or 0.97 percent of the person panels were determined to not be a duplicate and only 1,233, or 2.97 percent of the roster entries were determined to not be a duplicate by the Status Review process. The lower rate of change for person panels may indicate that without associated demographic characteristics, which roster entries lack, it is more difficult accurately to identify duplicates.

- The Status Review changed only a small percentage of pre-audit statues.

  - The Status Review process changed about 12 percent of the statuses for person panels with a pre-audit status of valid. The Status Review process changed about 13 percent of the statuses for person panels with a pre-audit status of invalid.

  - The Status Review process changed about 29 percent of the statuses for short form mail return roster entries with a pre-audit status of valid. The Status Review process changed less than 0.5 percent of the statuses for short form mail return roster entries with a pre-audit status of invalid.

  - The Status Review process changed about 10 percent of the statuses for long form mail return roster entries with a pre-audit status of valid. The Status Review process changed about 4 percent of the statuses for long form mail return roster entries with a pre-audit status of invalid.

- When the DCAR edit is applied to the post-DCAR data (i.e. the data after Status Review and Count Check edits), about one-third of the mail returns that originally failed the DCAR edit, meet the criteria to pass the DCAR edit. This is about 35 percent of those included in the Status Review process and about 32 percent of those included in the Count Check process. When the DCAR edit is applied to the post-DCAR data about 63.97 percent of the enumerator returns that originally included in the Count Check process, meet the criteria to pass the DCAR edit.

  The Count Check process only made changes only to the respondent filled or enumerator filled population counts. These results imply that the DCS2000 had much more success interpreting the numeric characters written by mail respondents than those written by Census enumerator

6. RECOMMENDATION

A process similar to the DCAR should be incorporated into the 2010 Census. DCAR corrected the data on a large number of cases that would have been included in the Coverage Edit Followup (CEFU) without the corrections made by the DCAR process. Without the DCAR process, CEFU in the Census 2000 would have included as many as 369,000 additional cases.

References

U.S. Census Bureau, 1999, "Preliminary Analysis of the Dress Rehearsal Data Capture Audit and Resolution (DCAR) Edit and Review and its Redefinition for Census 2000 [Revision 2]," Decennial Statistical Studies Division Census 2000 Procedures and Operations Memorandum Series #C-1, July 23, 1999.

Sheppard, Dave (2003), Coverage Edit Followup, Census 2000 Evaluation I.1, July, 2003.