



UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

February 28, 2011

ACSO Research and Evaluation Memorandum Series

No. ACS-R&E -11-03

MEMORANDUM FOR

Distribution List

From:

James Treat
Chief, American Community Survey Office

Subject:

Evaluation of the Process Review Operation

Contact Person:

Epaphrodite Uwimana
IT Specialist, American Community Survey Office

Attached is the final report from evaluation of the 2008 and 2009 process review operation.

The major findings include:

- About 3.5 percent of all mail returns were sent to the process review operation in 2008 and 2009.
- The combination of person order misalignment and duplicate names was the leading edit failure reason for cases that were sent to process review (almost 1/4 of all edit failures).
- Process review analysts didn't take action on over 94 percent of the cases that were sent to the process review operation because of the combination of duplicate name and person order misalignment.
- With the exception of this specific failure reason, no consistent actions were found for other edit failure reasons suggesting that automating process review would not be simple.
- While an automated process review operation is possible we do not believe that the effort required to develop that automation would be cost-effective.
- If all cases failing for the combination of person order misalignment and duplicate names were no longer sent to process review, the workload would be reduced by about 23 percent with limited impact on the outcome. We recommend this change and only this change.

Attachment: Evaluation of the Process Review Operation.

Distribution List:

Frank A. Vitrano (DMD)
Daniel H. Weinberg (DIR)
ACS Division Chiefs Forum List
Jerry Imel (DSSD)
Earl Letourneau (DSSD)
Robin Pennington (DSSD)
Mary Bucci (DMD)
Kevin Deardorff (DMD)
Adrienne Oneto (DMD)

Evaluation of the Process Review Operation

FINAL REPORT

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

Helping You Make Informed Decisions

Epaphrodite Uwimana
American Community Survey Office

Table of Contents

Executive Summary	3
Introduction.....	4
Background	4
Research Questions.....	5
Methodology	5
Results	5
<i>What are the monthly workloads for the process review operation?</i>	<i>5</i>
<i>Why are cases failing and what actions are process review analysts taking to fix them?</i>	<i>7</i>
<i>Are there specific edit failure reasons that have a consistent process review action suggesting they could be automated?</i>	<i>9</i>
<i>Can an automated process review operation be designed that produces the same results as the current manual process?</i>	<i>12</i>
Recommendations	12

Executive Summary

From the time that the ACS started using the clerical process review operation for cases that fail a data consistency edit during the automated data capture process, it was suggested that it might be possible to automate some or all aspects of this operation. The goal of this evaluation was to determine if any changes should be made to the current process review.

The research questions for this evaluation were:

- What are the monthly workloads for the process review operation?
- Why are cases failing and what actions are process review analysts taking to fix them?
- Are there specific edit failure reasons that have a consistent process review action suggesting they could be automated?
- Can an automated process review operation be designed that produces the same results as the current manual process?

Two different datasets were created and used to answer the research questions. The 2008 and 2009 unweighted data from the control system were used along with Microsoft Excel and SAS Data Mining tools.

The major findings include:

- About 3.5 percent of all mail returns were sent to the process review operation in 2008 and 2009.
- The combination of person order misalignment and duplicate names was the leading edit failure reason for cases that were sent to process review (almost 1/4 of all edit failures).
- Process review analysts didn't take action on over 94 percent of the cases that were sent to the process review operation because of the combination of duplicate name and person order misalignment.
- With the exception of this specific failure reason, no consistent actions were found for other edit failure reasons suggesting that automating process review would not be simple.
- While an automated process review operation is possible we do not believe that the effort required to develop that automation would be cost-effective.
- If all cases failing for the combination of person order misalignment and duplicate names were no longer sent to process review, the workload would be reduced by about 23 percent with limited impact on the outcome. We recommend this change and only this change.

Introduction

The American Community Survey (ACS) has shifted from key-from-paper technology (KFP) to imaging and a key-from-image (KFI) technology to capture data from mail returned questionnaires. The specific technology that the ACS uses is the Integrated Computer Assisted Data Entry (iCADE) system. There were many changes made in the automated data capture from questionnaires as a consequence of this shift. A new operation, “process review”, was added to ensure that questionnaires that were completed in error by respondents were processed correctly. This operation identifies records with specified characteristics that are believed to be associated with respondent completion errors. Once these records are identified, they are sent to a set of analysts for review. During this review process, changes can be made to the data capture records. From the time that ACS started this process review operation, it was suggested that it might be possible to automate some or all aspects of the operation. The goal of this evaluation is to determine if any changes should be made to the current process review operation.

Background

There are currently eight different reasons that cause a case to be flagged as process review eligible. These reasons are listed below and can occur alone or in combination:

1. The respondent provided count of the total number of persons is less than the number of names listed in the basic demographics section of the questionnaire.
2. The respondent provided count of the total number of persons is less than the number of Data Defined Persons (DDPs) included in the basic demographics section of the questionnaire.¹
3. The number of names listed in the basic demographics section of the questionnaire is less than the number of DDPs in the basic demographics section of the questionnaire.
4. The number of DDPs included in the basic demographics section of the questionnaire is less than the number of DDPs included in the detailed population characteristics section of the questionnaire.
5. The respondent name does not match any of the names listed in the basic demographics section of the questionnaire.
6. Duplicate names are detected.
7. Person order misalignment is detected.
8. The keyer notices a problem during keying and flags the case as eligible for process review.

¹ A Data Defined Person is a person record that includes sufficient data to qualify as a person. This means answers are provided for at least two questions in the basic section (not including name) or three questions in the detail section (not including name).

Once a case has been flagged as eligible and sent to the process review operation, there are five possible actions. A process review analyst can take more than one action on any case:

1. **No action** - This occurs when there is a doubt about what should be done on a case that was flagged for process review or when no error is found requiring correction.
2. **Cancel** - A person record that was created in error (for example when a respondent entered stray marks on the form) or a person determined to be a duplicate of another person is removed.
3. **Move name** – The respondent name is moved from the cover to the basic demographics section when the respondent name is listed in the detailed population section but not in the basic demographics section.
4. **Create** – The respondent name that should have been included in the basic and detail sections is added when it only appears on the cover.
5. **Connect data** – The basic demographic and detailed population data of a person are linked when names in the basic demographic and detailed population sections match but are listed in a different order.

Research Questions

The goal of this process review evaluation was to answer the following questions:

1. What are the monthly workloads for the process review operation?
2. Why are cases failing and what actions are process review analysts taking to fix them?
3. Are there specific edit failure reasons that have a consistent process review action suggesting they could be automated?
4. Can an automated process review operation be designed that produces the same results as the current manual process?

Methodology

Two datasets from two different years (2008, 2009) were created in order to address the research questions. Because the focus of this evaluation is on workloads, the data used are all unweighted. Besides a regular cross tabulation using Microsoft Excel, data mining techniques using SAS Data Mining software were also employed.

Results

What are the monthly workloads for the process review operation?

To assess the monthly workloads, the two datasets were analyzed by monthly sample panel. A total of 48,342 cases were sent to process review in 2008 (3.4 percent of all mail returns) and 46,746 cases were sent in 2009 panels (3.6 percent of all mail returns). Tables 1 and 2 summarize the monthly process review workloads in 2008 and 2009.

The average number of cases sent to the process review operation each month was 4,028 in 2008 and 3,895 in 2009. Failure rates are included in Tables 1 and 2 based on the percentage of the data captured mail returns identified for process review. While the monthly workloads have declined slightly over the two year period, the rate of failure has steadily increased from 3.3 percent to 3.6 percent. The drop in workloads can be explained by the drop in the total mail returns over the two year period. It's possible that many factors contribute to the slight increase in these edit failure rate and these numbers suggest that despite decreases in workloads; a slightly higher proportion of the mail returns were failing on monthly basis.

Table 1. Process Review Monthly Workloads - 2008

Panel	Monthly workload	Monthly mail returns	Percentage of mail returns
200801	4,058	122,482	3.3
200802	4,005	120,456	3.3
200803	3,942	119,719	3.3
200804	4,071	123,328	3.3
200805	4,093	120,102	3.4
200806	4,109	117,437	3.5
200807	4,123	117,435	3.5
200808	3,986	117,319	3.4
200809	3,980	122,313	3.3
200810	3,904	119,200	3.3
200811	3,958	116,824	3.4
200812	4,113	119,662	3.4
Total	48,342	1,436,277	3.4

Table 2. Process Review Monthly Workloads - 2009

Panel	Monthly workload	Monthly mail returns	Percentage of mail returns
200901	3,985	111,743	3.6
200902	4,096	110,369	3.7
200903	3,979	113,287	3.5
200904	3,854	111,480	3.5
200905	3,915	109,917	3.6
200906	3,860	108,038	3.6
200907	3,867	107,883	3.6
200908	3,786	107,690	3.5
200909	3,894	110,933	3.5
200910	3,897	107,867	3.6
200911	3,731	107,311	3.5
200912	3,882	108,583	3.6
Total	46,746	1,315,100	3.6

Why are cases failing and what actions are process review analysts taking to fix them?

Edit failure reasons

In 2008 and 2009, the following were the top 5 edit failure reasons that lead cases to be flagged for process review:

1. Combination of duplicate names and person order misalignment
2. Person order misalignment
3. Number of DDPs in basic demographics section is less than the number of DDPs in detailed population section
4. Respondent name does not match to any names listed in the basic demographics section.
5. A combination of 3 and 4.

Table 3 summarizes the frequencies of each of these top edit failure reasons and their occurrences of failures. The top five edit failure reasons account for over 2/3 of all failures. The top thirteen reasons for failure summarized in Table 3 include all reasons contributing 1 percent or more to the workload and describe over 93 percent of all cases that were sent to process review for both 2008 and 2009. Given that the overall workload is so small (about 3.5 percent of all mail returns), even the edit failure reason with the greatest number of failures (combination of duplicate names and person order misalignment) represented less than 1 percent of all mail returns.

Table 3. Process Review Failure Reason - 2008 and 2009²

Failure Reason (s)	Number of failures		Percent of all failures	
	2008	2009	2008	2009
Combination of duplicate names and person order misalignment	10,955	10,610	22.7	22.7
Person order misalignment	7,306	7,056	15.1	15.1
Number of DDPs (basic demographics) is less than the number of DDPs (detailed population)	6,894	6,963	14.3	14.9
Respondent name does not match any names	6,408	6,006	13.3	12.8
Combination of number of DDPs in basic section less than the number of DDPs in detail population and respondent name non match	3,337	3,598	6.9	7.7
The number of names listed in the basic demographics section is less than the number of DDPs in basic demographics	2,601	2,287	5.4	4.9
Respondent count of total number of people is less than the number of DDPs in basic demographics	2,515	2,255	5.2	4.8
Combination of the number of names on basic demographics is less than number of DDPs in the basic demographics and the respondent name doesn't match any names	1,202	1,118	2.5	2.4
Combination of respondent count of total number of persons less than the number of names listed in the basic demographics and respondent count of total number of people is less than the number of DDPs in basic demographics	1,180	1,026	2.4	2.2
Combination of number of DDPs (basic demographics) is less than the number of DDPs (detailed population), the respondent name doesn't match any names, and person order misalignment	968	951	2.2	2.0
Duplicate names	715	736	1.5	1.6
Combination of number of DDPs (basic demographics) is less than the number of DDPs (detailed population), respondent name does not match any names, duplicate names, and person order misalignment	631	607	1.3	1.3
Combination of respondent count of total number of people being less than the number of DDPs in basic demographics and the number of names listed in the basic demographics section being less than the number of DDPs in basic demographics	624	537	1.3	1.1

Process review actions

Table 4 summarizes the actions taken during the 2008 and 2009 process review operation. Process review analysts didn't take an action on 46 percent of the cases that were sent to process review in both 2008 and 2009. The analysts did however take different kinds of actions on the remaining cases where almost half of them resulted in a person being canceled. A total of 14 different actions or combinations of actions are summarized in Table 4.

² Only reasons of failure with 1 percent or more are summarized in this table.

Table 4. Process Review Actions - 2008 and 2009

Action taken	2008		2009	
	Number	Percentage of total actions	Number	Percentage of total actions
No action	22,135	45.8	21,291	45.5
Cancel	11,236	23.2	10,514	22.5
Move, connect data	5,047	10.4	4,852	10.4
Create	4,044	8.4	4,008	8.6
Connect data	3,080	6.4	3,064	6.6
Move	1,402	2.9	1,559	3.3
Cancel, Move, Connect data	616	1.3	616	1.3
Cancel, Connect data	331	0.7	382	0.8
Cancel, Create	241	0.5	238	0.5
Cancel, Move	88	0.2	89	0.2
Move, Create, Connect data	83	0.2	89	0.2
Cancel, Move, Create, Connect data	17	0.0	19	0.0
Create, Connect data	16	0.0	17	0.0
Cancel, Create, Connect data	6	0.0	8	0.0
Total	48,342	100.0	46,746	100.0

Are there specific edit failure reasons that have a consistent process review action suggesting they could be automated?

A cross tabulation of actions by edit failure reasons of in 2008 and 2009 indicated that process review analysts didn't take action on more than 94 percent of the cases that failed due to the combination of duplicate names and person order misalignment. Tables 5 and 6 summarize the edit failure reasons along with the actions that were associated with them for 2008 and 2009 respectively.

Although for many of the edit failure reasons process review analysts didn't take any action, there were edit failure reasons where more actions were taken compared to not taking action. For example process review analysts canceled a person on more than 93 percent of the cases where the edit failure reason was a discrepancy between the number of names in the basic and detailed sections. Process review analysts also moved and connected person's data on about 90 percent of the cases when this same edit failure reason was combined with person order misalignment and respondent name not matching any name. There were edit failure reasons for which there was no dominant action/inaction. For example when the edit failure reason was the combination of discrepancy between the number of names in the basic and detailed section along with respondent's name not matching any names, process review analysts created persons on 45 percent of cases, moved person's name on 40 percent of cases, and took no action on about 5 percent of cases.

Table 5. Process Review Edit Failure Reasons and Actions - 2008

Edit failure reason (s)	No action	Connect Data	Create	Move	Move, Connect	Cancel	Cancel, create	Other actions	Total
Duplicate names and person order misalignment	94.7	0.7	0.0	0.0	0.9	2.2	0	1.5	10,955
Person order misalignment	16.2	37.5	0.0	0.0	44.9	0.4	0.0	1.0	7,306
Number of DDPs (basic demographics) is less than the number of DDPs (detailed population)	23.0	1.0	1.2	0.8	1.4	70.5	0.1	2.1	6,894
Respondent name does not match any names	59.0	0.1	36.6	0.1	3.2	0.4	0.2	0.6	6,408
Combination of number of DDPs (basic demographics) is less than the number of DDPs (detailed population) and respondent name does not match any names	4.4	0.4	46.4	39.6	3.1	2.0	1.7	4.1	3,337
The number of names listed in the basic demographics section is less than the number of DDPs in basic demographics	73.8	0.2	0.0	0.0	0.3	25.3	0.0	0.4	2,601
Respondent count of total number of people is less than the number of DDPs in basic demographics	5.7	0.1	0.0	0.0	0.1	93.5	0.1	0.6	2,515
The number of names on basic demographics is less than number of DDPs in the basic demographics and the respondent name doesn't match any names	87.9	1.0	0.9	0.1	1.7	2.7	4.0	5.7	1,202
Combination of respondent count of total number of persons less than the number of names listed in the basic demographics and respondent count of total number of people is less than the number of DDPs in basic demographics	69.3	0.2	0.1	0.6	0.2	29.1	0.1	0.6	1,180
Combination of number of DDPs (basic demographics) is less than the number of DDPs (detailed population), the respondent name doesn't match any names, and person order misalignment	0.7	0.8	0.2	0.5	90.3	0.0	0.2	7.5	968
Duplicate names	88.0	.07	0.3	0.0	1.3	4.9	2.9	4.8	715
Combination of number of DDPs (basic demographics) is less than the number of DDPs (detailed population), respondent name does not match any names, duplicate names, and person order misalignment	2.7	0.1	0.3	0.3	0.5	74.1	9.5	22.0	631
Respondent count of total number of people is less than the number of DDPs in basic demographics and the number of names listed in the basic demographics section is less than the number of DDPs in basic demographics	9.8	0.0	0.0	0.0	0.3	88.3	0.6	1.6	624

Table 6. Process Review Edit Failure Reasons and Actions - 2009

Edit failure reason (s)	No action	Connect Data	Create	Move	Move, Connect	Cancel	Cancel, create	Other action	Total
Duplicate names and person order misalignment	94.1	0.9	0.0	0.0	0.9	2.8	0.2	1.3	10,610
Person order misalignment	15.8	38.1	0.0	0.0	44.3	0.5	0.0	1.3	7,056
Number of DDPs (basic demographics) is less than the number of DDPs (detailed population)	26.1	1.0	1.1	0.9	1.2	67.7	0.2	2.0	6,963
Respondent name does not match any names	58.1	0.0	37.1	0.3	3.6	0.4	0.3	0.5	6,006
Combination of number of DDPs (basic demographics) is less than the number of DDPs (detailed population) and respondent name does not match any names	4.9	0.6	45.3	40.9	2.4	1.9	1.7	4.0	3,598
The number of names listed in the basic demographics section is less than the number of DDPs in basic demographics	5.9	0.0	0.1	0.0	0.0	93.4	0.1	0.6	2,287
Respondent count of total number of people is less than the number of DDPs in basic demographics	78.2	0.1	0.0	0.0	0.2	20.9	0.1	0.6	2,255
The number of names on basic demographics is less than number of DDPs in the basic demographics and the respondent name doesn't match any names	69.1	0.1	0.3	0.0	0.2	29.5	0.0	0.8	1,118
Combination of respondent count of total number of persons less than the number of names listed in the basic demographics and respondent count of total number of people is less than the number of DDPs in basic demographics	86.4	1.5	1.6	0.0	1.8	3.4	3.5	5.3	1,026
Combination of number of DDPs (basic demographics) is less than the number of DDPs (detailed population), the respondent name doesn't match any names, and person order misalignment	0.7	0.6	0.1	0.1	89.6	0.1	0.0	8.8	951
Duplicate names	86.4	1.5	0.0	0.0	0.5	6.4	2.5	5.2	736
Combination of number of DDPs (basic demographics) is less than the number of DDPs (detailed population), respondent name does not match any names, duplicate names, and person order misalignment	2.8	0.6	0.1	0.1	1.0	74.5	6.7	20.9	607
Respondent count of total number of people is less than the number of DDPs in basic demographics and the number of names listed in the basic demographics section is less than the number of DDPs in basic demographics	14.1	0.2	0.2	0.0	0.6	81.9	1.0	3.0	537

From the data in Tables 5 and 6 there were limited edit failure reasons that we felt had a consistent enough response to automate.

Can an automated process review operation be designed that produces the same results as the current manual process?

Data mining and other statistical techniques were used to determine the associations between the different failure reasons (alone or in combination with others) and the actions taken during process review. The interest was on cases that were flagged for process review where process review analysts did not take any action. If these cases were found to have a common failure reason, they could be eliminated from the process review workload.

Rules generated from Decision Tree in SAS Data Mining using edit failure reasons as input variables and action as the target showed the same pattern as in Table 5 and 6 when it comes to edit failure reasons and actions that are taken in the process review operation. These rules indicated that in both 2008 and 2009 process review analysts didn't take action on over 93 percent of the cases where the combination of duplicate names and person order misalignment caused a case to be flagged. This category of duplicate name and person order misalignment constitute almost 20 percent of all cases that were sent to the process review operation in both 2008 and 2009. When each of these edit failure reasons was the sole cause for a case to be sent to process review, process review analysts took different kinds of actions. This indicated that the combination of both of these edit failure reasons leads to no action but it is not the case when each failure reason appears separately.

Although results from these research questions indicate that there are few consistent actions across edit failure reasons, a complete automation of the process review operation might not be advisable given the resources that would be required to develop the automation along with the fact we can't automate everything that process analysts do. A statistical analysis of the effects of edit failure reasons to action indicates that there are other factors that play a big role in the specific action taken during process review. Given the relatively small workload of the process review operation, we do not recommend that attempts be made to automate the complex decisions made by the analysts.

Recommendations

An analysis of 2008 and 2009 data indicate that eliminating cases where the reason for failure is a combination of duplicate names and person order misalignment would cut down the process review workload by 23 percent. This change would also eliminate roughly 50 percent of the cases on which process review analysts spent time but didn't take any action. Although there could be other areas of the process review operation to automate, focusing on the category where analysts don't take any action seems to be the safest and easiest way to introduce some automation into the process review operation. Process review analysts follow specific rules that are based on eight edit failure reasons in making decisions. It became apparent during observation of the process review operation that there are other important factors beyond the specified rules that contribute to actions taken during the operation. The data analysis also confirmed that the edit failure reasons are not the only factors contributing to actions that process review

analysts take. Some of the unwritten rules that process review analysts mentioned they rely on during decision making are people's age, people's suffixes (Jr., Sr., etc), career, military service, and others. Discounting this information and automating actions that process review analysts take might introduce error.