



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

07/31/2012

2012 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT  
MEMORANDUM SERIES #ACS12-RER-25

MEMORANDUM FOR ACS Research and Evaluation Steering Committee

From: James B. Treat (**signed 07/31/2012**)  
Chief, American Community Survey Office

Prepared by: Stephanie Baumgardner  
American Community Survey Office

Subject: An Evaluation of the Transition to the iCADE Data Capture System  
– Overview Report

Attached is the final American Community Survey Research and Evaluation report on the Evaluation of the Transition to the iCADE Data Capture System – Overview Report. We are conducting a comprehensive evaluation of the ACS experience with the iCADE system, the capture system first used by the ACS in 2007. This paper is the first in a series of papers on our findings. This report defines all of the steps involved in the iCADE system and presents the 2011 annual workloads associated with each of the steps requiring human intervention. This high-level view of the iCADE system provides basic information regarding the flow of forms through the data capture process.

If you have any questions about this report, please contact Stephanie Baumgardner at (301) 763-5893 or David Raglin at (301) 763-4226.

Attachment

cc:  
ACS Research and Evaluation Team



# An Evaluation of the Transition to the iCADE Data Capture System – Overview Report

FINAL REPORT

---

**Intentionally Blank**

# Table of Contents

- Introduction..... 1
- Background..... 1
- Research Questions..... 3
- Methodology..... 3
  - Data Sources ..... 3
  - Definitions..... 3
- Results..... 4
  - Laser sorter..... 5
  - Batching ..... 6
  - Scanning..... 7
  - Manual registration ..... 8
  - Exception Review ..... 9
  - Key From Image ..... 10
  - Key From Image Quality Assurance..... 11
  - Process Review ..... 12
  - Handoff to Sponsor ..... 13
- Conclusions..... 14

---

**Intentionally Blank**

## Introduction

The American Community Survey (ACS) uses three sequential modes of data collection, starting with mail. The data capture of ACS mail forms changed in 2007 from a key-from-paper (KFP) system to a key-from-image (KFI) system known as iCADE (Integrated Computer Assisted Data Entry). The iCADE system was originally developed for use in the Economic Directorate and has expanded to support many domestic and international surveys and censuses. The ACS transition to iCADE was necessary because the fully implemented ACS stressed the KFP system and resulted in a sizeable backlog in the keying of ACS questionnaires. In addition, the KFP system was becoming obsolete and could not be fully supported. While necessary, this transition was also expected to decrease data capture backlogs resulting in enhanced timeliness and lower costs.

We are currently conducting a comprehensive evaluation of the ACS experience with this new capture system. This paper is the first in a series of papers on our findings. This report defines all of the steps involved in the iCADE system and presents the 2011 annual workloads associated with each of the steps requiring human intervention. This high-level view of the iCADE system provides basic information regarding the flow of forms through the data capture process.

## Background

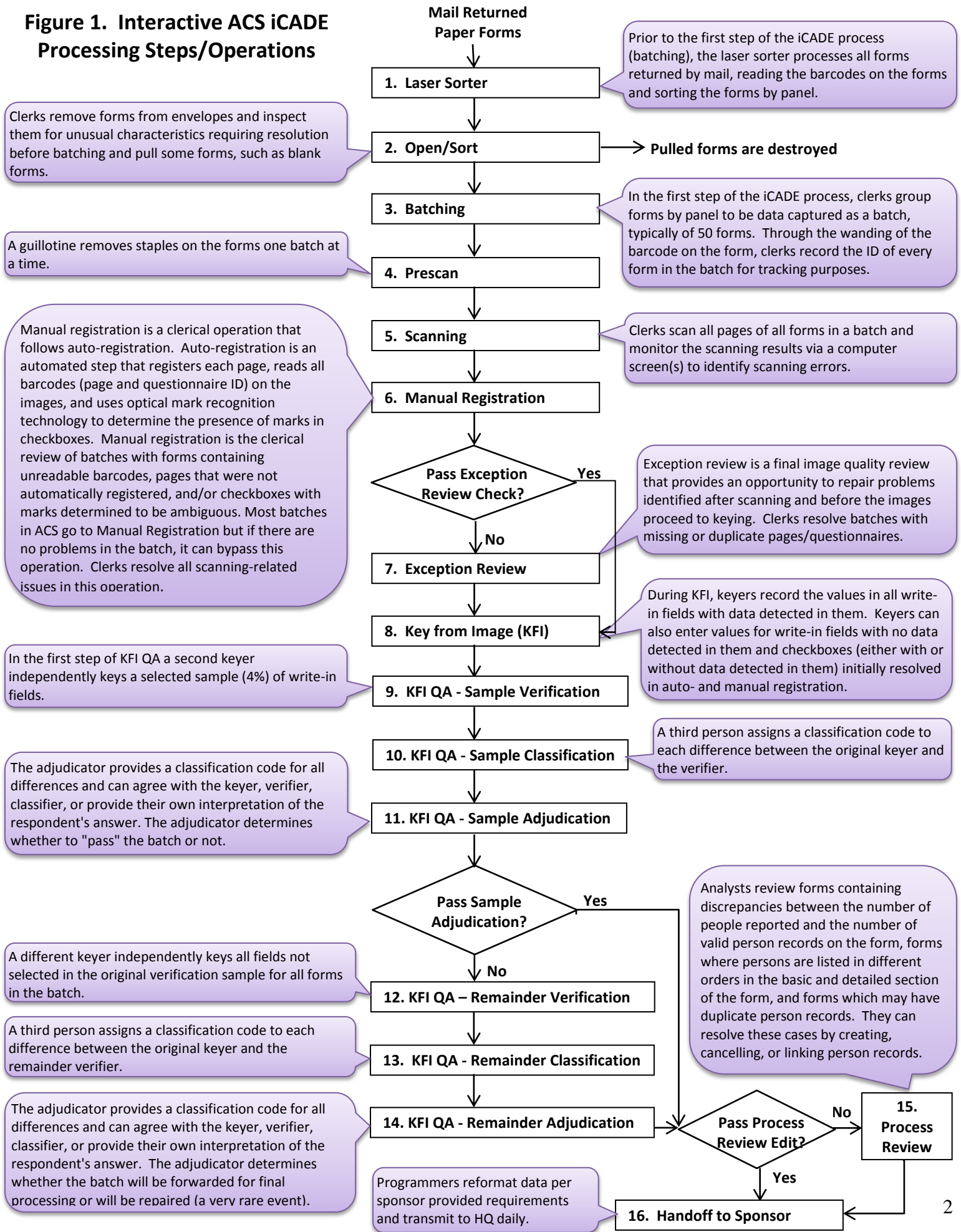
The iCADE system is comprised of 15 steps which require interaction with a clerk/analyst. They are Batching, Prescan, Scanning, Manual Registration, Exception Review, KFI, KFI Quality Assurance (QA) Sample components (verification, classification, adjudication), KFI QA Remainder components (verification, classification, adjudication), Process Review, and Handoff to Sponsor<sup>1</sup>. There are also automated processes that occur between many of the interactive steps such as Auto Registration and Batch Completeness Analysis that examine certain aspects of each batch and ready the batch for the next step of the process.

Figure 1 shows the basic flow of the paper mail forms through the interactive steps of the iCADE system. The first two steps in the process (laser sorter and open/sort) are not part of the iCADE system but are necessary for the ACS forms to be ready for the first step of iCADE, batching. Once batching is complete, all forms proceed with their respective batches through the remaining iCADE operations.

---

<sup>1</sup> The iCADE system includes an interactive step called OMA (Optical Mark Analysis) that occurs between Manual Registration and Exception Review in the iCADE system; this step is not currently in use for ACS.

**Figure 1. Interactive ACS iCADE Processing Steps/Operations**





## Research Questions

This research focuses on the data capture of housing unit questionnaires. For all interactive operations of the iCADE system plus the laser sorter step, this report answers the following questions using 2011 data:

- What were the total workloads for the year?
- What were the average daily workloads?
- What were the maximum daily workloads?
- How did the daily workloads vary throughout the year?

## Methodology

### Data Sources

Data used to answer the research questions come from the 2010 and 2011 ACS control files and special files compiled by the Economic Programming Division (EPD) for 2010 and 2011. We used the control files to determine the daily number of forms that were processed through the laser sorter in 2011 and of these forms, the daily number of forms for which data were received after being processed through iCADE (Handoff to Sponsor). We used the EPD files to determine the daily number of batches and forms processed by each of the iCADE operations.

### Definitions

To provide a basic overview of the flow of work through the iCADE system this report focuses on the **2011 mail return workloads**. The 2011 mail return workloads include only those mail returned forms processed by the laser sorter in 2011 which corresponds to the universe of mail forms used in the production of the 2011 ACS estimates. The 2011 mail return workloads provide a fixed universe which clearly shows the flow of forms through the iCADE system.

These workloads differ from the **2011 total iCADE workloads** in that they do not fully reflect the work being done in calendar year 2011. Forms that were processed by the laser sorter in 2010 and completed iCADE processing in 2011 are not part of the 2011 mail return workloads but are part of the 2011 total iCADE workload. Forms that were processed by the laser sorter in late 2011 and completed iCADE processing in 2012 are part of the 2011 mail return workloads but they are not part of the 2011 total iCADE workload for those operations that were completed in 2012. The total iCADE workloads by definition are better when analyzing costs, workloads and production and future reports will focus on the total iCADE workloads rather than these mail return workloads.

In this report for each of the iCADE operations, we chose to assign each batch to a single day's workload. This simplified calculations and resulted in more meaningful summaries. We defined daily workloads as the batches that started an operation on a specific day. Specifically, we did not account for batches that span over multiple days (we used the start date). For those batches that went through one or more operations multiple times, we used the date associated with the

last occurrence in each operation.

To calculate daily average workloads for each of the operations, we divided the total number of forms/batches processed in an operation by the number of days worked for that operation. We defined a work day as any day that had at least one batch assigned to the workload for a given operation. By this definition, if a single batch was keyed on a Saturday, that day would be included in the denominator since keying work occurred on that day even though it was not an official workday. In the 2011 data used for this report, the scanning operation was the only operation that processed batches on a Saturday or Sunday. Twenty-eight batches were scanned on two Sundays in 2011.

In 2011 the ACS sample increased from about 2.9 million to about 3.54 million addresses a year. This increase occurred starting with the June panel. To highlight this change we chose to display average daily workloads for January through May and June through December separately.

## Results

In 2011, the laser sorter processed more than 1.4 million forms. This is the total 2011 mail return workload and the universe for this analysis. Table 1 includes workloads by operation in terms of batches and forms. The table summarizes the total 2011 workloads, the average daily workloads (for January – May and June – December), and the maximum daily workloads for each operation. Note that the typical batch has 50 forms.

**Table 1 Workload Summary by iCADE Operation – 2011 Mail Return Workload**

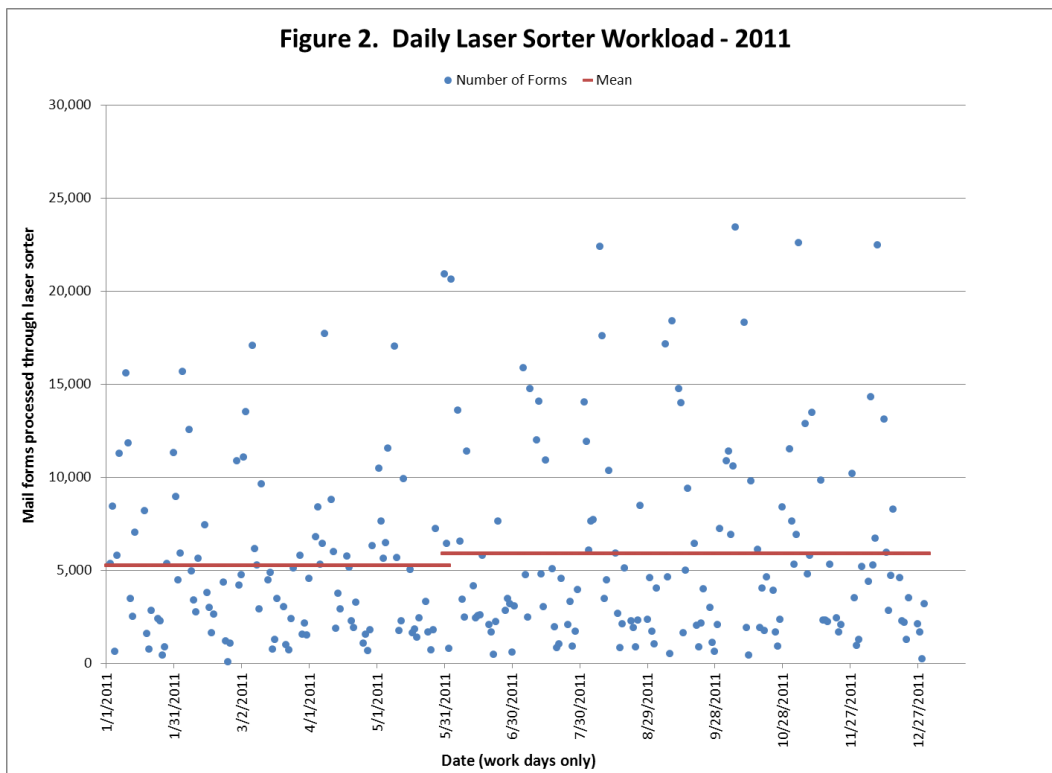
| Operation                   |                | Total Workload |         | Average Daily Workload (January-May) |         | Average Daily Workload (June-December) |         | Maximum Daily Workload |         |
|-----------------------------|----------------|----------------|---------|--------------------------------------|---------|--|---------|------------------------|---------|
|                             |                | Forms          | Batches | Forms                                | Batches | Forms                                  | Batches | Forms                  | Batches |
| Laser Sorter                |                | 1,412,042      | NA*     | 5,281                                | NA*     | 5,910                                  | NA*     | 23,448                 | NA*     |
| Batching                    |                | 1,399,673      | 28,436  | 5,017                                | 103     | 5,972                                  | 121     | 14,024                 | 283     |
| Scanning                    |                | 1,399,673      | 28,436  | 4,683                                | 96      | 6,166                                  | 125     | 14,309                 | 289     |
| Manual Registration         |                | 1,399,657      | 28,426  | 4,769                                | 97      | 6,212                                  | 126     | 15,806                 | 319     |
| Exception Review            |                | 73,543         | 1,517   | 262                                  | 5       | 315                                    | 6       | 1,250                  | 26      |
| Keying                      |                | 1,399,673      | 28,436  | 4,618                                | 94      | 6,119                                  | 124     | 10,120                 | 203     |
| KFI Sample QA Components    | Verification   | 1,399,673      | 28,436  | 4,605                                | 94      | 6,088                                  | 123     | 9,870                  | 198     |
|                             | Classification | 1,399,673      | 28,436  | 4,648                                | 95      | 6,090                                  | 123     | 7,600                  | 154     |
|                             | Adjudication   | 1,399,673      | 28,436  | 4,648                                | 95      | 6,090                                  | 123     | 7,600                  | 154     |
| KFI Remainder QA Components | Verification   | 91,813         | 2,249   | 327                                  | 8       | 389                                    | 9       | 874                    | 23      |
|                             | Classification | 91,813         | 2,249   | 326                                  | 8       | 387                                    | 9       | 818                    | 23      |
|                             | Adjudication   | 91,813         | 2,249   | 326                                  | 8       | 387                                    | 9       | 904                    | 25      |
| Process Review              |                | 54,963         | 23,993  | 183                                  | 80      | 237                                    | 104     | 351                    | 134     |
| Handoff to Sponsor          |                | 1,397,908      | 28,436  | 4,618                                | 95      | 6,048                                  | 123     | 10,791                 | 216     |

\* Not Applicable

Source: U.S. Census Bureau, 2011 American Community Survey Operational Data

## Laser sorter

Prior to the first step of the iCADE process (batching), the laser sorter processes all forms returned by mail by reading the barcodes on the forms and sorting the forms by panel. On average, the laser sorter processed about 5,300 forms daily in early 2011 and about 5,900 forms daily from June through December with a maximum of more than 23,000 forms processed on one day in 2011. Figure 2 shows the daily laser sorter workload throughout 2011. The laser sorter processes all mail returned forms on the day the forms are received in the processing center (forms received on Saturday are processed on Monday) so these daily workloads are a direct reflection of mail response patterns and/or the United States Postal Service delivery processes. While it appears that there are a few days each month where more than 10,000 forms are processed through the laser sorter, there are many days where 5,000 or fewer forms are processed by the laser sorter. The days where more than 10,000 forms are processed generally coincide with the first week of the month when the highest numbers of forms are typically returned. The effect of the sample expansion in June is evident by the increase in days where the number of forms processed exceeds 15,000.



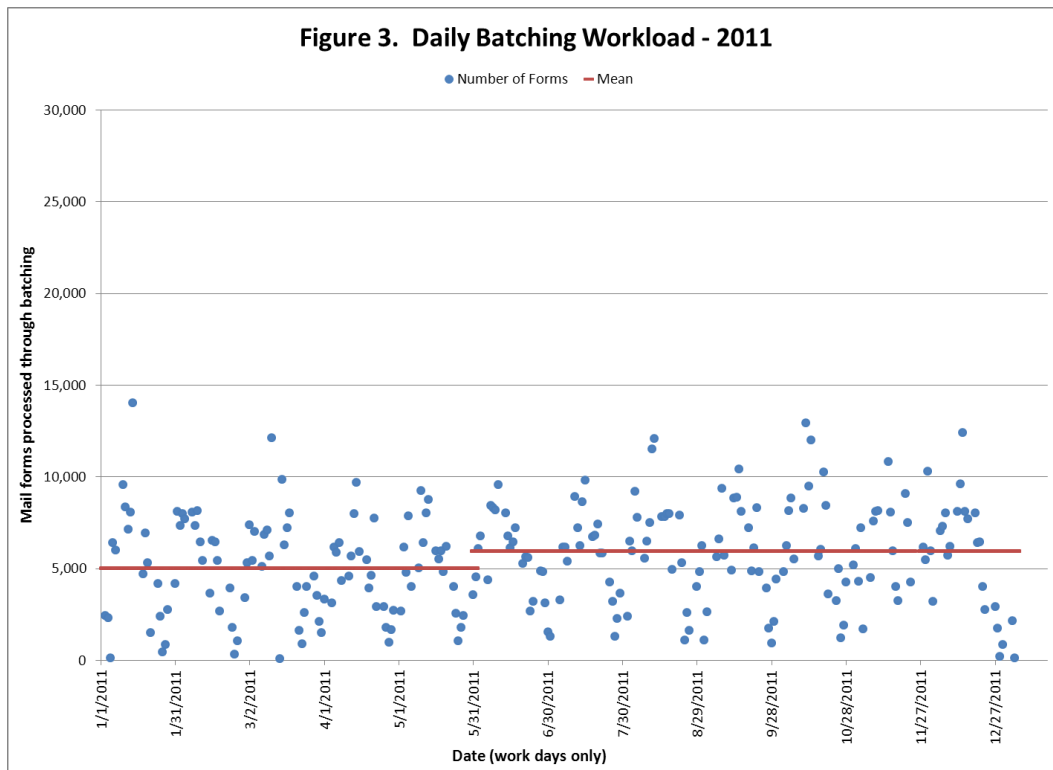
Source: U.S. Census Bureau, 2011 American Community Survey Operational Data

## Batching

In this first step of the iCADE process, clerks group forms by panel to be data captured as a batch, typically 50 forms. Through the wand of the barcode on the form, clerks record the identification number associated with every form in the batch for tracking purposes. Between the laser sorter and batching operations, the clerks in the open/sort operation pull forms for a variety of reasons (i.e., blank forms, blatantly fictitious data) and that is represented in the reduction (by about 12,400) in total forms batched.

In 2011 the batching operation created 28,436 mail return batches that moved through the remaining iCADE operations. Clerks batched an average of about 5,000 forms each day into 103 batches from January through May. After the sample expansion in June, clerks batched an average of about 6,000 forms a day into 121 batches. A daily high of 283 batches occurred in mid-January and may indicate a backlog in these 2011 batches as resources were focused on completing processing of all remaining 2010 forms. In addition, prior to May 2011, the batching date (not the laser sorter date) was used as the means of indicating the receipt a form. Therefore, prior to May 2011, the daily workloads for this operation generally increased a day or two before the cut for the second mailing and this monthly increase is most evident by the high number observed in mid-January.

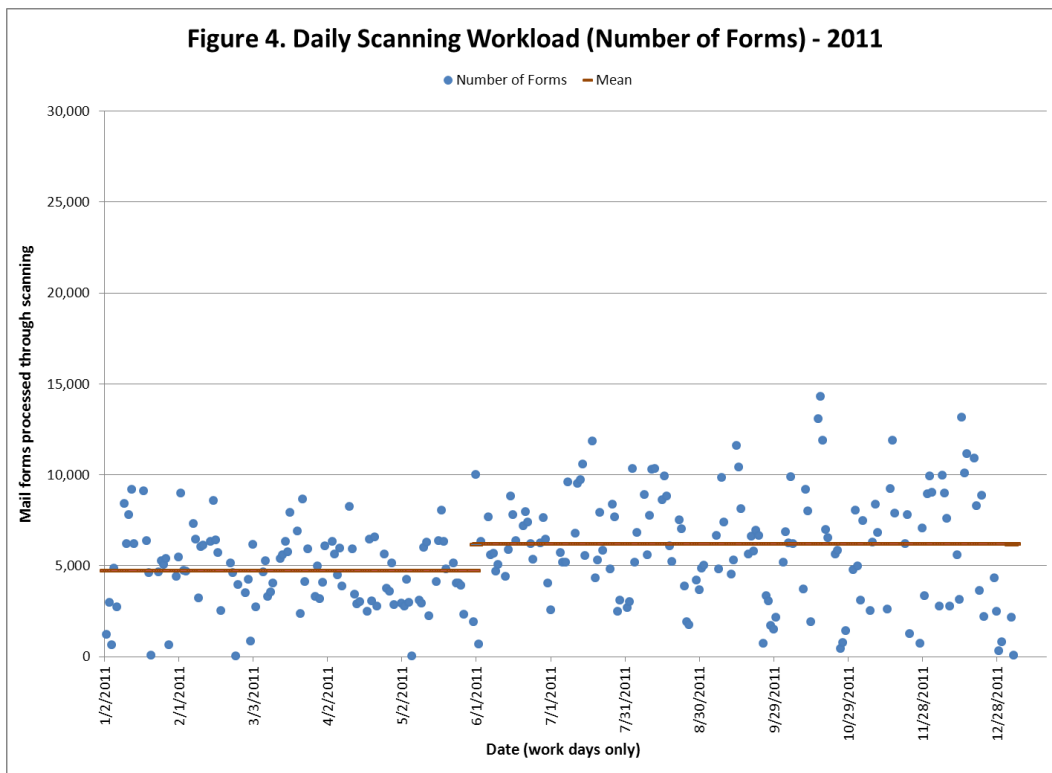
Figure 3 shows the daily variability in the batching workloads, which is a consequence of the workloads emerging from the laser sorter operation. Prior to the sample expansion in June, there appears to be a consistent pattern in the number of forms batched daily. After the sample expansion, this pattern is less apparent, especially from August through December.



Source: U.S. Census Bureau, 2011 American Community Survey Operational Data

## Scanning

After batching, a machine known as the guillotine removes the staples on the forms in the prescan operation. After prescan, the batch is ready for the scanning operation. Clerks scan all pages of all forms in a batch and monitor the scanning results via a computer screen(s) to identify scanning errors. Prior to the sample expansion in June, the average daily number of batches scanned (96) was lower than the average daily number of batches batched (103). This could be due to the push to batch forms prior to the cut for the second mailing when the batch date was the source used to indicate the receipt of the form. After the sample expansion, the average daily number of batches scanned (125) was higher than the average daily number of batches batched (121). This appears to be due to larger backlogs in the scanning operation for the first part of 2011. In addition, the change to using the laser sorter to indicate the receipt of a form likely contributed to a lower average daily number of forms batched as the batching clerks no longer had to be concerned about the cut date for the second mailing. Figure 4 shows how the daily scanning workload changed across time in 2011 and early 2012. The first 5 months of the year show a more controlled operation with a workload that remains somewhat consistent from day to day. After June, there is much more variability in the number of forms scanned on any given day and many more days with daily scanning workloads in excess of 10,000 forms.



Source: U.S. Census Bureau, 2011 American Community Survey Operational Data

## Manual registration

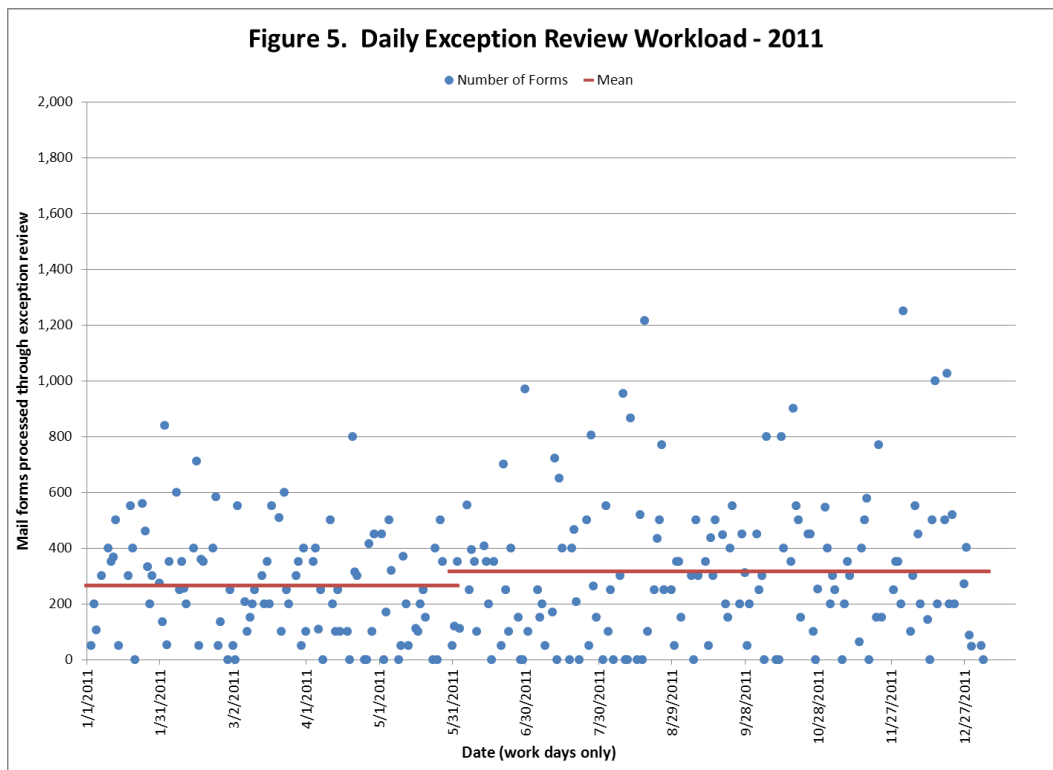
Manual registration is a clerical operation that follows auto-registration. Auto-registration is an automated step that registers each page, reads all barcodes (page and questionnaire ID) on the images, and uses optical mark recognition technology to determine the presence of marks in checkboxes. Manual registration is the first iCADE operation to examine image quality. This clerical review corrects issues not resolved in auto-registration including batches with forms containing unreadable barcodes, pages that were not automatically registered, and/or checkboxes with marks determined to be ambiguous. Most batches in ACS go to manual registration but if there are no problems in the batch, it can bypass this operation.

Table 1 shows that all but 10 of the 2011 batches went to this operation. The ten batches that did not go to this operation contained only 16 forms total. Given the large number of checkboxes on the ACS form (for a batch of 50 forms, there are 61,400 checkboxes), it is not surprising that for most batches at least some values are ambiguous and require human intelligence to determine the respondent's intent. While a chart depicting the daily workload for this operation is not shown here, it generally resembles the daily scanning workload chart in Figure 4 except for two days (in mid July and mid August) in which the manual registration operation processed more than 15,000 forms.

## Exception Review

Exception review is a final image quality review that provides an opportunity to repair problems identified after scanning and before the images proceed to keying. Batches require clerk intervention in the exception review operation only if the previous automated process indicates that there may be a duplicate or missing questionnaire or page. In 2011, only about 5 percent of the batches (1,517 of the 28,436) needed resolution by the exception review operation.

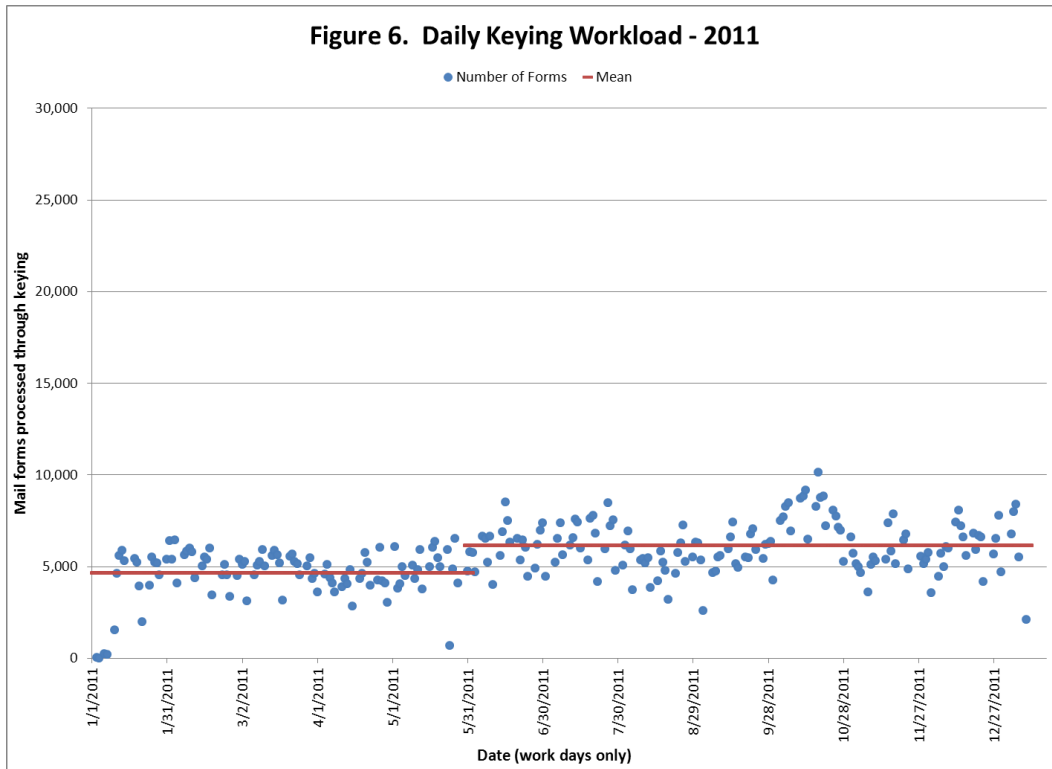
On average, this operation worked about five batches a day from January through May and about six batches a day after the sample expansion in June. For one day in 2011, this operation worked 26 batches, the highest daily number worked in 2011. Figure 5 shows the daily workloads for the exception review operation for 2011 and early 2012. They average about 300 forms. Please note the different scale for the y-axis as compared to previous figures. Similar to previous operations, the number of forms processed daily prior to the sample expansion appears to have less variability than the daily workloads after the sample expansion. Unlike the workloads for previous operations, these workloads vary dramatically with many high and many low daily workloads. Since this operation essentially performs a quality assurance of the scanning operation, the high workload days are likely due to problems with a scanner or scanner operator and the low workload days indicate that the scanning operation ran smoothly. There were 32 days in 2011 where exception review did not work any batches.



Source: U.S. Census Bureau, 2011 American Community Survey Operational Data

## Key From Image

After exception review, all batches then proceed onto the keying operation where keyers record the values in all write-in fields with data detected in them. Keyers can also enter values for write-in fields with no data detected in them and checkboxes (either with or without data detected in them) initially resolved in auto- and manual registration. On average in 2011, keyers completed 94 batches daily from January through May and 124 batches daily after the sample expansion in June. Figure 6 shows how the daily keying workloads changed across time in 2011 and early 2012. As was true with the figures on preceding operations, the keying workloads show more variability after the sample expansion in June. However, these daily workloads do show a lower level of variation from day to day than previous operations suggesting a more controlled operation. In October, there is a spike in the number of batches keyed daily with keyers completing over 200 batches on one day. This spike is the result of the keying operation clearing the backlog in preparation for the internet test that was to occur in November.



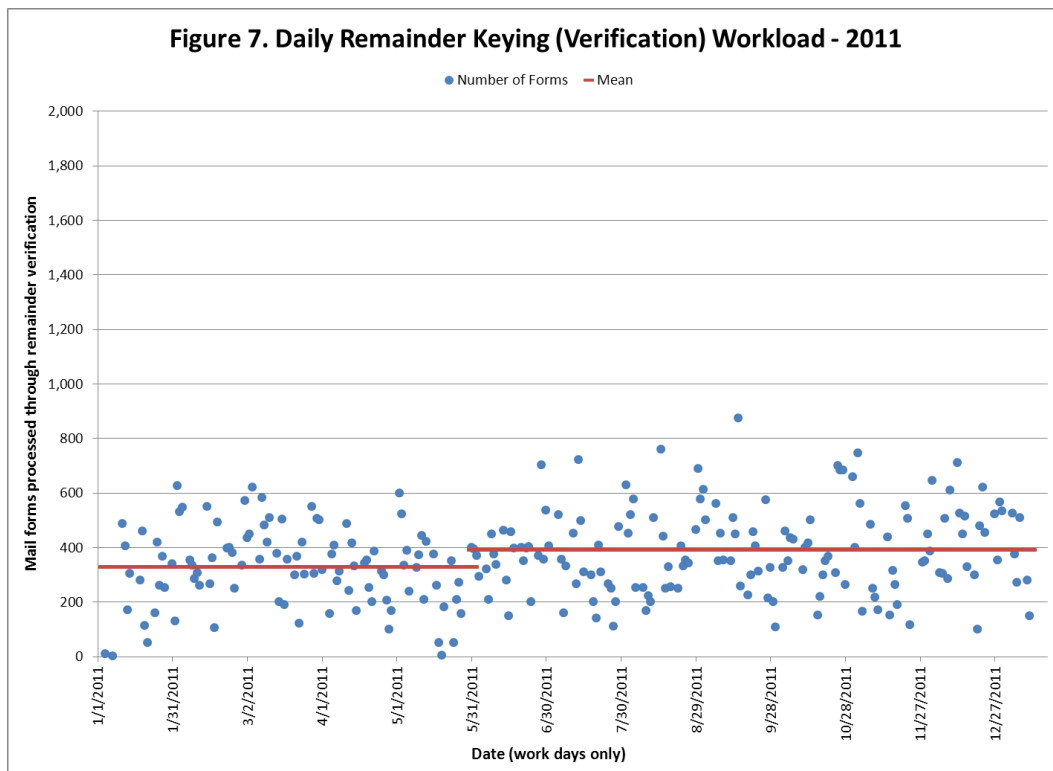
Source: U.S. Census Bureau, 2011 American Community Survey Operational Data



## Key From Image Quality Assurance

After keying, all batches progress onto the key from image (KFI) sample quality assurance (QA) components. In this series of steps, verifiers independently key a selected sample (4%) of write-in fields, classifiers assign classification codes to each difference between the original keyer and verifier, and then, based on the sample QA verification and classification operations, adjudicators determine whether the batch “passes”. On average, these operations completed about 95 batches daily from January through May and 123 batches daily from June through December. These averages are very much in line with the daily keying averages suggesting that keyers/clerks complete these sample QA components shortly after the initial keying of the batch.

The KFI sample QA adjudicator determines if a batch requires the remainder QA operations based on the accuracy of the original keyer’s entries. Similar to the sample QA components, the remainder QA operations consist of verification, classification, and adjudication steps. In the remainder QA operations, keyers/clerks key all fields not selected for the sample QA on all forms in the remainder verification step. After verification, remainder classifiers assign classification codes to each difference between the original keyer and remainder verifier, and then remainder adjudicators determine whether the batch is eligible for final processing. Of the 2011 mail return workload, about 8 percent of batches required KFI remainder QA operations. On average, 8 of the batches keyed daily pre-sample expansion and 9 of the batches keyed daily post expansion failed the sample QA and required re-keying of the batch. Figure 7 shows how the daily remainder verification keying workloads changed across time in 2011 and early 2012. Please note the different scale for the y-axis as compared to previous figures. In general, this operation shows the same trend of more variability after the sample expansion just as previous operations exhibited.

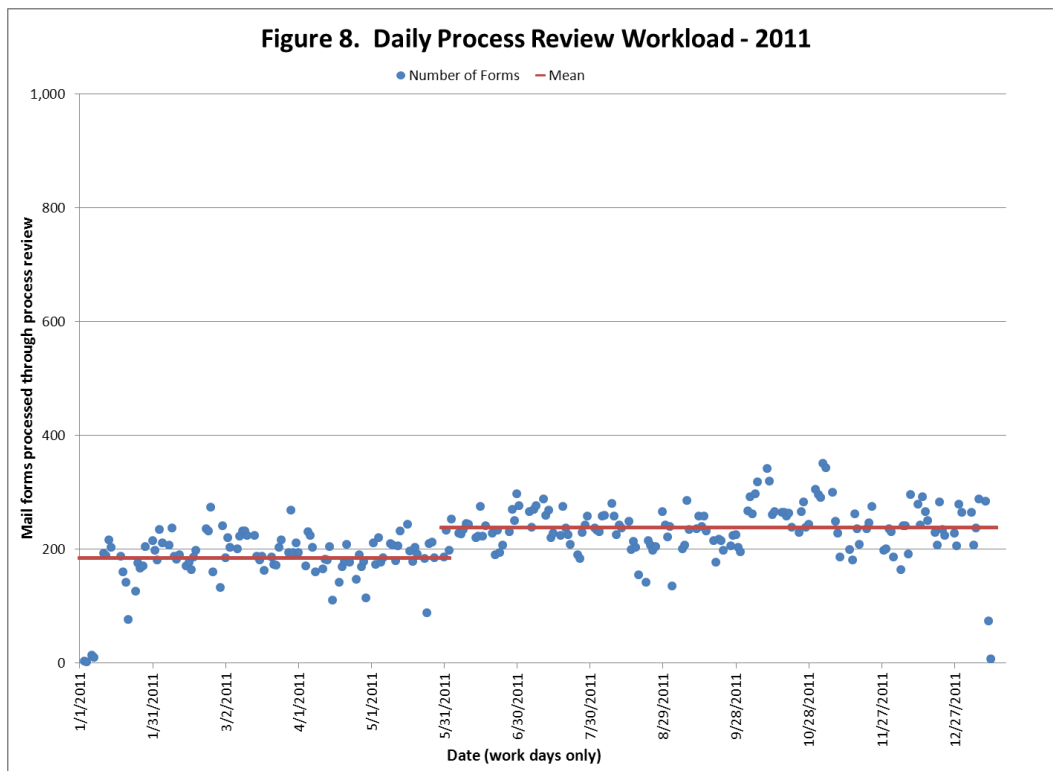


Source: U.S. Census Bureau, 2011 American Community Survey Operational Data

## Process Review

Once a batch has been through all keying QA operations, the next step of the iCADE system is an operation named process review. Process review is a computer-assisted clerical operation designed to look at cases flagged by an automated edit where the captured data suggest that the respondent may not have completed the form as expected. By examining the images of these forms, the process review analyst is able to clean up erroneous data, cancel duplicate person records, create person records, and connect 'basic' and 'detail' sets of data for the same person when they are not in the correct order.

For both the pre- and post sample expansion time periods in 2011, about 84 percent of the batches went to this operation but only about four percent of all forms required review. For this operation, it is better to look at the number of forms worked because generally the process review automated edit flags only a few forms per batch for review in the process review operation. Figure 8 shows how the daily process review workloads changed across time in 2011 and early 2012. Please note the different scale for the y-axis as compared to previous figures. Prior to the sample expansion in June, the process review operation processed an average of 183 forms daily. Following the sample expansion, this operation processed an average of 237 forms daily. The days where the workload is near zero correspond to when this operation first processed or finished processing forms contained within the 2011 mail return universe (forms processed through the laser sorter in 2011) and do not indicate the total iCADE workload for this operation on those days. Aside from the peaks seen in October (likely the result of the keying operation clearing their backlog), the daily workloads post sample expansion do not appear to be as varied as some of the other operations' daily workloads post sample expansion.

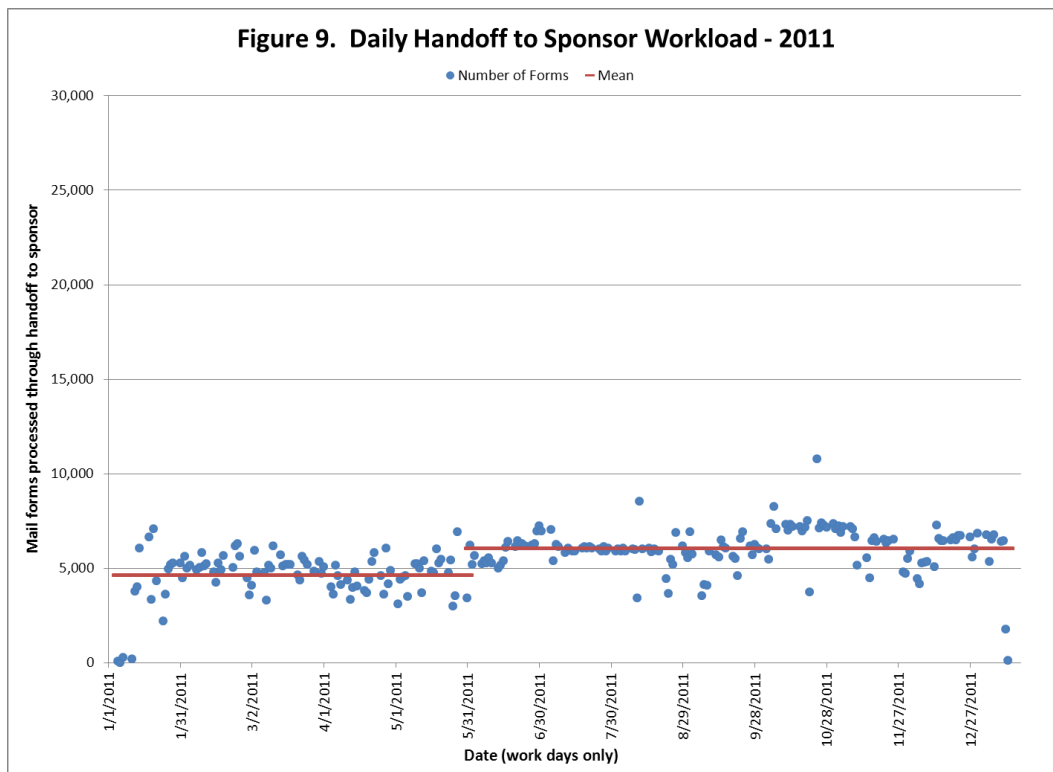


Source: U.S. Census Bureau, 2011 American Community Survey Operational Data

## Handoff to Sponsor

After a batch has been through all the applicable iCADE processes, programmers in NPC reformat the resulting data per sponsor provided requirements and transmit those data to headquarters daily. For this report, the control file is the data source that indicates the receipt of data at headquarters. The 2011 mail return workload consists of 28,436 batches that at the batching step contained 1,399,673 total forms. By the time the data were transmitted to headquarters, these 28,436 batches contained 1,397,908 total forms; 1,765 forms fewer than were batched in these batches. Forms that are 'killed' during the iCADE process can explain this difference. Forms are 'killed' due to missing pages, extra pages, pages out of order, forms that are batched but not scanned, and if annotations on the forms are considered vulgar. After consulting with an expert on the iCADE process, this number of 'killed' forms seems reasonable.

Prior to the sample expansion, the average daily number of batches received at headquarters was 95 while after the sample expansion, the average daily number of batches received was 123. Figure 9 shows the daily number of forms (in the 2011 mail return universe) transmitted to headquarters after completing the iCADE process in 2011 and early 2012. As in the process review operation, the days where the workload is near zero show when forms contained within the 2011 mail return universe are processed and do not indicate the total iCADE workload for this operation. The pre sample expansion time period shows a reasonable amount of variability in the daily number of forms transmitted. The post sample expansion time period shows an interesting pattern. For July and most of August, the number of forms transmitted daily varied very little. Then around October, the number of forms transmitted daily was consistently higher than the mean, which is likely the result of the keying operation clearing the backlog in preparation for the internet test that was to occur in November.



Source: U.S. Census Bureau, 2011 American Community Survey Operational Data

## Conclusions

Using the 2011 mail return workloads we described the flow of forms and batches through the iCADE system. Many of the iCADE operations process all batches while a few operations process just a subset of all batches/forms. This report provides some of the first summaries of failure rates for KFI QA, exception review, and process review. At first consideration, the workloads for the KFI remainder QA components seemed high at about 8 percent of all batches. Upon consultations with experts on QA, this rate is as expected and considered acceptable. These workload data also convey the subtleties of the process review operation in that 84 percent of batches go to the process review operation but analysts review only those forms flagged for this operation, about four percent of all forms.

The 2011 daily workload scatterplots for various operations show a common theme. The sample expansion that occurred in June resulted in about 50,000 more addresses a month receiving the ACS questionnaire. As expected, the sample increase translated to an increase in the average daily workloads for all iCADE operations. However, the daily workloads for most of the iCADE operations following the sample expansion show more variability than the workloads pre-sample expansion. This variability is most notable in the batching and scanning operations and since these two operations occur at the beginning of the iCADE process, the daily workloads for the later operations could just be a consequence of the variations in the earlier operations' workloads. Since the laser sorter daily workloads do not show a similar trend in daily workloads post sample expansion, it appears that the variability for the iCADE operations may be staffing related.

Starting in 2013, the ACS will offer respondents the ability to respond to the ACS via the Internet. As indicated from tests, the addition of the Internet option could reduce the mail returned paper forms by about one-half. The introduction of the Internet also introduces changes to the mailing schedule of paper forms that will result in fewer incoming forms in the beginning of the month and an expected peak of incoming forms later in the month, after the replacement questionnaire mailing. Given these considerable changes to ACS data collection, monitoring the processing of the paper forms and the transition to smaller workloads should be a priority to realize fully all cost savings. The scatterplots of the daily workloads shown in this report could be one tool to do this. The laser sorter processes essentially all mail returned forms on the day they are received in the processing center so one would expect that the timing change in the stream of incoming forms each month would be evident and the overall monthly workload of the laser sorter would decrease by about one-half. All other operations will have to adjust to managing smaller workloads. Producing charts depicting the daily workloads for some of the larger operations such as batching, scanning, and keying may provide one assessment into how each of the operations manage the change to the expected smaller workloads.