Independent Validation of USPS Service Performance Measurement Audit Design

Audit Compliance Review

FY2017 Quarter 1

May 26, 2017

Submitted to:
USPS

Submitted by:
ICF

Table of Contents

I.   Executive Summary

This report presents the results of our review of the USPS Internal Service Performance Measurement (SPM) for Quarter 1 (Q1) of fiscal year 2017 (FY17). This report presents the results of a compliance analysis of the audit measures designed to assess the accuracy, reliability, and representativeness of the sampling performance.

USPS continues its migration to the Informed Visibility (IV) Program, which enhances service performance measurement. IV provides comprehensive, consolidated data collection and monitoring of the main performance metrics. Unlike the old system of single piece measurement that relies on human interaction for recording when mail enters the mail stream and when it is delivered, IV provides barcoding-based random scan selection and sampling diagnostics on all mail. This new technology replaces the use of seeded mail to represent the full mail stream. Furthermore, IV utilizes census data for mail classes that utilized a manual seeding/recipient approach before, vastly increasing the volume in measurement and therefore the value of the diagnostics and scores available to the field on a daily basis.

To perform the analysis, ICF evaluated the accuracy, reliability, and representativeness of the sampling by assessing 32 audit measures. USPS provided information about each of the 32 audit measures, including the general criteria used to evaluate each measure; however, ICF used these criteria to develop more specific rules for evaluating compliance. ICF reviewed audit information to determine compliance with each measure and developed methods to examine the information provided by the USPS SPM team.

Of the 32 audit measures analyzed, ICF classified 8 measures as compliant, 12 measures as having possible issues with compliance, and 7 measures as non-compliant. We categorized five measures as not applicable because the data was not collected in the sampling scheme for FY17 Q1. Suggested remedies to bring the audit measures into compliance include the following:

- Provide useful data to determine whether the density data collected for the quarter is accurate, or if the quarter densities are proportional to the annual density measurements.
- Reduce margins of error below the maximum target levels.
- Provide transparent and supported explanations for cases of low compliance or declining compliance over time.
- Analyze first and last mile response rates by 3-digit ZIP code, delivery type, and carrier route type to determine differences in performance for under-represented groups.
- Provide, justify, and demonstrate compliance with a benchmark for First Mile or Last Mile imputation rates that defines a "limited volume for each mail type."

The body of this report provides additional detail and explanations of the compliance determinations; it also provides a prioritization of the changes we recommend for the short term and long term.

II.   Introduction

The U.S. Government Accountability Office (GAO) published a report in September 2015 that reviewed how the United States Postal Service (USPS) measures delivery performance and how the Postal Regulatory Commission (PRC) uses this information.[1] The GAO report provided several performance data findings, including:

- USPS and PRC reports on delivery performance do not provide sufficient analysis to hold the USPS accountable for meeting its statutory mission, including insufficient district-level analysis and a lack of reporting about rural delivery performance.
- The PRC has not fully assessed why USPS data are not complete and representative.
- Slightly more than half (55%) of market-dominant mail is included in the USPS measurement of on-time delivery performance.

In response to the GAO report, the PRC requested public comment on the quality and completeness of service performance data in Order No. 2791 (October 29, 2015).[2] In Order No. 3490 (August 26, 2016), the PRC provided an analysis of public comments received and required USPS to regularly provide descriptions of methodologies used to verify data accuracy, reliability, and representativeness for each service performance measure.[3]

In December 2016, the USPS issued a solicitation requesting a review of the audit plan for the Internal Service Performance Measurement (SPM) System. The audit plan emphasized three audit metrics—accuracy, reliability, and representativeness—and covered specific products, measurement phases, and major components of Internal SPM. In the spring of 2017, ICF reviewed Internal SPM results for Quarter 1 of fiscal year 2017 (FY17). This report presents the results of a compliance analysis of the audit measures designed to assess the accuracy, reliability, and representativeness of the sampling performance. The following sections detail the results of this audit review and ICF's recommendations.

III.   Evaluation Approach

ICF followed GAO standards for government auditing throughout the audit process, including those outlined in GAO's *Government Auditing Standards: 2017 Exposure Draft.* Our audit review focused on measurement results for letters and flats for the following products:

- Domestic First-Class Mail
    - Single-Piece letters and cards
    - Presort letters and cards
    - Single-Piece and Presort flats
- Standard Mail
    - High Density and Saturation letters

---

[1]U.S. Government Accountability Office, *Actions Needed to Make Delivery Performance Information More Complete, Useful, and Transparent*, September 2015, *http://www.gao.gov/products/GAO-15-756.*
[2]Postal Regulatory Commission, *Notice Establishing Docket Concerning Service Performance Measurement Data,* October 2015, *https://www.prc.gov/dockets/document/93660.*
[3]Postal Regulatory Commission, *Order Enhancing Service Performance Reporting Requirements and Closing Docket*, August 2016, *https://www.prc.gov/dockets/document/96994.*

- o High Density and Saturation flats
- o Carrier Route
- o Letters
- o Flats
- o Every Door Direct Mail-Retail flats
- Periodicals
- Package Services
  - o Bound Printed Matter flats.

The audit evaluated each of the following phases of internal measurement:

- **First Mile:** The time between the deposit of mail into a collection box or at a retail unit, for instance, and the first processing on postal equipment.
- **Processing Duration:** The time between initial processing and final processing for single-piece mail, and the time from the start-the-clock event (e.g., acceptance at a business mail entry unit) through final processing for commercial mail.
- **Last Mile**: The time between final processing and delivery for both single-piece and commercial mail.
- **Scoring and Reporting**: Review of Internal SPM processes for calculating service performance estimates and producing reports of market dominant product performance scores.
- **System Controls**: Review of business rules and administrative rights within the Internal SPM measurement processes and data recording and operating procedures for Postal personnel executing measurement processes.

The purpose of the audit is to ensure the accuracy, reliability, and representativeness of the sampling methodology and execution. These measures can be defined as follows:

- **Accuracy:** The closeness of computations of estimates to the "unknown" exact or true values.
- **Reliability:** The reproducibility and stability (consistency) of the obtained measurement estimates and scores.
- **Representativeness:** How well the sampled data reflects the overall population.

"Bias" combines accuracy and representativeness by evaluating the extent to which the performance estimates from the sample data tend to over- or under-estimate the population performance for all USPS mail.

1.   Compliance Approach

This section presents the approach that ICF followed to conduct the audit compliance review to evaluate the accuracy, reliability, and representativeness of the sampling. Specifically, ICF examined data and information describing 32 audit measures that are designed to ensure that the sampling is being conducted appropriately. USPS provided information about each of the 32 audit measures, including the general criteria used to evaluate each measure; however, ICF used these criteria to develop more specific rules for evaluating compliance.

The audit metrics are based on the following questions:

- Does the Internal SPM system produce results that are accurate?
- Does the Internal SPM system produce results that are reliable?
- Does the Internal SPM system produce results that are representative?

Following these high-level questions, the audit plan provides sub-questions and sub-sub-questions about specific Internal SPM processes to be examined. The audit criteria is used in answering sub-sub-questions, audit information to review or assess compliance, and methods to examine the information provided by USPS. Table 1 displays the audit questions, criteria, and information used by ICF to evaluate the compliance of the sampling process.

Table 1. Audit Plan Measures

Measure	Phase	Level 1	Level 2	Level 3	Audit Criteria (Yardstick)	Audit Information
1	First Mile	Is First Mile (FM) data Accurate?	Are Design (e.g. requirements, SOPs, business rules) and Execution of First Mile processes accurate?	Do carriers accurately complete First Mile sampling?	Procedures for sampling should be written and training provided regularly.	Validate that the sampling procedures are up-to-date and comprehensive.
2	First Mile	Is FM data Accurate?	Are Design (e.g. requirements, SOPs, business rules) and Execution of First Mile processes accurate?	Do carriers accurately complete First Mile sampling?	There should be processes to identify anomalies between expected and actual number of scans based on the collection box density.	Validate whether processes exist to verify the accuracy of the sampling responses.
3	First Mile	Is FM data Accurate?	Are Design (e.g. requirements, SOPs, business rules) and Execution of First Mile processes accurate?	Is the collection box density data accurate and complete?	Density tests should be performed on every active collection point annually and data collected should accurately reflect the volume in the boxes during the testing period.	Verify that there is a process to load/use Collection Point Management System (CPMS) density data.
4	Last Mile	Is Last Mile (LM) data Accurate?	Are Design (e.g. requirements, SOPs, business rules) and Execution of Last Mile processes accurate?	Do carriers accurately complete Last Mile sampling?	Procedures for sampling should be written and training provided regularly.	Validate that the sampling procedures are up-to-date and comprehensive.

Measure	Phase	Level 1	Level 2	Level 3	Audit Criteria (Yardstick)	Audit Information
5	Last Mile	Is LM data Accurate?	Are Design (e.g. requirements, SOPs, business rules) and Execution of Last Mile processes accurate?	Do carriers accurately complete Last Mile sampling?	There should be processes to identify anomalies between expected and actual number of scans based on the expected pieces in inventory.	Validate whether processes exist to verify the accuracy of the sampling responses.
6	Reporting/ Processing Duration Data	Is Reporting/ Data Accurate?	Are Design (e.g. requirements, SOPs, business rules) and Execution of Reporting processes accurate?	Are reporting procedures and requirements established and executed per design to produce accurate results?	Reporting requirements should be documented and align with regulatory reporting requirements.	Quarterly verification of requirements and report contents should occur.
7	Reporting/ Processing Duration Data	Is Reporting/ Data Accurate?	Are Design (e.g. requirements, SOPs, business rules) and Execution of Reporting processes accurate?	Are reporting procedures and requirements established and being executed per design to produce accurate results?	Exclusions, exceptions, and limitations should be documented in the Informed Visibility (IV) system and the final reports.	Validate whether Attachments A (Exclusion Reasons Breakdown) and B (Total Measured/ Unmeasured) are accurately produced for Internal SPM.
8	Reporting/ Processing Duration Data	Is Reporting/ Data Accurate?	Are Design (e.g. requirements, SOPs, business rules) and Execution of Reporting processes accurate?	Do non-automated exclusions and special exceptions (e.g. local holidays, non-certified mail, proxy data, low volume exclusions) create unbiased performance estimates.	Documented approval process should exist and be followed for all manual/special exclusions and exceptions and for adding and/or changing exclusions or other business rules.	Review approval process for all manual exclusions and special exceptions. Review process and decisions for any exclusions to confirm the focus is on measurement accuracy and not biased.

Measure	Phase	Level 1	Level 2	Level 3	Audit Criteria (Yardstick)	Audit Information
9	First Mile	Is FM data Reliable?	Are First Mile results designed and executed to produce reliable results?	Is use of imputations for FM Profile results limited to provide FM measurement that represents the district's performance?	All 67 districts should have a limited amount of volume for which imputed results are used within the quarter.	Review the volume of mail for which imputations are required.
10	First Mile	Is FM data Reliable?	Are First Mile results designed and executed to produce reliable results?	Is use of proxy data for FM Profile results limited to provide FM measurement that represents the district's performance?	All 67 districts should have a limited amount of volume for which proxy results are used within the quarter.	Review the volume of mail where proxy data are used.
11	Last Mile	Is Last Mile (LM) data Reliable?	Are Last Mile results designed and executed to produce reliable results?	Is use of imputations for LM Profile results limited to provide LM measurement that represents the district's performance?	All 67 districts should have a limited amount of volume for which imputed results are used within the quarter.	Review the volume of mail for which imputations are required.
12	Last Mile	Is LM data Reliable?	Are Last Mile results designed and executed to produce reliable results?	Is use of proxy data for LM Profile results limited to provide LM measurement that represents the district's performance?	All 67 districts should have a limited amount of volume for which proxy results are used within the quarter.	Review the volume of mail where proxy data are used.

Measure	Phase	Level 1	Level 2	Level 3	Audit Criteria (Yardstick)	Audit Information
13	Reporting/ Processing Duration Data	Is Reporting/ Data Reliable?	Does the Internal SPM system produce reliable results?	Are changes to SPM documented and available for reference?	Program and SPM changes are documented in an IV repository for reference.	Review documentation of systems' modifications and validate availability and robustness.
14	Reporting/ Processing Duration Data	Is Reporting/ Data Reliable?	Does the Internal SPM system produce reliable results?	Are changes to SPM documented and available for reference?	PRC Reports denote major methodology and process changes in quarterly results.	Review method and process changes as well as PRC Report narratives.
15	Reporting/ Processing Duration Data	Is Reporting/ Data Reliable?	Does the Internal SPM system produce reliable results?	Are changes to SPM documented and available for reference?	For each product measured, the on-time performance scores should have margins of error lower than the designed maximums for the quarter.	Review statistical precision by product and reporting level.
16	Reporting/ Processing Duration Data	Is Reporting/ Data Reliable?	Does the Internal SPM system produce reliable results?	Do processes exist to store and maintain official results reliably?	Processes should be established for storing final quarterly results.	Validate that vital scoring data are "frozen" for quarter close and that these data are maintained in accordance with data retention policy.
17	Reporting/ Processing Duration Data	Is Reporting/ Data Reliable?	Does the Internal SPM system produce reliable results?	Does the schedule allow for the production of reliable quarterly results given data and system constraints?	All critical defects and data repairs should be completed for the quarter prior to finalizing results. All data loading, ingestions, associations, consolidations, and aggregations should be completed.	Validate that there is a process to close the quarterly reporting period to include: 1) Review outstanding defects to determine impact or potential impact; 2) Review completed data repairs/defect repairs for comprehensiveness; and 3) Review data processing backlogs impacting the quarter.

Measure	Phase	Level 1	Level 2	Level 3	Audit Criteria (Yardstick)	Audit Information
18	First Mile	Is FM data Representative?	Do the design and business rules support representative results for First Mile Performance?	Does mail picked up from points outside the sample perform the same as mail included?	Carrier pickup procedures should be similar to First Mile for collection mail.	Standard Operating Procedures for the handling of mail picked up by carriers at delivery points should be similar to those for mail picked up from collection points.
19	First Mile	Is FM data Representative?	Does the execution of the First Mile measurement process yield results that are representative?	Do the sampling results indicate that all collection points were included (districts, ZIP codes, box types, box locations)?	Across the year, more than 98% of boxes should be selected for sampling at least one time. Across the year, more than 90% of all collection points should have at least one valid sampled piece.	Across 4 quarters, measure the total number of collection points which were selected for sampling and which resulted in valid samples to identify whether there is systematic non-coverage of boxes.
20	First Mile	Is FM data Representative?	Does the execution of the First Mile measurement process yield results that are representative?	Are the sampling response rates sufficient to indicate that non- response biases are immaterial? If no, does the data indicate differences in performance for under- represented groups?	All response rates should exceed 80% at a district level.	Calculate sampling response rate for each district.
21	First Mile	Is FM data Representative?	Does the execution of the First Mile measurement process yield results that are representative?	If the sampling response rates do not meet district threshold, are there differences in performance for under-represented	Coverage ratios should meet acceptable thresholds at the 3-digit ZIP code, box type, and box location levels.	For district response rates below thresholds, calculate coverage ratios for the 3-digit ZIP code, box type, and box location. For example, if a district falls below the response threshold, examine coverage

Measure	Phase	Level 1	Level 2	Level 3	Audit Criteria (Yardstick)	Audit Information
				groups?		ratios of ZIP codes and box types. For ratios outside the acceptable range, assess the impact of the missing/under-coverage.
22	First Mile	Is FM data Representative?	Does the execution of the First Mile measurement process yield results that are representative?	Are all valid collection points included in the collection profile (collection points, ZIP codes, and collection dates)?	All eligible collection points in CPMS should be measured in the profile.	Assemble full frame of collection points and assess whether all are represented in the profile. If not, determine the extent of missing points.
23	First Mile	Is FM data Representative?	Does the execution of the First Mile measurement process yield results that are representative?	Are all retail locations included in the final retail results for all shapes, dates, and ZIP codes?	All eligible retail locations should contribute data to the profile for some dates and mail types in the quarter.	Assemble a full frame of eligible retail locations and measure how many have at least one piece measured during the quarter.
24	Reporting/ Processing Duration Data	Is Processing Duration data Representative?	Do the execution of the Processing Duration and overall measurement process yield results that are representative?	How much of the population is included in measurement for each measured product?	At least 80% of the population is measured for each product.	Take the total measured volume for the quarter and the total population pieces from RPW for each product (PRC product reporting levels) and calculate the percent of mail in measurement.
25	Reporting/ Processing Duration Data	Is Processing Duration data Representative?	Do the execution of the Processing Duration and overall measurement	Does the percentage of mail not included in measurement represent a small	If the population in measurement is below 80%, the missing data should be "missing at random," that is, in such	Take the total unmeasured volume for the quarter and review the reason it was not included. Can performance be determined from any of the

Measure	Phase	Level 1	Level 2	Level 3	Audit Criteria (Yardstick)	Audit Information
			process yield results that are representative?	enough proportion of the mail as to have minimal impact on overall estimates of performance?	a way that the missing data is not likely to create measurement bias. If not, does the mail which is not included in measurement differ in performance from mail included?	segments and does it differ from the measured results?
26	Reporting/ Processing Duration Data	Is Processing Duration data Representative?	Do the execution of the Processing Duration and overall measurement process yield results that are representative?	Are all destinating ZIP codes and dates represented in the final data?	Each active ZIP code should have mail receipts for all products during the quarter.	Summarize the final data from the quarter by destination 5-digit ZIP code and product and assess against the full frame.
27	Reporting/ Processing Duration Data	Is Processing Duration data Representative?	Do the execution of the Processing Duration and overall measurement process yield results that are representative?	Are all originating plants and dates represented in the final data?	Each plant should have mail originating for all products.	Summarize the final data for a quarter by origin plant and product and assess against the full frame of plants/products to see if 100% complete.
28	Last Mile	Is LM data Representative?	Does the execution of the Last Mile measurement process yield results that are representative?	Are the sampling response rates sufficiently high to indicate that non-response biases are immaterial?	Response rates should be greater than 80% for each district for the quarter.	Measure the last mile sampling response rate by the district.

Measure	Phase	Level 1	Level 2	Level 3	Audit Criteria (Yardstick)	Audit Information
29	Last Mile	Is LM data Representative?	Does the execution of the Last Mile measurement process yield results that are representative?	If the sampling response rates do not meet district threshold, does the data indicate differences in performance for under-represented groups?	Response rates below 80% at the district level should be measured by 3-digit ZIP code, Delivery Point Type, and Carrier Route Type for the quarter.	Measure response rates by 3-digit ZIP code, delivery point type, and carrier route type.
30	First Mile	Is FM data Accurate?	Are Design (e.g. requirements, SOPs, business rules) and Execution of First Mile processes accurate?	Do carriers accurately complete the First Mile back bffice sampling?	There should be processes to identify anomalies between expected and actual number of back office sampling scans based on the volume expected.	Validate whether processes exist to verify the accuracy of the sampling responses.
31	First Mile	Is FM data Representative?	Does the execution of the First Mile measurement process yield results that are representative?	Do the back office sampling results indicate that all offices were included as appropriate?	All eligible back office sampling locations should contribute data to the profile for some dates and mail types in the quarter.	Across 4 quarters, measure the total number of offices selected for sampling and that resulted in valid samples to identify whether there is systematic non-coverage of back offices in sampling.
32	First Mile	Is FM data Representative?	Does the execution of the First Mile measurement process yield results that are representative?	Are all back office sampling locations included in the final results for all shapes, dates, and ZIP codes?	All eligible back office sampling locations should contribute data to the profile for some dates and mail types in the quarter.	Assemble a full frame of eligible back office sampling locations and measure how many have at least one piece measured during the quarter.

ICF requested data and information from the USPS SPM team to conduct the audit according to the criteria presented in Table 1.

ICF reviewed the data and information submitted and compared it to the audit criteria to determine compliance. When the FY17 Q1 data indicated possible issues with accuracy, reliability, or representatives, additional information was gathered and reviewed. Throughout this process, ICF documented results and flagged potential issues. After completing the compliance review, ICF quantified the impact or potential impact of compliance issues, as presented in Section IV.

IV. Audit Compliance Review Results

The following sections present the results of the audit compliance evaluation. ICF followed an evidence-based approach that evaluates whether or not the USPS SPM team performed the requisite steps to comply with the audit measures develop by USPS. That is, ICF requested certain data, calculations, and information that would demonstrate that the audit was performed in an appropriate fashion. ICF did not, however, perform the audit measure calculations or alter the audit metrics that were provided. Each section provides a summary of the audit measure, the activities required to conduct the audit review, and the request for information submitted by ICF to the USPS SPM team. Finally, each section concludes with a determination of compliant, possibly compliant, or non-compliant. Following this review of each measure, we present a summary of the audit compliance review (see Table 12). Appendix A presents the categorization scheme used to determine compliance.

This audit review was conducted with limited time, which narrowed the amount of data and information that could be obtained and discussed.

Measure 1: First Mile—Procedures for sampling should be documented and training provided regularly

Audit measure 1 evaluates a component of the First Mile sampling accuracy by analyzing the design and execution of First Mile sampling. Specifically, it is intended to assess whether the First Mile sampling procedures are being performed correctly by carriers. To conduct the review for audit measure 1, USPS must validate that sampling procedures and training materials are up-to-date and that training sessions are provided regularly to current and new employees.

To evaluate the compliance of this audit measure, ICF requested schedules and numbers of participants for the training sessions conducted for the FY17 Q1 data collection phase. In response, USPS submitted data showing the number of training completions by district, the number of trained personnel added or subtracted, the number of untrained personnel, and the percentage of training completed on two separate dates. The training information shows that 320,325 employees from 67 districts participated in the sampling training. The data does not provide information on the number of participants by training program, or the training schedule.

Overall, audit measure 1 is compliant.

Measure 2: First Mile—There should be processes to identify anomalies between expected and actual number of scans based on the collection box density

Similar to audit measure 1, audit measure 2 analyzes the First Mile sampling accuracy by evaluating the design and execution of First Mile sampling processes. Specifically, measure 2 assesses whether the First Mile sampling procedures are being performed correctly by carriers. To conduct the audit of audit measure 2, USPS must validate whether processes exist to verify the accuracy of the sampling responses.

To evaluate the compliance of this audit measure, ICF requested the expected and actual counts of scans by area and district. For areas/districts in which there are large percentage differences, ICF requested summaries describing the discrepancies and the potential for bias due to missing data. In response, IBM provided the FY17 Q1 Sampling Compliance Report presenting weekly compliance rates over time subdivided by area, district, and 3-digit ZIP code. The data does not, however, provide transparent or supported explanations for patterns of low compliance or decreasing compliance over time. For example, 3-digit ZIP 025 in the Greater Boston district had a weekly collection compliance rate of 92.86 percent for the week ending on 11/12/16, which drops sharply to 71.43 percent for the week ending on 11/19/16. The information provided does not indicate why such a sharp decrease in compliance occurred.

Therefore, audit measure 2 can be considered possibly compliant.

Measure 3: First Mile—Density tests should be performed on every active collection point annually and data collected should accurately reflect the volume in the boxes during the testing period

Audit measure 3 evaluates a component of the First Mile sampling accuracy by analyzing the design and execution of First Mile sampling processes. Specifically, it is intended to assess whether collection box density data is accurate and complete. To conduct the review of audit measure 3, USPS must verify that there is a process to load and use Collection Point Management System (CPMS) density data.

To evaluate the compliance of this audit measure, ICF asked clarifying questions, the answers to which implied that densities of collection boxes were not measured during the sampling period. ICF requested a table organized by area and district showing the total expected density of all sampled collection boxes, the total expected number of scanned usable pieces based on the annual density scans and the area/district usability rate, and the total actual number of scanned usable pieces. Based on the data and responses IBM provided, it appears that there are no useful data to determine whether the density data collected in Summer 2016 is accurate for FY2017 Q1, or if the Q1 densities are proportional to the annual density measurements.

This audit measure has two requirements. The first requirement is to ensure that density tests were performed on every active collection point annually. The second requirement is to ensure that the data collected in the annual density scans accurately reflect the volume in the boxes during the testing period. The first requirement is supported but no data were available to support the second requirement. Therefore, audit measure 3 can be considered possibly compliant. A sample of collection box densities should be taken each quarter.

Measure 4: Last Mile—Procedures for sampling should be written and training provided regularly

Audit measure 4 assesses the accuracy of Last Mile data by evaluating the design and execution of Last Mile processes. To do this, measure 4 specifically determines whether carriers are accurately completing the Last Mile sampling. The audit of this measure validates that the sampling procedures are up-to-date and comprehensive.

Similar to measure 1, ICF requested schedules and number of participants for the training sessions conducted for the FY17 Q1 data collection phase. In response, USPS submitted data showing the number of training completions by district, the number of trained personnel added or subtracted, the number of untrained personnel, and the percentage of training completed on two separate dates. The training information shows that 320,325 employees from 67 districts participated in the sampling training. The data does not provide, however, the number of participants by training program, or the schedule or frequency of training.

Overall, audit measure 4 is compliant.

Measure 5: Last Mile—There should be processes to identify anomalies between expected and actual number of scans based on the expected pieces in inventory

Audit measure 5 is another measurement of Last Mile data accuracy that focuses on the design and execution of Last Mile processes. Specifically, measure 5 asks whether carriers are accurately completing Last Mile sampling by assessing whether processes exist to verify the accuracy of sampling responses.

To evaluate compliance with this audit measure, ICF requested tables showing the expected and actual numbers of scans by area and district. ICF asked for explanations of large discrepancies and summaries of reasons for potential bias due to missing data. Additionally, ICF requested tables by area and district showing the number of mailpieces scanned at delivery points and the corresponding number where the mailpiece was matched to a scan in the processing system. In response, IBM provided the FY17 Q1 Sampling Compliance Report presenting weekly compliance rates over time subdivided by area, district, and 3-digit ZIP code.

The data does not, however, provide transparent or supported explanations for cases of low compliance or declining compliance over time. For example, 3-digit ZIP 908 in the Los Angeles district had a weekly collection compliance rate of 87.47 percent for the week ending on 10/1/16, which fell to 78.47 percent for the week ending on 12/24/16. The information provided does not indicate why such decreases or inconsistencies in compliance occurred.

Therefore, audit measure 5 can be considered possibly compliant.

Measure 6: Reporting—Reporting requirements should be documented and aligned with regulatory reporting requirements

Audit measure 6 is intended to assess the accuracy of reporting and data by examining the design and execution of reporting processes. Specifically, this audit measure asks whether reporting processes and requirements are established and executed per design to produce

accurate results. To determine compliance with measure 6, USPS must verify on a quarterly basis that the reporting requirements are documented and aligned with regulatory reporting requirements.

The USPS SPM team provided documentation of the sampling methodology (Requirements Documents) and the Excel spreadsheets of the Scores and Variance reports. The documents, however, can be improved upon.

Therefore, audit measure 6 can be considered possibly compliant.

Measure 7: Reporting—Exclusions, exceptions, and limitations should be documented in the Informed Visibility (IV) system and the final reports

Audit measure 7 is intended to assess the accuracy of reporting and data by looking at the design and execution of reporting processes. Specifically, this audit measure reflects whether reporting processes and requirements are established and executed per design to produce accurate results. To determine compliance with measure 7, USPS must validate whether Attachments A (Exclusion Reasons Breakdown) and B (Total Measured/Unmeasured) are accurately produced for Internal SPM. To this end, ICF requested the relevant documentation in the IV system and the relevant final reports for 2017 Q1.

Attachment A and Attachment B have not yet been produced for Internal SPM. The Internal SPM team was not required to produce these reports for FY17 Q1.

Therefore, audit measure 7 can be considered not applicable.

Measure 8: Reporting—Documented approval process should exist and be followed for all manual/special exclusions and exceptions and for adding and changing exclusions or other business rules

Like audit measure 7, audit measure 8 focuses on the accuracy of reporting and data by looking at the design and execution of reporting processes. Audit measure 8, however, asks whether non-automated exclusions and special exceptions (e.g., curtailments, local holidays, non-certified mail, proxy data, or special low volume exclusions) create unbiased performance estimates. To answer this, audit measure 8 requires a review of the approval process for all manual exclusions and special exceptions, and a review of the process and decisions for any exclusions to confirm a non-biased approach and an appropriate focus on measurement accuracy.

To confirm compliance, ICF requested documentation of the approval process, including processes required for manual and special exclusions, as well as exceptions for adding and changing exclusions or other business rules. In response, the system integrator submitted a document detailing the Internal SPM exclusion process, including reasons for exclusion, procedures from initiation through approval, implementation, and third-party validation. Additionally, the system integrator submitted a document describing job exclusions and mailer decertification request standard operating procedures (SOPs). This document includes reasons for possible exclusions, contact information for submitting requests in each region, information required to submit the request, and the time frame for processing the request.

Therefore, audit measure 8 is compliant.

Measure 9: First Mile—All 67 districts should have a limited amount of volume for which imputed results are used in the quarter

Audit measure 9 examines the reliability of First Mile data by focusing on whether design and execution produce reliable results. In particular, this measure asks if the use of imputed data for First Mile profile results is limited enough to provide the First Mile measurement that is representative of performance. To address this question, USPS must review the volume of mail for which imputations are required. To assess compliance with measure 9, ICF requested the volume of imputation for all 67 districts nationwide categorized by mail type. ICF also asked for clarification on the definition of "limited volume for each mail type." In response, IBM submitted an FY17 Q1 Total Volume analysis showing the imputation type code, total volume, First Processing Operation (FPO) volume, Origin Location Code (i.e., urban or rural) volume, and sample group volume, each organized by district.

The information provided by IBM does not provide a benchmark that defines whether imputation rates are within the "limited volume for each mail type," nor does it define this threshold. Therefore, ICF used a 10% criterion based on experience with similar imputed data. Developing a more exact criterion would require a detailed statistical analysis of the uncertainty from the use of imputed values. For the First Mile and Last Mile, missing data in one region are replaced by data from another region, which corresponds to the statistical method of single imputation. We can assume that the data are "missing at random," or that the probability that data are missing does not depend on actual values for the missing data but depends instead on one or more explanatory variables (e.g., the area or district). In a summary article, the University of Pennsylvania researcher and author J.L. Shafer[4] states "When the fraction of missing information is small (say, less than 5%) then single imputation inferences for a scalar estimand may be fairly accurate." Another researcher, Judi Scheffer[5], analyzed a data set with data missing at random using various imputation methods and concluded that single imputation methods "are fine" at the 10% level of imputation. Multiple imputation or other simulation methods could be applied to the SPM data to evaluate the impact of the imputed data on the estimated on-time performance and margin of error.

ICF excluded from the imputed volumes the cases where the Days Left categories were condensed because those are cases of collapsing categories rather than imputing data from other regions. The information in the "Q1 Sample Group Volume" tab of the spreadsheet provided shows that 10 districts have imputation percentages above 10% (see the table below for districts and associated percentages). All 10 districts have imputed data percentages between 10% and 20%.

---

[4] Schafer, J. S. Approximately 1999. "The Multiple Imputation FAQ Page." Available at: http://www.stat.ufl.edu/~athienit/STA6167/Missing%20Data/MI_FAQ.pdf
[5] Scheffer, J. 2002. "Dealing with Missing Data," *Research Letters in the Information and Mathematical Sciences* (2002) 3, 153-160.

Table 2. Districts with More than 10% Imputed First Mile Results

District	Imputed Percentage
Alaska	19%
Bay-Valley	12%
Long Island	11%
Los Angeles	12%
New York	11%
Philadelphia Metro	12%
South Florida	12%
South Jersey	10%
Suncoast	14%
Triboro	12%

Because these districts do not have a limited volume for which imputed results are used (based on a 10% threshold), audit measure 9 can be considered non-compliant.[6]

Measure 10: First Mile—All 67 districts should have a limited amount of volume for which proxy results are used in the quarter

As in audit measure 9, audit measure 10 focuses on the reliability of First Mile data by assessing design and execution reliability. Measure 10 specifically reflects whether the use of proxy data for First Mile profile results is limited enough to provide measurements representative of actual performance. This audit measure relies on a review of the volume of mail where proxy data are used. ICF requested data on the volume of proxy use for all 67 districts for each mail type, and for clarification regarding the definition of "limited volume for each mail type." In response, IBM provided an FY17 Q1 Total Volume analysis showing the imputation type code, total volume, FPO volume, Origin Location Code volume, and sample group volume, each organized by district.

The only district with any proxy data reported is the Caribbean, with proxy data accounting for 11% of its total results. However, the information provided by IBM does not provide a benchmark that defines whether imputation rates are within the "limited volume for each mail type," nor does it define this threshold. Therefore, ICF used 10% as a cutoff to determine whether the proxy results for each district were within a limited volume. ICF decided to use a 10% criterion as a rule of thumb based on experience with similar imputed data. Developing a more exact rule would require a detailed statistical analysis of the uncertainty due to the use of proxy values. The only district with proxy results reported only slightly exceeds the 10% threshold.

---

[6] Feedback from USPS to a draft of this report indicated that the above calculations combine letters and flats—which have different imputation rates—and thus should be assessed separately. The raw data provided to ICF for this measure does not allow for categorization by mail type, and it is not possible to revise the calculations due to time constraints. The classification for this measure as non-compliant, however, still holds. While it may be more informative to separate letters from flats in the imputation calculations, the number of districts not meeting the 10% threshold will be the same. That is, if a district fails to meet the 10% threshold, it may be that it does not meet it for letters, or for flats, or for both. Therefore, the overall conclusion of non-compliant does not change.

Therefore, audit measure 10 is compliant.

Measure 11: Last Mile—All 67 districts should have a limited amount of volume for which imputed results are used in the quarter

Audit measure 11 assesses the reliability of Last Mile data by considering whether design and execution lead to reliable results. In particular, measure 11 asks if the use of imputed data for Last Mile profile results is limited enough to provide a representative measurement of actual performance. Determining compliance with this measure requires a review of the volume of mail for which imputations are required.

ICF requested data on the volume of imputation for all 67 districts by mail type, and clarification on the definition of "limited volume for each mail type." In response, IBM provided Last Mile volumes for FY17 Q1, which show the volumes of single piece versus non-single piece mail for each district across multiple sample groups. The information indicates the volume of mail imputed by district, imputation type, and mail type. However, the information provided by IBM does not provide a benchmark that defines whether imputation rates are within the "limited volume for each mail type," nor does it define this threshold. Thus, ICF used a cutoff of 10% imputed data to determine whether each district was within a limited volume.

ICF excluded from the imputed volumes the cases where the Days Left categories were condensed because those are cases of collapsing categories rather than imputing data from other regions. Initially, ICF assessed the data by combining sample groups, but in response to feedback from USPS, ICF reanalyzed the data for each separate sample group. Sample group 1 had 57 districts above the 10% threshold for single piece mail, and 54 districts above the 10% threshold for non-single piece mail. Sample groups 2, 3, 5, 6, and 7 each had between 4 and 11 districts above the 10% threshold for imputed data. Table 3 shows the number of districts with more than 10% imputed data by sample group.

Because many districts have high percentages of imputed data, and because each sample group has multiple districts above the 10% threshold, audit measure 11 can be considered possibly compliant.

Table 3. Number of Districts with More than 10% Imputed Last Mile Results

Sample Group*	Number of Districts above 10% Imputed Data
Sample Group 1 (SP)	57
Sample Group 1 (non-SP)	54
Sample Group 2	11
Sample Group 3	5
Sample Group 5	5
Sample Group 6	5
Sample Group 7	4

* Note: Sample group 5 was reported for only SP, while sample groups 2, 3, 6, and 7 were reported for only non-SP. Sample group 1 was the only group reported for both SP and non-SP.

Measure 12: Last Mile—All 67 districts should have a limited amount of volume for which proxy results are used in the quarter

Audit measure 12 also focuses on the reliability of Last Mile data by evaluating the results of design and execution. Measure 12 looks at whether the use of proxy data for Last Mile profile results is limited enough in scope that results are still representative of actual performance. This audit measure involves a review of the volume of mail for which proxy data are used. ICF requested data on the volume of proxy use for all districts by mail type, and for clarification on the definition of "limited volume for each mail type." In response, IBM provided Last Mile volumes for FY17 Q1, which show the volumes of single piece versus non-single piece mail for each district. The information provided by IBM, however, does not provide a benchmark that defines whether imputation rates are within the "limited volume for each mail type," nor does it define this threshold. ICF assessed each district using a 10% cutoff for proxy results. One district has proxy results well above 10% (Chicago, at 17%). The sampling issues for Chicago were noted in the report narratives to make report users aware of the limitation.

Therefore, audit measure 12 can be considered possibly compliant.

Measure 13: Reporting—Program and SPM changes are documented in an IV repository for reference

Audit measure 13 analyzes the reliability of the Reporting and Processing Duration data by ensuring the SPM system produces reliable results. To evaluate reliability, measure 13 requires that changes to the SPM system are documented and available for reference. To perform the audit of measure 13, USPS must review documentation of systems' modifications and validate availability and robustness.

To evaluate the compliance of this audit measure, ICF requested information describing the process used to track the system's modifications and robustness. In response, the system integrator submitted a description of the modification tracking process along with business process management examples. Specifically, all requirement changes are documented in VersionOne IV project, while all defect items are tracked in ALM IV project.

Therefore, audit measure 13 is compliant.

Measure 14: Reporting—PRC Reports denote major methodology and process changes in quarterly results

Similar to audit measure 13, measure 14 analyzes the reliability of the reporting and Processing Duration data by ensuring the SPM system produces reliable results. Audit measure 14 requires that changes to the SPM system are documented and available for reference. To conduct the audit of measure 14, USPS must review the methodology and process changes as well as PRC Report narratives.

To evaluate the compliance of audit measure 14, ICF requested example PRC reports describing major methodological and process changes in quarterly results. In response, USPS submitted FY17 Q1 PRC reports and supporting data. The narratives accompanying the data describe substantive system deviations at a high level without much detail. While the requirements documents describe the methodology, they do not discuss deviations.

Therefore, audit measure 14 can be considered possibly compliant.

Measure 15: Reporting and Processing Duration—For each product measured, the on-time performance scores should have margins of error lower than the designed maximums for the quarter

Audit measure 15 analyzes the reliability of the Reporting and Processing Duration data by ensuring the SPM system produces reliable results. Similar to measures 13 and 14, audit measure 15 requires that changes to the SPM system are documented and available for reference. To conduct the review of audit measure 15, USPS must review the statistical precision by product and reporting level.

There are two general methods that can be used to evaluate this audit measure—the first method is to analyze First Mile and Last Mile target margins of error; the second method is to analyze the overall margins of error. ICF selected the former for the analysis below. To evaluate the compliance of audit measure 15, ICF requested the estimated precision for First Mile performance for each mail type and district to compare it with the target levels given in item 6.1 of the First Mile Sample Design Requirements document. ICF also requested the estimated precision for Last Mile performance to compare it with the target level given in item 3.1 of the Last Mile Sample Design Requirements document. Table 4 summarizes the target levels from the requirements documents.

Table 4. Target Margins of Error

First Mile Target		
Level of Aggregation	**Sampling Group**	**Quarterly Results Precision**
District	Single-Piece Letters	+/- 0.5%
District	Single-Piece Flats	+/- 1.0%
Last Mile Targets		
Level of Aggregation	**Sampling Group**	**Quarterly Results Precision**
District	Periodicals	+/- 1.0%
District	Standard Flats	+/- 1.0%
District	Presort FCM Letters	+/- 1.0%

ICF compared the First Mile margins of error and found that 130 of 268 districts exceeded the maximum target of error for the quarter, which amounts to 48.5% of all districts with data. Similarly, ICF compared the Last Mile margins of error and found that 82 of 547 districts exceeded the maximum target of error, which amounts to 15.0% of the total.

The second method involves analyzing the overall margins of error for on-time scores, which was the intent of the measure when the audit plan was designed. The criteria envisioned came from the Parallel Testing Success Criteria document and should be added to the Audit Plan. Below are the relevant portions of that document, defining the targeted maximums for on time scores.

Commercial Mail End-to-End Performance Success Criteria

Precision of end-to-end results falls within the maximum levels of precision targeted in the sample design at the overall quarterly district level for at least 90% of districts:

Table 5. Quarterly Results Precision—Commercial

End-to-End Results Category	Quarterly Results Precision Less than or Equal to:
Presort FCM Letters scores	+/- 1.0%
Presort FCM Flats scores	+/- 3.0%
Standard Letters scores	+/- 1.0%
Standard Flats scores	+/- 1.0%
Periodicals scores	+/- 2.0%

Single-Piece Mail End-to-End Performance Success Criteria

Precision of results falls within the maximum levels of precision targeted in the sample design at the overall quarterly district level for at least 90% of districts:

Table 6. Quarterly Results Precision—Single-Piece

End-to-End Results Category	Quarterly Results Precision Less than or Equal to:
Single-Piece FCM Letters scores	+/- 1.0%
Single-Piece FCM Flats scores	+/- 3.0%

Therefore, audit measure 15 can be considered non-compliant.

Measure 16: Reporting—Processes should be established for storing final quarterly results

Audit measure 16 analyzes the reliability of the reporting and Processing Duration data by ensuring the SPM system produces reliable results. Audit measure 16 requires processes to store and maintain official results. To conduct the review of audit measure 16, USPS must validate that essential scoring data are "frozen" for quarter close and that these data are maintained in accordance with the data retention policy.

To evaluate the compliance of audit measure 16, ICF requested an explanation of how data are frozen for quarter close and verification that those data are maintained and stored in

accordance with the data retention policy. In response, the system integrator provided a detailed description of how data are frozen.

Specifically, when pieces receive additional scans, the application tracks the change and accounts for it in all processing and reporting aggregates. Those pieces are included in a quarter based on the anticipated delivery date (ADD). At the quarter close, a control date in the application is changed such that the application will no longer trigger updates to an aggregate if the ADD of the changed piece is less than that date. For example, to close FY17Q2, the control date would be set to 4/1/2017. All aggregated data for service performance reporting is then unchanged after that point or "frozen." The reporting aggregates have different frequencies at which they refresh so there is about a two-week period following the quarter close date when data needs to be synced up without including the changes blocked by the control date. Once that is synced, a separate control date is adjusted so that no refresh of the aggregate data is attempted for ADDs before the now closed reporting quarter.

Therefore, measure 16 is compliant.

Measure 17: Reporting—All critical defects and data repairs should be completed for the quarter before finalizing results. All data loading, ingestions, associations, consolidations, and aggregations should be completed

Audit measure 17 analyzes the reliability of the Reporting and Processing Duration data by ensuring the SPM system produces reliable results. Audit measure 17 requires that the schedule allows for the production of reliable quarterly results given data and system constraints. To conduct the review of audit measure 17, USPS must validate that there is a process to close the quarterly reporting period which includes the following: (1) review outstanding defects to determine impact or potential impact; (2) review completed data repairs/defect repairs for comprehensiveness; and (3) review any data processing backlogs impacting the quarter.

To evaluate the compliance of audit measure 17, ICF requested a description of the processes used to close the quarter. The system integrator provided a detailed response regarding the various steps to close the quarter. Specifically, defects determined to have a significant impact on the reporting results for the quarter are scheduled for implementation no later than four days before the targeted quarter close date. As the items are deployed, the application and quality assurance teams validate and correct the problems until resolved. The processing backlog is also monitored daily throughout the quarter. Any potential backlog issues that may create a quarter close risk are escalated for resolution. Once all validations have completed successfully, and the backlog is resolved, the close of the quarter is executed by modifying the quarter close threshold.

Therefore, audit measure 17 is compliant.

Measure 18: First Mile—Carrier pickup procedures should be similar to First Mile for collection mail

Audit measure 18 analyzes the representativeness of the First Mile data by evaluating whether the design and business rules support representative results. Audit measure 18 tests whether mail picked up from points outside the sample performs the same as mail included. To conduct the review of audit measure 18, USPS must review Standard Operating Procedures for the handling of mail picked up by carriers at delivery points and ensure that those procedures are similar to those for mail picked up from sampling collection points.

To evaluate the compliance of audit measure 18, ICF requested an explanation of the review process for the Standard Operating Procedures, and a description of the verification process for the handling of mail for sampling collection points and non-collection points. In response, USPS clarified that handling of mail picked up at delivery points is identical to the handling of mail picked up at collection boxes. No information was provided, however, about the Standard Operating Procedures, and thus ICF cannot comment on the review process for the Standard Operating Procedures.

From the responses to the Clarification Questions, mail picked up by carriers at delivery points was not included in the First Mile Performance, although that mail may be measured in the Processing Duration and Last Mile phases.

Therefore, audit measure 18 can be considered possibly compliant.

Measure 19: First Mile—Across the year, more than 98% of boxes should be selected for sampling at least one time, and more than 90% of all collection points should have at least one valid sampled piece

Audit measure 19 evaluates whether First Mile data is representative by assessing if the execution of First Mile measurement processes yields representative results. Specifically, measure 19 asks if sampling results indicate that all collection points were included (districts, ZIP codes, box types, and box locations). To perform this audit, USPS must measure the total number of collection points selected for sampling across four quarters which result in valid samples. This allows for identification of any systematic non-coverage of boxes.

To evaluate compliance, ICF requested data on the total number of collection boxes in the US, the number not sampled in FY17 Q1, and the number that had no usable mail sampled in FY17 Q1. Also, ICF requested similar data on the total number of collection points in the United States, those that were not sampled in FY17 Q1, and those with no usable mail sampled in FY17 Q1. In response, IBM submitted documentation of collection boxes by location, ZIP code, district, and density group. The data showed the number of submitted requests and the usable mail volume for each collection point by quarter for FY16 Q2 through FY17 Q1. The information shows that over four quarters, more than 19,000 of the total 190,137 total collection boxes were not sampled (roughly 10%). Additionally, over 23,000 collection boxes that were sampled had no usable mail in the year (approximately 14%).

The intent of this measure was to examine sample coverage and identify any systematic exclusion that may be of concern. It may be, however, that the thresholds of 98% and 90% are too stringent or that additional audit criteria are needed.

Therefore, audit measure 19 can be considered non-compliant.

Measure 20: First Mile—All response rates should exceed 80% at a district level

Audit measure 20 analyzes the First Mile sampling representativeness by assessing whether sampling response rates meet district thresholds. Specifically, audit measure 20 is intended to determine whether the execution of the First Mile measurement process yields representative results by ensuring that non-response biases are immaterial. To conduct the review of audit measure 20, USPS must calculate the sampling response rate for each district and identify those districts with response rates less than 80%.

To evaluate the compliance of this audit measure, ICF proposed to analyze response rates for each district. IBM provided collection compliance information by the district for Q1 FY17. Based on these data, 57 of 67 districts did not meet the 80% compliance level for Q1.

Therefore, audit measure 20 can be considered non-compliant.

Measure 21: First Mile— Coverage ratios should meet acceptable thresholds at the 3-digit ZIP code, box type and box location levels

Audit measure 21 analyzes the First Mile sampling representativeness by assessing whether sampling response rates meet district thresholds and, for districts below thresholds, if the data indicate differences in performance for under-represented groups. To conduct the audit of measure 21, USPS must validate whether coverage ratios meet acceptable thresholds at the 3-digit ZIP code, box type, and box location levels. For example, if a district falls below response threshold, USPS must examine coverage ratios of ZIP codes and box types. For ratios outside the acceptable range, USPS must assess the impact of the missing data or under-coverage.

To evaluate the compliance of this audit measure, ICF requested coverage ratios at the 3-digit ZIP code level for all district response rates that did not meet the threshold. ICF further requested assessments performed to evaluate the impact of underrepresentation for these districts. In response, IBM provided coverage ratios at the 3-digit ZIP code level and by box type and box location. IBM further indicated that additional analyses of the impact of underrepresentation for districts below the threshold had not been conducted.

Therefore, audit measure 21 can be considered non-compliant because analysis of under-represented groups has not been conducted.

Measure 22: First Mile—All eligible collection points in CPMS should be measured in the profile

Audit measure 22 analyzes the First Mile sampling representativeness by assessing whether all valid collection points are included in the collection profile. To conduct the audit of measure 22, USPS must validate whether all eligible collection points in CPMS are measured in the profile. To do this, USPS must assemble a full frame of collection points and assess whether all are represented in the profile. If not, USPS must determine the extent of missing points.

To evaluate the compliance of this audit measure, ICF requested that USPS tabulate the total number of collection points in the United States and in the calculated First Mile profiles for FY17 Q1. In response, the system integrator provided Table 7, showing the number of collection points that have either received collection box scans in Q1 or received usable First Mile sample scans in Q1.

Table 7. Scans of Eligible First Mile Collection Points for FY17 Q1

# Eligible Collection Points	Points with Collection Box Scans		Points with Usable Sample Scans	
Count	Count	% of Eligible	Count	% of Eligible
190,910	188,614	98.80%	87,089	45.62%

Based on this information, audit measure 22 is compliant.

Measure 23: First Mile—All eligible retail locations should contribute data to the profile for some dates and mail types in the quarter

Audit measure 23 analyzes the First Mile sampling representativeness by assessing whether all retail locations are included in the final retail results for all shapes, dates, and ZIP codes. To conduct the review of audit measure 23, USPS must validate whether all eligible retail locations contributed data to the profile for some dates and mail types in the quarter. To do this, USPS should assemble a full frame of eligible retail locations and measure how many have at least one piece measured during the quarter.

To evaluate the compliance of this audit measure, ICF initially proposed to tabulate the total number of retail locations in the United States and the number of retail locations with some mail included in the calculated First Mile profiles for FY17 Q1.

ICF determined that this audit measure is not applicable for FY2017 Q1 because retail data was not used in Q1.

Measure 24: Processing Duration—At least 80 percent of the population is measured for each product

Audit measure 24 analyzes the Processing Duration sampling representativeness by assessing how much of the population is included in measurement for each measured product. To conduct the review of audit measure 24, USPS must validate whether at least 80% of the population is measured for each product. To do this, USPS must take the total measured volume for the quarter and the total population pieces from RPW for each product (PRC product reporting levels) and calculate the percent of mail in the measurement.

To evaluate the compliance of this audit measure, ICF requested the total measured volume for each quarter for each product and the total population pieces from RPW for each product. In response, IBM provided the information contained in Table 8.

Table 8. Processing Duration Measured Volumes

Mail Category	RPW Volume	Internal SPM Volume	Internal SPM Volume % of RPW Volume
Total First Class Mail	**15,880,158,171**	**11,448,790,435**	**72.1%**
Single-Piece Letters/Postcards	5,479,474,638	4,554,480,276	83.1%
Presort Letters/Postcards	9,998,687,034	6,525,940,760	65.3%
Single-Piece Flats	224,552,604	279,277,616	124.4%
Presort Flats	177,443,896	89,091,784	50.2%
Total Standard Mail	**22,300,194,697**	**14,522,132,407**	**65.1%**
High Density and Saturation Letters	1,918,192,344	1,416,493,367	73.8%
High Density and Saturation Flats/Parcels	3,114,482,052	1,132,989,164	36.4%
Carrier Route (Flats and Letters)	2,135,444,347	1,426,087,843	66.8%
Letters	13,079,744,231	9,276,402,282	70.9%
Flats	1,857,582,524	1,144,445,738	61.6%
Every Door Direct Mail	194,749,199	125,714,013	64.6%
Total Periodicals	**1,369,839,892**	**723,400,337**	**52.8%**
Total Package Services	**77,264,906**	**11,622,339**	**15.0%**
Bound Printed Matter Flats	77,264,906	11,622,339	15.0%

As presented in Table 8, the 80% coverage level was achieved for only two of the products (Single-Piece Letters/Postcards and Single-Piece Flats). Therefore, audit measure 24 can be considered non-compliant.

Measure 25: Processing Duration—If percentage of mail in measurement is below 80%, the missing data should be "missing at random"

Audit measure 25 analyzes the Processing Duration sampling representativeness by assessing whether the percentage of mail not included in measurement represents a small enough proportion of the mail as to have minimal impact on overall estimates of performance. To conduct the review of audit measure 25, USPS must validate whether there is measurement bias in the missing data if the population in measurement is below 80%. If the population in measurement is below 80%, the missing data should be "missing at random," that is, in such a way that the missing data is not likely to create measurement bias. If not, does the mail which is not included in measurement differ in performance from mail included? To assess this, USPS must take the total unmeasured volume for the quarter and review the reason it was not included. Can performance be determined from any of the segments and does it differ from the measured results?

To evaluate the compliance of this audit measure, ICF requested that USPS tabulate the measured volume and the estimated on-time performance by mail type, district, and month for

each case where the unmeasured volume is more than 20%. In response, IBM indicated that they have reviewed the data at the national level and not at lower levels because the information is only available at the national level. IBM noted some patterns identified through their analysis, including underrepresentation of End-to-End Periodicals and over-representation of Destination Entry Periodicals. IBM indicated that the cause might be a combination of the types of mailers choosing to use Full-Service Intelligent Mail (tending to be larger) and the process by which mail is eligible for measurement (e.g., having to get a start-the-clock). To address these trends, IBM indicated that the current reporting methodology uses national level weights to adjust the proportion for how much Destination Entry mail counts and how much End-to-End mail counts in the overall score. IBM also indicated that a similar process is used for Standard Mail products.

Based on IBM's analysis of trends in missing data, audit measure 25 can be considered possibly compliant. Data below the national level would allow for a more robust analysis of missing Processing Duration data but, as indicated above, this information is not available.

Measure 26: Processing Duration—Each active ZIP code should have mail receipts for all products during the quarter

Audit measure 26 analyzes the Processing Duration sampling representativeness by assessing whether all destination ZIP codes and dates are represented in the final Processing Duration data. To conduct the audit of measure 26, USPS must validate whether each active ZIP code has mail receipts for all products during the quarter. To assess this, USPS must summarize the final data from the quarter by destination 5-Digit ZIP code and product and assess against full frame.

To evaluate the compliance of this audit measure, ICF requested that USPS tabulate the volume of processed mail in the Processing Duration data by product, destination ZIP code, and date (including cases with zero processed mail). In response, the system integrator provided the processing volumes for each mail product and 5-Digit ZIP code. The system integrator noted that due to data retention, the Single Piece volumes are only available at the 5-Digit ZIP level from the beginning of December 2016 onward. They indicated that the commercial volumes are representative of the entire Q1. The system integrator did not provide date information, so ICF was not able to assess whether all dates are represented in Processing Duration data.

IBM analyzed the complete FY17 Q1 data for this metric by major product, using a full frame of 40,815 active 5-Digit ZIP codes and found the following:

Table 9. Processing Duration by Product

Metrics for Criteria	Single-Piece First-Class Mail Letters	Presort First-Class Mail Letters	First-Class Mail Flats	Standard Mail Letters	Standard Mail Flats	Periodicals	Bound Printed Matter Flats
Total Eligible ZIP codes With Measureable Mail Pieces	40,815	40,457	40,768	40,512	40,378	40,085	38,606

Metrics for Criteria	Single-Piece First-Class Mail Letters	Presort First-Class Mail Letters	First-Class Mail Flats	Standard Mail Letters	Standard Mail Flats	Periodicals	Bound Printed Matter Flats
Total Eligible ZIP codes Without Measurable Mail Pieces	0	358	47	303	437	730	2,209
Having Mail Pieces Processed from the Service Area of Every Destinating ZIP code	100.00%	99.10%	99.90%	99.30%	98.90%	98.20%	94.60%

Overall, all 5-Digit ZIP codes are covered in measurement and the vast majority are covered for all products. Therefore, audit measure 26 can be considered possibly compliant.

Measure 27: Processing Duration—Each plant should have mail originating for all products

Audit measure 27 analyzes the Processing Duration sampling representativeness by assessing whether all originating plants and dates are represented in the final data. To conduct the audit of measure 27, USPS must validate whether all mail is reflected in the origin plant data. To do this, USPS must summarize the final data for a quarter by origin plant and product, and assess against the full frame of plants and products to see if the data is 100% complete.

To evaluate the compliance of this audit measure, ICF requested that USPS tabulate the volume of processed mail in the Processing Duration data by product, origin plant, and date (including plants and dates with zero processed mail). In response, IBM provided origin plant volumes, including anticipated delivery date, mail class, mail shape, and measured volumes. IBM noted that blank cells for an Anticipated Date/Origin Facility combination indicate that no mail was delivered that day which originated at that facility.

ICF aggregated the origin plant data provided by IBM by product category across all origin plants and dates and compared it to the aggregated Processing Duration mail volumes presented in Table 8 above. This comparison indicates that the origin plant data is inconsistent with the Processing Duration mail volumes. Table 10 shows the total measured volume for each product from the aggregated data.

Table 10. Originating Plants by Product

Mail Category	Measured Volume
Total First Class Mail	**11,461,550,181**
Single-Piece Letters	4,554,480,278
Presort Letters/Postcards	6,538,632,504
Single-Piece Flats	279,277,617

Mail Category	Measured Volume
Presort Flats	89,159,782
Total Standard Mail	**13,669,914,808**
Letters	10,723,923,589
Flats	2,945,991,219
Total Periodicals	**759,047,831**
Total Package Services	**11,648,357**

IBM explained the Single-Piece First-Class data as follows: "For Single-Piece First-Class Mail the origin of Processing Duration is based on automation processing of the mail, but for Commercial mail, products are based on non-automation events (for example, unloading a pallet of mail from a truck sent by a customer to USPS). The first automation processing may occur at that same site or it may not. Our metric can only be pulled based on those non-automation events, and not the first automation processing event."

Because some of the volumes in Table 8 are higher than those in Table 10 (e.g., 14,522,132,407 for standard mail versus 13,669,914,808), it is possible that all originating plants and dates are not represented in the origin plant data. It is also possible that the data IBM provided for Measure 27 only reflects Internal SPM volume, and thus, the comparison of the data in Table 8 and Table 10 is imperfect.

To perform the audit as written, ICF needed a full frame of processing plants that are expected to process each type of mail, along with counts of measured pieces which originated in each plant. IBM provided a recent analysis of the data from FY17 Q1 showing the following plant coverage levels:

Table 11. Plant Coverage Levels

Coverage Criteria	Having Mail Pieces Originate/Process from the Service Area of Every Plant
Single-Piece First-Class Mail Letters	99.60%
Single-Piece First-Class Mail Flats	97.30%
Presort First-Class Mail Letters	96.40%
Presort First-Class Mail Flats	88.80%
Standard Mail Letters	95.10%
Standard Mail Flats	97.80%
Periodicals	92.90%
Bound Printed Matter Flats	91.60%

Therefore, audit measure 27 can be considered possibly compliant.

Measure 28: Last Mile—Response rates should be greater than 80 percent for each district for the quarter.

Audit measure 28 analyzes the Last Mile sampling representativeness by assessing whether sampling response rates are sufficiently high to indicate that non-response biases are immaterial. To conduct the audit of measure 28, USPS must validate whether response rates are greater than 80% for each district for the quarter. To do this, USPS must measure the last mile sampling response rate by district.

To evaluate the compliance of this audit measure, ICF proposed to analyze response rates for each district. IBM provided delivery compliance information by the district for Q1 FY17. Based on the data provided by IBM, 7 of 67 districts did not meet the 80% compliance level for Q1.

Therefore, audit measure 28 can be considered possibly compliant.

Measure 29: Last Mile—Response rates below 80 percent at the district level should be measured by 3-digit ZIP code, Delivery Point Type and Carrier Route Type for the quarter.

Audit measure 29 analyzes the Last Mile sampling representativeness by assessing whether, for districts below sampling response rate thresholds, the data indicate differences in performance for under-represented groups. To perform the audit of measure 29, USPS must validate whether response rates below 80% at the district level are measured by 3-digit ZIP code, Delivery Point Type, and Carrier Route Type for the quarter. To do this, USPS must measure response rates by 3-digit ZIP code, Delivery Point Type, and Carrier Route Type.

To evaluate the compliance of this audit measure, ICF requested response rates by 3-digit ZIP code, Delivery Point Type, and Carrier Route Type. In response, IBM provided information that categorized sampling requests into the following response types:

1. Scanned: at least one mailpiece was scanned (Response Rate = Yes)
2. No Pieces to Scan: carrier indicated there were no pieces to scan at sampling location
3. Passed Address: carrier reported that they had passed the sampling location
4. Denied: carrier indicated the sampling request was not on route
5. Ignored: sampling request was ignored
6. No Trigger: sampling request was not triggered
7. None: there were no responses recorded for the sampling request.

IBM indicated that if there were multiple responses per sampling request, they were prioritized according to this hierarchy. Based on the information provided by IBM, ICF can confirm that IBM tracks response rates by 3-digit ZIP code, delivery type, and carrier route type, but IBM has not indicated that they have analyzed this data to determine any differences in performance for under-represented groups.

Therefore, audit measure 29 can be considered non-compliant because the analysis has not been conducted.

Measure 30: First Mile—Processes exist to identify anomalies between expected and actual number of back office sampling scans based on the volume expected

Audit measure 30 analyzes the First Mile sampling accuracy by assessing whether carriers accurately complete First Mile back office sampling. To conduct the audit of measure 30, USPS must validate whether processes exist to identify anomalies between expected and actual number of back office sampling scans based on the volume expected. To do this, USPS must validate whether processes exist to verify the accuracy of the sampling responses.

To evaluate the compliance of this audit measure, ICF initially proposed to analyze the expected and actual number of scans by area and district and, in cases where there are large percentage differences, a summary of reasons for the discrepancies and the potential for bias due to missing data. ICF later determined that this audit measure is not applicable because back office sampling has not been conducted and is not currently planned. Due to the fact that both Back Office Flats Sampling and Retail measured the First Mile from the office to the first processing operation at the plant, they were deemed to be similar enough that both were not necessary as long as sufficient volume existed. When the processes and audit plan were being developed, the amount of volume that could be received from each method was an unknown. In the end, the amount of data received from the Retail Data Ingest was deemed more than sufficient to be statistically significant for flats, therefore allowing the USPS to eliminate the need for Back Office Flats Sampling.

Therefore, compliance with audit measure 30 can be considered non-applicable for FY17 Q1.

Measure 31: First Mile—All eligible back office sampling locations should contribute data to the profile for some dates and mail types in the quarter

Audit measure 31 analyzes the First Mile sampling representativeness by assessing whether back office sampling results indicate that all offices were included as appropriate. To conduct the audit of measure 31, USPS must validate whether all eligible back office sampling locations contribute data to the profile for some dates and mail types in the quarter. To ensure this, USPS must measure the total number of offices selected for sampling across four quarters and track which resulted in valid samples in order to identify whether there is any systematic non-coverage of back offices in sampling.

To evaluate the compliance of this audit measure, ICF initially proposed that USPS tabulate the total number of offices in the United States, the number that was not sampled in FY17 Q1, and the number that had no valid mail sampled in FY17 Q1. ICF later determined that this audit measure is not applicable because back office sampling has not been conducted and is not currently planned. Due to the fact that both Back Office Flats Sampling and Retail measured the First Mile from the office to the first processing operation at the plant, they were deemed to be similar enough that both were not necessary as long as sufficient volume existed. When the processes and audit plan were being developed, the amount of volume that could be received from each method was an unknown. In the end, the amount of data received from the Retail Data Ingest was deemed more than sufficient to be statistically significant for flats, therefore allowing the USPS to eliminate the need for Back Office Flats Sampling.

Therefore, compliance with audit measure 31 can be considered non-applicable for FY17 Q1.

Measure 32: First Mile—All eligible back office sampling locations contribute data for some dates and mail types in the quarter

Audit measure 32 analyzes the First Mile sampling representativeness by assessing whether back office sampling locations are included in the final results for all shapes, dates, and ZIP codes. To conduct the audit of measure 32, USPS must validate whether all eligible back office sampling locations contribute data to the profile for some dates and mail types in the quarter. This involves assembling a full frame of eligible back office sampling locations and measuring how many have at least one piece measured during the quarter.

To evaluate the compliance of this audit measure, ICF initially proposed that USPS tabulate the total number of back office locations in the United States and the number of back office locations with some mail included in the calculated First Mile profiles for FY17 Q1. ICF later determined that this audit measure is not applicable because back office sampling has not been conducted and is not currently planned. Due to the fact that both Back Office Flats Sampling and Retail measured the First Mile from the office to the first processing operation at the plant, they were deemed to be similar enough that both were not necessary as long as sufficient volume existed. When the processes and audit plan were being developed, the amount of volume that could be received from each method was an unknown. In the end, the amount of data received from the Retail Data Ingest was deemed more than sufficient to be statistically significant for flats, therefore allowing the USPS to eliminate the need for Back Office Flats Sampling.

Therefore, compliance with audit measure 32 can be considered non-applicable for FY17 Q1.

1.   Summary of Audit Compliance Review

Table 12 presents a summary of the results of the audit compliance review. ICF classified 8 measures as compliant, 12 measures as having possible issues with compliance, and 7 measures as non-compliant. We categorized five measures as not applicable because the data was not collected in the sampling scheme for FY17 Q1.

Table 12. Audit Compliance Review Summary

Measure	Audit Subject	Audit Criteria	Compliance Result
1	Is First Mile sampling accurately completed by carriers?	Procedures for sampling should be written and training provided regularly.	Compliant
2	Is First Mile sampling accurately completed by carriers?	There should be processes to identify anomalies between expected and actual number of scans based on the collection box density.	Possible issues with compliance
3	Is the collection box density data accurate and complete?	Density tests should be performed on every active collection point annually and data collected should	Possible issues with compliance

Measure	Audit Subject	Audit Criteria	Compliance Result
		accurately reflect the volume in the boxes during the testing period.	
4	Is Last Mile sampling accurately completed by carriers?	Procedures for sampling should be written and training provided regularly.	Compliant
5	Is Last Mile sampling accurately completed by carriers?	There should be processes to identify anomalies between expected and actual number of scans based on the expected pieces in inventory.	Possible issues with compliance
6	Are reporting procedures and requirements established and being executed per design to produce accurate results?	Reporting requirements should be documented and align with regulatory reporting requirements.	Possible issues with compliance
7	Are reporting procedures and requirements established and being executed per design to produce accurate results?	Exclusions, exceptions, and limitations should be documented in the IV system and the final reports.	Not applicable
8	Do non-automated exclusions and special exceptions (e.g. curtailments, local holidays, non-certified mail, proxy data, special low volume exclusions) create unbiased performance estimates?	Documented approval process should exist and be followed for all manual/special exclusions and exceptions and for adding and/or changing exclusions or other business rules.	Compliant
9	Is use of imputations for FM Profile results limited to provide FM measurement that represents the district's performance?	All 67 districts should have a limited amount of volume for which imputed results are used within the quarter.	Non-compliant
10	Is use of proxy data for FM Profile results limited to provide FM measurement that represents the district's performance?	All 67 districts should have a limited amount of volume for which proxy results are used within the quarter.	Compliant
11	Is use of imputations for LM Profile results limited to provide LM measurement that	All 67 districts should have a limited amount of volume for which imputed results are used within the quarter.	Possible issues with compliance

Measure	Audit Subject	Audit Criteria	Compliance Result
	represents the district's performance?		Possible issues with compliance
12	Is use of proxy data for LM Profile results limited to provide LM measurement that represents the district's performance?	All 67 districts should have a limited amount of volume for which proxy results are used within the quarter.	Possible issues with compliance
13	Are changes to SPM documented and available for reference?	Program and SPM changes are documented in an IV repository for reference.	Compliant
14	Are changes to SPM documented and available for reference?	PRC Reports denote major methodology and process changes in quarterly results.	Possible issues with compliance
15	Are changes to SPM documented and available for reference?	For each product measured, the on-time performance scores should have margins of error lower than the designed maximums for the quarter.	Non-compliant
16	Do processes exist to store and maintain official results reliably?	Processes should be established for storing final quarterly results	Compliant
17	Does the schedule allow for the production of reliable quarterly results given data and system constraints?	All critical defects and data repairs should be completed for the quarter prior to finalizing results. All data loading, ingestions, associations, consolidations, and aggregations should be completed.	Compliant
18	Does mail picked up from points outside the sample perform the same as mail included?	Carrier pickup procedures should be similar to First Mile for collection mail.	Possible issues with compliance
19	Do the sampling results indicate that all collection points were included (districts, ZIP codes, box types, box locations)?	Across the year, more than 98% of boxes should be selected for sampling at least one time. Across the year, more than 90% of all collection points should have at least one valid sampled piece.	Non-compliant
20	Are the sampling response rates sufficient to indicate that non-response biases are immaterial? If not, does the data indicate	All response rates should exceed 80% at a district level.	Non-compliant

Measure	Audit Subject	Audit Criteria	Compliance Result
	differences in performance for under-represented groups?		
21	If the sampling response rates do not meet district threshold, does the data indicate differences in performance for under-represented groups?	Coverage ratios should meet acceptable thresholds at the 3-digit ZIP code, box type, and box location levels.	Non-compliant
22	Are all valid collection points included in the collection profile (collection points, ZIP codes and collection dates)?	All eligible collection points in CPMS should be measured in the profile.	Compliant
23	Are all retail locations included in the final retail results for all shapes, dates, ZIP codes?	All eligible retail locations should contribute data to the profile for some dates and mail types in the quarter.	Not applicable
24	How much of the population is included in measurement for each measured product?	At least 80% of the population is measured for each product.	Non-compliant
25	Does the percentage of mail not included in measurement represent a small enough proportion of the mail as to have minimal impact on overall estimates of performance?	If the population in measurement is below 80%, the missing data should be "missing at random," that is, in such a way that the missing data is not likely to create measurement bias. If not, does the mail which is not included in measurement differ in performance from mail included?	Possible issues with compliance
26	Are all destinating ZIP codes and dates represented in the final data?	Each active ZIP code should have mail receipts for all products during the quarter.	Possible issues with compliance
27	Are all originating plants and dates represented in the final data?	Each plant should have mail originating for all products.	Possible issues with compliance
28	Are the sampling response rates sufficiently high to indicate that non-response biases are	Response rates should be greater than 80% for each district for the quarter.	Possible issues with compliance

Measure	Audit Subject	Audit Criteria	Compliance Result
	immaterial?		
29	If the sampling response rates do not meet district threshold, does the data indicate differences in performance for under-represented groups?	Response rates below 80% at the district level should be measured by 3-digit ZIP code, Delivery Point Type and Carrier Route Type for the quarter.	Non-compliant
30	Is First Mile back office sampling accurately completed by carriers?	There should be processes to identify anomalies between expected and actual number of back office sampling scans based on the volume expected.	Not applicable
31	Do the back office sampling results indicate that all offices were included as appropriate?	All eligible back office sampling locations should contribute data to the profile for some dates and mail types in the quarter.	Not applicable
32	Are all back office sampling locations included in the final results for all shapes, dates, ZIP codes?	All eligible back office sampling locations should contribute data to the profile for some dates and mail types in the quarter.	Not applicable

V.   Conclusion

USPS continues its migration to the Informed Visibility (IV) Program, which enhances service performance measurement. The methodology involves collecting and merging performance data for the three phases of mail delivery—First Mile, Processing Duration, and Last Mile. The USPS SPM team has developed new calculation and statistical methods to estimate and combine the performance in each phase. The calculations required the processing of large amounts of data, including the use of physical samples.

1.   Areas of Improvement

This report presents the results of the audit compliance review to evaluation of the accuracy, reliability, and representativeness of the sampling. To perform the audit compliance review, ICF examined data and information describing 32 audit measures that are designed to ensure that the sampling process is being conducted in an appropriate fashion.

As summarized in Table 12, ICF classified 8 measures as compliant (measures 1, 4, 8, 10, 13, 16, 17, and 22), 12 measures as having possible issues with compliance (measures 2, 3, 5, 6, 11, 12, 14, 18, 25, 26, 27, and 28), and 7 measures as non-compliant (measures 9, 15, 19, 20, 21, 24, and 29). The remaining five measures were classified as non-applicable because the relevant sampling data was not collected in FY17 Q1 (measures 7, 23, 30, 31, and 32). Please

refer to Section IV: Audit Compliance Review Results above for a more detailed discussion of the classification rationale for each measure.

2.   Get-well Plan

We recommend changes to improve the compliance of the audit measures. Table 13 summarizes our audit-specific recommendations.

Table 13. Audit Measure-Specific Guidance

Measure	Compliance Status	Steps to Achieve Compliance
Measure 2	Possibly compliant	Provide transparent or supported explanations for cases of low compliance or declining compliance over time.
Measure 3	Possibly compliant	Provide useful data to determine whether the density data collected for the quarter is accurate, or if the quarter densities are proportional to the annual density measurements.
Measure 5	Possibly compliant	Provide transparent or supported explanations for cases of low compliance or decreasing compliance over time.
Measure 6	Possibly compliant	Sampling methodology should be written in a manner to align with regulatory reporting requirements, and the documentation should be comprehensive and well documented.
Measure 9	Non-compliant	Provide, justify, and demonstrate compliance with a benchmark for first-mile imputation rates that defines a "limited volume for each mail type."
Measure 11	Possibly compliant	Provide, justify, and demonstrate compliance with a benchmark for last mile imputation rates that defines a "limited volume for each mail type."
Measure 12	Possibly compliant	Provide, justify, and demonstrate compliance with a benchmark for last mile proxy rates that defines a "limited volume for each mail type."
Measure 14	Possibly compliant	Describe substantive system deviations in detail.
Measure 15	Non-compliant	Margins of error must be reduced to lower than maximum target levels.
Measure 18	Possibly compliant	Provide information about the Standard Operating Procedures.
Measure 19	Non-compliant	Adjust sampling method so that across the year, more than 98% of boxes are selected for sampling at least one time, and more than 90% of all collection points have at least one valid sampled piece.
Measure 20	Non-compliant	Response rates must exceed 80% for all districts.
Measure 21	Non-compliant	Conduct additional analyses of the impact of underrepresentation for districts below the threshold.
Measure 24	Non-compliant	Achieve 80% coverage for each product.

Measure	Compliance Status	Steps to Achieve Compliance
Measure 25	Possibly compliant	Data below the national level should be collected to allow for an analysis of missing Processing Duration data; otherwise, this audit measure should be dropped.
Measure 26	Possibly compliant	Provide date information, so representation in Processing Duration data can be assessed.
Measure 27	Possibly compliant	Provide an explanation for why many plants do not have mail originating for all products or provide alternate volume data for analysis/comparison.
Measure 28	Possibly compliant	Response rates must exceed 80% for all districts.
Measure 29	Non-compliant	Analyze response rates by 3-digit ZIP code, delivery type, and carrier route type to determine differences in performance for under-represented groups.

3. Study Limitations

This study was limited in scope, specifically in the analysis of raw data. For this audit compliance review, ICF only analyzed summary data provided by USPS and its SPM contractors. We formulated questions to solicit data and information from the USPS SPM team to evaluate whether the audit was conducted appropriately. We did not, however, perform the various analyses to ensure that the calculations were done correctly.

4. Next Steps

This section provides a list of action items that prioritize the sampling and audit-related issues discussed in this report. We categorize the action items into those that USPS should start addressing immediately and those that can be addressed over time.

Address Immediately
- No new data collection required
    - Non-compliant.
        - Measure 9: Investigate and explain First Mile high imputation rates for 10 districts.
            - Follow-up: Analyze impact of imputed data on estimated performance and margin of error.
        - Measure 21: Explain patterns of First Mile low response rates inside some districts. This might indicate bias or possible data manipulation.
        - Measure 29: Explain patterns of Last Mile low response rates inside some districts. This might indicate bias or possible data manipulation.
    - Possibly compliant.
        - Measure 2: Explain patterns of First Mile low compliance. This might indicate bias or possible data manipulation.
        - Measure 5: Explain patterns of Last Mile low compliance. This might indicate bias or possible data manipulation.

- Measure 11: Investigate and explain Last Mile high imputation rates for 5 districts.
  - Follow-up: Analyze impact of imputed data on estimated performance and margin of error.
- Measure 12: Investigate and explain Last Mile high proxy rate for 1 district.
  Follow-up: Analyze impact of proxy data on estimated performance and margin of error.
- New data collection required
  - Possibly compliant.
    - Measure 3: Quarterly sampling of collection box densities.
  - Non-compliant.
    - Measure 15: Collect more data for First Mile and Last Mile to reduce margins of error: Increase number of collection boxes or delivery points sampled and increase number of mail pieces sampled per collection box or delivery point, which will increase the amount of usable data
    - Measure 19: Ensure First Mile data are representative: Increase number of collection boxes sampled and increase number of mail pieces sampled per collection box, which will increase the amount of usable data.
    - Measure 20: Improve First Mile response rates by improving training and by reducing technical issues in receiving data requests that might cause low response rates.
  - Possibly compliant.
    - Measure 28: Improve Last Mile response rates by improving training and by reducing technical issues in receiving data requests that might cause low response rates.

Address Over Time
- No new data collection required
  - Possibly compliant.
    - Measure 6: Better documentation; align with regulatory requirements.
    - Measure 14: Documentation of major methodological changes.
    - Measure 18: Prepare and review SOPs for mail picked up at delivery points.
      - Follow-up: Include these data in the First Mile sample.
    - Measure 25: Investigate patterns of potential bias due to lack of representativeness in internal SPM volumes.
    - Measure 27: Evaluate origination plants and dates to ensure representation in Processing Duration data.
- New data collection required
  - Possibly compliant.
    - Measure 26: Tabulate and review Processing Duration data volumes by delivery point ZIP code and date.

Appendix A. Compliance Categorization Scheme

Measure	Audit Criteria	Compliance Determination Cutoff		
		Compliant	**Possibly Compliant**	**Non-Compliant**
1	Procedures for sampling should be written and training provided regularly.	Training is provided regularly, materials are up-to-date.	Training is provided regularly or materials are up-to-date.	Training is not provided regularly; materials are out of date.
2	There should be processes to identify anomalies between expected and actual number of scans based on the collection box density.	Clear processes exist to identify anomalies between expected and actual number of scans	Discrepancies in compliances rates calculated but not explained	Discrepancies in compliance rates are not explained
3	Density tests should be performed on every active collection point annually and data collected should accurately reflect the volume in the boxes during the testing period.	Density tests are performed on every active collection point annually and data accurately reflects volume in boxes	Data does not indicate accuracy of volume measurements *or* sample of collection box densities is not taken each quarter.	Data does not indicate accuracy of volume measurements *and* sample of collection box densities is not taken each quarter.
4	Procedures for sampling should be written and training provided regularly.	Training is provided regularly, materials are up-to-date.	Training is provided regularly or materials are up-to-date.	Training is not provided regularly; materials are out of date.
5	There should be processes to identify anomalies between expected and actual number of scans based on the expected pieces in inventory.	Clear processes exist to identify anomalies between expected and actual number of scans	Processes exist to identify anomalies between expected and actual number of scans but are insufficiently explained	Processes do not exist to identify anomalies between expected and actual number of scans

Measure	Audit Criteria	Compliance Determination Cutoff		
		Compliant	**Possibly Compliant**	**Non-Compliant**
6	Reporting requirements should be documented and align with regulatory reporting requirements.	Documentation of sampling methodology is provided and Excel spreadsheets of Scores and Variance reports are provided and are complete	Either sampling methodology documentation *or* Scores and Variance reports are not provided, or documentation is poor or incomplete	Neither sampling methodology documentation nor Scores and Variance reports are provided, and/or documentation is incomplete or missing
7	Exclusions, exceptions and limitations should be documented in the IV system and in the final reports.	Exclusions, exceptions, and limitations are well documented	Exclusions, exceptions, and limitations are documented but not comprehensively	Exclusions, exceptions, and limitations are not documented
8	Documented approval process should exist and be followed for all manual/special exclusions and exceptions and for adding and/or changing exclusions or other business rules.	Documented approval process exists and is followed for manual/special exclusions and exceptions for adding and/or changing exclusions for other business rules	Documented approval process exists but does not sufficiently explain the manual/special exclusions and exceptions for adding and/or changing exclusions for other business rules	Approval process lacks documentation
9	All 67 districts should have a limited amount of volume for which imputed results are used within the quarter.	0 districts above 10% imputed or proxy data	1-2 districts above 10%	3+ districts above 10%
10	All 67 districts should have a limited amount of volume for which proxy results are used within the quarter.	0 districts above 10% imputed or proxy data	1-2 districts above 10%	3+ districts above 10%

Measure	Audit Criteria	Compliance Determination Cutoff		
		Compliant	Possibly Compliant	Non-Compliant
11	All 67 districts should have a limited amount of volume for which imputed results are used within the quarter.	0 districts above 10% imputed or proxy data	1-2 districts above 10%	3+ districts above 10%
12	All 67 districts should have a limited amount of volume for which proxy results are used within the quarter.	0 districts above 10% imputed or proxy data	1-2 districts above 10%	3+ districts above 10%
13	Program and SPM changes are documented in an IV repository for reference.	Program and SPM changes are documented in an IV repository for reference	Changes are documented but incompletely	Insufficient documentation provided
14	PRC Reports denote major methodology and process changes in quarterly results.	PRC Reports denote major methodology and process changes in quarterly results	Reports document methodology but do not sufficiently describe deviations	Insufficient documentation provided
15	For each product measured, the on-time performance scores should have margins of error lower than the designed maximums for the quarter.	Each margin of error in the Scores Report is less than or equal to the target unsigned margin of error from the corresponding Requirements Document item.	10% or less of the margins of error are less than or equal to the target unsigned margin of error from the corresponding Requirements Document item.	More than 10% of the margins of error are less than or equal to the target unsigned margin or error from the corresponding Requirements Document item.
16	Processes should be established for storing final quarterly results	A well-defined process is described for storing final quarterly results while adhering to data retention policy.	A process is described for storing final quarterly results but does not adhere to the data retention policy or is insufficiently documented.	Little to no information is provided about the process for storing final quarterly results and doing so in accordance with data retention policy.

Measure	Audit Criteria	Compliance Determination Cutoff		
		Compliant	**Possibly Compliant**	**Non-Compliant**
17	All critical defects and data repairs should be completed for the quarter prior to finalizing results. All data loading, ingestions, associations, consolidations and aggregations should be completed.	A detailed response regarding the various steps to close the quarter is provided. The steps are reasonable and robust.	An incomplete response is provided that does not account for all of the steps necessary to close out the quarter, or is insufficiently documented.	An incomplete response is provided that does not account for all of the steps necessary to close out the quarter.
18	Carrier pickup procedures should be similar to First Mile for collection mail.	Carrier pickup procedures are similar to First Mile for collection mail with supporting documentation.	Carrier pickup procedures are described as similar to First Mile for collection mail but with no supporting information.	Carrier pickup procedures are not described as similar to First Mile for collection mail and/or provide no supporting information.
19	Across the year, more than 98% of boxes should be selected for sampling at least one time. Across the year, more than 90% of all collection points should have at least one valid sampled piece.	Across the year, more than 98% of boxes are selected for sampling at least one time. Across the year, more than 90% of all collection points should have at least one valid sampled piece.	Less than 98% of boxes are selected *or* less than 90% of collection points have at least one valid sampled piece.	Less than 98% of boxes are selected *and* less than 90% of collection points have at least one valid sampled piece.
20	All response rates should exceed 80% at a district level.	All response rates exceed 80% at district level. A compliant response means that an "eligible" sampling request was correctly responded to by the carrier.	50% of response rates exceed 80% at district level.	Less than 50% of response rates exceed 80% at district level

Measure	Audit Criteria	Compliance Determination Cutoff		
		Compliant	**Possibly Compliant**	**Non-Compliant**
21	Coverage ratios should meet acceptable thresholds at the 3-digit ZIP code, box type and box location levels.	All coverage ratios meet acceptable thresholds at the 3-digit ZIP code level, box type, and location levels	Most coverage ratios meet acceptable thresholds but deviations are not explained	Most coverage ratios do not meet acceptable thresholds and deviations are not explained
22	All eligible collection points in CPMS should be measured in the profile.	All eligible collection points in CPMS are measured in the profile.	More than 50% of eligible collection points in CPMS are measured in the profile.	Less than 50% of eligible collection points in CPMS are measured in the profile.
23	All eligible retail locations should contribute data to the profile for some dates and mail types in the quarter.	N/A -- Retail Sampling Not Performed		
24	At least 80% of the population is measured for each product.	All products achieve 80% or greater processing duration data measurement	50% or more of products exceed 80% coverage level	Less than 50% of products achieve 80% coverage level
25	If population in measurement is below 80%, the missing data should be "missing at random", that is, in such a way that the missing data is not likely to create measurement bias. If not, does the mail which is not included in measurement differ in performance from mail included?	Data is assessed at the national level, but sub-national data is not available so gaps in trends could exist but cannot be evaluated		

Measure	Audit Criteria	Compliance Determination Cutoff		
		Compliant	**Possibly Compliant**	**Non-Compliant**
26	Each active ZIP code should have mail receipts for all products during the quarter.	Processing duration data for all ZIP codes and requested dates is provided	Data is provided for 50% or more of ZIP codes and dates	Data is provided for less than 50% of the ZIP codes and dates requested
27	Each plant should have mail originating for all products.	All originating plants and dates are represented in the processing duration data	50% or more of plants and dates are represented in the processing duration data	Less than 50% or plants are represented in the processing duration data
28	Response rates should be greater than 80% for each district for the quarter.	All districts meet 80% response rate threshold	50% or more of districts meet 80% response rate threshold	Less than 50% of districts meet 80% response rate threshold.
29	Response rates below 80% at the district level should be measured by 3-digit ZIP code, Delivery Point Type, and Carrier Route Type for the quarter.	Response rates below 80% at the district level are measured by 3-digit ZIP code, Delivery Point Type, and Carrier Route Type for the quarter.	Response rates below 80% at the district level are measured by 3-digit ZIP code, Delivery Point Type, and Carrier Route Type for the quarter but are incomplete.	Response rates below 80% at the district level are not measured by 3-digit ZIP code, Delivery Point Type, and Carrier Route Type for the quarter.
30	There should be processes to identify anomalies between expected and actual number of back office sampling scans based on the volume expected.	N/A -- back office sampling not performed		
31	All eligible back office sampling locations should contribute data to the profile for some dates and mail types in the quarter.	N/A -- back office sampling not performed		

Measure	Audit Criteria	Compliance Determination Cutoff		
		Compliant	**Possibly Compliant**	**Non-Compliant**
32	All eligible back office sampling locations should contribute data to the profile for some dates and mail types in the quarter.	N/A -- back office sampling not performed		