Postal Regulatory Commission
Submitted 5/29/2020 3:50:20 PM
Filing ID: 113302
Accepted 5/29/2020

Independent Validation of USPS Service Performance Measurement Audit Design

Audit Compliance Review

FY2020 Quarter 2

May 18, 2020

Submitted to:
USPS

Submitted by:
ICF

Table of Contents

Executive Summary **3**

I. Introduction **6**

II. Evaluation Approach **6**

1. Compliance Approach 7

III. Audit Compliance Review Results **8**

IV. Summary of Audit Compliance Review **35**

V. Conclusion **37**

1. Areas of Improvement 37
2. Improvement Plan 38
3. Study Limitations 38
4. Next Steps 38

Appendix A. Compliance Categorization Scheme **39**

Executive Summary

This report presents the results of ICF's review of the USPS Internal Service Performance Measurement (SPM) for Quarter 2 (Q2) of Fiscal Year 2020 (FY20). ICF completed similar compliance analyses for Q1, Q3, and Q4 of FY17; FY18; FY19; and Q1 of FY20. This report presents the results of a compliance analysis for FY20 Q2 of a set of audit measures designed to assess the accuracy, reliability, and representativeness of the sampling performance.

USPS has completed its migration to the Internal Service Performance Measurement (SPM) system, which enhances service performance measurement. The Internal SPM system became the official data source on October 1, 2018. Internal SPM provides comprehensive, consolidated data collection and monitoring of the service 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, Internal SPM 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, Internal SPM uses census data for mail classes that previously used a manual seeding/recipient approach, which vastly increases the volume in measurement and the value of the diagnostics and scores available to the field on a daily basis.

ICF evaluated the accuracy, reliability, and representativeness of the sampling activities by assessing a set of 26 audit measures. ICF reviewed audit information to determine compliance of each measure and developed methods to examine the information provided by the USPS SPM team.

Table ES-1 summarizes the results of the compliance analysis. For FY20 Q2, ICF classified 19 measures as achieved, 6 measures as partially achieved, and 1 measure as not achieved.

Table ES-1. Audit Compliance Review Summary

Measure	Phase	Audit Subject	Audit Criteria	FY20 Q2 Result
1	First Mile	Is First Mile sampling accurately completed by carriers?	Procedures for sampling should be written and training provided to employees responsible for performing sampling.	Achieved
2	First Mile	Is First Mile sampling accurately completed by carriers?	Carrier sampling weekly compliance rates should consistently exceed 80 percent for most districts	Partially Achieved
3	First Mile	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.	Achieved
4	Last Mile	Is Last Mile sampling accurately completed by carriers?	Procedures for sampling should be written and training provided to employees responsible for performing sampling.	Achieved
5	Last Mile	Is Last Mile sampling accurately completed by carriers?	Carrier sampling weekly compliance rates should consistently exceed 80 percent for most districts	Achieved

Measure	Phase	Audit Subject	Audit Criteria	FY20 Q2 Result
6	Reporting	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.	Achieved
7	Reporting	Are reporting procedures and requirements established and being executed per design to produce accurate results?	Exclusions, exceptions, and limitations should be documented in the Internal SPM system and the final reports.	Achieved
8	Reporting	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?	A documented approval process should be in place and be followed for all manual/special exclusions and exceptions and for adding and/or changing exclusions or other business rules.	Achieved
9	First Mile	Is use of imputations for FM Profile results limited to provide FM measurement that represents the district's performance?	Most districts should have a limited volume for which imputed results are used within the quarter.	Achieved
10	First Mile	Is use of proxy data for FM Profile results limited to provide FM measurement that represents the district's performance?	Most districts should have a limited volume for which proxy results are used within the quarter.	Achieved
11	Last Mile	Is use of imputations for LM Profile results limited to provide LM measurement that represents the district's performance?	Most districts should have a limited volume for which imputed results are used within the quarter.	Not Achieved
12	Last Mile	Is use of proxy data for LM Profile results limited to provide LM measurement that represents the district's performance?	Most districts should have a limited volume for which proxy results are used within the quarter.	Achieved
13	Reporting	Are changes to SPM documented and available for reference?	Program and SPM changes are documented in an Internal SPM repository for reference.	Achieved
14	Reporting	Are changes to SPM documented and available for reference?	PRC Reports denote major methodology and process changes in quarterly results.	Achieved
15	Reporting/ Processing Duration	Does the Internal SPM system produce reliable results?	For each product measured, the on-time performance scores should have margins of error lower than the designed maximums for the quarter.	Achieved
16	Reporting	Do processes exist to store and maintain official results reliably?	Processes should be established for storing final quarterly results	Achieved
17	Reporting	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.	Achieved
18	First Mile	Do the sampling results indicate that all collection points were included (districts, ZIP codes, box types, box locations)?	Between the first quarter and the end of the current quarter, the percentage of boxes selected for sampling at least one time should be more than the quarterly target percentage.	Achieved
19	First Mile	Are the sampling response rates sufficient to indicate that non-response biases are immaterial? If not, does the data indicate differences in performance for under- represented groups?	Most response rates should exceed 80% at a district level.	Partially Achieved

Measure	Phase	Audit Subject	Audit Criteria	FY20 Q2 Result
20	First Mile	If the sampling response rates do not meet the 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 levels for districts with poor coverage.	Partially Achieved
21	First Mile	Are all valid collection points included in the collection profile (collection points, ZIP codes, and collection dates)?	Most eligible collection points in CPMS should be measured in the profile.	Achieved
22	First Mile	Are all retail locations included in the final retail results for all shapes, dates, and ZIP codes?	Most eligible retail locations should contribute data to the profile for some dates and mail types in the quarter.	Achieved
23	Processing Duration	How much of the volume is included in the measurement for each measured product?	At least 70% of the volume is measured for each product.	Partially achieved
24	Processing Duration	Are all destinating ZIP codes and dates represented in the final data?	Most active ZIP codes should have mail receipts for all products during the quarter.	Achieved
25	Last Mile	Are the sampling response rates sufficiently high to indicate that non-response biases are immaterial?	Most response rates should exceed 80% at a district level.	Partially Achieved
26	Last Mile	If the sampling response rates do not meet the 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 levels for districts with poor coverage.	Partially Achieved

Based on the results of the 26 audit measures, ICF recommends changes to improve the compliance of the partially or not achieved audit measures. Table ES-2 summarizes our audit measure-specific recommendations following the results of the FY20 Q2 audit compliance review.

Table ES-2. Audit Measure-Specific Recommendations to Achieve Compliance

Measure	Compliance Status	Recommendation
Measure 2	Partially achieved	Achieve consistent 80% weekly compliance rates for collection.
Measure 11	Not achieved	Reduce the need for imputed data for First Class Flats in Last Mile.
Measure 19	Partially achieved	Increase collection response rates at the district level.
Measure 20	Partially achieved	Increase collection coverage ratios for districts with poor coverage.
Measure 23	Partially achieved	Achieve 70% coverage for most products.
Measure 25	Partially achieved	Increase delivery response rates at the district level.
Measure 26	Partially achieved	Increase delivery coverage ratios for districts with poor coverage.

The body of this report provides additional details (including specific metrics for each of the 26 audit measures) and explanations of the compliance determinations; it also provides a prioritization of the changes we recommend for the short and long terms.

I. 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 findings, including:

- USPS and PRC reports on delivery performance provide insufficient analysis to hold the USPS accountable for meeting its statutory mission, including insufficient district-level analyses and a lack of reporting about rural delivery performance.
- 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, PRC requested public comments on the quality and completeness of service performance data in Order No. 2791 (October 29, 2015).[2] In Order No. 3490 (August 26, 2016), 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 of each service performance measure.[3]

USPS developed a proof-of-concept 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. ICF reviewed Internal SPM results for Q1, Q3, and Q4 of FY17; FY18; FY19; and Q1 of FY20. This report presents the results of a follow-on analysis to evaluate the compliance determinations for Q2 of FY20. The following sections detail the results of this audit review and ICF's recommendations.

II. 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 the following products:

- Domestic First-Class Mail
  - Single-Piece letters and cards
  - Presort letters and cards
  - Single-Piece and Presort flats
- USPS Marketing 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*.

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

"Standard Mail" was renamed "USPS Marketing Mail" in January 2017.

The audit evaluated 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 was to ensure the accuracy, reliability, and representativeness of the sampling methodology and execution. These terms are 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 reflect the overall volume.

"Bias" combines accuracy and representativeness by evaluating the extent to which the performance estimates from the sample data tend to over- or underestimate the volume performance of all USPS mail.

1. Compliance Approach

This section presents the approach 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 26 audit measures designed to ensure that the sampling is being conducted appropriately. USPS provided information about each of the 26

audit measures, including the general criteria used to evaluate each measure. 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 secondary and tertiary questions about specific Internal SPM processes to be examined. The audit criteria are used in answering tertiary questions, audit information to review or assess compliance, and methods to examine the information USPS provided. Table 1 displays the audit questions, criteria, and information ICF used in FY20 Q2 to evaluate the compliance of the sampling process.

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 submitted data and information and compared it to the audit criteria to determine compliance. When the FY20 Q2 data indicated possible issues with accuracy, reliability, or representativeness, ICF requested clarification and additional information. 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.

III. Audit Compliance Review Results

The following sections present the results of the audit compliance evaluation for FY20 Q2. ICF followed an evidence-based approach that evaluated whether the USPS SPM team performed the requisite steps to comply with the audit measures USPS developed and ICF redesigned. That is, ICF requested certain data, calculations, and information that would demonstrate that the audit was performed appropriately. ICF did not, however, perform the audit measure calculations or alter the audit metrics after USPS approved them. Each section begins with a summary of the audit measure for Q2, activities required to conduct the audit review, and the requests for information ICF submitted to the USPS SPM team. Finally, each section concludes with a determination of achieved, partially achieved, or not achieved for FY20 Q2. Following this review of each measure, we present a summary of the audit compliance review. Appendix A presents the categorization scheme used to determine compliance in Q2.

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 to employees responsible for performing sampling.	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?	Carrier sampling weekly compliance rates should consistently exceed 80 percent for most districts.	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 to employees responsible for performing sampling.	Validate that the sampling procedures are up-to-date and comprehensive.
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?	Carrier sampling weekly compliance rates should consistently exceed 80 percent for most districts.	Validate whether processes exist to verify the accuracy of the sampling responses.
6	Reporting/ Processing	Is Reporting/ Data Accurate?	Are Design (e.g., requirements, SOPs, business rules) and	Are reporting procedures and requirements	Reporting requirements should be documented and aligned with	Quarterly verification of requirements and report contents should occur.

Measure	Phase	Level 1	Level 2	Level 3	Audit Criteria (Yardstick)	Audit Information
	Duration Data		Execution of Reporting processes accurate?	established and executed per design to produce accurate results?	regulatory reporting requirements.	
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 Internal Service Performance Measurement (SPM) 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, and low volume exclusions) create unbiased performance estimates?	A documented approval process should be in place and be followed for all manual/special exclusions and exceptions and for adding 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.
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?	Most districts should have a limited 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?	Most districts should have a limited 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	Most districts should have a limited volume for which imputed results are used within the quarter.	Review the volume of mail for which imputations are required.

Measure	Phase	Level 1	Level 2	Level 3	Audit Criteria (Yardstick)	Audit Information
				represents the district's performance?		
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?	Most districts should have a limited volume for which proxy results are used within the quarter.	Review the volume of mail where proxy data are used.
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 Internal SPM 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?	Does the Internal SPM system produce reliable results?	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	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

Measure	Phase	Level 1	Level 2	Level 3	Audit Criteria (Yardstick)	Audit Information
					aggregations should be completed.	3) Review data processing backlogs impacting the quarter.
18	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)?	Between the first quarter and the end of the current quarter, the percentage of boxes selected for sampling at least one time should be more than the quarterly target percentage.	Across the fiscal year, 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.
19	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?	Most response rates should exceed 80% at a district level.	Calculate sampling response rate for each district.
20	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 the district threshold, are there differences in performance for under-represented groups?	Coverage ratios should meet acceptable thresholds at the 3-digit ZIP Code levels for districts with poor coverage.	For district response rates below thresholds, calculate coverage ratios for the 3-digit ZIP codes.
21	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)?	Most 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.
22	First Mile	Is FM data Representative?	Does the execution of the First Mile measurement process yield results	Are all retail locations included in the final retail results for all shapes, dates, and ZIP codes?	Most eligible retail locations should contribute data to the profile for some dates	Assemble a full frame of eligible retail locations and measure how many have at least one piece measured during the quarter.

Measure	Phase	Level 1	Level 2	Level 3	Audit Criteria (Yardstick)	Audit Information
			that are representative?		and mail types in the quarter.	
23	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 volume is included in the measurement for each measured product?	At least 70% of the volume is measured for each product.	Take the total measured volume for the quarter and the total population pieces for each product (PRC product reporting levels) and calculate the percent of mail in measurement.
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?	Are all destinating ZIP codes and dates represented in the final data?	Most active ZIP codes 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.
25	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?	Most response rates should exceed 80% at a District level.	Measure the last mile sampling response rate by the district.
26	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 the 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 levels for districts with poor coverage.	For district response rates below thresholds, calculate coverage ratios for the 3-digit ZIP codes.

Measure 1: First Mile—Procedures for sampling should be documented and training provided to employees responsible for performing sampling

Quarter 2 Result: Achieved.

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 to current and new employees responsible for performing sampling. USPS provides training to employees at the time of onboarding, when there are significant changes to the sampling methodology, and when sampling issues are identified.

To evaluate the compliance of audit measure 1, ICF requested schedules and numbers of participants in the training sessions conducted for the FY20 Q2 data collection phase. In response, USPS submitted data showing the number of training completions by Performance Cluster (PFC). The training information shows that over 384,000 employees completed the mandatory trainings out of over 443,000 active employees across 67 PFCs (86.58%).

Audit measure 1 evaluates whether sampling procedures and training materials are up to date, and whether—in at least 80% of districts—training is provided to at least 75% of participants responsible for performing sampling. The sampling procedures and training materials were up to date. Training completion rates of 75% or higher were achieved in 64 of the 67 (95.52%) PFCs.

Therefore, audit measure 1 can be considered achieved for FY20 Q2.

Measure 2: First Mile—Carrier sampling weekly compliance rates should consistently exceed 80 percent for most districts

Quarter 2 Result: Partially Achieved.

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 audit measure 2, ICF requested the expected and actual counts of scans by area and district. For areas/districts having large percentage differences, ICF requested summaries describing the discrepancies and the potential for bias due to missing data. In response, USPS provided data from the FY20 Q2 Sampling Compliance Report presenting weekly compliance rates over time subdivided by area and district.

The Sampling Compliance Report defines First Mile compliance as cases for which (1) at least one mail piece was scanned or (2) the carrier indicated No Pieces to Scan and the minimum expected density for the collection box is less than or equal to 35 based on reference information on collection box density. The compliance rate calculation includes only eligible requests, which are for a specific date and collection box for which a sampling request was

generated and triggered by at least one device. The calculation excludes requests for which the address was passed before the carrier arrived or the request was denied because it was not on the carrier's route.

Audit measure 2 evaluates the percentage of districts that maintain weekly First Mile compliance rates of at least 80% for all weeks and whether reasons for lower compliance were investigated in cases of lower compliance. Of the 67 districts, 51 (76.1%, i.e., less than 80% of districts) had weekly compliance rates that were all at least 80%. The Q2 collection compliance was consistently very low in three of the other 16 districts, which leads to uncertainty in the estimated on-time rates. USPS did not find any special circumstances, such as extreme weather events, that would explain the low compliance patterns.

As outlined in Appendix A, audit measure 2 is considered not achieved if less than 50% of districts achieve compliance rates of at least 80% across all 13 weeks of the quarter. Audit measure 2 is considered partially achieved if between 50% and 80% of districts achieve compliance rates of at least 80% across all 13 weeks of the quarter. Audit measure 2 is considered achieved if more than 80% of districts achieve compliance rates of at least 80% across all 13 weeks of the quarter.

Therefore, audit measure 2 can be considered partially achieved for FY20 Q2.

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

Quarter 2 Result: Achieved.

Audit measure 3 evaluates a component of 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 are accurate and complete. To conduct the review of audit measure 3, USPS must verify a process is in place to load and use Collection Point Management System (CPMS) density data.

Audit measure 3 stipulates that density tests be performed on every active collection point annually, and data collected should accurately reflect the volume in the boxes during the testing period. ICF requested data on the number of active collection points in the quarter and the number of those collection points that had a density scan in the past 12 months. USPS conducts an annual density scan. The most recent density scan was conducted in September 2019, during FY19 Q4, and thus the total number of scanned boxes, 231,659, for measure 3 will remain constant from FY20 Q1 through FY20 Q4.

The data show that 96.29% of collection boxes—231,659 scanned boxes of the 240,585 total active collection boxes—had density data over the past 12 months. The 240,585 total active boxes included 239,096 with a box record and another 1,489 scanned boxes that did not have a box record. The 231,659 scanned boxes included 230,170 scanned boxes with a box record and another 1,489 scanned boxes that did not have a box record.

Audit measure 3 is considered achieved if density tests were performed in the last 12 months on at least 95% of the active collection points in the audited quarter. Otherwise, audit measure 3 is

partially achieved if density tests were performed in the last 12 months on at least 80% of the active collection points in the audited quarter.

Therefore, audit measure 3 can be considered achieved for FY20 Q2.

Measure 4: Last Mile—Procedures for sampling should be written and training provided to employees responsible for performing sampling

Quarter 2 Result: Achieved.

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 Last Mile sampling. The audit of this measure validates the sampling procedures are up to date and comprehensive. USPS provides training to employees at the time of onboarding, when there are significant changes to the sampling methodology, and when sampling issues are identified.

Similar to measure 1, ICF requested schedules and number of participants for the training sessions conducted for the FY20 Q2 data collection phase. In response, USPS submitted data showing the number of training completions by PFC. The training information shows that over 384,000 employees completed the mandatory trainings (86.58%) out of over 443,000 active employees across 67 PFCs.

Audit measure 4 evaluates whether sampling procedures and training materials are up to date, and whether—in at least 80% of districts—training is provided to at least 75% of participants responsible for performing sampling. The sampling procedures and training materials were up to date. Training completion rates of 75% or higher were achieved in 64 of the 67 (95.52%) PFCs.

Therefore, audit measure 4 can be considered achieved for FY20 Q2.

Measure 5: Last Mile—Carrier sampling weekly compliance rates should consistently exceed 80 percent for most districts

Quarter 2 Result: Achieved.

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 the compliance of audit measure 5, 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 mail pieces scanned at delivery points and the corresponding number for which the mail piece was matched to a scan in the processing system. In response, USPS provided data from the FY20 Q2 Delivery District Compliance Report presenting weekly compliance rates over time subdivided by area and district.

The Delivery District Compliance Report defines Last Mile compliance as cases for which (1) at least one mail piece was scanned or (2) the carrier indicated No Pieces to Scan and the expected number of pieces for the delivery point is less than or equal to 2 based on Last Mile

inventory information. The compliance rate calculation includes only eligible requests, which are for a specific date and delivery point for which a sampling request was generated and triggered by at least one device. The calculation excludes requests for which the address was passed before the carrier got there or the request was denied because it was not on the carrier's route.

Audit measure 5 evaluates the percentage of districts that maintain weekly Last Mile compliance rates of at least 80% for all weeks and whether reasons for lower compliance were investigated in cases of lower compliance. Of the 67 districts, 60 (89.55%, i.e., more than 80% of districts) had weekly compliance rates that were all at least 80%. The Q2 delivery compliance was consistently very low in three of the other 7 districts, which leads to uncertainty in the estimated on-time rates. USPS did not find any special circumstances, such as extreme weather events, that would explain the low compliance patterns.

As outlined in Appendix A, audit measure 5 is considered not achieved if less than 50% of districts achieve compliance rates of at least 80% across all 13 weeks of the quarter. Audit measure 5 is considered partially achieved if between 50% and 80% of districts achieve compliance rates of at least 80% across all 13 weeks of the quarter. Audit measure 5 is considered achieved if more than 80% of districts achieve compliance rates of at least 80% across all 13 weeks of the quarter.

Therefore, audit measure 5 can be considered achieved for FY20 Q2.

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

Quarter 2 Result: Achieved.

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 according to their 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 FY20 Q2 Excel spreadsheets of the Scores and Variance reports. The Requirements documents were previously reorganized for FY18 Q1 and include a glossary, an overview, and an Excel file listing the values of the various configurable values used in the Internal SPM sample design. For FY20 Q2, USPS provided an updated set of the Requirements Documents, including the most recent document, which is dated 03/01/2019.

Therefore, audit measure 6 can be considered achieved for FY20 Q2.

Measure 7: Reporting—Exclusions, exceptions, and limitations should be documented in the Internal SPM system and the final reports

Quarter 2 Result: Achieved.

Audit measure 7 is intended to assess the accuracy of reporting and data by examining the design and execution of reporting processes. Specifically, this audit measure reflects whether

reporting processes and requirements are established and executed according to their design to produce accurate results.

To evaluate the compliance of audit measure 7, ICF requested documentation in the form of attachments, exclusions, exceptions, and limitations in the Internal SPM system. In response, USPS produced Attachments A (Exclusion Reasons Breakdown) and B (Total Measured/ Unmeasured). Attachment A describes the exclusions for Presort First-Class Mail, USPS Marketing Mail: Letters and Flats, Periodicals, and Bound Printed Matter Flats. Within each mail category, the exclusions were allocated to 15 different exclusion reasons. Attachment A also gives the percentages of the total exclusions for that category attributed to each exclusion reason for the quarter. The Exclusion reasons are: No Piece Scan, Non-Compliant, No Start-the-Clock, Invalid Entry Point for Discount Claimed, Other, Inaccurate Scheduled Ship Date, Non-Unique Intelligent Mail barcode (IMb), Postal Automated Redirection System (PARS), Long Haul, Incorrect Entry Facility, Inconsistent Service Performance Measurement Data, Orphan Handling Unit, FAST Appointment Irregularity, Non-Unique Physical IMb, and Excluded ZIPs.

Attachment B details the total number of pieces from Revenue, Pieces, and Weight: Origin-Destination Information System (RPW-ODIS) for First Class Mail, USPS Marketing Mail, Periodicals, and Bound Printed Matter Flats. Additionally, Attachment B provides the count of pieces included and excluded in the quarterly service measurement reports, the number of pieces eligible for Full Service IMb, the number of Full Service IMb pieces, the number of Full Service IMb pieces included in measurement, and the number of Full Service IMb pieces excluded from measurement.

For Measure 7 to be considered achieved, exclusions, exceptions, and limitations must be well-documented. The exclusions are well-documented. USPS defines exclusions and exceptions interchangeably. Limitations are documented in the PRC Scores and Variance reports.

Therefore, audit measure 7 can be considered achieved for FY20 Q2.

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

Quarter 2 Result: Achieved.

Similar to audit measure 7, audit measure 8 focuses on the accuracy of reporting and data by examining the design and execution of reporting processes. Audit measure 8 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 evaluate the compliance of audit measure 8, ICF requested documentation of the approval process, including processes required for manual and special exclusions, and exceptions for adding and changing exclusions or other business rules. In June 2016, USPS submitted a document detailing the Internal SPM exclusion process, including reasons for exclusion, procedures from initiation through approval, implementation, and third-party validation.

Additionally, in FY20 Q1, USPS submitted an updated document describing job exclusions and mailer decertification request standard operating procedures (SOPs). This document included reasons for possible exclusions, contact information for sending requests in each region, information required to submit the request, and the time frame for processing the request. No changes to the processes described in these documents have been reported by USPS.

ICF obtained five Decertification Request Forms describing Q2 exclusions.

The measure is achieved if documented approval processes are in place and are followed for manual/special exclusions and exceptions and for adding and changing exclusions or other business rules. If documented approval processes are in place but do not explain the requests sufficiently—or were not followed for some requests—the measure is partially achieved. This analysis assumes that the provided forms are a comprehensive representation of all requests for exclusions or exemptions in Q2. Documented approval processes were in place and followed for manual/special exclusions and exceptions and for adding and changing exclusions for FY20 Q2.

Therefore, audit measure 8 can be considered achieved for FY20 Q2.

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

Quarter 2 Result: Achieved.

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 a First Mile measurement representative of performance. To address this question, USPS must review the volume of mail for which imputations are required.

To evaluate the compliance of audit 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, USPS provided a data file showing the imputation rates organized by district and by sample group.

The information provided by USPS 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% or 20% criterion based on experience with similar imputed data. 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 assume the data are "missing at random" or 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, University of Pennsylvania researcher and author, J.L. Shafer, 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."[4] Another

[4] Schafer, J.S. Approximately 1999. "The Multiple Imputation FAQ Page." Available at: http://hsta559s12.pbworks.com/w/file/fetch/51964826/jShafer.multip.Imputationprimer99.pdf *pdf*

researcher, Judi Scheffer, 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.[5] 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 margins of error.

The imputation rate is the imputed volume divided by the total volume. The data received from USPS had already been analyzed by USPS to show the imputation rate and included the percentage of imputation in each district. USPS also provided a separate file with raw data showing the underlying volumes used to calculate each percentage.

Two districts had imputation rates greater than 10% for Single-Piece First-Class Letters/Cards and Flats (Flats exclude Retail District imputation type[6]), which represents the total across both sample groups. Similarly, two districts had imputation rates greater than 10% for Single-Piece First-Class Letters/Cards. One district had imputation rates above 20% for these two mail types. For Single-Piece First Class Flats (excluding Retail District), no districts had an imputation rate above 10%.

Compliance is achieved when for every sample group, no more than one district falls above a 20% imputed data cutoff, and no more than five districts exceed a 10% imputed data cutoff. Otherwise, compliance is partially achieved when for every sample group, two or three districts fall above a 20% imputed data cutoff and no more than five districts exceed a 10% imputed data cutoff.

As for every sample group no more than five districts exceed a 10% imputed data cutoff and no more than one district exceeds a 20% imputed data cutoff, audit measure 9 can be considered achieved for FY20 Q2.

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

Quarter 2 Result: Achieved.

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 mail volume for which proxy data are used.

ICF requested data on the volume of proxy use for all 67 districts for each mail type and clarification regarding the definition of "limited volume for each mail type." In response, USPS provided a data file showing the percent of proxy data by district and by sample group. USPS also provided a separate file with raw data showing the underlying volumes used to calculate each percentage.

---

[5] Scheffer, J. 2002. "Dealing with Missing Data," *Research Letters in the Information and Mathematical Sciences* (2002) 3, 153-160.

[6] The imputation counts for Flats excluded the Retail District imputation type because that imputation is the expected behavior for most Flats mail.

A proxy is designed to be used only if a technical failure prevents the sample requests for a day from being delivered in time for the carriers to complete sampling. A proxy may also be used if changes in the sample request volume required daily is increased or decreased by a factor large enough to introduce a bias in sample results at the onset of the change. It is not a part of the standard reporting methodology, but instead, it is a supplement to the methodology when needed.

The information USPS provided does not include a benchmark that defines whether proxy rates are within the "limited volume for each mail type," nor does it define this threshold. Therefore, ICF used 10% and 20% as cutoffs to determine whether the proxy results for each district were within a limited volume. ICF decided to use a 10% or 20% criterion as a rule of thumb based on experience with similar imputed data. The proxy rate is the proxy volume divided by the total volume. The data provided by USPS shows that in FY20 Q2, for all mail types and all districts, the proxy data imputation rate was below 10%.

Compliance is achieved when for every sample group, no more than one district falls above a 20% proxy data cutoff and no more than five districts exceed a 10% proxy data cutoff. Otherwise, compliance is partially achieved when for every sample group, two or three districts fall above a 20% proxy data cutoff and no more than five districts exceed a 10% proxy data cutoff.

Therefore, audit measure 10 can be considered achieved for FY20 Q2.

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

Quarter 2 Result: Not Achieved.

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 mail volume for which imputations are necessary.

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, USPS provided the Last Mile imputations analysis for FY20 Q2, which shows the percentages of imputed mail for each district across multiple sample groups. The sample groups are First-Class Flats, Periodicals, Presort First-Class Letters/Cards, Single Piece First-Class Letters/Cards, USPS Marketing Mail Flats and Bound Printed Matter, USPS Marketing Mail Letters/Cards, All Mail, and All Mail Excluding Presort and Single Piece First-Class Flats (Sample Group 1). USPS also provided a separate file with raw data showing the underlying volumes used to calculate each percentage. The information USPS provided does not include 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 cutoffs of 10% and 20% imputed data based on experience with similar imputed data.

Table 2 shows the number of districts with more than 10% or 20% imputed data by sample group and overall. When combining all sample groups (the "All Mail" sample group), one district (Caribbean) was above the 10% threshold for imputed data, but it was not above the 20%

threshold. For each of the sample groups "All Mail excluding Presort and Single-Piece First-Class Flats" and USPS Marketing Mail Letters/Cards, one district (Caribbean) was above the 10% threshold for imputed data, but it was not above the 20% threshold. For Presort First-Class Letters/Cards zero districts exceeded the 10% threshold for imputed data. For First-Class Flats, six districts (Caribbean, Alaska, Greater Michigan, Mississippi, Salt Lake City, and Honolulu) exceeded the 10% threshold, and two districts (Caribbean and Alaska) exceeded the 20% threshold. For Periodicals, two districts (Caribbean and Alaska) exceeded the 10% threshold, and one district (Caribbean) exceeded the 20% threshold. For Single-Piece First-Class Letters/Cards and for USPS Marketing Mail Flats and Bound Printed Matter, one district (Caribbean) exceeded both the 10% and 20% thresholds.

Table 2. Number of Districts by Percentage of Imputed Last Mile Results

Sample Group	Number of Districts Above 10% Imputed Data	Number of Districts Above 20% Imputed Data
First-Class Flats	6	2
Periodicals	2	1
Single-Piece First-Class Letters/Cards	1	1
USPS Marketing Mail Flats and Bound Printed Matter	1	1
USPS Marketing Mail Letters/Cards	1	0
All Mail	1	0
All Mail Excluding Presort and Single-Piece First Class Flats (Sample Group 1)	1	0
Presort First-Class Letters/Cards	0	0

Compliance is achieved when, for each sample group other than "All Mail," no more than one district falls above a 20% imputed data cutoff and no more than five districts exceed a 10% imputed data cutoff. Otherwise, compliance is partially achieved when, for each sample group other than "All Mail," no more than three districts fall above a 20% imputed data cutoff and no more than five districts exceed a 10% imputed data cutoff.

Due to the fact that the First-Class Flats sample group has six districts exceeding a 10% imputed data cutoff, audit measure 11 can be considered not achieved for FY20 Q2.

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

Quarter 2 Result: Achieved.

Audit measure 12 also focuses on the reliability of Last Mile data by evaluating the results of design and execution. Measure 12 evaluates whether the use of proxy data for Last Mile profile results is limited enough in scope to yield results that are still representative of actual performance. This audit measure involves a review of the mail volume for which proxy data are used.

ICF requested data on the volume of proxy use for all 67 districts by mail type and clarification on the definition of "limited volume for each mail type." In response, USPS provided the Last Mile imputations analysis, which shows the percentages of proxy mail for each district for the following sample groups: First-Class Flats, Periodicals, Presort First-Class Letters/Cards, Single

Piece First-Class Letters/Cards, USPS Marketing Mail Flats and Bound Printed Matter, USPS Marketing Mail Letters/Cards, All Mail, and All Mail Excluding Presort and Single Piece First-Class Flats (Sample Group 1). USPS also provided a separate file with raw data showing the underlying volumes used to calculate each percentage. The information USPS provided does not include 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% or 20% cutoff for proxy results.

The table below shows a summary of the number of districts above the 10% or 20% threshold by sample group. Within each sample group and across all sample groups, zero districts exceeded the 10% threshold for proxy data.

Table 3. Number of Districts by Percentage of Proxy Last Mile Results

Sample Group	Urban/Rural	Number of Districts Above 10% Proxy Data	Number of Districts Above 20% Proxy Data
First-Class Flats	All	0	0
Periodicals	All	0	0
Presort First-Class Letters/Cards	All	0	0
Single Piece First-Class Letters/Cards	All	0	0
USPS Marketing Mail Flats and Bound Printed Matter	All	0	0
USPS Marketing Mail Letters/Cards	All	0	0
All Mail	All	0	0
All Mail Excluding Presort and Single Piece First-Class Flats (Sample Group 1)	All	0	0

As noted for audit measure 10 above, a proxy is designed to be used only if a technical failure prevents the sample requests for a day from being delivered in time for the carriers to complete sampling. A proxy may also be used if changes in the sample request volume required daily is increased or decreased by a factor large enough to introduce a bias in sample results at the onset of the change. It is not a part of the standard reporting methodology, but instead, it is a supplement to the methodology when needed.

Compliance is achieved when, for each sample group other than "All Mail," no more than one district falls above a 20% proxy data cutoff and no more than five districts exceed a 10% proxy data cutoff. Otherwise, compliance is partially achieved when, for each sample group other than "All Mail," no more than three districts fall above a 20% proxy data cutoff and no more than five districts exceed a 10% proxy data cutoff.

Therefore, audit measure 12 can be considered achieved for FY20 Q2.

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

Quarter 2 Result: Achieved.

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 be 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 audit measure 13, ICF requested information describing the process used to track the system's modifications and robustness. In response, the Postal Service's system integrator submitted a description of the modification tracking process along with business process management examples. Specifically, all requirement changes are documented in Postal data systems, such as VersionOne and ALM. ICF also requested copies of the Internal SPM repository of documented changes in Q2. The number of changes documented for reference in a repository was 6.

Therefore, audit measure 13 can be considered achieved for FY20 Q2.

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

Quarter 2 Result: Achieved.

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 be documented and available for reference. To conduct the audit of measure 14, USPS must review the methodology and process changes and the PRC Report narratives.

To evaluate the compliance of audit measure 14, ICF requested the PRC reports describing major methodological and process changes in quarterly results. In response, USPS submitted FY20 Q2 PRC reports and supporting data. The narratives accompanying the data describe substantive system deviations at a high level.

Therefore, audit measure 14 can be considered achieved for FY20 Q2.

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

Quarter 2 Result: Achieved.

Audit measure 15 analyzes the reliability of the Reporting and Processing Duration data by ensuring the SPM system produces reliable results. To conduct the review of audit measure 15, USPS must review the statistical precision by product and reporting level.

Audit measure 15 stipulates for each product measured, the end-to-end on-time performance scores should have margins of error lower than the designed maximums for the quarter. ICF requested data with the end-to-end margins of error by sampling group and district. The target criteria for the margins of error came from the Parallel Testing Success Criteria document. Below are the relevant portions of that document, defining the targeted maximums for the end-to-end margins of error for on-time scores.

Commercial Mail End-to-End Performance Success Criteria

The 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 4. Quarterly Results Precision—Commercial

End-to-End Results Category	Quarterly Results Precision Less than or Equal to
Presort First Class Mail (FCM) Letters and Cards scores	+/- 1.0%
Presort FCM Flats scores	+/- 3.0%
USPS Marketing Letters scores	+/- 1.0%
USPS Marketing Flats scores	+/- 1.0%
Bound Printed Matter Flats scores	+/- 1.0%
Periodicals scores	+/- 2.0%

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

The 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 5. Quarterly Results Precision—Single-Piece

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

For the analysis of audit measure 15, ICF excluded mail types with very low volumes. Because the volumes are so low, the impact on overall mail performance is also low. Based on the Revenue Piece Weight (RPW) volumes tabulated in Table 8 below, three mail types from Tables 4 and 5 each contribute less than 1% of the total mail volume (i.e., less than 1% of 32.3 billion mail pieces = 323 million mail pieces). Thus, the following three mail types were excluded from the analysis: Single Piece First Class Mail (FCM) Flats, Presort FCM Flats, and Bound Printed Matter Flats.

Zero districts exceeded the precision target levels for Single Piece FCM flats (excluded category), USPS Marketing Letters, and Presort FCM Flats. One district (1.49%), not always the same district, exceeded the precision target levels for USPS Marketing Flats, Presort FCM Letters and Cards, and Periodicals. Two districts (2.99%) exceed the precision target levels for Single-Piece FCM Letters and Cards.

Only 1 category exceeded the precision target level for more than 10% of districts: the excluded mail type Bound Printed Matter Flats. Bound Printed Matter Flats exceeded the targeted level (1%) in 40 districts (40/66 = 60.61%, since the Caribbean District was not applicable).

Audit measure 15 is considered achieved if—for each mail type—10% or less of the district end-to-end margins of error are greater than or equal to the target unsigned margins of error presented in Tables 4 and 5 above. To be considered partially achieved, no more than 20% of the district margins of error can be greater than or equal to the target unsigned margins of error for each mail type.

For Bound Printed Matter Flats, more than 20% of districts exceeded the target level, but this mail type is excluded from the analysis due to its low volume. For the remaining sample groups, at most 2.99% of the districts exceeded the targeted level.

Therefore, audit measure 15 can be considered achieved for FY20 Q2.

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

Quarter 2 Result: Achieved.

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 Postal Service’s system integrator provided a detailed description of how data are frozen.

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). Once we get to the day which we want to close the quarter, 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 FY20 Q2 the control date would be set to 03/31/2020. All aggregated data for service performance reporting is then unchanged after that point or "frozen." The reporting aggregates have varying 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 prior to the now closed reporting quarter. For verification purposes, these “frozen” data can then be accessed via a query of the reporting data within the retention period.

Therefore, audit measure 16 can be considered achieved for FY20 Q2.

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

Quarter 2 Result: Achieved.

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 a process to close

the quarterly reporting period is in place, which involves 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 Postal Service's system integrator provided a detailed response regarding the various steps to close the quarter. Throughout the quarter, system defects are tracked in ALM. Defects determined to have significant impact to 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 QA teams validate and fix until resolved. 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 as described in audit measure 16.

Therefore, audit measure 17 can be considered achieved for FY20 Q2.

Measure 18: First Mile—Between the first quarter and the end of the current quarter, the percentage of boxes selected for sampling at least one time should be more than the quarterly target percentage

Quarter 2 Result: Achieved.

Audit measure 18 evaluates whether First Mile data are representative by assessing if the execution of First Mile measurement processes yields representative results. Specifically, measure 18 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 over the fiscal year that results in valid samples. This measurement allows for the identification of any systematic non-coverage of boxes.

To evaluate the compliance of audit measure 18, ICF requested data on the total number of collection boxes in the United States and the number sampled in FY20 Q2. In response, USPS submitted data showing the number of submitted requests and the percentage of eligible boxes requested by month for FY20.

The target percentage levels for audit measure 18 were previously calculated for the audit of FY17 Q4 using a simulation model. Using the sample targets for each ZIP3 and urban/rural combination together with the measured collection box densities, the simulation model simulated the process of daily sampling of collection boxes. For each simulated year of 303 delivery days, the model was used to calculate the number of times each collection box was sampled during each of the four quarters. The simplified simulation model used the average density across a week instead of using separate densities for each day. The model used the average daily number of sample requests across a week instead of using separate numbers of sample requests for each day. Finally, the model used a full static set of collection boxes instead of accounting for some boxes going in and out of service during the year.

The results for one simulated year were the following: 62% of boxes sampled at least once during Q1; 75% of boxes sampled at least once during Q1 and Q2; 81% of boxes sampled at

least once during Q1, Q2, and Q3; and 84% of boxes sampled at least once during the entire year. These percentages did not vary substantially across simulations. To account for the variability across simulations and for the simplifying assumptions, the target values for this measure were chosen to be the following: (1) 58% of boxes sampled at least once during Q1; (2) 70% of boxes sampled at least once during Q1 and Q2; (3) 76% of boxes sampled at least once during Q1, Q2, and Q3; and (4) 80% of boxes sampled at least once during the entire year. For the analysis of FY20 Q2, the applicable target percentage is thus 70%.

As noted in the PRC Scores Reports narratives for FY17, the impacts of Hurricanes Irma and Maria impeded service in the Caribbean district and Service Performance for mail originating from or destined to that district was suspended starting on September 16, 2017 and restarted in Q3. Therefore, the Caribbean district was excluded from the simulation modeling. However, a "what-if" analysis of the original simulation data shows that if the Caribbean district coverage rates for the entire year had equaled the lowest or highest of the coverage rates for the other districts, the national coverage rate for the year would have been between 84.2% and 84.3%. Thus, the impact on the annual target values for this measure, and by extension on the quarterly target values, is negligible.

Audit measure 18 is achieved in FY20 Q2 if—up to the second quarter of the year—more than 70% of boxes were selected for sampling at least once. The measure is partially achieved in FY20 Q2 if between 56% and 70% of boxes (i.e., 80% to 100% of the target percentage, 70%) were selected for sampling at least once.

The information shows that for FY20 Q2 of the 179,253 boxes eligible for sampling, 138,764 boxes (77.41%) were sampled during FY20 Q1 and Q2. Therefore, audit measure 18 can be considered achieved for FY20 Q2.

Measure 19: First Mile—Most response rates should exceed 80% at a district level

Quarter 2 Result: Partially Achieved.

Audit measure 19 analyzes the First Mile sampling representativeness by assessing whether sampling response rates meet district thresholds. Specifically, audit measure 19 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 19, 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 audit measure 19, ICF requested response rates for each district based on the response rate definition described above for audit measure 2. In response, USPS provided the Collection District Compliance Report for FY20 Q2. Based on these data, 60 of 67 districts (roughly 90%) met the 80% compliance threshold.

Therefore, audit measure 19 can be considered partially achieved for FY20 Q2 (defined as having between 50% and 95% of district response rates meeting the 80% threshold).

Measure 20: First Mile—Coverage ratios should meet acceptable thresholds at the 3-digit ZIP Code levels for districts with poor coverage

Quarter 2 Result: Partially Achieved.

Audit measure 20 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 underrepresented groups.

To evaluate the compliance of audit measure 20, ICF requested coverage ratios at the 3-digit ZIP code level for all district response rates that did not meet the 80% threshold. ICF also requested assessments performed to evaluate reasons for low response rates. In response, USPS provided the Collection Failed Compliance Report that included coverage ratios at the district and 3-digit ZIP code level using the same definitions as in audit measure 2. USPS did not provide any additional information about reasons for low response rates.

USPS analyzed the compliance rates at the 3-digit ZIP Code level for the 7 districts identified in audit measure 19 as not meeting the 80% district threshold. Table 6 shows the percentage of ZIP Codes meeting the 80% and 60% coverage levels within each of these 7 districts.

Table 6. Percentage of ZIP Codes Meeting 80% and 60% Thresholds for Districts with Poor First Mile Coverage for FY20 Q2

District	Percentage of ZIP3s Meeting 80% Threshold	Percentage of ZIP3s Meeting 60% Threshold
Chicago	66% (2/3)	100% (3/3)
Gateway	65% (13/20)	90% (18/20)
Philadelphia Metropo	40% (2/5)	100% (5/5)
Long Island	80% (4/5)	100% (5/5)
Lakeland	50% (9/18)	89% (16/18)
New York	50% (1/2)	100% (2/2)
Los Angeles	43% (3/7)	86% (6/7)

Audit measure 20 is achieved if: (1) at least 95% of district response rates exceed 80% or (2) for each district with a response rate below 80%, either the response rate is at least 80% for 80% of ZIP3s, or reasons for the low response rates are provided. Audit measure 20 is partially achieved if (1) less than 95% of district response rates exceed 80% and (2) for each district with a response rate below 80%, either the response rate is at least 60% for 60% of ZIP3s or reasons for the low response rates are provided. Audit measure 20 is not achieved if (1) less than 95% of district response rates exceed 80%, and (2) for at least one district with a response rate below 80%, the response rate is at least 60% for less than 60% of the ZIP3s, and the reasons for low response rates are not provided.

Of the 67 districts, 60 (90%) had response rates above 80% as described in audit measure 19. For some of the 7 districts with response rates below 80%, less than 80% of the ZIP3s had response rates at or above 80%, and for each district with a response rate below 80%, the response rate was at least 60% for 60% (or more) of the ZIP3s. Reasons for the low response rates were not provided.

Therefore, audit measure 20 can be considered partially achieved for FY20 Q2.

Measure 21: First Mile—At least 95% of eligible collection points in CPMS should be measured in the profile

Quarter 2 Result: Achieved.

Audit measure 21 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 21, USPS must validate whether all or the vast majority of 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 or most are represented in the profile. If not, USPS must determine the extent of missing points.

To evaluate the compliance of audit measure 21, ICF requested that USPS tabulate the total number of collection points in the United States and the calculated First Mile profiles for FY20 Q2. In response, the Postal Service's system integrator provided the information presented in Table 7, which is the number of collection points that received collection box scans in FY20 Q2 and the number of those collection points that provided usable First Mile sample scans in Q2. Note that the numbers of eligible collection points for measures 18 and 21 are not the same because different eligibility criteria are applicable.

Table 7. Scans of Eligible First Mile Collection Points for FY20 Q2

Eligible Collection Points	Points with Collection Box Scans		Points with Usable Sample Scans	
Count	Count	% of Eligible	Count	% of Eligible
181,247	180,248	99.45%	77,896	42.98%

If at least 95% of eligible collection points in CPMS are measured in the profile, the measure is achieved. Otherwise, if at least 50% of eligible collection points in CPMS are measured in the profile, the measure is partially achieved.

Therefore, audit measure 21 can be considered achieved for FY20 Q2.

Measure 22: First Mile—Most eligible retail locations should contribute data to the profile for some dates and mail types in the quarter

Quarter 2 Result: Achieved.

Audit measure 22 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 22, 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 audit measure 22, ICF requested a table summarizing the number of retail locations in the United States and the number with some mail included in the calculated First Mile profiles. USPS provided a table that showed there were 15,881 retail facilities identified as having a point of sale (POS) system, of which 15,691 (98.80%) provided retail scan data to the First Mile profiles.

Audit measure 22 is achieved if at least 95% of eligible retail facilities provided retail mail scans to the First Mile profiles. Otherwise, audit measure 22 is partially achieved if at least 50% of eligible retail facilities provided retail mail scans to the First Mile profiles.

Therefore, audit measure 22 can be considered achieved for FY20 Q2.

Measure 23: Processing Duration—At least 70% of the volume is measured for each product

Quarter 2 Result: Partially Achieved.

Audit measure 23 analyzes the Processing Duration sampling representativeness by assessing how much of the volume is included in measurement for each measured product. To conduct the review of audit measure 23, USPS must validate whether at least 70% of the volume is measured for each product. To do this, USPS must take the total measured volume for the quarter and the total population pieces from the Revenue Piece Weight (RPW) for each product (i.e., PRC product reporting levels) and calculate the percentage of mail in the measurement.

To evaluate the compliance of audit measure 23, ICF requested the total measured volume for each quarter for each product and the total pieces for each product using all mail or full-service mail only. In response, USPS provided the information presented in Table 8.

ICF divided the total measured volume for each quarter for each product by the total pieces from RPW for each product to evaluate the compliance of this audit measure for Single-Piece First Class Mail, Every Door Direct Mail, and High Density and Saturation Flats/Parcels Destination Entry Two-Day. For all other product types, ICF divided the Full-Service total measured volume for each quarter for each product by the Full-Service pieces from *PostalOne!* for each product. Table 8 presents the results of this calculation for each product. Only the applicable ratios for this measure are shown in the last two columns.

As presented in Table 8, the 70% coverage level was achieved for 7 of the 13 products (54%).

Table 8. Processing Duration Measured Volumes, including Full-Service Volumes

Mail Category	RPW Volume (A)	*PostalOne!* Full-Service Volume (B)	Internal SPM Volume (C)	Full-Service Internal SPM Volume (D)	C/A (%)	D/B (%)
Total First Class Mail	**14,205,849,076**	**9,394,938,749**	**9,621,301,667**	**6,550,321,309**	**67.7%**	
Single-Piece Letters/Postcards	3,854,130,219	N/A	2,898,463,156	N/A	75.2%	
Presort Letters/Postcards	10,038,099,122	9,280,550,234	6,474,023,588	6,474,023,588		69.8%
Single-Piece Flats	180,432,374	N/A	172,517,202	N/A	95.6%	
Presort Flats	133,187,361	114,388,515	76,297,721	76,297,721		66.7%
Total USPS Marketing Mail	**17,039,835,081**	**13,603,453,040**	**12,220,972,114**	**11,421,918,375**	**71.7%**	**84.0%**
High Density and Saturation Letters	1,686,342,488	1,588,421,159	1,325,328,853	1,325,328,853		83.4%
High Density and Saturation Flats/Parcels excl. Destination Entry Two-Day	1,359,824,947	469,811,399	314,851,839	314,851,839		67.0%
High Density and Saturation Flats/Parcels Destination Entry Two-Day	1,451,023,549	N/A	711,533,478	N/A	49.0%	
Carrier Route (Flats and Letters)	1,303,481,985	1,256,290,014	1,023,331,461	1,023,331,461		81.5%
Letters	10,268,035,568	9,566,501,918	8,177,795,394	8,177,795,394		85.5%
Flats	832,512,826	722,428,550	580,610,828	580,610,828		80.4%
Every Door Direct Mail	138,613,718	N/A	87,520,261	N/A	63.1%	
Total Periodicals	**1,020,817,587**	**773,678,075**	**564,399,655**	**564,399,655**		**73.0%**
Total Package Services	**56,416,770**	**11,621,453**	**5,624,411**	**5,624,411**		**48.4%**
Bound Printed Matter Flats	56,416,770	11,621,453	5,624,411	5,624,411		48.4%
Total for All Mail Types	**32,322,918,514**					

Audit measure 23 can be considered partially achieved for FY20 Q2 because more than 50% of the products—but not 100% of the products—achieved the 70% coverage level.

Measure 24: Processing Duration—Most active ZIP codes should have mail receipts for all products during the quarter

Quarter 2 Result: Achieved.

Audit measure 24 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 24, 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 the full frame.

To evaluate the compliance of audit measure 24, ICF requested that USPS tabulate the volume of processed mail in the Processing Duration data by product and destination ZIP code. In response, USPS provided the processing volumes for each mail product and 5-Digit ZIP code.

USPS analyzed the complete FY20 Q2 data for this metric by major product using the full frame of 40,684 active 5-Digit ZIP codes and provided the results presented in Table 9.

Table 9. Processing Duration by Product

Metrics for Criteria	Single-Piece First-Class Letters	Presort First-Class Letters	First-Class Flats	USPS Marketing Mail Letters	USPS Marketing Mail Flats	Periodicals	Bound Printed Matter Flats
Total Eligible ZIP Codes with Measurable Mail Pieces	40,684	40,396	40,666	40,387	40,281	39,947	36,880
Having Mail Pieces Processed from the Service Area of Every Destination ZIP Code	100.00%	99.29%	99.96%	99.27%	99.01%	98.19%	90.65%

Audit measure 24 is considered achieved if at least 95% of ZIP codes provide measured data for all products other than Bounded Printed Matter Flats (BPMF) and if at least 85% of ZIP codes provide measured data for BPMF. Because of the make-up of the BPMF product and how this product is handled in processing, the achievement threshold for BPMF has been set at the lower value of 85%. This mail is inducted as flat mail pieces but is often not eligible to run on the AFSM machines. When this happens, decisions are made in processing to manually process this product directly to the 5 digits with no scanning taking place, instead of running this on Small Parcel Bundle Sorters (SPBS), and therefore those mail pieces are not included in the measurement. Audit measure is considered partially achieved if at least 50% of ZIP codes provide measured data for every product.

Based on the information provided by USPS, audit measure 24 can be considered achieved for FY20 Q2 because at least 95% of ZIP codes provided measured data for every product except BPMF, while more than 85% of ZIP codes provided measured data for BPMF.

Measure 25: Last Mile—Most response rates should exceed 80% at the District level

Quarter 2 Result: Partially Achieved.

Audit measure 25 analyzes Last Mile sampling representativeness by assessing whether sampling response rates meet district thresholds. Specifically, audit measure 25 is intended to determine whether execution of the Last Mile measurement process yields representative results by ensuring that non-response biases are immaterial. To conduct the review of audit

measure 25, 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 audit measure 25, ICF requested response rates for each district based on the response rate definition given above for audit measure 5. In response, USPS provided the Delivery District Compliance Report for FY20 Q2. Based on these data, 63 of 67 districts (roughly 94%) met the 80% compliance level for FY20 Q2.

Therefore, audit measure 25 can be considered partially achieved for FY20 Q2 (defined as having between 50% and 95% of district response rates meeting the 80% threshold).

Measure 26: Last Mile—Coverage ratios should meet acceptable thresholds at the 3-digit ZIP Code levels for districts with poor coverage

Quarter 2 Result: Partially Achieved.

Audit measure 26 analyzes Last 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 underrepresented groups.

To evaluate the compliance of audit measure 26, ICF requested coverage ratios at the 3-digit ZIP code level for all district response rates that did not meet the 80% threshold. ICF also requested assessments be performed to evaluate reasons for low response rates. In response, USPS provided the Delivery District Compliance Report for FY20 Q2 that included coverage ratios at the district and 3-digit ZIP code level using the same definitions as in audit measure 5. USPS provided no additional information about reasons for low response rates.

USPS analyzed the compliance rates at the 3-digit ZIP Code level for the 4 districts identified in audit measure 25 as not meeting the 80% district threshold. Table 10 shows the percentage of ZIP Codes meeting the 80% and 60% coverage levels within each of these 4 districts.

Table 10. Percent of ZIP Codes Meeting 80% and 60% Thresholds for Districts with Poor Last Mile Coverage for FY20 Q2

District	Percentage of ZIP3s Meeting 80% Threshold	Percentage of ZIP3s Meeting 60% Threshold
Baltimore	66% (6/9)	100% (9/9)
Chicago	67% (2/3)	100% (3/3)
New York	50% (2/4)	100% (4/4)
Triboro	57% (4/7)	100% (7/7)

Audit measure 26 is achieved if: (1) at least 95% of district response rates exceed 80%, or (2) for each district with a response rate below 80%, either the response rate is at least 80% for 80% of ZIP3s or reasons for the low response rates are provided. Audit measure 26 is partially achieved if: (1) less than 95% of district response rates exceed 80%, and (2) for each district with a response rate below 80%, either the response rate is at least 60% for 60% of ZIP3s or justifications for the low response rates are provided.

Of the 67 districts, 63 (94.0%) had response rates above 80%. For some (in fact, all) of the 4 districts with response rates below 80%, less than 80% of the ZIP3s had response rates at or above 80%, and for all of the 4 districts with response rates below 80%, at least 60% of the

ZIP3s had response rates at or above 60%. Reasons for the low response rates were not provided.

Therefore, audit measure 26 can be considered partially achieved for FY20 Q2.

IV. Summary of Audit Compliance Review

Table 11 summarizes the results of the audit compliance reviews for FY20 Q2. For the FY20 Q2 analysis, ICF classified 19 measures as achieved, 6 measures as partially achieved, and 1 as not achieved.

Table 11. Audit Compliance Review Summary

Measure	Phase	Audit Subject	Audit Criteria	FY20 Q2
1	First Mile	Is First Mile sampling accurately completed by carriers?	Procedures for sampling should be written and training provided to employees responsible for performing sampling.	Achieved
2	First Mile	Is First Mile sampling accurately completed by carriers?	Carrier sampling weekly compliance rates should consistently exceed 80 percent for most districts	Partially Achieved
3	First Mile	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.	Achieved
4	Last Mile	Is Last Mile sampling accurately completed by carriers?	Procedures for sampling should be written and training provided to employees responsible for performing sampling.	Achieved
5	Last Mile	Is Last Mile sampling accurately completed by carriers?	Carrier sampling weekly compliance rates should consistently exceed 80 percent for most districts	Achieved
6	Reporting	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.	Achieved
7	Reporting	Are reporting procedures and requirements established and being executed per design to produce accurate results?	Exclusions, exceptions, and limitations should be documented in the Internal SPM system and the final reports.	Achieved
8	Reporting	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?	A documented approval process should be in place and be followed for all manual/special exclusions and exceptions and for adding and/or changing exclusions or other business rules.	Achieved
9	First Mile	Is use of imputations for FM Profile results limited to provide FM measurement that represents the district's performance?	Most districts should have a limited volume for which imputed results are used within the quarter.	Achieved

Measure	Phase	Audit Subject	Audit Criteria	FY20 Q2
10	First Mile	Is use of proxy data for FM Profile results limited to provide FM measurement that represents the district's performance?	Most districts should have a limited volume for which proxy results are used within the quarter.	Achieved
11	Last Mile	Is use of imputations for LM Profile results limited to provide LM measurement that represents the district's performance?	Most districts should have a limited volume for which imputed results are used within the quarter.	Not Achieved
12	Last Mile	Is use of proxy data for LM Profile results limited to provide LM measurement that represents the district's performance?	Most districts should have a limited volume for which proxy results are used within the quarter.	Achieved
13	Reporting	Are changes to SPM documented and available for reference?	Program and SPM changes are documented in an Internal SPM repository for reference.	Achieved
14	Reporting	Are changes to SPM documented and available for reference?	PRC Reports denote major methodology and process changes in quarterly results.	Achieved
15	Reporting/ Processing Duration	Does the Internal SPM system produce reliable results?	For each product measured, the on-time performance scores should have margins of error lower than the designed maximums for the quarter.	Achieved
16	Reporting	Do processes exist to store and maintain official results reliably?	Processes should be established for storing final quarterly results	Achieved
17	Reporting	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.	Achieved
18	First Mile	Do the sampling results indicate that all collection points were included (districts, ZIP codes, box types, box locations)?	Between the first quarter and the end of the current quarter, the percentage of boxes selected for sampling at least one time should be more than the quarterly target percentage.	Achieved
19	First Mile	Are the sampling response rates sufficient to indicate that non-response biases are immaterial? If not, does the data indicate differences in performance for under-represented groups?	Most response rates should exceed 80% at a district level.	Partially Achieved
20	First Mile	If the sampling response rates do not meet the 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 levels for districts with poor coverage.	Partially Achieved
21	First Mile	Are all valid collection points included in the collection profile (collection points, ZIP	Most eligible collection points in CPMS should be measured in the profile.	Achieved

Measure	Phase	Audit Subject	Audit Criteria	FY20 Q2
		codes and collection dates)?		
22	First Mile	Are all retail locations included in the final retail results for all shapes, dates, and ZIP codes?	Most eligible retail locations should contribute data to the profile for some dates and mail types in the quarter.	Achieved
23	Processing Duration	How much of the volume is included in the measurement for each measured product?	At least 70% of the volume is measured for each product.	Partially achieved
24	Processing Duration	Are all destinating ZIP codes and dates represented in the final data?	Most active ZIP codes should have mail receipts for all products during the quarter.	Achieved
25	Last Mile	Are the sampling response rates sufficiently high to indicate that non-response biases are immaterial?	Most response rates should exceed 80% at a district level.	Partially Achieved
26	Last Mile	If the sampling response rates do not meet the 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 levels for districts with poor coverage.	Partially Achieved

V. Conclusion

USPS has completed its migration to the Internal Service Performance Measurement (SPM) system, 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 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 of the evaluation of the accuracy, reliability, and representativeness of the sampling. To perform the audit compliance review, ICF examined data and information describing 26 audit measures designed to ensure that the sampling process is conducted appropriately.

As summarized in Table 11 for FY20 Q2, ICF classified 19 measures as achieved (measures 1, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 21, 22, and 24), 6 measures as partially achieved (measures 2, 19, 20, 23, 25, and 26), and 1 measure as not achieved (11).

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

2. Improvement Plan

ICF recommends changes to improve the compliance of the audit measures. Table 12 summarizes our audit-specific recommendations following the results of the FY20 Q2 audit compliance review.

Table 12. Audit Measure-Specific Recommendations to Achieve Compliance

Measure	Compliance Status	Recommendation
Measure 2	Partially achieved	Achieve consistent 80% weekly compliance rates for collection.
Measure 11	Not achieved	Reduce the need for imputed data for First Class Flats in Last Mile.
Measure 19	Partially achieved	Increase collection response rates at the district level.
Measure 20	Partially achieved	Increase collection coverage ratios for districts with poor coverage.
Measure 23	Partially achieved	Achieve 70% coverage for most products.
Measure 25	Partially achieved	Increase delivery response rates at the district level.
Measure 26	Partially achieved	Increase delivery coverage ratios for districts with poor coverage.

3. Study Limitations

For this audit compliance review, ICF only analyzed summary data USPS provided. 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
  - Partially achieved
    - Measure 2: Investigate ways to achieve consistently high collection compliance rates.
    - Measures 19 and 20: Investigate ways to increase collection compliance rates.
    - Measures 25 and 26: Investigate ways to increase delivery compliance rates.
  - Not achieved
    - Measure 11: Reduce the need for imputed data for First Class Flats in Last Mile.
- New data collection required
  - Partially achieved
    - Measure 23: Investigate ways to increase Internal SPM volumes relative to RPW or *PostalOne!* volumes.

Appendix A. Compliance Categorization Scheme

Measure	Phase	Audit Criteria	Compliance Determination Cutoff		
			Achieved	Partially Achieved	Not Achieved
1	First Mile	Procedures for sampling should be written and training provided to participants responsible for performing sampling.	1. In at least 80% of districts, training is provided to at least 75% of participants responsible for performing sampling. 2. Written sampling plans and training materials are up-to-date and consistent. Both 1 and 2.	1. In at least 80% of districts, training is provided to at least 75% of participants responsible for performing sampling. 2. Written sampling plans and training materials are up-to-date and consistent. Either 1 or 2 but not both.	1. In at least 80% of districts, training is provided to at least 75% of participants responsible for performing sampling. 2. Written sampling plans and training materials are up-to-date and consistent. Neither 1 nor 2.
2	First Mile	`Carrier sampling weekly compliance rates should consistently exceed 80 percent for most districts.	For at least 80% of districts, either all the weekly compliance rates are at least 80%, or the reasons for low compliance are investigated.	For between 50 and 80% of districts, either all the weekly compliance rates are at least 80%, or the reasons for low compliance are investigated.	For less than 50% of districts, either all the weekly compliance rates are at least 80%, or the reasons for low compliance are investigated.
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.	Density tests were performed in the last 12 months on at least 95% of the active collection points in the audited quarter.	Density tests were performed in the last 12 months on between 80 and 95% of the active collection points in the audited quarter.	Density tests were performed in the last 12 months on less than 80% of the active collection points in the audited quarter.
4	Last Mile	Procedures for sampling should be written and training provided to participants responsible for performing sampling.	1. In at least 80% of districts, training is provided to at least 75% of participants responsible for performing sampling. 2. Written sampling plans and training materials are up-to-date and consistent. Both 1 and 2.	1. In at least 80% of districts, training is provided to at least 75% of participants responsible for performing sampling. 2. Written sampling plans and training materials are up-to-date and consistent. Either 1 or 2 but not both.	1. In at least 80% of districts, training is provided to at least 75% of participants responsible for performing sampling. 2. Written sampling plans and training materials are up-to-date and consistent. Neither 1 nor 2.
5	Last Mile	Carrier sampling weekly compliance rates should consistently exceed 80 percent for most districts.	For at least 80% of districts, either all the weekly compliance rates are at least 80%, or the reasons for low compliance are investigated.	For between 50 and 80% of districts, either all the weekly compliance rates are at least 80%, or the reasons for low compliance are investigated.	For less than 50% of districts, either all the weekly compliance rates are at least 80%, or the reasons for low compliance are investigated.

Measure	Phase	Audit Criteria	Compliance Determination Cutoff		
			Achieved	Partially Achieved	Not Achieved
6	Reporting	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	Reporting	Exclusions, exceptions, and limitations should be documented in the Internal SPM system and the final reports.	Exclusions, exceptions, and limitations are well documented.	Exclusions, exceptions, and limitations are documented but poorly or incompletely.	Exclusions, exceptions, and limitations are not documented.
8	Reporting	A documented approval process should be in place and be followed for all manual/special exclusions and exceptions and for adding and/or changing exclusions or other business rules.	A documented approval process is in place and is followed for manual/special exclusions and exceptions for adding and/or changing exclusions for other business rules.	A documented approval process is in place but does not sufficiently explain the manual/special exclusions and exceptions for adding and/or changing exclusions for other business rules or was not followed for some requests.	Approval process lacks documentation.
9	First Mile	Most districts should have a limited volume for which imputed results are used within the quarter.	At most 1 district above 20% imputed data and at most 5 districts above 10% imputed data.	2-3 districts above 20% imputed data and at most 5 districts above 10% imputed data.	4+ districts above 20% imputed data or 6+ districts above 10% imputed data.
10	First Mile	Most districts should have a limited volume for which proxy results are used within the quarter.	At most 1 district above 20% proxy data and at most 5 districts above 10% proxy data.	2-3 districts above 20% proxy data and at most 5 districts above 10% proxy data.	4+ districts above 20% proxy data or 6+ districts above 10% proxy data.
11	Last Mile	Most districts should have a limited volume for which imputed results are used within the quarter.	At most 1 district above 20% imputed data and at most 5 districts above 10% imputed data.	2-3 districts above 20% imputed data and at most 5 districts above 10% imputed data.	4+ districts above 20% imputed data or 6+ districts above 10% imputed data.
12	Last Mile	Most districts should have a limited volume for which proxy results are used within the quarter.	At most 1 district above 20% proxy data and at most 5 districts above 10% proxy data.	2-3 districts above 20% proxy data and at most 5 districts above 10% proxy data.	4+ districts above 20% proxy data or 6+ districts above 10% proxy data.
13	Reporting	Program and SPM changes are documented in an Internal SPM repository for reference.	Program and SPM changes are documented in an Internal SPM repository for reference.	Changes are documented but incompletely.	Insufficient documentation provided.
14	Reporting	PRC Reports denote major methodology and	PRC Reports describe the major methodology and	Reports document methodology but do not sufficiently describe deviations.	Insufficient documentation provided.

Measure	Phase	Audit Criteria	Compliance Determination Cutoff		
			Achieved	Partially Achieved	Not Achieved
		process changes in quarterly results.	process changes in quarterly results.		
15	Reporting/ Processing Duration Data	For each product measured, the on-time performance scores should have margins of error lower than the designed maximums for the quarter.	For each mail type, either sufficient reasons for excluding that mail type are provided or 10% or less of the district margins of error are greater than or equal to the target unsigned margin of error.	For each mail type, either sufficient reasons for excluding that mail type are provided or at most 20% of the district margins of error are greater than or equal to the target unsigned margin of error. For at least one mail type, more than 10% is greater than or equal to the target unsigned margin of error.	For at least one mail type, sufficient reasons for excluding that mail type are not provided, and more than 20% of the district margins of error are greater than or equal to the target unsigned margin of error.
16	Reporting	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.
17	Reporting	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	First Mile	Between the first quarter and the end of the current quarter, the percentage of boxes selected for sampling at least one time should be more than the quarterly target percentage.	Between the first quarter and the end of the current quarter, the percentage of boxes selected for sampling at least one time is more than the quarterly target percentage.	Between the first quarter and the end of the current quarter, the percentage of boxes selected for sampling at least one time is more than 80%, but less than 100%, of the quarterly target percentage.	Between the first quarter and the end of the current quarter, the percentage of boxes selected for sampling at least one time is less than 80% of the quarterly target percentage.
19	First Mile	Most response rates should exceed 80% at a District level.	At least 95% of response rates exceed 80% at District level. A response means that the carrier responded correctly to an "eligible" sampling request.	Between 50% and 95% of response rates exceed 80% at District level.	Less than 50% of response rates exceed 80% at District level.
20	First Mile	Coverage ratios should meet acceptable	1. At least 95% of district response rates exceed 80%	1. Less than 95% of district response rates exceed 80% and 2. For each	1. Less than 95% of district response rates exceed 80%

Measure	Phase	Audit Criteria	Compliance Determination Cutoff		
			Achieved	Partially Achieved	Not Achieved
		thresholds at the 3-digit ZIP Code levels for districts with poor coverage.	or 2. For each district with a response rate below 80%, either the response rate is at least 80% for 80% of ZIP3s, or reasons for the low response rates are provided.	district with a response rate below 80%, either the response rate is at least 60% for 60% of ZIP3s or reasons for the low response rates are provided.	and 2. For at least one district with a response rate below 80%, the response rate is at least 60% for less than 60% of ZIP3s, and the reasons for the low response rates are not provided.
21	First Mile	Most eligible collection points in CPMS should be measured in the profile.	At least 95% of eligible collection points in CPMS are measured in the profile.	At least 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.
22	First Mile	Most eligible retail locations should contribute data to the profile for some dates and mail types in the quarter.	At least 95% of retail locations are measured in the profile.	At least 50% of retail locations are measured in the profile.	Less than 50% of retail locations are measured in the profile.
23	Processing Duration	At least 70% of the volume is measured for each product.	All products achieve 70% or greater processing duration data measurement.	50% or more of products exceed 70% coverage level.	Less than 50% of products achieve 70% coverage level.
24	Processing Duration	Most active ZIP Codes should have mail receipts for all products during the quarter.	For every product, at least 95% (85% for Bounded Printed Matter Flats) of destination ZIP5s provided some measured data.	For every product, at least 50% of destination ZIP5s provided some measured data, but for some products, less than 95% (85% for Bounded Printed Matter Flats) provided some measured data.	For every product, less than 50% of destination ZIP5s provided some measured data.
25	Last Mile	Most response rates should exceed 80% at a District level.	At least 95% of response rates exceed 80% at District level. A response means that an "eligible" sampling request was correctly responded to by the carrier.	Between 50% and 95% of response rates exceed 80% at District level.	Less than 50% of response rates exceed 80% at District level.
26	Last Mile	Coverage ratios should meet acceptable thresholds at the 3-digit ZIP Code levels for districts with poor coverage.	1. At least 95% of district response rates exceed 80% or 2. For each district with a response rate below 80%, either the response rate is at least 80% for 80% of ZIP3s, or reasons for the low response rates are provided.	1. Less than 95% of district response rates exceed 80% and 2. For each district with a response rate below 80%, either the response rate is at least 60% for 60% of ZIP3s or reasons for the low response rates are provided.	1. Less than 95% of district response rates exceed 80% and 2. For at least one district with a response rate below 80%, the response rate is at least 60% for less than 60% of ZIP3s, and the reasons for the low response rates are not provided.