

BEFORE THE
POSTAL REGULATORY COMMISSION
WASHINGTON, D.C. 20268-0001

SERVICE PERFORMANCE MEASUREMENT
SYSTEMS FOR MARKET DOMINANT PRODUCTS

Docket No. PI2015-1

**RESPONSES OF THE UNITED STATES POSTAL SERVICE TO
QUESTIONS 1-16 OF COMMISSION INFORMATION REQUEST NO. 1**
(June 12, 2017)

The United States Postal Service hereby provides its responses to the above-listed questions of Commission Information Request No. 1, issued on May 12, 2017.

Each question is stated verbatim and followed by the response.

Respectfully submitted,

UNITED STATES POSTAL SERVICE
By its attorneys:

Anthony Alverno
Chief Counsel, Global Business & Service
Development

B.J. Meadows III
Jeffrey A. Rackow
Attorneys

475 L'Enfant Plaza West, S.W.
Washington, D.C. 20260-1101
(202) 268-3009
June 12, 2017

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

1. On February 17, 2017, the Postal Service filed Library Reference USPS-LR-PI2015-1/7, which includes the Postal Service's audit plan for the proposed internal service performance measurement system for market dominant products. Within the audit plan, the Postal Service states that it will contract the audit task to a third party. Please provide the following information:
 - a. The contractual statement of work describing the auditing and reporting tasks required of the contractor.
 - b. All documentation developed by the contractor explaining how it intends to perform audits and develop reports.
 - c. The "trial run" audit report provided by the contractor.
 - d. The first official audit report encompassing at least one quarter's data provided by the contractor.

RESPONSE:

- a) As noted in this question, on February 17, 2017, the Postal Service filed an audit plan dated February 10, 2017, which indicated that the Postal Service plans to procure a third party vendor to conduct the internal SPM audit. The requested contractual statement of work for that audit is not yet completed. If the Commission would like to review the latest Postal Service draft of the statement of work for that audit, the Postal Service would be happy to file the draft under seal upon request. As noted below in the response to Question 1(c), the Postal Service is providing the Commission with the proof of concept audit report performed by a third-party contractor for Quarter 1 of FY 2017.
- b) The contractor for that audit has not yet been selected; see response to Question 1(a).
- c) The proof of concept audit report (CIR.1.Q.1c.Audit.Report.Concept.pdf) is being filed as an electronic attachment to this response.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

- d) The contractor for that audit has not yet been selected; see response to Question 1(a).

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

2. The Postal Service has alerted the Commission to the possibility that service performance measurements under the proposed system and service performance measurements obtained under the current system may not match. Where the actual measurements obtained under the proposed system fall outside the margin of error of the actual measurements obtained under the legacy system, please provide the following:
- a. An analysis and discussion of whether or not the differences are statistically significant for each individual item under measurement.
 - b. An analysis and discussion addressing why the results may be different for each individual item under measurement (including a discussion of the representativeness and accuracy of measurements obtained using each measurement system).

RESPONSE:

- a. The analysis of the results for FY 2017 Quarter 2 compared the following origin/destination combined on-time performance scores:
 - District and national Single-Piece First-Class Mail Letters/Cards for Two-Day and Three-to-Five-Day service standard groups
 - District, area, and national First-Class Mail Flats for Overnight, Two-Day, and Three-to-Five-Day service standard groups
 - District, area, and national Presort First-Class Mail Letters/Cards for Overnight, Two-Day, and Three-to-Five-Day service standard groups
 - District, area, and national USPS Marketing Mail High Density and Saturation Letters for all service standards and entry types combined
 - District, area, and national USPS Marketing Mail High Density and Saturation Flats for all service standards and entry types combined

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

- District, area, and national USPS Marketing Mail Carrier Route for all service standards and entry types combined
- District, area, and national USPS Marketing Mail Letters for all service standards and entry types combined
- District, area, and national USPS Marketing Mail Flats for all service standards and entry types combined
- Area and national Periodicals for all service standards combined
- District, area, and national Bound Printed Matter Flats for all service standards and entry types combined

The following table shows the results of the comparisons.

Quarter 2 FY2017	# Comparisons		# IV outside the legacy Margin of Error		# Significantly Different	
	District	Area/Nation	District	Area/Nation	District	Area/Nation
Presort First-Class Letters	200	3	157	3	156	3
First-Class Mail Flats	200	24	105	16	105	16
USPS Marketing Mail Carrier Route	67	8	34	4	30	4
USPS Marketing Mail Flats	67	8	44	7	39	7
USPS Marketing Mail High Density & Saturation Flats	67	8	40	6	27	4
USPS Marketing Mail High Density & Saturation Letters	67	8	53	8	53	8
USPS Marketing Mail Letters	67	8	56	8	56	8
Periodicals	0	8	0	6	0	6
Bound Printed Matter Flats	67	8	33	6	22	6
Total	802	83	522	64	488	62

In total, 802 district level score comparisons were made, and 522 (65 percent) internal SPM on-time scores fell outside the margin of error of the

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

legacy system. Further comparisons showed that 488 of these 522 internal SPM scores, or 93 percent, had statistically significant differences. For the 83 area and national scores compared, 64 internal SPM scores fell outside the margins of error of the legacy system, with 62 of the 64 significantly different.

- b. Of importance to note is that the margins of error are generally quite small for both internal SPM and the legacy systems, resulting in small score differences identified as statistically significant using the standard testing for differences between means. For example, the national scores fell within +/- 0.2 for several products and yet were statistically different due to the small margins of error. Small differences may not be meaningful even if they are statistically significant when comparing the sampling estimates using only sampling error. Neither the legacy system nor the internal SPM system is measuring every piece of mail from collection/mail acceptance to delivery; thus, estimates from both are subject to error, some of which is due to sampling and some due to non-sampling. Sources of non-sampling error exist in both the legacy and internal SPM systems, and while not quantifiable, surely contribute to some differences.

It is also important to remember that there are substantive differences between Internal SPM and the current legacy measurement system.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

Therefore, there is no expectation that the service scores can or will be identical for each product between the two systems.

Parallel testing and comparison analyses to date indicate that there are differences in both the Processing Duration phase as well as First Mile and Last Mile which contribute to the differences. Processing Duration differences for commercial products stem from differences between the legacy system and internal SPM in processing the data. While national level differences tend to be small, some products have larger Processing Duration differences for product and service standard combinations. Differences in Last Mile represent differences in methodology as well as differences in the sampling data collected. There are more substantive measurement methodology differences between EXFC and internal SPM for measuring Single-Piece First-Class Mail.

For Presort First-Class Mail Letters/Cards, the national scores differed by approximately 1 point for each service standard for FY2017 Quarter 2. The national level last mile delivery factors differed by about 1 point for each service standard, with the internal SPM system indicating longer times in last mile than for the legacy system.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

USPS Marketing Mail Letters showed a similar pattern of the national legacy score being 0.71 point higher than internal SPM. The difference in delivery factors between the two systems was approximately the same, indicating that at the national level, the primary contributing factor is longer times observed in Last Mile in internal SPM than in the legacy system. Similarly, USPS Marketing Mail High Density and Saturation Letters had a legacy score that was 0.85 point higher than internal SPM, with the delivery factor difference being 0.8 point larger for internal SPM.

USPS Marketing Mail High Density and Saturation Flats, Carrier Route Flats, and Flats, Bound Printed Matter Flats, and Periodicals all had national level differences that were quite small for Quarter 2, each less than 0.3.

Differences for First-Class Mail Single-Piece Letters/Cards and Flats were larger than those for commercial products at national, area, and district levels. Differences can be seen across all three transit legs, and are discussed in the response to Question 4 below.

Differences for some district level and granular service standard level scores were larger than the national level differences. Analyses into the systematically larger internal SPM First Mile and Last Mile differences and into the largest differences between district level scores are ongoing. These

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

analyses are focusing on identifying the sources of differences. For example, it appears that some of the reason for differences in scores is due to longer times in Last Mile for sampled pieces to Post Office Box addresses in Internal SPM combined with the over-representation of those address types in the sample as a result of differing response rates. Further explanation of these issues can be found in the response to Question 9.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

3. The Postal Service prefaces each set of quarterly data generated under the proposed system with a narrative which identifies possible limitations or concerns with the reported data. Please provide:
- a. A current list of limitations, concerns, or unresolved issues for each item under measurement.
 - b. A discussion of the impact on the accuracy, reliability, and representativeness of measurements for each issue identified in a.
 - c. An estimation of when each issue identified in a. will be resolved.

RESPONSE:

- a) The following is a list of limitations, concerns, or unresolved issues:
 - 1. Limited flats available in collection for SPFC measurement
 - 2. Calculating First Mile profiles for pieces first observed in incoming mail processing operations rather than expected outgoing operations
 - 3. Calculating Last Mile profiles using unweighted profiles where all sampled mailpieces were weighted equally
 - 4. Incomplete processing inventory
- b) The following addresses the impact on accuracy, reliability and representativeness of measurements for issues identified in response to Question 3a:
 - 1. In FY2017 Quarter 2, the Postal Service began using retail data for SPFC flats to vastly increase the amount of flats in measurement in the first mile. There are enough retail flats with Special Services to meet statistical significance for SPFC flats overall. In FY2017 Quarter 3, the

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

Postal Service will implement a new imputation process that allows collection flats data to be imputed from retail flats data when not enough collection flats exist. This imputation will be done at the district level, which will be a more accurate representation of the first mile performance of that location than imputing collection profiles from the Area or National data.

2. In FY2017 Quarter 3, the Postal Service will complete its process for handling pieces first observed in incoming mail operations. Step 1 was ensuring there was enough volume in the first mile measurement, which adding the retail data ingest accomplished. Step 2 is to begin using those data in Quarter 3 to map pieces first seen in incoming operations back to their actual origin. This will create unique first mile profiles for this volume, and classify the volume into its correct service standard (since much of this volume is first seen at destination, rather than at origin). In addition, the Postal Service is adding logic in Quarter 3 to utilize more barcodes (including special services and IBI) to create unique pieces in situations where mail processing equipment did not spray a barcode on first processing. This will remove pieces improperly classified as first seen on incoming operations, and put them back into the normal flow, improving the accuracy of first mile profiles.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

3. Calculating Last Mile profiles using unweighted data was done due to problems in following the standard operating procedure for P.O. Box Last Mile Sampling. The process, when not followed, will produce false failures in situations where employees are scanning mail that was delivered the day prior but not yet picked up by the customer. The significantly higher last mile failure rates in P.O. Box Sampling is largely due to this process execution issue, and USPS Customer Service Operations is actively engaged in trying to improve the execution of this process. The impact, when weighting this volume, is that the P.O. Box Sampling errors end up with outsized weights, since they often have lighter volumes. This results in having an outsized negative impact on the service due to this negative bias, and therefore was not representative of true performance. The Postal Service expects to resume weighting in FY2017 Quarter 3, and monitors the improvement in the P.O. Box Sampling process.
4. Processing inventory for P.O. Box Sampling has been calculated three hours earlier than on street sampling due to some technical limitations. Incomplete inventory affects the weights that were discussed in item 3 above. The Postal Service has resolved one of the technical limitations and is working towards moving the inventory cutoff time to as close to the same as on street sampling as possible in FY2017

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

Quarter 3. This will improve inventory accuracy, as well as improve the ability to implement the weighted Last Mile profiles in Quarter 3.

c) The Internal SPM will undergo continuous improvement similar to legacy measurement systems to ensure that the system is accurate, reliable and representative. Expected resolution timelines for issues identified in Question 3a are as follows:

1. The issue of limited flats available in collection was resolved in FY2017 Quarter 2 via the utilization of the retail Special Services data.
2. Calculating first mile profiles for pieces first observed in incoming operations will be resolved in FY2017 Quarter 3.
3. The Postal Service's goal is to utilize weighted Last Mile profiles in FY2017 Quarter 3.
4. Incomplete processing inventory will be resolved in FY2017 Quarter 3.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

4. On August 11, 2016, the Postal Service filed Library Reference USPS-LR-PI2015-1/3, which consists of a disaggregation of the Single-Piece First-Class Mail service performance data broken down by First Mile, Processing Duration, and Last Mile.
- a. Please provide equivalent information for FY 2016 Quarter 4, FY 2017 Quarter 1 and Quarter 2.
 - b. Identify and explain all statistically significant differences between the first mile, last mile, and processing times derived from the proposed measurement system and EXFC data.

RESPONSE:

- a) The spreadsheet that is being electronically filed as CIR.1.Q.4a.SPM SPFC Scores.xlsx with this response contains requested data for FY16Q4, FY17Q1, and FY17Q2.

- b) The Library Reference USPS-LR-PI2015-1/3 provides information on the on-time performance of single-piece mail in Internal SPM, disaggregated into First Mile, Processing Duration, and Last Mile segments. This information was provided by shape, aggregated across the Two-Day and Three-To-Five-Day service standards groups. There are differences in volumes and on-time performance for Single-Piece letters and flats between service standard groups in the measurement systems that affect performance when it is viewed across service standards. These differences are due to natural volume distribution and performance differences, as well as in part due to design issues for Internal SPM which are currently being addressed, for example as noted in the response to Question 12. When comparing the

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

Internal SPM data to legacy EXFC system performance, it is more appropriate, therefore, to conduct the comparisons separately for Two-Day and Three-To-Five-Day service standard groups, as was done for the purposes of this response.

The data in the spreadsheet that is being electronically filed with this response as CIR.1.Q.4b.Supporting Analyses.xlsx contain the national level data for Quarter 4 FY2016 and Quarters 1 and 2 of FY2017 for Internal SPM and EXFC. These data show that there are differences in all three transit legs for both letters/cards and flats. There is no methodology defined to calculate margins of error for each transit leg component in EXFC. However, the overall margins of error are sufficiently small and the differences are sufficiently large to indicate that nearly all differences are likely significant. For example, Processing Duration scores for Two-Day letters/cards differed by 1.37 in Quarter 4, with First Mile Impact differing by 0.54 and Last Mile Impact differing by 1.02 to account for the 2.9 point overall score difference. In Quarter 1, Last Mile differences were smaller, resulting in smaller differences in the overall scores.

Detailed analyses of Processing Duration differences are still ongoing, but it is anticipated that the EXFC design may contribute to some differences in this area. As an example, EXFC test mail volumes are relatively even throughout

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

the weeks of a quarter based on the sample design, while actual mail volumes vary significantly in some time periods, for example, in the peak holiday season. To the extent that service performance differs during such periods, overall estimates may be impacted. Differences in mail volumes between origin and destination district pairs as measured in internal SPM versus how they are designed in EXFC may also contribute to differences in scores. Further analyses are planned following the incorporation of the design changes described in the response to Question 12.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

5. The Postal Service provided service performance data from the proposed system for Quarter 2 and Quarter 3 of FY 2016. Please provide the statistical data that normally accompanies service performance reports (e.g., margins of error, confidence intervals, *etc.*). If this information is still under development, please indicate when the Commission might expect the data to be filed.

RESPONSE:

Requested data are not available for FY16Q2 and FY16Q3, as the Postal Service did not calculate the margin of error/confidence intervals for this time period, and it is not possible for the Postal Service to generate the requested data.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

6. For each quarter provided to date, please explain why the margins of error for some products are greater in the proposed system than in the legacy system.

RESPONSE:

There are a large number of service standard groupings used for USPS Marketing Mail™, Bound Printed Matter Flats, and Periodicals. However, the Margins of Error analyses below are based on one metric which represents the overall margins of error for that product, rather than individually assessing up to six service standards per product for all 67 Districts. Periodicals margin of error results were analyzed at the Area level.

When comparing the FY2017 margins of error, a majority of the proposed Internal Service Performance Measurement (SPM) system margins of error are less than or equal to the Legacy SPM system. This is generally expected because the sampling rates are higher for internal SPM than for the legacy system. However, there are cases where the proposed system’s margins of error are greater than the legacy system.

Number of Districts where Internal SPM System Margins of Error are greater than in the Legacy SPM (where District level differences are greater than or equal to 0.1)

	USPS Marketing Mail™							First-Class® Flats			Bound Printed Matter Flats Overall	Periodicals Within County
	Carrier Route Overall	Flats Overall	HD & Sat. Flats Overall	HD & Sat. Letters Overall	Letters Overall	Mixed Flats Overall	Mixed Letters Overall	Overnight	Two-Day	Three-to-Five-Day		
FY17 Q1 Total	2	1	21	2	0	45	17	1	0	1	30	0
FY17 Q2 Total	2	4	38	6	2	54	29	1	0	0	33	0

In both Quarters 1 and 2, only one product, USPS Marketing Mail™ Mixed Flats Overall had a majority of Districts where the proposed system margins of error were greater than the legacy system, with 45/67 and 54/67 respectively. Bound Printed Matter Flats

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

Overall had the next highest rate of Districts with Internal SPM ranges greater than Legacy SPM with 30/67 and 33/67 respectively. USPS Marketing Mail™ High Density and Saturation Flats Overall (21/67 and 38/67) and Mixed Letters Overall (17/67 and 29/67) were the only other products with double digit results for both Q1 and Q2.

Furthermore, most of these differences are very small, with only a small number greater than or equal to 0.5, as shown in the table below:

Number of Districts where Internal SPM System Margins of Error are greater than in the Legacy SPM (where District level differences are greater than or equal to 0.5)

	USPS Marketing Mail™							First-Class® Flats			Bound Printed Matter Flats Overall	Periodicals Within County
	Carrier Route Overall	Flats Overall	HD & Sat. Flats Overall	HD & Sat. Letters Overall	Letters Overall	Mixed Flats Overall	Mixed Letters Overall	Overnight	Two-Day	Three-to-Five-Day		
FY17 Q1 Total	1	0	1	1	0	15	1	1	0	0	6	0
FY17 Q2 Total	1	1	4	0	0	15	1	1	0	0	11	0

There are two primary reasons why the Internal SPM margins of error are greater than the Legacy SPM results by 0.5 or more: sample size and difference in methodology.

Sample size:

Some districts have small sampling volumes in the proposed system due to the district size and mail volumes there and/or challenges sampling in the local area. Many of the individual instances, where only one district out of 67 is greater than 0.5 in the proposed system, are due to low sampling volume in that district. Specifically, the Caribbean District is the only district flagged in six out of the nine groups above. There are three instances where only one district has a margin of error which is greater than 0.5 or more in the proposed system, and those three districts (Alaska, Honolulu, and Mississippi)

RESPONSE OF THE UNITED STATES POSTAL SERVICE TO COMMISSION INFORMATION REQUEST NO. 1

also have a small sample size comparatively. All of these districts are generally known as having lower mail volumes than the other 63 districts. Likewise, Caribbean has had more technical challenges with their delivery sampling, resulting in lower sampling volumes than targeted in each quarter.

Differences in Methodology:

A difference in methodology between the legacy and proposed system is the second reason why some products have larger margins of error in the proposed system than in the legacy system. Specifically, in IMAPS the Last Mile Profile is calculated using all collected Last Mile data for the quarter and then applied to the available Processing Duration data, whereas the Internal SPM system calculates Last Mile Profiles for each day of the quarter, and applies to the Processing Duration for that specific date. As a result of this difference, for a product such as Mixed Flats, with very limited volumes in Full Service overall or for a specific service standard group, the internal SPM system may have lower sampled volume and greater variability in the estimates when many days of the quarter are not represented in the measured volume. USPS Marketing Mail Mixed Flats and Mixed Letters are measurement categories to account for very small mailings submitted without the electronic documentation that allows for placement into a product category. These volumes present well under 1 percent of measured USPS Marketing Mail volume. Bound Printed Matter Flats is also a low volume group which tends to have sporadic volumes, not necessarily encompassing every day of the quarter for every district.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

7. Please refer to Library References USPS-LR-PI2015-1/2, filed on August 10, 2016, and USPS-LR-PI2015-1/3, filed on August 11, 2016. The narrative tabs from First-Class Mail Single-Piece Letters and Postcards, First-Class Mail Flats, and Standard Mail state that the last mile methodology changed substantially from the initial statistical design, primarily by using unweighted pieces to construct the last mile profile.
- a. Please explain how the revised sampling methodology that uses an unweighted sample produces sufficiently representative results.
 - b. How has the measurement bias produced from unweighted samples in the last mile profile referenced in the narrative affect the accuracy, reliability, and representativeness of the internal service performance measurement results?

RESPONSE:

- a) It is the Postal Service's intention to utilize weighted last mile profiles. The Postal Service had to utilize unweighted profiles for two main reasons:
 - i. The first issue pertains to P.O. Box Sampling. Customer behavior with P.O. Boxes differs from street delivery in that P.O. Boxes are picked up less often. This creates more situations where "stale" mail is in the box. If the sampling process is not followed explicitly, an employee could very easily sample mail that was delivered a day or two before the actual sample date, thereby creating a statistically biased (and negative) last mile impact number. This has been the case at least through Quarter 2.
 - ii. The second reason is incomplete P.O. Box inventory. Due to technical reasons, through Quarter 2 the Postal Service has had to complete inventory for P.O. Box sampling three hours earlier than on-street

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

sampling, thereby creating situations where inventory business logic determines there is little to no volume in a box, while in actuality it might be quite heavy. As of Quarter 3, the technical issue has been resolved, and the Postal Service is working to move the P.O. Box inventory to as close to the on-street sampling inventory time as possible.

- b) These two issues combine to create a situation where P.O. Box Last Mile impacts are overstated, and then they are assigned higher weights in Last Mile profile calculations due to the inventory count that underrepresents their true volume. When applying the weighting methodology, where lighter volumes and samples are weighted more heavily, the negative bias in P.O. Box Sampling results is amplified and disproportionately affects overall performance. As stated above, the P.O. Box inventory issue will be addressed in Quarter 3. In addition, significant effort has been invested in correcting the issues in our sampling process so that the process accurately measures performance in Last Mile. The Postal Service is focused on correcting this in Quarter 3.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

8. Please refer to Library Reference USPS-LR-PI2015-1/6, filed February 16, 2017. The narrative tabs from First-Class Mail Single-Piece Letters and Postcards state that some technical issues with identifying and tracking unique mailpieces through automation based on processing scans existed during the quarter. Discuss the impact of issues associated with inaccuracies in processing inventories on the accuracy, reliability, and representativeness of service performance results produced by the internal Service Performance Measurement system.

RESPONSE:

Tracking mailpieces through automation involves having business logic to determine which barcodes, either those already on the mailpieces or those that are applied to the mailpieces, constitute unique pieces, and therefore the beginning of the Processing Duration measurement. In some cases, a barcode is not applied on the first scan in mail processing, and therefore may appear to be a first mile failure during sampling or may become a FPO2 piece if the next hit is on a destinating operation. In Quarter 3, the Postal Service is adding logic that utilizes either a Special Services barcode from retail or an IBI by itself to create a unique piece. This helps resolve the issue where the first scan does not always produce a barcode, and makes the measurement more accurate.

When the first scan of a mailpiece is missed, it increases first mile impact and/or shifts more volume into FPO2 when it truly was an FPO1 piece. This fix helps minimize that problem. Ensuring the volumes are correctly categorized as FPO1 or FPO2 has been a focus of continuous improvement efforts and the Postal Service has been successful in improving the accuracy, reliability and representativeness of service measurement.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

9. Library Reference USPS-LR-PI2015-1/2, Preface at 2, and Library Reference USPS-LR-PI2015-1/4 at 39, filed on August 26, 2016, state that both measurement systems (EXFC and the proposed internal system) are subject to sampling and non-sampling error.
- a. Please discuss the sources of sampling error in the proposed system and the legacy system, and the magnitudes of these errors in terms of their impact on the accuracy, reliability, and representativeness of results.
 - b. Please describe any additional steps taken to mitigate the non-sampling errors identified in the statistical design plan.
 - c. Have any additional sources of non-sampling error been identified other than those identified in the statistical design plan? If so, please describe them, what steps are being taken to mitigate them, and their magnitude on the accuracy, reliability, and representativeness of results.

RESPONSE:

- a. Sampling errors are the result of taking measurements from less than the full population. For internal SPM, random sampling is employed in two areas: for measurement of the transit time from collection boxes/chutes to initial processing for single-piece First-Class Mail, and for measurement of the transit time from final processing to delivery for all of the measured market dominant letters and flats. These measurements from sampling are applied to data taken from all available measured mailpieces observed in postal processing, using a model based on certain assumptions about independence between time spent in each transit leg. The service performance estimation methodology and the associated margins of error calculations were derived based upon the sampling models employed. The margins of error displayed in the quarterly reports reflect the sampling error of

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

the performance estimates. The magnitude of these errors varies across the products measured and by geography, largely as a result of differing sampled volumes. For example, the sampling error is higher for First-Class Mail Flats, which represent about 1 percent of total Market Dominant products, than for Single-Piece First-Class Mail Letters/Cards, which represent nearly 13 percent of the total mail because fewer flats were sampled than letters/cards.

The External First-Class (EXFC) is based on test mailpieces which have been designed to emulate certain characteristics of the population being measured. The systems employ random sampling techniques for the design of the test mail sample, and just as in the internal SPM system, the performance estimation methodology and margins of error calculations were derived from the underlying sampling model used. One aspect of variability is ignored in the estimation approach. Historical mail volumes sent between districts which are used as part of the EXFC design and performance estimate calculations are based upon sampling data from the ODIS-RPW system. The variability of those volume estimates is not captured in the reported EXFC sampling error. Instead, these volume estimates are created using 12 quarters of volume data to reduce the variability and then treated as constants within the service performance and variance estimation calculations. Just as for internal SPM, the magnitude of the margins of error depends primarily on the sample size and the service performance for the specific estimates being made. In EXFC,

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

origin/destination service performance estimates have varying sampling errors because, while destinating test mail volumes are similar for each district, the origin test mail volumes vary significantly in line with population mail flows.

The legacy measurement system for commercial mail products relies on estimates of time from final processing to delivery for mail going to a panel of anonymous households and small businesses participating in the measurement program as reporters. These estimates are applied to data taken from all available measured commercial mailpieces observed in postal processing, using a model based on certain assumptions about the independence between time spent in processing and final delivery. The service performance estimation methodology and the associated margins of error calculations were derived based upon the sampling models employed. Sampling error is estimated through the margin of error calculations.

For all the estimates in both internal SPM and the legacy measurement systems, the magnitude of the sampling error impacts the reliability of estimates, which is the category where the statistical precision was captured, but will not impact the accuracy or representativeness of the estimates as those measures were defined. Performance estimates with larger margins of

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

error indicate estimates with greater uncertainty/less reliability than those with smaller margins of error.

- b. There were two types of non-sampling errors described in the *Statistical Design Plan for Internal Service Performance Measurement*. The first was that the sampled population includes only mailpieces identified as measurable through available processing scans. It is not possible to evaluate the extent to which the exclusion of un-scanned or non-measured mail introduces error in the service performance estimates. The second type of potential non-sampling error identified in the plan was the use of mail with Special Services requested to represent transit time for all mail sent over the retail counter. The methodology assumes that duration in First Mile for the pieces is the same for mail with and without Special Services. There is no method to evaluate or estimate the error that this assumption introduces, though it is assumed to be very small.

Efforts to reduce the impact of the first issue, the exclusion of un-scanned or un-measured mail, include encouraging the adoption of Full Service Intelligent Mail by mailers, reducing the volume of un-scanned mail through scanning processes such as the bundle visibility initiative, and initiatives to reduce the volume of measurement exclusions. In addition, where patterns of lower volume in measurement are clearly correlated with service performance

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

differences, population data are used as weights to adjust the estimates. This methodology is employed for Periodicals estimates, for example, where Destination Entry mail represents a significantly higher volume of mail in measurement than in the full population. Estimates of performance are made for Destination Entry and End-to-End mail separately and then are weighted together using population data rather than the measured volume to improve the accuracy of the overall service performance estimates. Similar processes are used for USPS Marketing Mail products and are the same in both legacy and internal SPM systems.

The second issue, potential differences in transit time in pieces inducted at retail with and without Special Services, has not specifically been addressed. Through reporting for Quarter 1 FY 2017, the retail data were not included at all in the First Mile estimates. The focus has been on developing the business rules necessary to accurately measure the time in First Mile for both letters and flats with Special Services. In Quarter 2 FY 2017, the rules were refined for flats to provide more stable estimates than were available with the limited data available in collection. In Quarter 3 FY 2017, the focus will be to address the remaining issues so that the information for letters can also be included.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

c. Non-sampling errors can stem from a variety of situations. Thus far in internal SPM the following additional types of potential non-sampling errors have been identified:

- Coverage errors may occur from failure to accurately represent population units in the sample. Un-scanned or un-measured pieces are one such example. Other types include systematic issues of excluding ZIP Code areas or address types from the sample. As issues have been identified, defects have been logged so that they can be prioritized and addressed. For example, ZIP Code 969 had been excluded from internal SPM sampling prior to Quarter 3 FY 2017 pending software updates needed to account for the time zone difference.
- There is an inability to obtain information about some sample cases. When there are high instances of non-response which are not randomly distributed in the sample, non-response patterns may lead to over or under coverage of segments of the population. For example, challenges with the accuracy of geographic coordinate data can prevent scanners from triggering at requested sampling addresses and may lead to the under-representation of those address types and the over-representation of other types with higher response rates, impacting representativeness and potentially the accuracy of the measurement results.

RESPONSE OF THE UNITED STATES POSTAL SERVICE TO COMMISSION INFORMATION REQUEST NO. 1

- Response errors may occur when the data collected are not accurate. For example, if postal personnel scan mail that had been delivered on a prior day along with mail being delivered on the sampled day, data collected would inaccurately estimate longer times in Last Mile, leading to lower estimates of on-time performance. Retraining on proper scanning procedures has been the focus to reduce these errors which are primarily impacting mail to P.O. Boxes.
- The use of equally weighted sampled data for Last Mile Profiles, even though the sampled addresses were selected with unequal probabilities, creates a non-sampling error that impacts the accuracy of service performance estimates. The impact of this issue is generally to underestimate time in Last Mile, which would lead to higher service performance scores. Equally weighted data were used to mitigate the impact of very large sampling weights on reliability. A small change to the sample design was implemented to avoid very large weights, and the results of those changes are currently being studied.
- Another type of non-sampling error occurs when there are significant volumes of mail which require the use of imputed values when sampling data are missing or insufficient. For example, very limited volume of flats in collection boxes resulted in high levels of imputation of First Mile duration. The levels of imputation required and the

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

potential impact on the accuracy will be assessed quarterly at the district and product levels to identify situations which warrant revisions to the imputation logic, the sampling targets, or other aspects of the system. These types of non-sampling errors may impact the accuracy of estimates but do not necessarily result in systematically higher or lower service performance estimates. The implementation of the Retail Profile is expected to help mitigate the impact of having limited collection flats in the first mile duration.

It is not possible to directly quantify the impact of the non-sampling errors because the service performance of the entire population is unknown. However, by generating metrics around the measures, it is possible to better understand the issues. The Postal Service has conducted analyses around each non-sampling error, leading to actions to address the underlying root causes as described above. The Postal Service continues to look for opportunities to improve Internal SPM as part of its continuous improvement efforts.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

- 10.** In various narrative sections, which discuss the limitations of measurement for products, the Postal Service states that there were several days where no sampling occurred due to “null transmissions.” Please quantify the impact of null transmissions on accuracy, reliability, and representativeness. What progress has been made on resolving this issue?

RESPONSE:

Null transmissions can occur for various reasons. In order to mitigate the impact of null transmissions on service performance measurement, business rules were created leveraging sampling data from the impacted Districts from prior weeks. These District data from prior weeks are used as a proxy for the performance profiles when null transmissions occur. This approach ensures the accuracy, reliability, and representativeness of Internal SPM on days where no sampling occurred due to null transmissions.

Each specific reason for null transmissions that have occurred, the areas impacted, and the status of resolution are listed below.

1. USPS mobile scanners currently cannot download sampling requests for a sampling day when they are sent prior to 12 AM CT of the sampling day. The scanners are programmed to connect and download at 2 AM local time. Therefore, scan requests have to be available by connection time for successful transmission.
 - a. This impacts only the Guam (UTC+10) and Atlantic (UTC-4) time zones

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

- b. This impacted only seven (7) dates during FY17Q1 for collection sampling in the Atlantic time zone. This was resolved by sending sample requests to align with the connection time to ensure successful transmission. Since implementing the change, no requests have been reported as missing from downloads.
 - c. This impacts all sampling in Guam. Changes to mobile scanner software and infrastructure are currently in progress to resolve the Guam time zone issue. These changes will enable downloads to occur for Guam prior to the 12 AM CT connection time. These changes are planned for implementation in Q4 FY17.
2. Internal SPM database failures that occur during the sampling request send timeframe can cause null transmission.
- a. Service tickets are created, logged, and addressed by the software vendors to minimize impact. All issues have been resolved the same day to minimize impact on future sampling dates.
 - b. This impacted one date during FY17Q1 and two dates during FY17Q2.
 - i. FY17Q1: 12/6 Delivery Sampling - Arizona, Colorado/Wyoming, Salt Lake City
 - ii. FY17Q2: 1/3 Delivery Sampling - Alaska, Arizona, Bay-Valley, Colorado/Wyoming, Los Angeles, Nevada Sierra, Portland,

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

Sacramento, Salt Lake City, San Diego, San Francisco, Santa Ana,
Seattle, Sierra Coastal

- iii. FY17Q2: 1/13 Delivery Sampling - Arizona, Colorado/Wyoming,
Salt Lake City

3. Sampling request generation failed during processing due to system memory utilization issues.

- a. This impacted one district, Chicago, for three days during FY17Q1.
- b. The Postal Service has identified the root cause and deployed software changes to resolve it. No issues have been observed after changes were deployed.

4. Internal SPM software defect caused a null transmission for the Caribbean district on 2/1. The defect was identified and resolved on the same day. No issues have been observed since the defect has been resolved.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

11. Please describe and explain the changes in methodology or business rules that were made to account for collection and processing on Sundays and holidays.

RESPONSE:

In FY17 Quarter 1, processing plants processed outgoing mail on a number of Sundays in December in order to advance volume from what would otherwise have been very heavy Mondays. In addition, some plants process outgoing mail on certain non-widely observed holidays to advance volume from the next day's outgoing volume. The Postal Service's business rules did not take that into account. Volumes processed on Sundays and holidays were always assumed to be from the prior processing day, which means they were assumed to be delayed. This was not accurate in Q1, nor is it accurate around the year during some other holidays when Operations advances mail. As a result, business rules were established to exclude Sunday/Holiday FPO1 volume from processing duration measurement due to the ambiguity. However, this does not eliminate the possibility for this mail having first mile failure from Saturday collections.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

12. Please refer to Library Reference USPS-LR-PI2015-1/6, Single-Piece First-Class Mail, "Narratives" tab. What revisions to the methodology or business rules are being considered to adjust measurement when pieces "are first observed in incoming mail processing operations rather than in the expected outgoing operation?"

RESPONSE:

There are a number of revisions that will impact measurement of pieces first observed in incoming processing operations rather than in the expected outgoing operation, referred to as FPO2 mail.

Currently, the census of single-piece mail volume observed in processing duration data is mapped back to an estimated origin location based on the assumption that pieces originate from the local processing area of the plant where they are first observed in processing. This is a reasonable assumption when first processing occurs in an expected outgoing operation, referred to as FPO1 mail. However, when a piece is first observed in incoming processing, this assumption is likely inaccurate in the majority of cases. For EXFC flats, for example, it was observed based on FY17 Quarter 1 data that only roughly 33 percent of FPO2 flats originated from the local processing area of the plant in which they were first observed. The current mapping process results in a less diverse range of estimated origin locations for FPO2 mail. Also, because FPO2 mail is first observed closer to destination, it places a disproportionate amount of FPO2 mail in the Two-Day service standard group, as opposed to Three-To-Five-Day. Because FPO2 mail tends to experience longer durations in First Mile, this causes estimates of

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

Two-Day performance to be lower overall and Three-To-Five-Day performance estimates to be higher.

Beginning in FY17 Quarter 3, historical data from retail and collection sampled pieces that are FPO2 mail will be used to create a more representative mapping of origin 3-digit ZIP Codes to first observed processing locations, including origins that are not local, which are a significant part of the FPO2 population. By applying this new mapping to the FPO2 mail observed in the census of single-piece processing duration, volumes will be mapped to a more realistic distribution of origin locations and service standards compared to the current process. The current mapping process will continue to be used for FPO1 mail.

An additional concern with respect to FPO2 pieces is that the duration in First Mile varies systematically between local Two-Day mail and non-local mail. For example, an FPO2 sampled piece traveling from Northern Virginia to Maryland, first observed in the Maryland area, will likely show a shorter duration in First Mile than an FPO2 piece traveling from Northern Virginia to Los Angeles that is first observed in the Los Angeles area. Analysis of sampled data separated into local versus non-local FPO2 pieces confirms this assumption. Because First Mile profiles are calculated as an average across service standards, an additional adjustment will be introduced to processing duration calculations for FPO2 pieces to account for the local versus non-local impact. For FPO2 pieces, true duration in processing is unknown because origin processing is not observed. To address this issue, the service standard from the assumed origin to

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

destination will be used as a proxy for the unknown plant processing duration. This assumption was tested and was observed to be a reasonable approximation based on an analysis of FPO1 mail. This processing duration adjustment normalizes the First Mile duration for local versus non-local sampled pieces, and the remainder of the designed methodology will remain in place to develop complete estimates of service performance with the application of First Mile and Last Mile profiles.

There have also been a number of enhancements that have been implemented to improve the accuracy of volume being included in the FPO2 group. In Quarter 2, internal SPM began ingesting sort plan data to capture the mail class of the sort plan of the first processing scan. Often, FPO2 volume was actually determined to be unbarcoded or unread commercial volumes that were incorrectly identified as single piece. Using the sort plan class to identify the class of the mailpiece helped eliminate those pieces.

Logic was also added to map ID Tag scans back to the sort plan that applied the ID Tag. This was an extension of the previous item, and allows the class of the sort plan for all first processing scans to be captured, and volumes to be classified appropriately. All of these enhancements are part of Postal Service efforts to continuously improve the Internal SPM to ensure it is accurate, reliable, and representative.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

- 13.** Please refer to Library Reference USPS-LR-PI2015-1/6, Single-Piece First-Class Mail, "Narratives" tab. Please discuss any technical issues associated with tracking mailpieces through automation processing, and describe the steps that are being taken to resolve these issues. Please explain whether these issues are related to the issues with last mile weights, and if so how? If not, please explain how these issues affect the accuracy, reliability, and representativeness of service performance results.

RESPONSE:

Tracking mailpieces through automation involves having business logic to determine which barcodes that are either already on the mailpieces or that are applied to the mailpieces constitute unique pieces, and therefore mark the beginning of the Processing Duration measurement. In some cases, a barcode is not applied on the first scan in mail processing, and therefore may appear to be a first mile failure during sampling or may become a FPO2 piece if the next hit is on a destinating operation. In Quarter 3, the Postal Service is adding logic that utilizes either a Special Service barcode from retail or an IBI by itself to create a unique piece. This helps resolve the issue where the first scan does not always produce a barcode, and makes measurement more accurate.

When the Postal Service misses the first scan of mailpiece, it increases first mile impact and/or shifts more volume into FPO2 when it truly was an FPO1 piece. This fix helps minimize that problem.

Ensuring the volumes are correctly categorized as FPO1 or FPO2 has been a focus of continuous improvement efforts and has been successful in improving the accuracy, reliability, and representativeness of service measurement.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

Tracking unique mailpieces is not related to using a weighted Last Mile profile.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

- 14.** Several narratives for Standard Mail reference a lack of end-to-end coverage for certain districts. Please explain what caused the gaps in Standard Mail end-to-end coverage, and how they are being addressed.

RESPONSE:

Volumes at certain uncommon service standards for End To End Standard Mail do not exist. As this is not a problem in measurement, there is no plan to address it as such.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

- 15.** Please refer to the narrative tab for Standard Mail Flats (FY 2017 Quarter 1) in Library Reference USPS-LR-PI2015-1/6. Please discuss:
- a. any progress that has been made on the development of non-deviant margins of error,
 - b. the obstacles to developing non-deviant margins of error,
 - c. how the Postal Service is addressing these obstacles, and
 - d. an estimate of when these obstacles will be resolved.

RESPONSE:

- a. In the reports submitted for FY 2017 Quarter 2, this limitation has been removed because the quality assurance testing indicated that the software defects previously identified for the margins of error around the on-time estimates had been repaired.
- b. The margin of error calculations are quite complex and require significant focus to address each scenario which arose in the treatment of imputed values. Furthermore, the calculations are computationally intensive and require significant processing time, resulting in limited opportunities to repair the software and rerun if issues are found.
- c. The Postal Service is managing the process through defect identification, prioritization, resolution, and independent quality assurance. The identified software defects for estimating the margins of error for estimates which do not involve a First Mile sampling component were addressed for FY2017 Quarter 2. Currently, there is one open defect which impacts the calculation of margins of error for Single-Piece First-Class Mail.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

- d. The open defect for First Mile margins of error is scheduled for repair and retesting during FY2017 Quarter 3. Testing is also underway on the margins of error for service variance with a target for completion and resolution of any issues for FY2017 Quarter 3 reporting.

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

- 16.** In its narratives, the Postal Service states that samples for USPS Marketing Mail and Bound Printed Matter Flats were combined to develop service performance estimates.
- a. Please explain what steps, if any, are being taken to develop sufficiently unique samples for each.
 - b. If no action is being taken, please explain how the aggregation of these samples results in sufficiently unique estimates, and how these estimates are generated.
 - c. Please explain how the aggregation of these samples result in estimates that accurately represent each product, and how these estimates are generated.
 - d. If this aggregation does not result in estimates that accurately represent each product, please explain what steps are being taken to develop representative estimates for USPS Marketing Mail and Bound Printed Matter Flats.
 - i. When does the Postal Service anticipate completing these actions?
 - ii. How will the Postal Service assess whether the individual estimates are sufficiently unique?

RESPONSE:

- a) There is not sufficient volume for Bound Printed Matter Flats to leave them in their own Last Mile sampling group, and therefore it is not feasible to develop sufficiently unique samples for each product. The design decision to combine the Last Mile samples was made based on the similar mail preparation and processing between Marketing Mail flats and Bound Printed Matter Flats, and was done so that sufficient sampled pieces are available for Last Mile profile calculations to generate robust estimates of Last Mile performance.
- b) The Processing Duration is measured separately for Marketing Mail Flats and Bound Printed Matter Flats. This results in unique volumes and processing

**RESPONSE OF THE UNITED STATES POSTAL SERVICE
TO COMMISSION INFORMATION REQUEST NO. 1**

scores for both Marketing Mail Flats and Bound Printed Matter Flats. The processing duration performance is then combined with the composite Last Mile profile for Marketing Mail and Bound Printed Matter flats, which creates unique estimates for both product types.

- c) The measurement described in part B above accurately measures each product if the performance for each product in last mile is sufficiently similar to each other. The Postal Service expects that the performance in last mile for each product is similar enough to not make an appreciable difference in the overall performance, and is more reliable than relying on sampling groups that have insufficient samples to be statistically valid.
- d) The Postal Service believes its current measurement is sufficient to accurately represent each product type. Bound Printed Matter Flats performance is primarily driven by the Processing Duration. This Processing Duration performance estimate is based solely on Bound Printed Matter Flats mail. With the application of Last Mile profiles based on time remaining after processing to meet service standard, the differences in processing between USPS Marketing Mail flats and Bound Printed Matter Flats are accounted for. Last Mile performance for Marketing Mail and Bound Printed Matter Flats is measured the same way in the legacy system, with sampled pieces for these two product types combined in Last Mile Profile calculations.