Metropolitan Transportation Authority – New York City Transit

New Customer-Focused Subway Metrics

Report 2018-S-72 | January 2020

OFFICE OF THE NEW YORK STATE COMPTROLLER
Thomas P. DiNapoli, State Comptroller
Division of State Government Accountability

Audit Highlights

Objective

To determine whether New York City Transit's (Transit) new metrics provide customers with better information about their trips and provide a more accurate picture of Transit's performance. The audit covers the period between July 1, 2017 and November 30, 2018.

About the Program

The Metropolitan Transportation Authority (MTA) is a public benefit corporation chartered by the New York State Legislature in 1965. Transit is an agency of the MTA that operates New York City's subways and the majority of its bus service.

Under the Public Authorities Law, the MTA is required to issue an annual report on its mission statement, measurements, and performance indicators. One of the goals the MTA cites in its annual report is to provide on-time and reliable service to customers.

In July 2017, the MTA prepared a Subway Action Plan (Plan) to improve service. One of the major changes in the plan was an agreement to implement new customer-focused performance measures with a goal of making these metrics more relevant and easier to understand by subway customers. In September 2017, Transit introduced the new metrics[1]: Additional Platform Time (APT), the average time that customers wait at a station beyond their scheduled wait time; Additional Train Time (ATT), the average time customers spend on board a train beyond their scheduled travel time; and the sum of these, Additional Journey Time (AJT). AJT is the key component of Customer Journey Time Performance (CJTP), the percentage of customer trips completed within five minutes of the scheduled time. It was not reported as a performance measure on the MTA's Subway Performance Dashboard until November 2019. Section 1276-f of the Public Authorities Law, enacted as part of the fiscal year 2019-20 State Budget, requires the MTA to publicly report APT, ATT, AJT, and CJTP.

Key Findings

- Transit's new customer-focused measures do not appear to meet the Plan's goals. For a metric to be relevant, it should be closely connected to the goals, easily understood, and straightforward.
- Transit does not disclose the significant assumptions made during its calculation of subway APT, ATT, and AJT or the limitations of the data used.
- APT and ATT depend on knowing where and when a customer entered and exited the system. However, Transit's automated fare collection system does not require customers to swipe out of the system, so Transit does not know where and when each customer's trip ends. Without this information, the new measures rely heavily on a series of assumptions, each of which introduces uncertainty and complexity to the model.

1 It is the MTA's practice to refer to performance measures interchangeably in its documentation as measures, indicators, or metrics.

- APT, ATT, and AJT represent the average additional time for each leg of a trip. A customer who transfers multiple times would have to add the times for each leg of their trip together to obtain the total APT, ATT, or AJT of their trip. However, based on the published definition, this is not clear.

Key Recommendation

- Evaluate whether APT, ATT, and CJTP meet the goals of the Plan and disclose the assumptions and margin of error for each assumption.

Office of the New York State Comptroller
Division of State Government Accountability

January 17, 2020

Mr. Patrick J. Foye
Chairman and Chief Executive Officer
Metropolitan Transportation Authority
2 Broadway
New York, NY 10004

Dear Mr. Foye:

The Office of the State Comptroller is committed to helping State agencies, public authorities, and local government agencies manage their resources efficiently and effectively. By doing so, it provides accountability for tax dollars spent to support government operations. The Comptroller oversees the fiscal affairs of State agencies, public authorities, and local government agencies, as well as their compliance with relevant statutes and their observance of good business practices. This fiscal oversight is accomplished, in part, through our audits, which identify opportunities for improving operations. Audits can also identify strategies for reducing costs and strengthening controls that are intended to safeguard assets.

Following is a report of our audit of Metropolitan Transportation Authority – New York City Transit entitled *New Customer-Focused Subway Metrics*. The audit was performed pursuant to the State Comptroller's authority as set forth in Article X, Section 5 of the State Constitution and Section 2803 of the Public Authorities Law.

This audit's results and recommendations are resources for you to use in effectively managing your operations and in meeting the expectations of taxpayers. If you have any questions about this report, please feel free to contact us.

Respectfully submitted,

Division of State Government Accountability

Contents

Glossary of Terms

Term	Description	Identifier
AFC	Automated fare collection	*Key Term*
AJT	Additional Journey Time	*Key Term*
APT	Additional Platform Time	*Key Term*
ATT	Additional Train Time	*Key Term*
Board	MTA Board of Directors	*Key Term*
CJTP	Customer Journey Time Performance	*Key Term*
Committee	MTA Board's Transit and Bus Committee	*Key Term*
Division	System Data and Research Division (within Transit's Division of Operations Planning)	*Division*
MTA	Metropolitan Transportation Authority	*Auditee*
OP	Transit's Division of Operations Planning	*Division*
Plan	Subway Action Plan	*Key Term*
Transit	New York City Transit	*Agency*

Background

The Metropolitan Transportation Authority (MTA) is a public benefit corporation chartered by the New York State Legislature in 1965. New York City Transit (Transit), an agency of the MTA, operates New York City subways and the majority of bus service. Transit reported carrying about 1.7 billion subway riders in 2018, with an average weekday total of about 5.4 million and an average weekend (Saturday and Sunday) total also of about 5.4 million. As of September 2018, the MTA had a 17-member Board of Directors (Board), and most issues regarding bus and subway service were addressed at meetings of the Board's Transit and Bus Committee (Committee).

Transit's Division of Operations Planning (OP) is responsible for designing effective and efficient subway service. OP's System Data and Research Division (Division) calculates key performance measures related to Transit's goal to provide on-time and reliable service. Performance measure results are often reviewed by the Board, which provides feedback, guidance, and recommendations to Transit on the success of its operations. On July 25, 2017, in response to the Board's demand for performance improvements, the MTA Chair released the Subway Action Plan (Plan). Subsequently, a set of new performance measures were implemented, including: Additional Platform Time (APT) and Additional Train Time (ATT), the sum of which is referred to as Additional Journey Time (AJT). AJT, the key component of Customer Journey Time Performance (CJTP), was not reported as a performance measure on the MTA's Subway Performance Dashboard until November 2019. Public Authorities Law Section 1276-f, enacted as part of the fiscal year 2019-20 State Budget, requires the MTA to publicly report APT, ATT, AJT, and CJTP. The MTA defines the new performance measures as follows:

APT – "The average added time that customers spend waiting on the platform for a train, compared with their scheduled wait time. APT is measured using a combination of customers' MetroCard entry data into stations and train departure times from those stations, using information from the real-time train tracking technologies that provide train arrival information."

ATT – "The average additional unanticipated time customers spend onboard the train due to various service issues. Additional Train Time is measured using a combination of customers' MetroCard entry data into their starting stations and customers' arrival times at their destination stations, using information from the real-time train tracking technologies that provide train arrival information."

CJTP – "The percentage of customers whose journeys (waiting and travel time) are completed within five minutes of their scheduled journey time." This can also be understood as the percentage of trips where AJT is less than five minutes.

Audit Findings and Recommendations

We concluded that the MTA's new customer-focused statistics as reported to its Board and Committee do not meet their intended goal as public performance measures. We found that these measures rely on significant assumptions that cannot be verified and thus may not reflect the actual customer experience.

APT and ATT depend on knowing where and when a customer entered and exited the system. However, Transit's automated fare collection (AFC) system does not require customers to swipe out of the system, so Transit does not know where and when each customer's trip ends. Without this information, the new measures rely heavily on a series of assumptions, each of which introduces uncertainty and complexity to the model.

In response to our preliminary findings, Transit officials disagreed with criticism of the new metrics for reporting customer travel time. They cited several globally recognized transit organizations that claim the MTA's new metrics are a significant step forward, improving the way it measures and monitors service from the customer point of view. However, some of the systems cited (e.g., London) record both passenger entry and exit data and, unlike the MTA, do not rely on significant assumptions.

New Performance Metrics

The MTA's model for determining performance metrics is based on MetroCard swipe data, which is maintained by an AFC database, and consists of:

- MetroCard number
- Date
- Time (in six-minute intervals) – 240 daily intervals
- Booth number (station swiped) – 768 subway booths

The limitations of this data are:

- It is incomplete. The MTA estimates that 4 percent of riders do not use MetroCards but, nonetheless, includes these trips in its model despite having no data for these riders.
- The swipe data does not record actual time of entry but rather the six-minute interval someone entered the system.
- Neither passenger destinations nor arrival times are recorded.

Additionally, OP's practice is to select one day per month (model date) to assume the travel time at every station and time intervals for the following

month. The model assumes that the date selected is representative of rider experience for the entire month. (The limitations of this practice are explored later in this report.)

Calculation of the new performance measures is based on three other primary assumptions:

- **Destination Assumption**: Transit has no data on trip destinations and, therefore, Transit officials assume the destination station based on the next swipe of the same MetroCard. (The next swipe must occur within 48 hours.) Once an origin and a destination have been assigned, Transit classifies them as linked trips. Trips for which there is no second swipe are considered "irrational" and are handled as exceptions.
- **Route Assumption**: Transit assumes that riders select the most efficient route to their destination.
- **Interval Assumption**: According to Transit officials, MetroCard swipe data is received in blocks of six-minute intervals. Swipe times are allocated randomly by the model into 30-second intervals within the six-minute period. For example, a swipe that appears at 12:00:00 p.m. (noon) in the AFC data may have swiped in at any point from 12:00:00–12:05:59, so each swipe is randomly assigned to a 30-second interval within this period.

Tables 1 and 2 illustrate how the APT, ATT, and AJT performance measures are calculated, as well as the limited data available for calculating the metric.

Table 1 – APT Calculation

Known: MetroCard Swipe Time Range	**Between 7:24 and 7:30 a.m.**
Interval Assumption: Customer Starting Wait Time	7:25:30 a.m.
Destination and Route Assumptions: Scheduled Board Time	7:26:30 a.m.
Scheduled Waiting Time	**1 minute**
Destination and Route Assumptions: Actual Train Departure	7:27:44 a.m.
Actual Waiting Time	**2 minutes and 14 seconds**
	APT = 1:14

Table 2 – ATT Calculation

Destination and Route Assumptions: Scheduled Board Time	7:26:30 a.m.
Destination and Route Assumptions: Scheduled Arrival at Destination Time	7:30:45 a.m.
Scheduled Train Travel Time	**4 minutes and 15 seconds**
Destination and Route Assumptions: Actual Board Time	7:27:44 a.m.
Destination and Route Assumptions: Actual Arrival at Destination	7:29:50 a.m.
Actual Train Travel Time	**2 minutes and 6 seconds**
	ATT = -2:09

AJT is calculated based on combining both APT and ATT, which, in this example, equals negative 55 seconds (1:14 + -2:09). This trip passed CJTP because AJT did not exceed five minutes.

Destination Assumption and Scaling

The Destination Assumption was based, in part, on studies of the Transport for London and Massachusetts Bay Transportation Authority; the latter system does not require riders to swipe out. According to a Transit official, this assumption was tested over 17 years ago (in 2002) using survey data and was accurate 90 percent of the time. Transit has not revisited it since, which raises questions regarding its relevancy due to changes in travel patterns, demographics, and the way that work is performed (e.g., telecommuting).

Transit also uses a "scaling up" process to account for customers whose destination and path cannot be predicted, referred to as "irrational trips." The process involves counting MetroCard swipes during one-hour testing intervals and dividing the MetroCard swipe data into two groups of "paired" swipes (MetroCard swipes where the destination can be predicted) and "unpaired" swipes (irrational trips). Each paired swipe is increased or "scaled up" to account for irrational trips. For example, on the model date, there were 5,756 MetroCard swipes at booth R138 (agent booth at Penn Station in Manhattan) from 7:00 a.m. to 8:00 a.m., of which 612 swipes were deemed irrational. The remaining 5,144 swipes with predicted destinations were each given a value of 1.119 (612/5,144) riders per trip in order to maintain the original volume of passengers (5,144 * 1.119 = 5,756).

For the model date, there were 5,928,982 subway MetroCard swipes, of which Transit officials could predict the destination for 5,086,755 based on the

location of subsequent swipes. The remaining 842,227 swipes (14 percent) were deemed irrational because a destination could not be predicted. A destination cannot be predicted if there is only one MetroCard swipe on the model date and: there is no subsequent swipe on the following day to be used as the destination; the subsequent MetroCard swipe is made at the same station or deemed too close to the origin station to be a rational destination; or one-trip MetroCards are used. All 842,227 irrational trips are thus still included in the APT and ATT calculations using the scaling-up process, under the assumption that these riders' behavior is similar to that of riders (whose destination can be predicted) who board at approximately the same time and station.

Furthermore, Transit increases the volume of trips by approximately 4 percent every month to account for estimated fare evaders, including legal non-AFC passengers, such as children under 44 inches tall and employees who enter through the gates. There is no information for these subway riders' trip origin, destination, or train line.

Route Assumption

For customers who enter a station with multiple subway lines available to reach their destination, Transit cannot tell which line they actually used. In these cases, it assigns a train line based on the assumption that they each took the most efficient route. The route is determined using Dijkstra's algorithm, which is a sequence of steps designed to determine the shortest travel path to a specified destination based on the line the rider is assigned to for the destination.

Use of the algorithm notwithstanding, there is no guarantee that Transit predicted the correct train line for each passenger. On many routes, riders could take multiple train lines to get from origin to destination or could need to transfer.

For example, using Transit's trip destination assumptions for the October 17, 2018 model date, just 2.1 million of the 5.9 million subway riders had a trip with only one train line to their destination. As shown in Table 3, the remaining 3.8 million subway riders (64 percent) could choose from multiple train lines to their destination or could need to transfer at another station.

Table 3 – Percent of Model Date Trips With Multiple Train Line Options or Transfer Required

Travel Options From Origin to Destination	Model Date Passenger Volume
Only one direct line available	2.1 million
Multiple lines available	1.2 million
No direct line available (rider must transfer)	2.6 million
Total	**5.9 million**
Percent with multiple line options or no direct line	64.4%

If the model determines a trip will require a transfer, the trip is broken into legs, where each different train taken is considered a separate step for calculation of APT and ATT. Trips broken into legs are not re-assembled for calculation, so the cumulative time for a complex trip is not considered in the metrics; therefore, the metrics are artificially lower. Consequently, a customer using the Subway Performance Dashboard would have to add the time for each leg of their trip to arrive at a total APT and ATT. However, this is not clear from the definition published on the MTA's website.

Interval Assumption

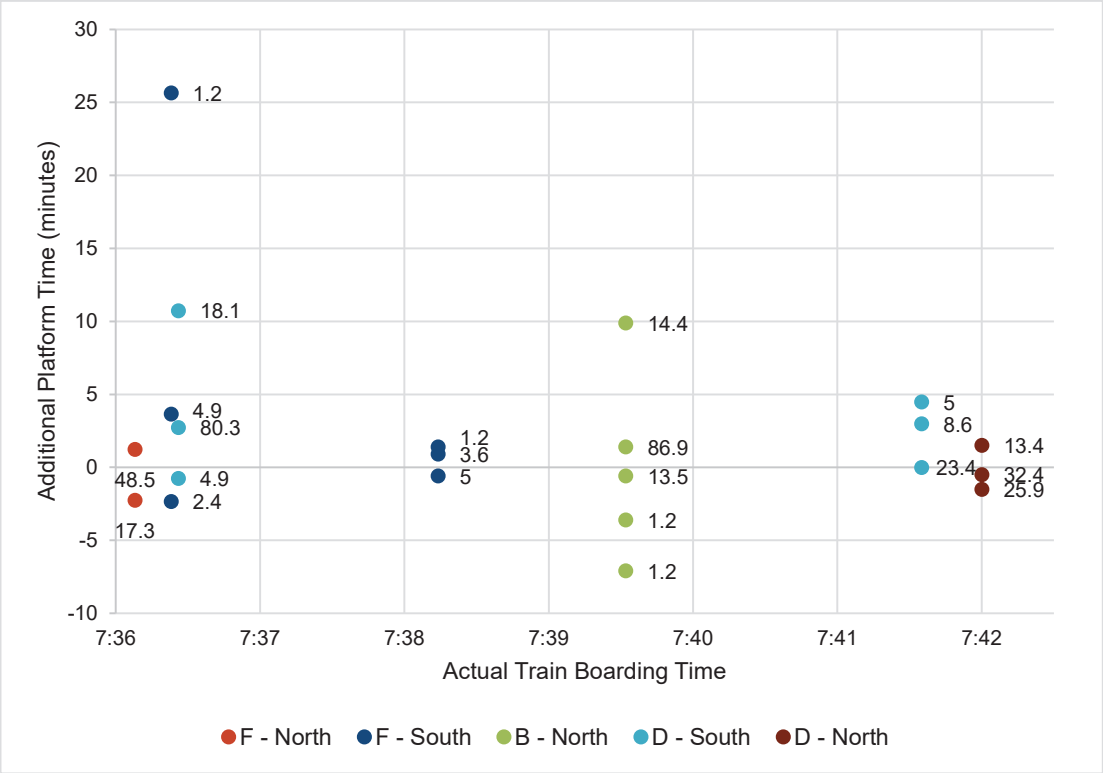
Additionally, the start wait time assigned to each customer can have a significant impact on APT, ATT, AJT, and CJTP because it is a key component for determining the scheduled train to board. As shown in Table 4, although there were 12 different possibilities, Transit assigned the individual to a start wait time of 3:37:00 p.m. (highlighted row). Based on the assigned time, the calculated APT was -1:54. However, had this individual been assigned to another of the 12 start wait times, APT would have ranged from -1:54 to 3:06.

Table 4 – Sample of APT Depending on Start Wait Time Assigned by Transit

Start Wait Time	Rider's Scheduled Boarding Time	Scheduled Wait Time	Actual Train Departure Time	Actual Wait Time	APT
3:34:00 p.m.	3:35:30 p.m.	01:30	3:38:36 p.m.	04:36	03:06
3:34:30 p.m.	3:35:30 p.m.	01:00	3:38:36 p.m.	04:06	03:06
3:35:00 p.m.	3:35:30 p.m.	00:30	3:38:36 p.m.	03:36	03:06
3:35:30 p.m.	3:35:30 p.m.	00:00	3:38:36 p.m.	03:06	03:06
3:36:00 p.m.	3:40:30 p.m.	04:30	3:38:36 p.m.	02:36	-01:54
3:36:30 p.m.	3:40:30 p.m.	04:00	3:38:36 p.m.	02:06	-01:54
3:37:00 p.m.	3:40:30 p.m.	03:30	3:38:36 p.m.	01:36	-01:54
3:37:30 p.m.	3:40:30 p.m.	03:00	3:38:36 p.m.	01:06	-01:54
3:38:00 p.m.	3:40:30 p.m.	02:30	3:38:36 p.m.	00:36	-01:54
3:38:30 p.m.	3:40:30 p.m.	02:00	3:38:36 p.m.	00:06	-01:54
3:39:00 p.m.	3:40:30 p.m.	01:30	3:41:36 p.m.	02:36	01:06
3:39:30 p.m.	3:40:30 p.m.	01:00	3:41:36 p.m.	02:06	01:06

Furthermore, to combine all assumptions, we reviewed the travel options customers had after swiping at a specific station during a six-minute window. Figure 1 shows the 413 customers who swiped at a specific booth at 34th Street – Herald Square between 7:36 a.m. and 7:42 a.m. on the October 17, 2018 model date. On November 2, 2018, there were seven different trains that ran on three lines (B, D, and F trains). Transit allocated the 413 customers to different scheduled and actual train combinations, which resulted in 22 different APT results. Due to the multiple assumptions, these results cannot be verified.

Figure 1 – Interval Assumptions Calculated by the MTA
November 2, 2018, 7:36 to 7:42 a.m., 34th Street – Herald Square*



*Assumed percentage volume labeled next to each point.

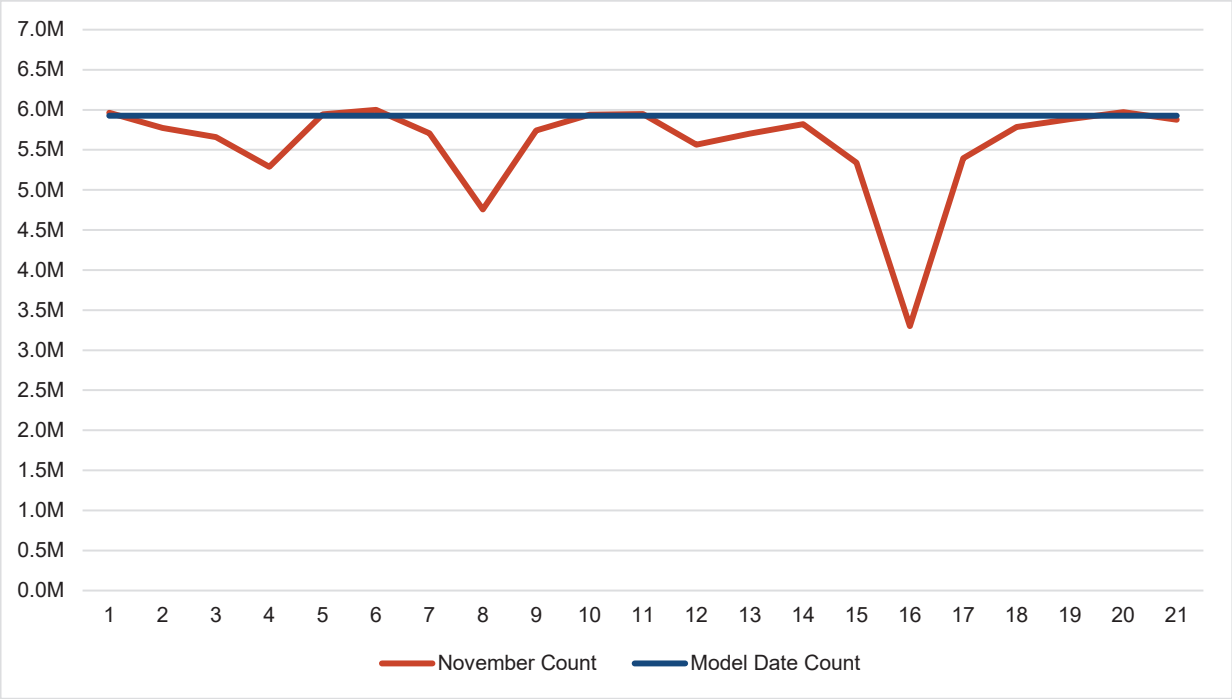
Other Factors Affecting Performance Metric Calculations

OP's practice is to select a model date to assume the travel patterns at every station and time interval for the month. The model assumes that the date selected is representative of riders' experience for the entire following month.

According to OP officials, the model date is chosen by looking back at the nine most recent (available) weekdays and taking the day with the median number of subway swipes. They added that this decision is also affected by other factors including holidays, storms, and cordon count day (i.e., an annual compilation of Subway and Bus Ridership within the Manhattan Central Business District). These other factors introduce subjectivity and inconsistency within the model, making the data less comparable (e.g., one month selected based on median, the next month selected based on different criteria).

Because date selection is judgmental, it may not be representative of the majority of days. For instance, Transit selected October 17, 2018 as its model date for November 2018. We calculated the number of possible station and time intervals for each November weekday and assessed the relationship between the model date count and November's daily count. We found that only about 7.5 percent of November's booth and time intervals matched exactly with the model date. This implies that the model date did not reflect the swipe patterns of 92.5 percent of booth and time intervals. In response to our preliminary findings, Transit officials provided a table with the number and percent of stop-weekdays in November to show that there is minimal variability between overall daily ridership at each station as well as the overall ridership at each time interval systemwide. However, some day-to-day variability still exists. For example, the model date had 5.9 million swipes, while only 6 of the 21 weekdays in November had a volume that met or exceeded that amount. Some days had significantly fewer swipes as shown in Figure 2.

Figure 2 – Subway MetroCard Swipes on October 17, 2018 (Model Date) Compared to November 2018*



*Includes all weekdays except Thanksgiving.

In addition to the main assumptions cited, APT and ATT calculations exclude or do not use all available data, including information pertaining to:

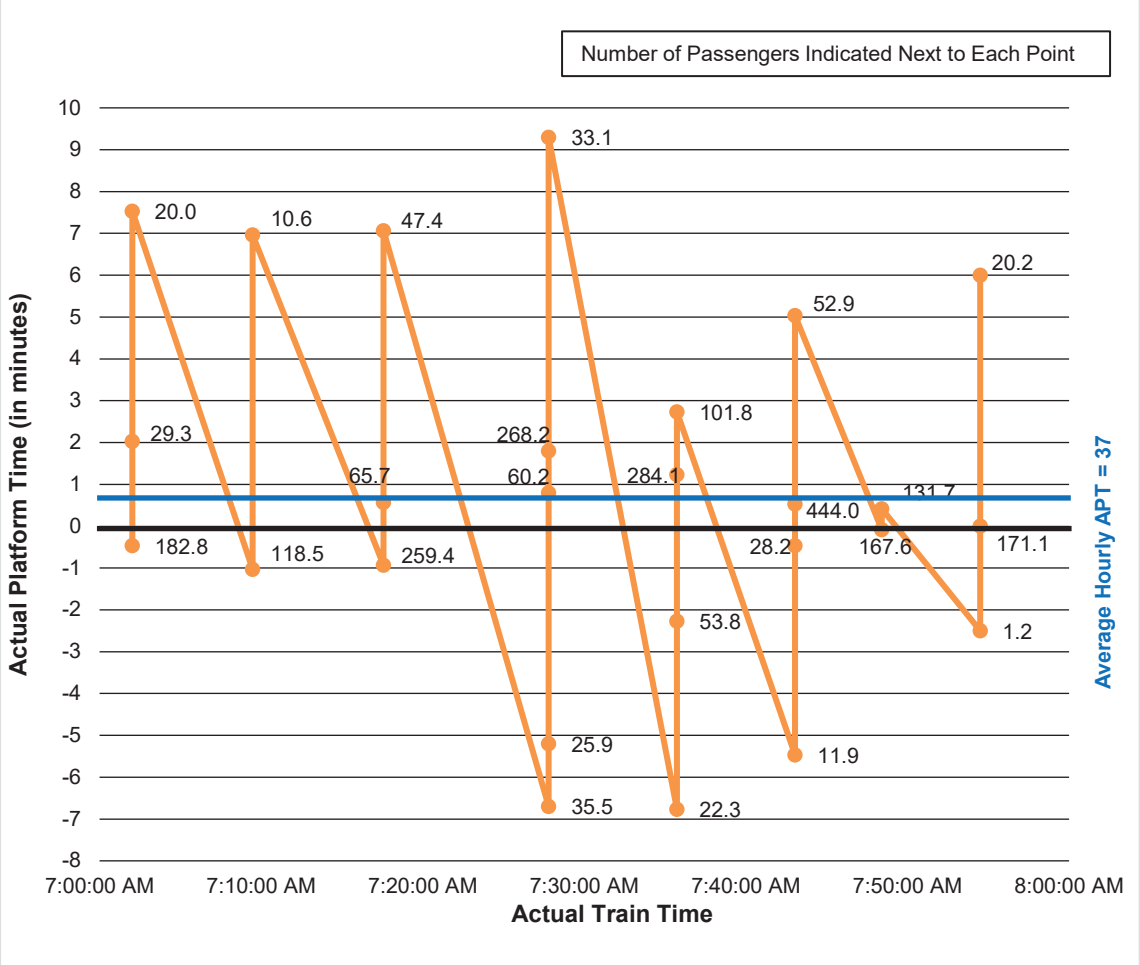
- **Weekend and Overnight Subway Riders**: The new subway metrics provide no information to customers about travel during late nights and on weekends.
- **Non-Subway Options**: We noted that 816,432 subway swipes in the data indicated that part of the customer's trip involved the use of a bus or a non-subway option (e.g., Port Authority Trans-Hudson trains). However, this information is not used in determining the customer's subway destination when the model is applied. As a result, Transit is making an assumption when it would otherwise know where the customer is getting off the subway and onto a bus. Transit officials stated they do not use bus swipes because the data comes from the AFC system and is also in six-minute intervals. Thus, using the data would not improve the reliability and accuracy of the APT and ATT calculation. However, Transit has not formally assessed whether there is a benefit to using this additional data; we note that excluding information about large groups of riders could skew results.
- **Station or Booth Closures**: Trips beginning or ending at a station or booth that was closed at the time and trips beginning or ending at a stop experiencing data issues on the model date are excluded because they are considered irrational. Transit explained that it only checks for data for 450 total stations instead of 472 because certain stations are often closed for repairs.

Additionally, as a result of its methodology for calculating new performance measures, Transit assigns negative additional (platform and train) time to customers who board a train or arrive at their destination earlier than they were scheduled. The following example shows the results of this methodology:

Transit assigns two customers to a start wait time for a train on a platform at 7:00 a.m. and 7:05 a.m., respectively. A train is scheduled to arrive every 10 minutes (6:50 a.m., 7:00 a.m., and 7:10 a.m.). Transit determines that these two customers' scheduled wait times are 0 and 5 minutes, respectively. However, if the 7:00 a.m. train arrived 6 minutes late, then one customer would have waited 6 minutes and the other would have only waited 1 minute rather than the expected 5 minutes. The APT for these two customers would be 6 minutes (7:00 a.m.–7:06 a.m.) and -4 minutes (7:10 a.m.–7:06 a.m.), respectively. However, in the averaging of APT, the -4 minutes will offset customers with +6 minutes.

Figure 3 shows Transit's calculation of APT for the northbound 1 train at one point of entry at Penn Station from 7:00 a.m. to 8:00 a.m. The net APT for this hour is only 37 seconds. This low number may have been attributed to only 184 of the 2,647 riders being assigned an APT greater than 5 minutes. While it is possible that more than 184 riders waited more than 5 minutes, the actual volume of passengers cannot be confirmed. However, the overall averages are dependent on these estimates. Thus, due to the multiple assumptions that went into this calculation, it is possible that the actual customer experience differed from what was reflected by the calculation for this performance metric.

Figure 3 – APT for IRT 1 Train Northbound at Penn Station November 2, 2018 7 a.m. to 8 a.m.



Note: The trends of APT for each train from 7 a.m. to 8 a.m. compared to the actual weighted average APT for the same time period. The marks are labeled by Volume Sum.

Although Transit claims that its practices for determining these performance metrics are common practice industrywide, we reviewed other cities that use similar methodologies and found that origin–destination was not heavily based on assumptions. In Boston, Chicago, Washington, DC, and London, additional factors are used to more accurately portray customers' experience. For example, continuous surveys are conducted and used to validate the assumptions and results of their predictions. In addition, some of these systems require swiping in and out, which eliminates the need to make assumptions about origin–destination and how long the trip took.

Transit could not provide any evidence of how the metrics were used by customers. The customer-focused metrics were originally developed as part of the Plan, but we found that, due to the assumptions and predictions used, the metrics do not offer any useful information to customers. Transit officials explained that the new metrics are regular topics of discussion at Board meetings as well as internal leadership meetings; however, apart from the metrics being included as part of the statistics reported, they did not provide evidence of such. They provided several charts and graphs that are prepared and added that the measures are reviewed by subway leadership and managers; however, no decisions directly result from these reviews. They added that the measures are available to the public.

Transit has no specific or established goals for APT and ATT. According to Transit officials, these metrics are used in concert with other measures to provide information but do not individually drive any decision making. Transit officials could not provide us with any specific documentation or benchmark that would indicate a particular percentage is unacceptable or too low, thus requiring investigation and corrective action.

Recommendation

1. Evaluate whether the APT, ATT, and CJTP meet the goals of the Plan and disclose the assumptions and margin of error for each assumption.

Audit Scope, Objective, and Methodology

Our audit objective was to determine whether Transit's new metrics provide customers with better information about their trips and provide a more accurate picture of Transit's performance. The audit covers the period between July 1, 2017 and November 30, 2018.

To accomplish our objective and evaluate the relevant internal controls, we reviewed MTA-Transit's related policies and procedures, as well as regulations and laws. We interviewed officials and employees of Transit's OP and Department of Subways to obtain an understanding of the processes followed to calculate the new customer-focused metrics.

Statutory Requirements

Authority

This audit was performed pursuant to the State Comptroller's authority as set forth in Article X, Section 5 of the State Constitution and Section 2803 of the Public Authorities Law.

We conducted our performance audit in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objective. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objective.

In addition to being the State Auditor, the Comptroller performs certain other constitutionally and statutorily mandated duties as the chief fiscal officer of New York State. These include operating the State's accounting system; preparing the State's financial statements; and approving State contracts, refunds, and other payments. In addition, the Comptroller appoints members to certain boards, commissions, and public authorities, some of whom have minority voting rights. These duties may be considered management functions for purposes of evaluating organizational independence under generally accepted government auditing standards. In our opinion, these functions do not affect our ability to conduct independent audits of program performance.

Reporting Requirements

We provided a draft copy of this report to MTA-Transit officials for their review and formal comments. Those comments were considered in preparing this final report and are attached in their entirety at the end of the report.

In response to our draft report, MTA officials did not agree with our finding that the new customer-focused metrics do not meet the Plan's goal of implementing performance measures that are more relevant to subway customers and easier to understand. They added that the new metrics were well received by an international transit benchmarking community, the MTA Board, numerous press outlets, and external advocates. However, the response does not reflect whether the everyday rider of the subway has actually used these metrics or whether they have access to the information required to evaluate if service has improved. Transit officials stated they have already acted on OSC findings by revising and expanding the information on the dashboard and additional changes are planned. Our responses to certain MTA comments are included in the report's State Comptroller's Comments.

Within 180 days after the final release of this report, as required by Section 170 of the Executive Law, the Chairman of the Metropolitan Transportation Authority shall report to the Governor, the State Comptroller, and the leaders of the Legislature and fiscal committees advising what steps were taken to implement the recommendation contained herein, and if the recommendation was not implemented, the reasons why.

Agency Comments

2 Broadway
New York, NY 10004
212 878-7000 Tel

Patrick J. Foye
Chairman and Chief Executive Officer

Metropolitan Transportation Authority
State of New York

December 30, 2019

Ms. Carmen Maldonado
Audit Director
The Office of the State Comptroller
Division of State Government Accountability
59 Maiden Lane, 21st Floor
New York, NY 10038

Re: Draft Report #2018-S-72 (New Customer-Focused Subway Metrics)

Dear Ms. Maldonado:

This is in reply to your letter requesting a response to the above-referenced draft report.

I have attached for your information the comments of Andy Byford, President, MTA New York City Transit, which address this report.

Additionally, I will be working with staff to ensure that management is following up on and enforcing the audit's recommendations, where appropriate, and requesting regular, interim reports to that effect.

Sincerely,

Patrick J. Foye
Chairman and Chief Executive Officer

c: Anni Zhu, Acting Chief of Staff to the MTA Chairman & Chief Executive Officer
Michele Woods, Acting Auditor General, MTA Audit Services

2 Broadway
New York, NY 10004-2207
646 252-5800 Tel
646 252-5815 Fax

Andy Byford
President

New York City Transit

Date December 30, 2019

To Patrick J. Foye, Chairman, Metropolitan Transportation Authority

From Andy Byford, President, MTA New York City Transit

Re **New York State Comptroller's Request for Response to the Audit Report "New Customer-Focused Subway Metrics" 2018-S-72**

New York City Transit has reviewed the Office of the State Comptroller's Draft Audit Report on New Customer-Focused Subway Metrics (2018-S-72). NYCT welcomes outside review of its performance metrics and is continuously seeking to improve metric calculation methodologies by integrating new data sources and ideas.

Response to Key Findings

Comptroller Key Finding 1: Transit's new customer-focused measures do not appear to meet the Plan's goals. For a metric to be relevant, it should be closely connected to the goals, easily understood, and straightforward.

The OSC report asserts that the new customer-focused metrics do not align with the goals of the Subway Action Plan, but fails to back up this assertion. The Subway Action Plan promised to deliver a public subway performance dashboard, including measures of the customer experience. NYCT delivered on this promise on an aggressive timeline, including the introduction of innovative new customer-focused metrics. The dashboard and the new metrics were well received. Transit experts at TransitCenter and the international transit benchmarking Community of Metros (CoMET) praised the metrics as global best practices. The metrics also received accolades from members of the MTA Board, while numerous press outlets highlighted the advances and explained the metrics to their readers. NYCT designed these metrics to depict average customer experience in an easy to use measurable metric, and feedback from external advocates and benchmarking agencies confirms we hit the mark.

When discussing how easily understood Additional Platform Time (APT), Additional Train Time (ATT), and Customer Journey Time Performance (CJTP) are, it is important to distinguish between how easy it is to understand the meaning of the metric values from how easy it is to understand the calculation process. The metric definitions and values are straightforward and relate to the passenger experience. It is easy to look for trends in APT, ATT, and CJTP and draw conclusions about improving passenger experiences

Comment 1

(e.g., the average time passengers spend riding the 1 train is now X seconds closer to schedule than it was this time last year, while still remaining Y seconds slower than the scheduled time). This simplicity of metric interpretation is valuable, while complexity of calculation is an unfortunate necessity due to constraints NYCT operates under.

The OSC report correctly highlights numerous limitations of the input data available, which make the process of calculating the metrics more complex than it could be in other systems (e.g., those where users "tap out" as they exit the system). NYCT overcame these challenges by drawing on (and building on) best practices from other systems, as well as academic research. The result is a sophisticated model which integrates advanced concepts from network optimization as well as transportation engineering and planning. NYCT recognizes that this sophistication can make the metric calculation less straightforward. Nonetheless, NYCT strives to make the process transparent.

NYCT is actively working to improve the input data available for metric calculations. For example, the OSC report notes that the granularity of MetroCard swipe timestamps is at the 6-minute level, which necessitates the entry count for each 6-minute period be distributed randomly across that period. OMNY, which is rolling out quickly across the subway system, provides timestamps with precision down to the second. Other aspects of the input data are unlikely to change in the near future (e.g., the lack of "tap out" data to designate trip destinations, as is available in other cities). This will be discussed in more detail in response to the 3rd key finding. Given that the NYS legislature legally mandated NYCT to produce APT, ATT, and CJTP, and that key constraints on input data will remain, the current methodology of modeling customer trips remains the best approach available.

Comment 2

Comptroller Key Finding 2: Transit does not disclose the significant assumptions made during its calculation of subway APT, ATT, and AJT or the limitations of the data used.

NYCT is committed to transparency and has consistently made efforts to share the methodology behind its metrics, including in meetings with leadership and members of the Board, as well as through posted definitions and a section on metric calculation within the subway dashboard help page (http://dashboard.mta.info/Help). NYCT also subjected its methodology to peer review by presenting at the Transportation Research Board conference.

Comment 3

Nonetheless, NYCT is open to further increasing the documentation available to the public. NYCT has already acted on OSC findings by revising and expanding the methodology explanations on the dashboard, and additional changes are planned. More detail is available below in the response to the OSC report's key recommendation.

Comptroller Key Finding 3: APT and ATT depend heavily on knowing where and when a customer entered and exited the system. However, Transit's automated fare collection system does not require customers to swipe out of the system, so Transit does not know where and when each customer's trip ends. Without this information, the new measures rely heavily on a series of assumptions, each of which introduce uncertainty and complexity into the model.

The OSC report is correct that users do not swipe or tap as they exit the NYCT subway system, which does make tracing an individual customer's trip challenging. NYCT implemented a well-established methodology of using patterns of swipes over multiple days to identify likely destinations. As the OSC report confirms, this approach infers destinations for the vast majority of trips, and the other trips are accounted for with by using similar origin-destination patterns. This does add complexity to the methodology, but that complexity is necessary to compute APT, ATT, and CJTP as mandated. The alternative would be to require all customers to swipe or tap as they leave the NYCT subway, which would be a major deviation from our current fare policy, and would require substantial capital expenses to implement. This would likely result in long lines forming at exits during peak times, and would be a high burden to place on customers just to simplify calculations and audits. This type of system is generally implemented in association with zone-based or distance-based variable fares. New York City Transit does not plan to implement zone-based or distance-based variable fares at this time.

Comment 4

Comptroller Key Finding 4: APT, ATT and AJT represent the average additional time for each leg of a trip. A customer who transfers multiple times would have to add the times for each leg of their trip together to obtain the total APT, ATT, or AJT of their trip. However, based on the published definition, this is not clear.

The calculation of APT and ATT separately for each leg of a trip (as opposed to lumped together for a complete trip including all transfers) is an intentional feature of the metrics. This separation of the legs allows NYCT to produce values of APT and ATT by line, which is generally how most subway metrics are produced and interpreted. Dashboard users can simply add APT or ATT values for whatever combinations of lines they choose, in order to represent their actual (or theoretical) trips.

The OSC report makes a valid point that this aspect of the methodology may not have been clear to all users of the metric. NYCT acted on this feedback when revising the explanations in the help section of the dashboard. It now contains the sentences "All estimates are for each individual train a customer uses in their journey, not all trains in their journey combined---in transit terminology, each 'unlinked trip'. Put another way, each 'leg' of a subway journey involving transfers is counted separately in calculating these metrics." Shorter versions of this note were added in other locations, such as the text which comes up when hovering over the information icon above a graph on the dashboard.

Comment 5

Response to Key Recommendations

Comptroller Key Recommendation: Evaluate whether APT, ATT, and CJTP meet the goals of the plan and disclose the assumptions and margin of error for each assumption.

NYCT has evaluated APT, ATT, and CJTP and determined that they meet the goals of the Subway Action Plan. Furthermore, as the Comptroller report noted, NYCT is now legally required to report these metrics, with the definitions written into law. NYCT will focus on continuing to improve performance, while measuring that progress with accurate and meaningful performance metrics.

With respect to disclosing assumptions, NYCT has acted on feedback from the Comptroller by revising and expanding material explaining customer-focused metrics. Definitions on the subway performance dashboard (http://dashboard.mta.info/) have been rewritten to include distinctions the OSC report highlighted as important (e.g., the separate calculation of metrics for each leg of a trip). The help page of the subway performance dashboard has a greatly expanded section under the Frequently Asked Questions, entitled "How are each of the indicators calculated?", which provides much more detail on assumptions than could fit in a definition. NYCT Committee materials are already voluminous so there is a need to avoid adding large blocks of explanatory text. Nonetheless, NYCT is expanding and clarifying the definitions in the NYCT Committee materials within space constraints, and plans to start using the new definitions in early 2020.

While NYCT can outline assumptions and processes, NYCT will not be able to include a margin of error corresponding to each assumption. Margins of error are commonly used to quantitatively describe the uncertainty resulting from using a sample to represent a population (e.g., surveying a small random group of likely voters to represent everyone who will vote on election day). Many of the assumptions used in this model are procedural (e.g., the rules for inferring trip destinations) and hence are not compatible with traditional statistical methods for computing margins of error. Whenever metrics are generated for a given day, it involves simulating a complete day of passenger trips.

Comment 6

It is possible, however, to perform a sensitivity analysis to assess whether the day selected to generate the set of unlinked passenger trips heavily influences results. Toward that end, NYCT will rerun APT, ATT, and CJTP calculations for example day(s) with differing sets of unlinked trips, and measure the effect on performance metrics. This will provide quantitative evidence of the sensitivity to this model input.

Response to Other Sections of the Report

The OSC report states multiple times that CJTP was not available until November 2019 (see page 1 and page 6). CJTP was introduced in September 2018, as is documented in

the September 2018 NYCT Committee materials, which are publicly available on the MTA website (http://web.mta.info/mta/news/books/archive/180924_1030_Transit.pdf).

Comment 7

The OSC report presents a figure on page 14 displaying ridership by day for weekdays in November. There are two clear dips in ridership, but the axis does not make it clear which days they are. It is possible that they represent holidays, and if so this should be explicitly stated. In the interest of transparency, it would help to label the x-axis by the actual date, as opposed to the number of the weekday, so readers can more easily look up the dates. It is important to note that APT, ATT, and CJTP were not calculated for Thanksgiving, so the ridership not being similar on Thanksgiving is irrelevant. For more minor holidays, NYCT will generally overestimate ridership by using ridership from a typical day from the previous month. Because denied boardings due to train capacity are factored into APT, ATT, and CJTP, erring on the side of using a higher ridership than actually rode on a particular day will generally tend to lead to slightly worse values for customer-focused metrics. Therefore, the current approach is slightly conservative in that it errs on the side of understating performance.

Comment 8

More generally, the OSC report focused heavily on how variability of ridership and various assumptions which go into generating the origins and destinations of trips could influence metric results. More specifically, the OSC rightly pointed out that the 6-minute granularity of AFC data means that the precise start time of each trip is not generally known. Station entries in each 6-minute period are randomly distributed across that period. This does mean that for any specific trip, the start time could easily be off by several minutes, making it impossible to say for certain what the APT/ATT of a given trip was. The limitations of destination inference (e.g., that destinations cannot be inferred for MetroCards which are used only once) has a similar effect in that it prevents tracing the exact APT/ATT for specific trips.

The OSC report focused heavily on how NYCT's subway passenger performance metrics are built upon a series of assumptions—necessary assumptions, due to the limitations of available MetroCard data—that introduce a large amount of uncertainty into the calculation of APT, ATT, and CJTP for each individual rider. The OSC report then implies, without substantiation, that the reported aggregate line- and system-level APT, ATT, and CJTP must also be subject to a very large degree of uncertainty.

This criticism ignores one of the fundamental properties of aggregate passenger metrics: while the uncertainty for an individual trip may be high, if this uncertainty is effectively random (in a statistical sense), the uncertainty of the aggregate metric, calculated over hundreds of thousands, or even millions, of trips, is far smaller, small enough to make the aggregate metrics meaningful measures of system performance.

Comment 9

Just as there is a high uncertainty in the outcome of a single coin flip, but low uncertainty in the outcome of 10,000 coin flips, there is low uncertainty in the average value of APT

(or ATT, or CJTP) calculated over 5+ million trips, even if the APT for each individual is highly uncertain. For the OSC to claim that NYCT's customer metrics "do not offer any useful information to customers", they would have had to show that these assumptions produce errors that affect APT/ATT/CJTP *systemically*, i.e. in a single direction, across millions of trips—and this they did not do.

Comment 10

Conclusion

NYCT has carefully reviewed the findings and recommendations in the draft OSC report and has already begun to revise and expand the publicly available explanations of the calculation process to increase transparency. NYCT recognizes that the sophisticated network modeling which makes the calculation of APT, ATT and CJTP possible can be difficult to follow, even though the concept of the metrics is elegantly simple. The complexity is necessary to overcome the limitations of available input data, and NYCT will strive to be as transparent as possible about the methodology and assumptions. Toward that end, NYCT will continue to update and refine these explanations as the methodology evolves.

NYCT will also continue to monitor performance with a portfolio of complimentary performance indicators. These metrics, many of which are calculated independently from each other, tell a coherent and consistent story of service improvement. Since a dropping to a low of 58.1% in January 2018, On-Time Performance has increased dramatically and is now consistently over 80%. This aligns with the reduction in actual running times, and is also seen clearly in ATT, which has improved by roughly a minute over the same time period. As the OSC report pointed out, that minute of savings is the average for each leg of the trip, meaning that riders making transfers could easily be seeing substantially more than one minute of time savings over their entire trip. This outcome is indicative of the success of Fast Forward, Save Safe Seconds, and the Subway Action Plan.

Sincerely,

Andy Byford
President, New York City Transit

State Comptroller's Comments

1. The new customer-focused metrics are not easy to understand. They require that customers understand what is being measured and how, so that they can correctly evaluate them against their experience. For instance, in the report we point out that the definitions of the new metrics during the audit period did not make it clear that they were not for entire trips, but instead for each leg of a trip. Customers view their experience from a trip perspective and, therefore, likely misunderstood what the metrics represented.

2. MTA indicates the new OMNY system will address the MetroCard six-minute interval because the MetroCard would be phased out and replaced by a contactless fare payment system using Apple Pay, Google Wallet, Samsung Pay, debit cards, or credit cards with near-field communication enabled. However, this does not take into account that many customers who use the subway system do not have access to this technology or use credit or debit cards to pay their fares.

3. In response to our draft report, the MTA implies that, because it subjected its methodology to peer review at the Transportation Research Board conference, its methodology is transparent to the riding public. However, this is apples and oranges logic. Whether the MTA peer reviewed its methodology is unrelated to whether the methodology and what the metrics represent was transparent to the riding public.

4. For the record, we did not recommend that the MTA implement a system where customers swipe or tap in when they enter and leave the subway. We disclose this information because Transit used the London system as an example in its response to our preliminary findings.

5. Although Transit revised its explanations on the dashboard to state that estimates are for each individual train in the customer's journey, it is only a partial disclosure because there is no information as to how the total journey time can be determined. Moreover, the dashboard does not provide information on the number of trips that involve more than one train and the total time.

6. It is unclear how Transit validates its assumptions if it has no methodology to evaluate their accuracy. For example, the Destination Assumption was validated by the Massachusetts Bay Transportation Authority 17 years ago using survey data and was found to be accurate 90 percent of the time. Similar to Transit, this authority does not require riders to swipe out. Without a similar analysis, Transit may not know its model is valid.

7. As stated in the report, AJT, a key component of CJTP, was not on the dashboard, along with the other metrics, until November 2019. While information was publicly available, it does not mean that it was transparent to the riding public.

8. The chart includes all the weekdays in November 2018, excluding Thanksgiving.

9. Transit's reply indicates that the report ignores that, while the uncertainty of one ride may be high, the uncertainty of the aggregate is far smaller, making it a meaningful measure. However, the MTA Dashboard Help indicates that the assumptions are not strictly correct and a person's next swipe doesn't always predict their destination. It also states that the metrics are not calculated for late nights and weekends because the assumptions do not hold as well, but an average total of about 5.4 million weekend riders do not have metrics that are relevant and easier to understand.

10. As stated in the report, Transit could not provide any evidence regarding how the metrics were used by customers. In addition, there is no goal for APT and ATT and there was no documentation indicating if any percentage would prompt review or corrective action. Prior to the new metrics, Transit had goals for On-Time Performance and Wait Assessment that were straightforward, monitored, and more transparent.

Contributors to Report

Executive Team

Tina Kim - *Deputy Comptroller*
Ken Shulman - *Assistant Comptroller*

Audit Team

Carmen Maldonado - *Audit Director*
Robert C. Mehrhoff - *Audit Manager*
David DiNatale - *Examiner-in-Charge*
Jim Cherian - *Senior Examiner*
Celedonia Deaver - *Senior Examiner*
Lillian Fernandes - *Senior Examiner*
Susan Gordon - *Senior Examiner*
Andrea Majot - *Senior Editor*

Contact Information
(518) 474-3271
StateGovernmentAccountability@osc.ny.gov
Office of the New York State Comptroller
Division of State Government Accountability
110 State Street, 11th Floor
Albany, NY 12236

Like us on Facebook at facebook.com/nyscomptroller
Follow us on Twitter @nyscomptroller
For more audits or information, please visit: www.osc.state.ny.us/audits/index.htm