State of New York

Carolyn Pokorny
Inspector General

Office of the Inspector General
Metropolitan Transportation Authority
One Penn Plaza, 11th Floor, Suite 1110
New York, NY 10119
212-878-0000

July 2, 2019

Timothy Mulligan
Executive Vice President
MTA New York City Transit
2 Broadway, 30th Floor
New York, NY 10004

Re: NYC Transit Fare Evasion Surveys:
OIG Concerns
MTA/OIG #2019-12

Dear Mr. Mulligan:

In view of the heightened attention to fare evasion on both subways and buses and the associated revenue losses, the Office of the Metropolitan Transportation Authority Inspector General (OIG) examined the methods used by MTA New York City Transit or Transit to estimate the levels of fare evasion in its system. The use of sound, statistically valid methods is essential for producing reliable estimates of both the current levels of evasion and the trends over succeeding quarters and years.

We note that NYC Transit has recently taken multiple measures to control and reduce fare evasion. The impacts of these initiatives would be difficult to assess without reliable data on evasion levels and trends. We also note Transit has recently begun reporting its fare evasion estimates to the MTA Board, thus bringing much greater visibility to these figures. To our knowledge, these estimates previously had been internally held.

OIG reviewed Transit's Subway Fare Evasion (SFE) survey and Bus Fare Evasion (BFE) survey. Our review focused on the survey sampling methodologies: e.g. statistical assumptions, sample sizes, sample selection, confidence intervals, and margin of error formulas. We also learned about data collection practices: Traffic Checker instructions, scheduling process, and the actual data collected. We identified areas of concern in the decade-old design of both surveys, and in their respective data collection practices.

Working with Transit staff, we reviewed six quarterly SFE and BFE reports–from mid-2017 through the end of 2018–along with related information. We presented the concerns covered in this letter in a meeting with Transit Operations Planning staff on April 2, 2019. The staff recently informed us that many changes are under consideration in regards to the methodologies for both surveys.

As one broad point, we note that the SFE and BFE surveys each require Traffic Checkers to distinguish among numerous categories of system entry. Although some of this detail is useful for management decision-making, both surveys may benefit from simplification to make data collection less burdensome and thereby improve the accuracy of the overall fare evasion estimates.

This letter summarizes our major concerns and questions about each of the surveys that were shared orally with members of NYC Transit Operations Planning earlier this year.

BACKGROUND

Transit's Operations Planning Division (OP) conducts quarterly fare evasion surveys that produce statistical estimates of Bus Fare Evasion and Subway Fare Evasion. The Transit Office of Management and Budget (OMB) distributes these estimates in quarterly reports along with breakdowns by numerous subcategories. The OMB quarterly reports also provide several years of prior survey results along with the percentage point changes for recent periods. For each of the two surveys, Traffic Checkers collect the data using a stratified simple random sample.[1]

As with any statistical survey, the evasion surveys produce results that are only point estimates of the true population values. Of course, for fare evasion the true value is unknowable without doing an impossible 100 percent sample. And even then the results would be subject to data collection errors (non-sampling errors). To address the statistical uncertainty of a sample estimate, statisticians calculate the margin of error, using a formula based on the type of sample, the sample size, and the variability within the survey results. This error margin is then used to provide a confidence interval around the point estimate, usually tied to a 95 percent probability. Critically, a statistically valid margin of error formula is the basis for concluding whether the change from a previous period is statistically significant. In the fourth quarter of 2018, Transit reported that the quarterly subway fare evasion rate of 3.4% had a "standard 95% confidence interval of $\pm$ 1.22%." If the error margin is accurate, this means that the there is a 95 percent chance that the true value actually fell between 2.18% and 4.62%.

We will now discuss the two fare evasion surveys separately, beginning with the subway survey.

OBSERVATIONS

The Subway Fare Evasion (SFE) Survey

The SFE survey began around 2008, after MTA Audit Services found that fare evasion was much higher than the estimates produced at the time by Station Agents (2.1 percent vs. 0.32 percent). OP conducted a pilot study in which auditors measured fare evasion at randomly

[1] To stratify means to divide the population into subgroups according to characteristic(s) believed significant and/or for which separate results are desired. Random samples are selected from each subgroup.

selected fare control areas (i.e., contiguous groups of turnstiles/gates within a station).

After analyzing the pilot study results, OP created a methodology for measuring fare evasion with a simple random sample of fare control areas stratified by 20 levels of activity (average ridership across fare control area and hour of day), two median income levels, and three parts of the week (weekday, Saturday and Sunday). Transit staff described the process in a 2011 paper.[2] OP later modified the stratification to use 40 levels of activity and the three parts of the week. (The sample is no longer stratified by median income level). OP randomly selects 280 location-time combinations to survey each quarter. It is important to note that in the six quarters OIG reviewed through the end of 2018, OP completed an average of only about 168 of those surveys (60 percent).

The SFE survey has not had specific, designated Traffic Checkers who perform the work. On a given day, the responsible manager told us he assigns one available Checker from the team of 26 to perform the scheduled SFE surveys. This uniformed Checker monitors sampled fare control areas for 30-minute periods, recording observations on a computer tablet or paper form. The Checker is responsible for accurately tallying 20 types of entry, including:

- Illegal entries (e.g. opportunistic entries through an open exit gate);
- Questionable entries (e.g. passengers entering through a gate with questionable agent authorization);
- Legal entries not captured by the automated fare collection (AFC) system (e.g. agent-authorized entry by fare-paying persons with bulk items).

The Checker manager told us that at fare control areas with multiple exit gates, Checkers are instructed to pick one of the gates to observe and to ignore the other gate(s).

One analyst involved in the original survey design suggested that OP saw the project as an internal management tool to get an estimate of the problem, not an official analysis whose results would be formally reported to the MTA Board or considered a robust and reliable estimate of revenue loss. He said OP made some methodological decisions based on practical limitations and available resources, but in hindsight it appears that some of the statistical assumptions about the nature of fare evasion should have been validated more thoroughly. While the initial process may well have been sufficient for the needs of the time, the SFE survey has become a project with larger implications. Moreover, current OP staff "inherited" the sampling methodology and were not fully aware of the rationale behind it.

OIG has a number of questions and concerns about the SFE survey. These are as follows.

[2] AV Reddy, J Kuhls, A Lu. (2011). Measuring and Controlling Subway Fare Evasion. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2216. Transportation Research Board of the National Academies: Washington, D.C. pp. 85–99. DOI: 10.3141/2216-10.

Concerns About Survey Design:

1. Through the third quarter of 2018, the margin of error cited in OMB's SFE quarterly reports (±0.2 percent) came from the stratified simple random sampling method described in the 2011 methodology paper. This approach viewed fare evasion as a phenomenon represented by a binomial distribution.[3] During our review, OP told us it had recently begun to consider whether it would be more appropriate to assume that the proportion of evaders is *clustered* rather than *uniform*, the latter being an assumption of the binomial distribution. In February 2019, OMB released the 4th Quarter 2018 SFE report with new and larger error margins based on a revised formula that is used for clustered sampling methods. However, OP still uses the same simple random sampling method, which may require adjustments to fit the requirements of a clustered sampling method. More review and validation of the sampling method and margin of error formula may be necessary.
2. As noted, OP completed only about 60 percent of surveys in the period of our review. In the fourth quarter 2018 Survey, OP completed less than 50 percent.[4] Important questions include: What is the impact of this shortfall on the accuracy of the estimates, and thus their representativeness of the entire population of riders? Also, do the missed surveys bias the results (e.g. by disproportionately occurring in certain areas or times of day)?

Concerns about Data Collection Practices:

1. Certain fare control areas have multiple banks of turnstiles and/or visual obstructions that make it difficult to see all gates/turnstiles from any given vantage point. Is it reasonable to expect reliable data from a single Traffic Checker at such complex control areas, especially given the detailed categorization required?
2. The instruction to Checkers to watch a single gate in a fare control area with multiple gates would typically lead to undercounting illegal gate entries. This approach appears arbitrary and we also note that the methodology does not include a factor to compensate for this.
3. OP has not created a manual with instructions for SFE Traffic Checkers. It seems clear that such an important and relatively complex survey should have a standard set of written instructions for training and reference.

[3] A binomial distribution is used to estimate the probable outcomes of a series of events such as coin tosses when each event has two possible outcomes. However, the binomial distribution is based on assumptions--such as the independence of each event--that are questionable when applied to fare evasion. In fact, it appears likely that passenger decisions are not always independent but are subject to influence by the behavior of other passengers.
4 We based these percent-completed figures on the reported *"Total Number of Six Minute Periods Observed,"* dividing by five to estimate the number of 30-minute periods.

4. SFE Traffic Checkers are required to wear MTA uniforms (BFE checkers are not.). Has OP analyzed the impact of the presence of uniformed personnel on evasion behavior? [5]
5. Currently, SFE Traffic Checkers have limited availability, are shared with other surveys, and often can't be substituted for in their absence. If a Checker leaves the job or is out long-term, his/her position cannot be filled until the following pick. These constraints have a direct impact on the number of checks completed, which affects the sample's statistical validity when applied to the population as a whole.

The Bus Fare Evasion (BFE) Survey

The BFE survey began around 2008, as part of OP's effort to report ridership and daily bus passenger-miles to the Federal Transit Administration (FTA). At the time, NYCT was seeking FTA's approval for a way of estimating bus passenger-miles using AFC data. Because AFC data only captures paying customers, OP designed a survey to estimate what proportion of riders do not interact with the farebox (e.g. fare evaders, children riding free with an adult, uniformed officials using a flash-pass). Using the results of this survey, OP was able to estimate a system-wide fare evasion rate. OP staff described the process in a paper published in 2011.[6]

Around 2011, OP revised the methodology in order to produce borough-level estimates of fare evasion. In its current form, the BFE methodology uses a simple random sample stratified by day type (weekday, Saturday and Sunday) and by borough. Transit's BFE survey covers only Regular Bus Service (RBS), excluding certain types of RBS (e.g. school trips, very late night trips). The survey also excludes Express Bus Service, Select Bus Service (SBS), and MTA Bus. Transit's Eagle Teams have a distinct method for estimating fare evasion on SBS. This Eagle Team effort is separate from the BFE survey.[7] MTA Bus Operations Planning does its own BFE survey for MTA Bus local buses. MTA Bus officials told us they follow Transit's BFE survey procedures.

Traffic Checkers ride on the sampled buses and record each boarding under one of 13 different categories of legal or illegal entries. In addition to three categories of MetroCard/coin payments, legal entries include multiple types of flash passes, badges, and passengers in uniform. Illegal entries include front door evaders, rear door boardings, and multiple types of incomplete payments.

---

[5] In April 2019, OP told us it has a plan to create a separate cadre of Checkers who would work only on fare evasion surveys. OP said this would eliminate the requirement for SFE Checkers to wear uniforms.

[6] A Lu, A Reddy. (2011). Algorithm to Measure Daily Bus Passenger Miles Using Electronic Farebox Data for National Transit Database Section 15 Reporting. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2216. Transportation Research Board of the National Academies: Washington, D.C. pp. 19–32. DOI: 10.3141/2216-03.

[7] For the first time, OMB's BFE report for the 4th quarter of 2018 included a separate line for the Eagle Team's SBS evasion estimate as well as merged figures for RBS and SBS.

At the beginning of each Bus Operator pick, OP selects a random sample of bus trips to survey. Because calendar quarters do not usually align with the periods of Bus Operator picks, OP aggregates surveys from multiple picks to produce quarterly estimates of fare evasion. OP says the sample it draws is about double what it needs to produce a statistically valid estimate. The methodology for determining the sample size in each pick has undergone changes during the time of our review.

Again, OIG has a number of questions and concerns to raise about the BFE survey. These are described below.

Concerns About Survey Design:

1. The sample selection is not fully random because, for practical reasons, each sampled trip is paired with a return trip on the same route. Also, the sample is built around Bus Operator picks, which do not correspond to the quarterly reporting period. The results from one quarter, therefore, actually include observations from two or three different random samples. OP should assess what implications these factors have on the margin of error.
2. OP told us that it completed 321 samples in the 4th Quarter 2018. In quarters when OP does not complete its full sample, is there statistical bias (day of the week, time, geography, etc.) in which surveys are omitted? Such bias would affect the sample's representativeness of the full population of riders.
3. The margin of error calculation (used for many years to determine the target sample size) assumed that fare evasion events are independent and identically distributed within each stratum, e.g. the method implicitly assumed that all weekday trips in Manhattan have the same evasion rate regardless of route or time of day. This assumption seems questionable given that, empirically, certain routes are known to have significantly more fare evasion than others within the same borough and day of week. This may have led OP to draw an insufficient number of samples and/or overstate the accuracy of the estimates. OP recently changed the BFE's margin of error formula to follow the SFE group's new cluster formula described above. Is the new margin of error formula now compatible with the longstanding approach to sampling and data collection, which OP has not revised? More review and validation of the sampling method and margin of error formula may be necessary.
4. In order to determine how many bus trips must be surveyed to obtain the target number of boardings, OP uses estimates of the average number of passengers per bus trip ("load factor") in each borough. However, OP does not calculate separate load factors for weekday/Saturday/Sunday, even though the 15 strata are so differentiated. Furthermore, OP bases the load factors on results from past BFE surveys, which represent a relatively small number of trips. Has OP considered calculating load factors with AFC data covering all bus trips in a given stratum (adjusted for average fare evasion)?

5. Express bus service is not covered by any of the fare evasion surveys. OP excluded express bus routes from the BFE survey in 2014 (because express bus fare evasion had been relatively low) but OMB did not footnote the change or adjust the prior-year estimates shown in the subsequent quarterly reports.

Concerns about Data Collection Practices:

1. Traffic Checker availability issues make it difficult to cover all sampled trips: OP has only six dedicated BFE Checkers, whose shifts can only include only one weekend day. OP expects around 20 percent Checker unavailability. If a Checker leaves or is out for a prolonged period, her/his position cannot be filled until the following pick. At the end of the last 2018 pick, only four BFE Checkers remained out of the six. The Checkers' schedules also prevent sampling any bus trips occurring very late at night.
2. Can one Checker accurately monitor all-door evasion, especially on more crowded buses, and given the detailed categorization required? OP acknowledged that it can be difficult to monitor both front and rear doors, and three-door articulated buses are particularly challenging. In some situations, it can be hard for a Checker to hear the distinct tone the farebox produces to indicate an invalid MetroCard or incomplete fare payment. In addition, "short drops" (the payment of less than full fare) can be hard to detect because the Operator does not always press the reset button.

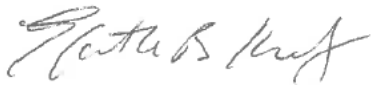
Question Raised by Operation Planning

An OP manager raised the concern that methodological changes to the surveys might make new results not comparable to past estimates. He said this could pose difficulties when there is so much focus on the trends and how they relate to changes in enforcement and other new strategies. Other OP staff responded that methodological changes might primarily affect the error margins, and the new methods could still produce point estimates that are generally comparable to past results. OIG believes that if Transit makes sampling and data collection changes that substantially affect the point estimates, it would still be beneficial to have more accurate results. In any case, future reports could reference the changes, and in the long run it would be preferable to have more confidence in the reported trends.

To that end, we encourage MTA New York City Transit to undertake a thorough review of the survey design methodologies and data collection practices used in its bus and subway fare evasion surveys and revise them as necessary to provide reliable estimates and trends. Such a review may require appropriate consultant resources due to its complexity, but the need to have confidence in fare evasion statistics is critical.

As always, we appreciate the courtesy and cooperation afforded to us by NYC Transit staff. If you have any questions, please do not hesitate to call me at (212) 878-0022.

Very truly yours,



Elizabeth Keating
Executive Deputy Inspector General