

Appendix I: Detailed Methodology

How We Developed Our Profiles and MetroCard Ratings

This appendix describes in detail the methodology used by the NYPIRG Straphangers Campaign to develop our profiles of New York City subway lines and our MetroCard Ratings.

Essentially, we reviewed six measures of transit performance compiled by MTA New York City Transit, presented them in concise, one-page, rider-friendly profiles, and then gave a MetroCard Rating based on their overall performance. We chose these six measures — which are fully described below — for several reasons. This included their importance to riders, as reflected in New York City Transit's polling of riders and in our own survey of 38 transit experts. Additional factors included the availability, reliability and comparability of the data.

The MetroCard Rating was developed in two steps, explained more fully below. First, we decided how much weight to give each of the six measures of transit services in our profiles. Then we placed each line on a scale that permits fair and consistent comparisons. Under that formula, a line whose 2010⁷ scores fell on average at the 50th percentile of lines for all six performance measures would earn a MetroCard Rating of \$1.15. Those scoring at the 95th percentile would receive a rating of \$2.25, the current base fare.

1. Presenting New York City Transit Data to Riders

We decided to report data in the form of concise one-page profiles for each subway line. That met our goal of presenting the information in a way that would be easily understandable to the riding public.

Below is a description of each of six measures of transit performance that we used. We have listed the published sources of the data; if no published source is listed, we received the data from MTA New York City Transit in electronic form. In 1997, New York City Transit officials reviewed a draft version of the profiles and concluded:

Although it could obviously be debated as to which indicators are most important to the transit customer, we feel that the measures that you selected for the profiles are a good barometer in generally representing a route's performance characteristics...Further, the format of your profiles...is clear and should cause no difficulty in the way the public interprets the information.⁸

An advance summary of the findings for the 2011 State of the Subways report was provided to MTA New York City Transit.

⁷ Data on scheduled service, service regularity, breakdown rate, interior cleanliness and in-car announcements was taken from MTA New York City Transit 2010 sources. On crowding, we cite 2009 MTA NYCT cordon count and 2010 passenger loading guidelines — the most recent data available at the time of this report's preparation. For the sake of brevity, we refer to data from sources cited in this report as '2010 data.'

⁸ The draft included the same six measures of service as the final version. Transit officials did note that for some lines, "it may be more useful to present the profiles by corridor rather than individual route...especially for such high-volume corridors as the Lexington Avenue express." (Source: Letter, to Gene Russianoff, staff attorney, Straphangers Campaign from Lois Tandler, Acting Chief of Operations Planning, MTA New York City Transit, April 17, 1997.) Since all the data we use is broken down by line, we felt the profiles should reflect this.

A. Scheduled headways

We measured amount of service based on the scheduled "headways" between trains for weekday morning rush, afternoon rush and midday hours. Headways are the number of minutes scheduled between train arrivals. For example, the 4 line is scheduled to arrive every four minutes during the weekday morning rush. Because virtually all subway lines operate at the same interval — 20 minutes — during late night hours, we did not include overnight headways in our analysis. This approach allowed us to include the B and C—two train lines that do not regularly operate during overnight hours.

For our profiles, we decided to have the morning and afternoon rush hour intervals each contribute 40% to the overall headways measurement; midday headways account for the remaining 20%. We felt that this fairly reflected the relative use of service. For any line which has different scheduled intervals for northbound and southbound trains, the average headway was considered. Due to changes in the way MTA New York City Transit reports its headway data, the amount of scheduled service figures cited in this report may not be comparable to those published in our thirteen previous reports. System average data was calculated by averaging data by time period from the 20 lines measured in this report. (Source: Subway Service Frequency (Headway in Minutes): By Route and Time of Day, A and B Divisions, December 12, 2010, MTA New York City Transit Subway Schedules Division of Operations Planning, Department of the Executive Vice President.)

B. Regularity of Service

Regularity of service measures the adherence of *actual* intervals to *scheduled* intervals between trains. A line with a *low* regularity, for example, would show either gaps in train service during some portion of the day, and/or train bunching at others.

In 2001, MTA New York City Transit created a new measure of this indicator, called *wait assessment*:

The percentage of service intervals is no more than the scheduled interval plus 2 minutes during the hours of 6 a.m. to 9 a.m. and 4 p.m. to 7 p.m., or plus 4 minutes during the hours of 9 a.m. to 4 p.m. and 7 p.m. to 9 p.m.

In 2008, transit officials modified this definition:

[Wait assessment] is the percent of instances that the time between trains does not exceed schedule by more than 2 minutes (peak) or 4 minutes (off-peak). The reporting time is 6:00 a.m. to midnight.

This reporting time (6 a.m. to midnight) represents a departure from the reporting time used by transit officials between 2001 and 2007. Further, 2009 data was published as 12-month, rather than 6-month, rolling averages.

In 2010, transit officials changed the definition again:

Wait assessment (WA), which is measured weekdays between 6:00 a.m. - midnight is defined as the percent of actual intervals between trains that are no more than the scheduled interval plus 25%.

Transit officials continue:

Beginning in July 2010 the 1 thru 6 line results reflect the current month (Dec. 2010) data for all train trips derived from ATS-A. All results for prior year and the current results for the 7, S 42nd St., BMT and IND divisions are based on a sample methodology and reported as a 12 month rolling average. Systemwide results reflect 12 month rolling average for all lines...IRT Division results reflects the 12 month weighted average of sample data and all ATS-A train trips for the 1 thru 6 lines averaged with the 12 month sample data for the 7 and S 42nd St.

(Source: Operations Performance Summary, Subway Weekday Wait Assessment (6 am – midnight) 12 Month Averages, January-December 2010, MTA New York City Transit Committee Meeting Agenda, February 2011, p. 3.4-3.5.)

C. Mean distance between failures (MDBF)

MTA New York City Transit states that MDBF measures subway car reliability and "is calculated as revenue car miles divided by the number of delay incidents attributed to car-related causes." In this report we cited data for the 12-month rolling average for the period ending December 2010. The system average quoted is the "fleet average" published by MTA New York City Transit. (Source: Car Reliability — Mean Distance between Failures by Line (Miles), MTA New York City Transit Committee Meeting Agenda, February 2011, p. 3.15.)

D. Chance of getting a seat

We developed a formula to calculate the chance of being able to get a seat at the most crowded point on each line. First, we identified each line's "instance of greatest crowding" using New York City Transit's 2009 *Weekday Cordon Count*. We did this by isolating for each line the most crowded 1-hour interval at the most crowded point entering or exiting Manhattan's Central Business District (CBD). Then we divided the number of seats on all cars on each line by the number of passengers during that 1-hour interval. For example, the 1 line was at its most crowded point exiting the CBD, uptown at West 66th Street, between 6:00 p.m. and 7:00 p.m. on the day the count was taken; the average number of passengers counted was 85.813 per car. Cars on the 1 line are of the class R62-A, a 51-foot A-subdivision car with 44 seats. Thus the ratio of the number of seats to the total number of passengers per car would be 44/85.813 or 52%. This figure, 52%, represents the chance that a rider will be able to get a seat on a train at the 1 line's most crowded point entering/exiting the CBD, during the most crowded 1-hour weekday interval. In cases where more than one car class was used on a line, we evaluated crowding based on the seating guidelines for the predominant type of car used on the line. As the G line does not enter the CBD, passenger loading data is not included in MTA New York City Transit's cordon count. For this reason, we report no crowding data for the G line in this report, and subsequently, calculate no MetroCard Rating for the G. In addition, no passenger crowding data has been made available for the M line following substantial 2010 changes to the route pattern. Therefore we report no crowding data, and subsequently, give no MetroCard Rating for the M line in this report. System average data was calculated by averaging the 'chance of getting a seat' scores of 19 lines cited in the cordon

count. (Source: New York City Transit Subdivision 'A' and 'B' Car Assignments December 12, 2010, and *Year 2009 Weekday Cordon Count*, MTA New York City Transit Division of Operations Planning, Department of the Vice President.)

E. Passenger Environment Surveys: Cleanliness and Announcements

New York City Transit conducts a periodic "Passenger Environment Survey" (PES) to measure the quality of the transit environment experienced by riders. It does this for subway cars, stations and buses and releases the results semi-annually. The PES is performed by "surveyors who are specifically trained for this function and who have no direct association with the departments affected by the survey evaluations. The surveying of...subway cars is conducted throughout each quarterly recording period to the extent necessary to depict a 'representative' sample of NYC Transit's vehicles."⁹ Our profiles represent the first time that PES findings have been presented to the public on a line-by-line basis. We included PES ratings on cleanliness and announcements because New York City Transit's own surveys of rider opinion show both are of major concern to riders.

(1) Interior cleanliness

The PES includes a rating on the cleanliness of the interior of subway cars while in passenger service. It defines a car with a light degree of dirtiness as one with "occasional 'ground in' spots, but generally clean." Interior cleanliness in our profile was measured as the sum of the total percentage of cars with "no dirtiness of car floors and seats," and those with "a light degree of dirtiness of car floors and seats" during the last six months of 2010. System averages calculated by Operations Planning were cited in this report. (Source: MTA New York City Transit's Department of Operations Planning — Systems Data and Research, Passenger Environment Survey data for the second half of 2010.)

(2) Adequacy of routine in-car announcements

In-car announcements are also monitored in the Passenger Environment Survey. Our profiles note the average percentage of cars with adequate "routine in-car announcements" for the last six months of 2010. PES rates announcements as adequate if they are "understandable" and "correct." Such announcements would include at least four of the five following items:

- next station announced;
- transfer options, if applicable;
- route designation announced (either letter or number corresponding to a train line);
- next station announced; and
- "stand clear of the closing doors" announced.

This represents a change in PES methodology made in the third quarter of 2000. System averages calculated by Operations Planning were cited in this report. (Source: Department of Operations Planning — Systems Data and Research, Passenger Environment Survey data for the second half of 2010.)

⁹ Source: MTA New York City Transit *Passenger Environment Survey*, Third Quarter, 1996, p. 1.

2. Developing a Straphangers Campaign MetroCard Rating

We decided to include one overall rating for each of 18 subway lines. The rating is intended as a shorthand tool to summarize all of the information reported in the profiles and to allow for comparisons among lines.

As described below, under the formula used, a line whose 2010 scores fell on average at the 50th percentile of all lines for all six performance measures would receive a MetroCard Rating of \$1.15. A line which matched the 95th percentile of this range would receive a line rating of \$2.25. However, some lines which ranked high on some measures of performance may have received only an average MetroCard rating due to poor relative performance in other areas. (See Figure 1.)

Some riders may find this scale too generous, believing that performance levels should be far better than they are now. Other riders, who value transit service over other ways to travel in New York City, may believe the subways and buses to be a bargain.¹⁰

The MetroCard rating does not seek to make a subjective value judgment of the worth of subway service. It is not based on economic factors, such as the cost of providing service or comparisons to the costs of other modes of transportation. Instead, it is only a yardstick that permits a simple and direct ranking of subway lines.

A. Ranking Subway Performance Indicators

We used two sources of information to formulate a scale of the relative importance of various subway line performance indicators. First, the Straphangers Campaign conducted a poll of 38 transit experts, activists and members of the riding public. We asked them to rank eight indicators of subway performance that opinion polls indicated were of major concern to riders. Second, we examined MTA New York City Transit's own rider opinion surveys. One performance indicator, crime, ranked high in both the Straphangers Campaign's poll and in the MTA rider surveys, but could not be included in the profile project because applicable data was not available on a line-by-line basis. A second measurement, "enroute schedule adherence," (commonly referred to as "on-time performance") was dropped from consideration because New York City Transit made changes to its definition and to the time periods surveyed.

Three lines — the Grand Central, the Franklin Avenue and the Rockaway Shuttle — were dropped from consideration because not all six measures of service were available for these lines. Due to the lack of reliable crowding data for the G and M lines, no MetroCard Rating was calculated for those lines.

The remaining 18 subway lines were evaluated on the basis of six indicators, all regularly measured by New York City Transit on a line-by-line basis. Each measure was assigned a percentage weight based on the priorities expressed by those polled as follows; the measures are grouped by the type of indicator:

¹⁰ For example, in critiquing an earlier draft of our profile, transit officials said: "After all, what alternative mode of transportation along Lexington Avenue can even remotely compare at a cost of \$1.50 to the speed, frequency, and we dare say, reliability of the subway service." (Source: Letter, April 17, 1997, to Gene Russianoff, staff attorney, Straphangers Campaign from Lois Tendler, Acting Chief of Operations Planning, MTA New York City Transit.)

Amount of service	(total: 30%)
<i>Scheduled headways</i>	30%
Dependability of service	(total: 35%)
<i>Regularity of service</i>	22.5%
<i>Mean distance between failures</i>	12.5%
Comfort/usability	(total: 35%)
<i>Chance of getting a seat</i>	15%
<i>Interior cleanliness</i>	10%
<i>Adequacy of in-car announcements</i>	10%
	(total = 100%)

Data and/or specific calculations on three of these indicators — chance of getting a seat, interior cleanliness and in-car announcements — have never before been released to the public on a line-by-line basis.

B. Calculating the MetroCard Rating

In our first five surveys, 1996 performance data served as a baseline. As we said in our 1997 report, Line Ratings "will allow us to use the same formula for ranking service on subway lines in the future. As such, it will be a fair and objective barometer for gauging whether service has improved, stayed the same, or deteriorated over time."

However, in 2001, 2003, 2004, 2005, 2008, 2009 and 2010, transit officials made changes in how performance indicators are measured and/or reported. The Straphangers Campaign unsuccessfully urged MTA New York City Transit to re-consider its new methodologies, because of our concerns about the fairness of these measures and the loss of comparability with past indicators. Transit officials also rejected our request to re-calculate measures back to 1996 in line with their adopted changes. As a result, in this report we were forced to redefine our baseline with current data, and considerable historical comparability was lost.

Due to these changes by New York City Transit in the measurement and reporting of data, we have established a new baseline this year — largely 2010 data published in this report — against which we hope to be able to compare future performance.

C. 2011 MetroCard Ratings

18 lines were rated on a linear scale for each of the six measurements used. A line in 2010 equaling the system best would receive a score of 100 for that indicator, while a 2010 line matching the system worst would receive a score of 0. Thus all lines in this report received a score for each measurement between 0 and 100. The lines' scores were then multiplied by the respective weights assigned each indicator. The six adjusted scores were then added up (see Figures 1 and 2), which use the 4 line as an example.

The summed totals were then placed on a logarithmic scale. This scale emphasizes the relative differences between scores nearest the top and bottom of the scale, and ensures that a small difference in performance between any two lines translates to a small difference between scores.

Our method reflects the odds, rather than the percentage chance, of any train on a given line meeting a basic level of satisfaction.^{11,12}

Finally, we converted the scale to a dollar-based line rating, to offer riders a simple basis for comparisons among lines.¹³ We calibrated this scale so that a line whose 2010 scores fell on average at the 50th percentile of all lines for all six performance measures would receive a rating of \$1.15. A line which matched the 95th percentile of this range would be rated \$2.25. Each figure was then rounded to the nearest 5¢. The actual scale used to convert summed raw scores to line ratings is shown on the following page, with the 4 line as an example. (See Figures 2 and 3.)

Because of changes in data reporting at New York City Transit, 2011 Straphangers MetroCard Ratings cannot be compared to ratings given in previous State of the Subways reports.

¹¹ J. H. Aldrich and F. D. Nelson, *Linear Probability, Logit and Probit Models*. Sage Publications, Beverly Hills, CA., 1984.

¹² G. Henderson, H. Adkins and P. Kwong, "Subway Reliability and the Odds of Getting There on Time," *Transportation Research Record 1297: Public Transit Research: Management and Planning*, 1991, p. 10-13.

¹³ This rating method is similar to the "hedonic" method of ranking items based on user satisfaction, as originally developed by Sherwin Rosen. (Source: S. Rosen, 'Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition,' *Journal of Political Economy* 82, p. 34-55.)

Figures 2 and 3: Sample Straphangers Campaign MetroCard Rating

Indicator	Highest rating in system—2010	Lowest rating in system—2010	4 line value	4 line score	Percentage weight	4 line adjusted score
Scheduled Service	6 line	C line	AM rush — 4 min; noon — 8 min; PM rush — 4 min	71	30%	21
Service Regularity	85% regular	66% regular	68% regular	12	22.5%	3
Breakdown Rate	every 843,598 miles	every 54,838 miles	every 167,534 miles	14	12.5%	2
Crowding	70% seated	28% seated	33% seated	13	15%	2
Cleanliness	96% clean	87% clean	91% clean	44	10%	4
Announcements	100% adequate	72% adequate	97% adequate	89	10%	9
Summed adjusted score		4 line -- 41 pts (after rounding of all weighted indicator scores) -- 14th best in system				

