Appendix I: Detailed Methodology
How We Developed Our Profiles and MetroCard Ratings

This appendix describes in detail the methodology used by 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 2004 scores fell on average at the 50th percentile of 21 lines for all six performance measures would earn a MetroCard Rating of $1.00. Those scoring at the 95th percentile would receive a rating of $2.00.

1. Presenting New York City Transit Data to Riders

Our first step was to gather information on transit performance. 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 diskette 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.[9]

[9] 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 Tendler, Acting Chief of Operations Planning, MTA New York City Transit, April 17, 1997.) Since all the data we use is compiled by New York City Transit by line, we felt the profiles should reflect this.

An advance summary of the findings for the 2005 State of the Subways report was provided to MTA New York City Transit.

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 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 train lines which do not regularly operate during overnight hours, like the 3, B, C, V and W lines.

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 reflected the relative use of service. For any line which has different scheduled intervals for northbound and southbound trains, the average headway was reported. Due to changes in the way NYCT reports its headway data, the amount of scheduled service figures cited in this report are not comparable to those published in our seven previous reports. System average data was calculated by averaging data by time period from the 22 lines measured in this report. (Source: Frequency of service: *approximate scheduled intervals between trains in minutes* (effective February, 2004), *weekdays*.)

B. Regularity of Service

Regularity of service measures the adherence of *actual* intervals to *scheduled* intervals between trains. In 2001, MTA New York City Transit created a new measure of this indicator, called Wait Assessment:

> The percentage of [actual] service intervals 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.

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. (Source: "Service Quality Indicators, Second Half 2004 and Second Half 2003," *MTA New York City Transit Committee Meeting Agenda,* February 22, 2005, p.182-5.)

C. Mean distance between failures (MDBF)

MTA New York City Transit states that MDBF measures subway car reliability, and "is determined by dividing the total number of subway car miles traveled in revenue service by the total number of mechanical failures that result in train delays."[10] In this

[10] "Subway Service Performance—Mean Distance Between Failures," *MTA 1269(d) Appendix to the Strategic Business Plan*, July 1998, p. 9.

report we cited data for the 12-month moving average which covers performance during the 2004 calendar year. (Source: Department of Subways Division of Car Equipment Projects and Operations Monthly MDBF By-Line Report data for the year 2004 and "Operations Performance Summary: Subway Service," *MTA New York City Transit Committee Meeting Agenda,* March 29, 2004, p. 4.)

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 2003 *Weekday Cordon Count*. We did this by isolating the most crowded 1-hour interval for each line, at the most crowded point of the route listed.[11] Then we divided the number of seats on cars in a line by the number of passengers at that 1-hour interval. For example, the 1/9 line was at its most crowded point at 59th Street in Manhattan, heading uptown between 6:00 and 7:00 p.m. on the day the count was taken; the average number of passengers counted was 87.7 per car. Cars on the 1/9 line are of the class R62A-WH, a 51-foot A division car with 40 seats. Thus the ratio of the number of seats to the total number of passengers per car would be 40/87.7, or 46%. This figure, 46%, represents the chance that a rider will be able to get a seat on a train at the 1/9 line's most crowded point, during the most crowded 1-hour weekday rush hour 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. Crowding data reported in our first five State of the Subways report is not comparable to the data we calculated for this report because of changes in the cordon count methodology used by MTA New York City Transit Operations Planning. System average data was calculated by averaging the 'chance of getting a seat' scores of 21 lines cited. (Source: *New York City Transit Subdivision 'A' and 'B' Car Assignments* and *Year 2003 Weekday Cordon Count*.)

E. Passenger Environment Surveys: Cleanliness and Announcements

New York City Transit conducts a semiannual "Passenger Environment Survey" (PES) to measure the quality of the transit environment experienced by riders. It does this for subway cars, stations and buses. 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. . .recording period to the extent necessary to depict a 'representative' sample of NYC Transit's vehicles."[12] Our profiles represent the first time that PES findings have been presented to the public on a line-by-line basis. We

[11] The *Weekday Cordon Count* does not measure crowding on the G lines. In its 2002 report, the V line was not included.
[12] Source: MTA New York City Transit *Passenger Environment Survey,* Third Quarter, 1996, p. 1.

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 average of the total percentage of cars with "no dirtiness of car floors and seats," or "a light degree of dirtiness of car floors and seats" during the last six months of 2004. 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 2004.)

(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 adequate "routine in-car announcements" for the last six months of 2004. PES rates the "percent of number of correct announcements vs. total potential announcements expected," including:

- next station announced (while enroute or in a station);
- transfer options, if applicable;
- route designation announced (either letter or number corresponding to a train line);
- route designation announced (borough/terminal);
- next station announced (while at or when leaving a station);
- "stand clear of the closing doors" announced.

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 2004.)

2. Developing a Straphangers Campaign MetroCard Rating

We decided to include one overall rating for each of 21 subway lines.[13] 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 2004 scores fell on average at the 50th percentile of all lines for all six performance measures would receive a

[13]The G line was included in the profile project, but not assigned a Straphangers Campaign MetroCard Rating because crowding data for this route originates from a source different than the one used for the other lines in the system and is not considered comparable by MTA New York City Transit.

MetroCard Rating of $1.00. A line which matched the 95th percentile of this range would receive a line rating of $2.00. 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.[14]

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 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 does not record this for rush hour and midday trains.

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. The G was included in the overall charts summarizing our findings, but did not, as noted above, receive a MetroCard rating.

The remaining 21 subway lines were evaluated on the basis of six indicators. All the indicators are 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:

Amount of service	(total: 30%)
Scheduled headways	*30%*

[14]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.)

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

Three of these indicators—the mean distance between failures, 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—1997-2001

In our first six 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, that was not possible in 2003. As we noted in our 2001 Report Card: "In May 2001, transit officials made major changes in how several of the indicators are derived. The Straphangers Campaign unsuccessfully urged 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. Since transit officials rejected our request to re-calculate measures back to 1996 in line with their adopted changes, some historical comparability may be lost in future State of the Subways reports." That has, in fact, become the case. As a result, we were not able to compare the performance of lines on four of the six measures.

New York City Transit procedures in the measurement and reporting of data have further changed in the past two years. For this reason, we established a new baseline this year—2004 data published in our 2005 report—against which to compare future performance.

C. Calculating the MetroCard Rating—2004

21 lines were rated on a linear scale for each of the six measurements used. A line in 2004 equaling the system best would receive a score of 100 for that indicator, while a 2004 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 afforded each indicator. The six adjusted scores were then added up, as shown in Figure 1, which uses 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. This method reflects the odds, rather than the percentage chance, of any train on a given line meeting a basic level of satisfaction.[15,16] (See Figure 2.)

Finally, we converted the scale to a dollar-based line rating, to offer riders a simple basis for comparisons among lines.[17] We calibrated this scale so that a line whose 2004 scores fell on average at the 50th percentile of all lines for all six performance measures would receive a rating of $1.00. A line which matched the 95th percentile of this range would be rated $2.00. 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, 2005 Straphangers MetroCard Ratings cannot be compared to ratings given in previous State of the Subways reports.

[15] J. H. Aldrich and F. D. Nelson, *Linear Probability, Logit and Probit Models*. Sage Publications, Beverly Hills, CA., 1984.
[16] 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," p. 10-13, Washington, D.C., 1991.
[17] 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, Vol. 82, No. 1 (Jan 1974), pp. 34-55.)