

## **The Subways Seeded the Massive Coronavirus Epidemic in New York City**

Jeffrey E. Harris\*

Department of Economics  
Massachusetts Institute of Technology  
Cambridge MA 02139 USA  
[jeffrey@mit.edu](mailto:jeffrey@mit.edu)

Updated April 24, 2020

*National Bureau of Economic Research Working Paper No. 27021*, April 19, 2020

*Social Science Research Network No. 3574455*, April 13, 2020

*Abstract.* New York City's multipronged subway system was a major disseminator – if not the principal transmission vehicle – of coronavirus infection during the initial takeoff of the massive epidemic that became evident throughout the city during March 2020. The near shutoff of subway ridership in Manhattan – down by over 90 percent at the end of March – correlates strongly with the substantial increase in the doubling time of new cases in this borough. Subway lines with the largest drop in ridership during the second and third weeks of March had the lowest subsequent rates of infection in the zip codes traversed by their routes. Maps of subway station turnstile entries, superimposed upon zip code-level maps of reported coronavirus incidence, are strongly consistent with subway-facilitated disease propagation. Reciprocal seeding of infection appears to be the best explanation for the emergence of a single hotspot in Midtown West in Manhattan.

\*The comments of the following individuals are greatly appreciated: Robin Bell, Jay Bhattacharya, Marlin Boarnet, Ken Boynton, Gil Brodsky, Peggy Cardone, Lee Cohen-Gould, Philip Cooley, Mike Cragg, Peter Diamond, Denise Everett, Richard Florida, Michael Fulginiti, Mariana Gerstenblüth, Daniel Geselowitz, Ray Girouard, Beatriz González López-Valcarcel, Michael Grovak, Joseph Guernsey, Robert Hanlon, Ali Harris, Barry Harris, Dena Harris, Jarrett Harris, Neil Harrison, Bill James, Paul Joskow, Thomas Kalb, Stuart Katz, Karl P. Keller, Ronald Klempner, Moritz Kraemer, Ronald Laporte, Kathryn Blackmond Laskey, Ken Laskey, Zoe Lizarre, John Lowell, Marylee Maendler, Mark Mandell, Melissa Oppenheim Margolis, Andrea Lubeck Moskowitz, Sean X. Luo, Heide O'Connell, David Posnett, Andrew Racine, Thomas Reichert, June Blender Rogers, Ron Rogers, George Rutherford, Brina Sedar, Todd W. Schneider, Susan Goldberg Simon, Tim Sullivan, Kieran Smith, Rivana Cohen Stadlander, Peter Temin, Pat Tracy, Patricia Triunfo, Shuang Troy, Mark Weinstein, William Welch, William Wheaton, and Delbert Yoder. The opinions expressed here are solely those of the author and do not represent the views of the Massachusetts Institute of Technology, Eisner Health, the National Bureau of Economic Research, or any other individual or organization. The author has received no direct or indirect remuneration for this article, and has no conflicts of interest to declare. This is the second article in a series. For the first article, see [Harris \(2020\)](#).

## Introduction

The starting point for this study is the singular nature of the COVID-19 epidemic in New York City. By the third week in April 2020, total confirmed coronavirus infections topped 145 thousand, or about one-sixth of all reported cases in the U.S. This cumulative total was considerably greater than the combined number of reported cases in the counties comprising Chicago, Detroit, Los Angeles, Miami, Boston, Philadelphia, New Orleans, Seattle and Houston combined. The New York City tally has exceeded total cases in the Lombardy region of Italy, the Community of Madrid and the Province of Tehran combined.

How could the epidemic have spread so extensively in such a relatively small space and in just a few weeks? To address this question, we focus here on the city's pervasive, multipronged subway system. Based largely on observational data, we conclude that in all likelihood, the subway system was a major disseminator – if not the principal transmission vehicle – of coronavirus infection during the initial exponential takeoff of the epidemic during the first two weeks of March 2020. Moreover, the ensuing marked decline in subway use was the main vehicle by which the public's growing perception of risk was translated into reduced transmission of the virus.

Our study does not exclude other mechanisms for the primary spread of coronavirus infection. The public schools, for example, may have played a significant role. This hypothesis is indirectly supported by the key roles played by the closures of public schools and the subsequent vaccination of young schoolchildren in blunting outbreaks of influenza in mid-twentieth century Japan (Reichert et al. 2001). While the New York City public school system has educated over 1.1 million students in more than 1,700 public schools, the city's public subway system, we shall soon see, has typically chauffeured more than 5 million rides per working day – from Eighth Avenue in Manhattan to Euclid Avenue in Brooklyn, from Lexington Avenue in the Bronx, with just one transfer, to Forest Hills–71<sup>st</sup> Avenue in Queens.

## Reported COVID-19 Cases and Subway Turnstile Entries

Figure 1 simultaneously tracks the daily movements of two variables from March 1 through April 3, 2020. The pink-filled circles show the numbers of new coronavirus infections reported each day by the New York City Department of Health (New York Department of Health and Mental Hygiene 2020). For this variable, the vertical axis on the left is rendered on a logarithmic scale. That way, a straight-line trend would represent the exponential growth

typically seen during the initial upsurge of an epidemic where everyone in the population is naïve to the infectious agent (Harris 2020).

For the same variable of newly reported cases, the horizontal axis at the bottom ticks off the date that the coronavirus test was *performed*. By contrast, in Figure 1 of the first article in this series (Harris 2020), we tracked newly reported infections in relation to the date the test results were *received*. The new reporting convention, which has been recently adopted by the city’s health department, has the advantage that it cuts out the delay between the date that a healthcare worker swabbed a sample from a patient’s nose or throat and the date that the laboratory notified the department of the test result. It has the disadvantage, however, that the most recent daily counts are unreliable because the department is still waiting for the lab reports to come in.

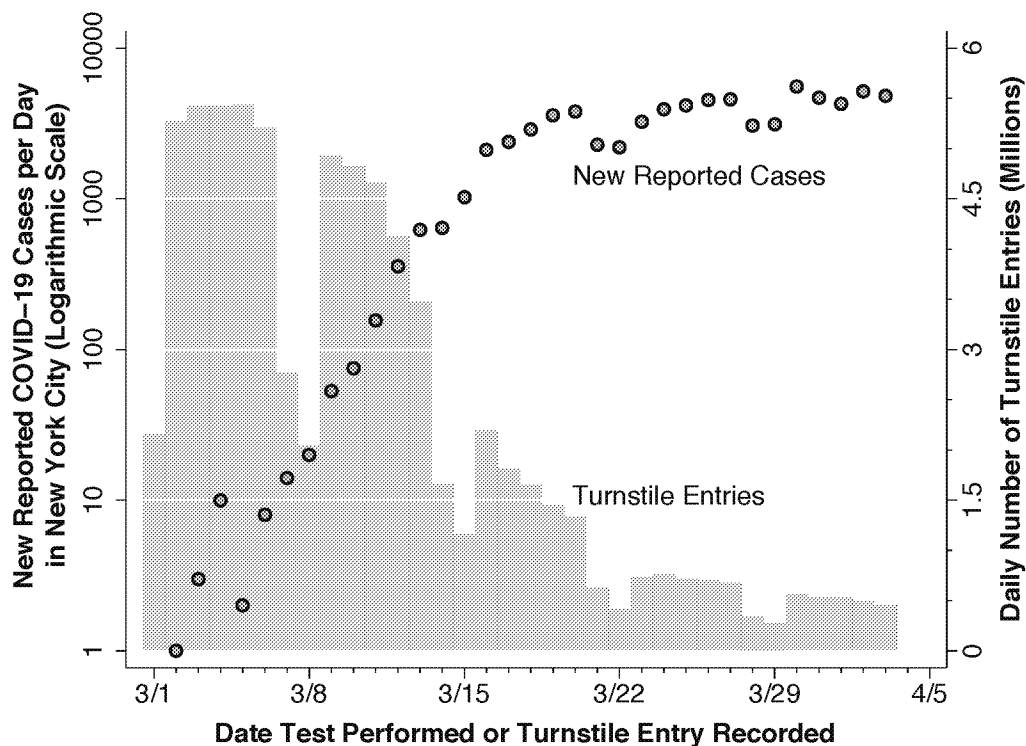


Figure 1. Numbers of Newly Diagnosed COVID-19 Cases (Pink Data Points, Left Axis) and Millions of Subway Turnstile Entries (Blue Bars, Right Axis), New York City, March 1–April 3, 2020.

No matter what convention is employed to mark off the calendar on the horizontal axis, the trend in the daily reported incidence of new COVID-19 cases tells the same story. There is a rapid upswing during the first half of the month, with a doubling time in Figure 1 of just 1.4

days, a rate of acceleration consistent with a basic reproductive number  $R_0$  exceeding than 3 (Harris 2020). This was followed by a marked slowing with a doubling time of 19 days. As we've earlier discussed, there are a number of valid reasons why the numbers of reported cases understate the total number of coronavirus infections. Still, when all of the indicators are viewed together, the conclusion that the epidemic curve in New York City has been flattening is inescapable (Harris 2020).

The second variable tracked in Figure 1 above represents the total numbers of entries every day into any of the approximately 4,600 turnstiles located throughout New York City's 496 subway stations. These counts are reported each week by the Metropolitan Transportation Authority (MTA) (Metropolitan Transportation Authority (MTA) 2020b, c, Whong 2020, Wellington 2020). This variable is represented as sky-colored vertical bars, measured in millions of entries tallied along the vertical axis on the right side of Figure 1. For this variable, the horizontal axis measures the dates on which riders passed through the system's turnstiles. While the MTA also reports turnstile exits, the data do not allow an analyst to link a particular rider's station of entry with that rider's station of exit.

Figure 1 shows only the volume of rides from March 1 onward. Still, the counts shown during the first full week of the month – from Sunday March 1 through Saturday March 6 – are quite typical of the pattern for prior weeks, peaking during mid-week at about 5.5 million rides per day and dropping during the weekends (Schneider 2020). During the second week of March, however, we begin to see a slight decline in subway usage, overall about 19 percent lower than the previous week. This decline in subway use accelerates markedly beginning on Monday March 16, the day that New York City Mayor de Blasio issued an order limiting gatherings and closing numerous places of congregation. By the third week overall, subway usage is down 68 percent from the first week in March, and by the fourth week, it's down 86 percent.

Simple comparison of the two trends in Figure 1 cannot by itself answer questions of causation. Still, the parallel between the continued high ridership on MTA subways and the rapid, exponential surge in infections during the first two weeks of March supports the hypothesis that the subways played a role. The subsequent plummeting of ridership appears likewise to parallel the flattening of the reported incidence curve. The steep fall in the heights of the blue bars undoubtedly represent the public's response to widespread publicity about the ferocity of the outbreak that had been gathering storm for two weeks. As economists say, the

precipitous drop in subway ridership was *endogenous*. Even so, the temporal pattern in Figure 1 is compatible with the conclusion that the subway system was the vehicle by which the public's response was translated into reduced transmission of the virus.

### Subway Ridership by Borough

Figure 2 focuses more sharply on the trends in subway turnstile entries, breaking down the trends by the borough in which the subway station of entry was located. We have included the Staten Island railway, which connects to Manhattan via the Staten Island Ferry. The vertical axis now measures turnstile entries as a percentage of the volume recorded on Monday, March 2, 2020. To better appreciate the proportional changes in ridership, the vertical axis is rendered on a logarithmic scale.

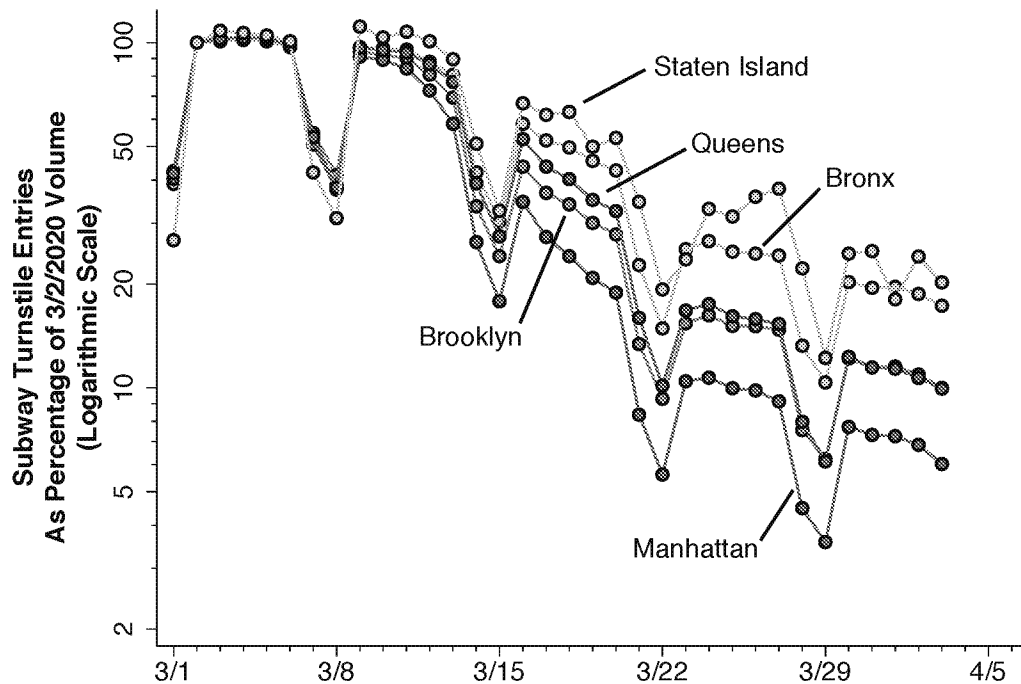
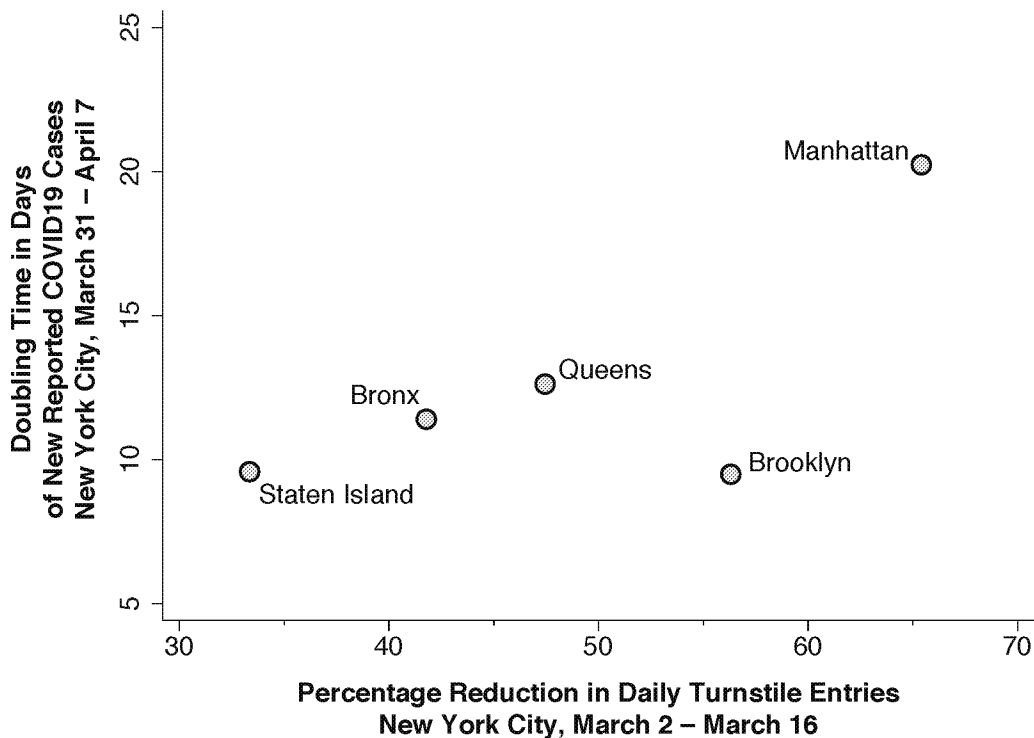


Figure 2. Daily Numbers of Turnstile Entries for the Five Boroughs of New York City, Computed on a Logarithmic Scale as a Percentage of Peak Ridership on March 2, 2020 (Corrected from original April 13 version).

During the first week of March, the ridership volumes in the five boroughs, calculated in percentage terms, are indistinguishable, except for a greater weekend dropdown in Staten Island. As the second calendar week comes to a close, we can begin to see a divergence among boroughs, which becomes increasingly prominent over time. By Monday March 23, Manhattan ridership has fallen to 10.5 percent of its March 2 volume, as shown by the purple data points,

and by Monday March 30, it's down to 7.8 percent of peak. By contrast, Bronx, represented by the sky-blue data points, was down to 25.2 percent of peak volume by Monday March 23 and 20.3 percent of peak by Monday March 30. Staten Island, represented by the mango data points, experienced an even smaller drop in volume.

For each of the five boroughs, Figure 3 compares the percentage decline in turnstile entries from March 2 through March 16, shown on the horizontal axis, against the estimated doubling times of new reported COVID-19 cases 15 days later during the week starting on March 31. The borough of Manhattan stands out from the other four. By March 16, Manhattan turnstile entries had fallen to 65 percent of their March 2 peak. About two weeks later, the trend in the number of new reported infections was virtually flat, with a doubling time of 20 days. From formulas developed in our earlier report (Harris 2020), it is likely that the reproductive number  $R$  in Manhattan as a whole is now less than 1. That is, the number of individuals coming down with a new coronavirus infection during any given day is outweighed by the number of previously infected individuals who lost their infectivity during that same day.






*Figure 3. Percentage Reduction in Daily Turnstile Entries from March 2 to March 16 Versus the Estimated Doubling Time of New Reported COVID-19 Cases During the Subsequent Week from March 31 to April 7. Five Boroughs, New York City.*

The finding that a 65-percent drop subway ridership is associated with a subsequent reversal of the COVID-19 epidemic in the borough of Manhattan hardly proves causation. It could be that the decline in ridership is no more than an indicator – what economists call a *proxy* – for other concurrent social distancing activities that ultimately contributed to the observed decline in reported infections. In any event, it would be inappropriate to draw firm conclusions from what would amount to a Manhattan-versus-the-rest study.

Still, the findings in Figure 3 help resolve a puzzle posed by more than a few observers. In one study, researchers hypothesized that the lower rate of coronavirus infection in Manhattan had something to do with income or social class (Wellington 2020). Manhattan was dominated by “stay-at-home professionals with more job security and benefits,” while the remaining boroughs were populated principally by “low-paid front-line workers” such as grocery clerks, delivery workers, transit workers, and cleaning and maintenance workers (Florida 2020). Figure 3 shows a clear temporal relation between the accelerated evacuation of the subways in Manhattan and the subsequent leveling off of the COVID19 incidence curve in that borough. Manhattanites could afford to stay off the subway, while many inhabitants of the other four boroughs could not.

### **Diversity of COVID-19 Incidence by New York City Zip Code**

Figures 4 and 5, respectively, map the cumulative numbers of COVID-19 cases per 10,000 population in New York City zip codes at two points in time: March 31 and April 8. In each map, we use the same fixed three-class color scheme to characterize the cumulative incidence. Light green  signifies a cumulative incidence rate less than 70 cases per 10,000. Medium green  signifies a rate of at least 70 but less than 100 cases per 10,000. Dark green  stands for a rate of at least 100 per 10,000, which is equivalent to saying that at least 1 percent of the population has been infected as of the specified date. These maps were modified from published maps depicting the numbers of positive tests, but not incidence (New York Department of Health and Mental Hygiene 2020). For an animated GIF, click [here](#).

Comparison of the two maps, depicting the evolution of the coronavirus epidemic over just 9 days, shows the initial seeding and subsequent spread from several distinct hotspots: Borough Park (11219) and Midwood (11230) in Brooklyn; Morris Park–Westchester Square (10461) in the Bronx; a swath of contiguous zip codes extending eastward from East Elmhurst (11370) in Queens; and a hotspot centered around Midtown West (10018) in Manhattan. By

April 8, the zip code with the highest cumulative incidence was East Elmhurst (11370) with 180 cases per 10,000 population.

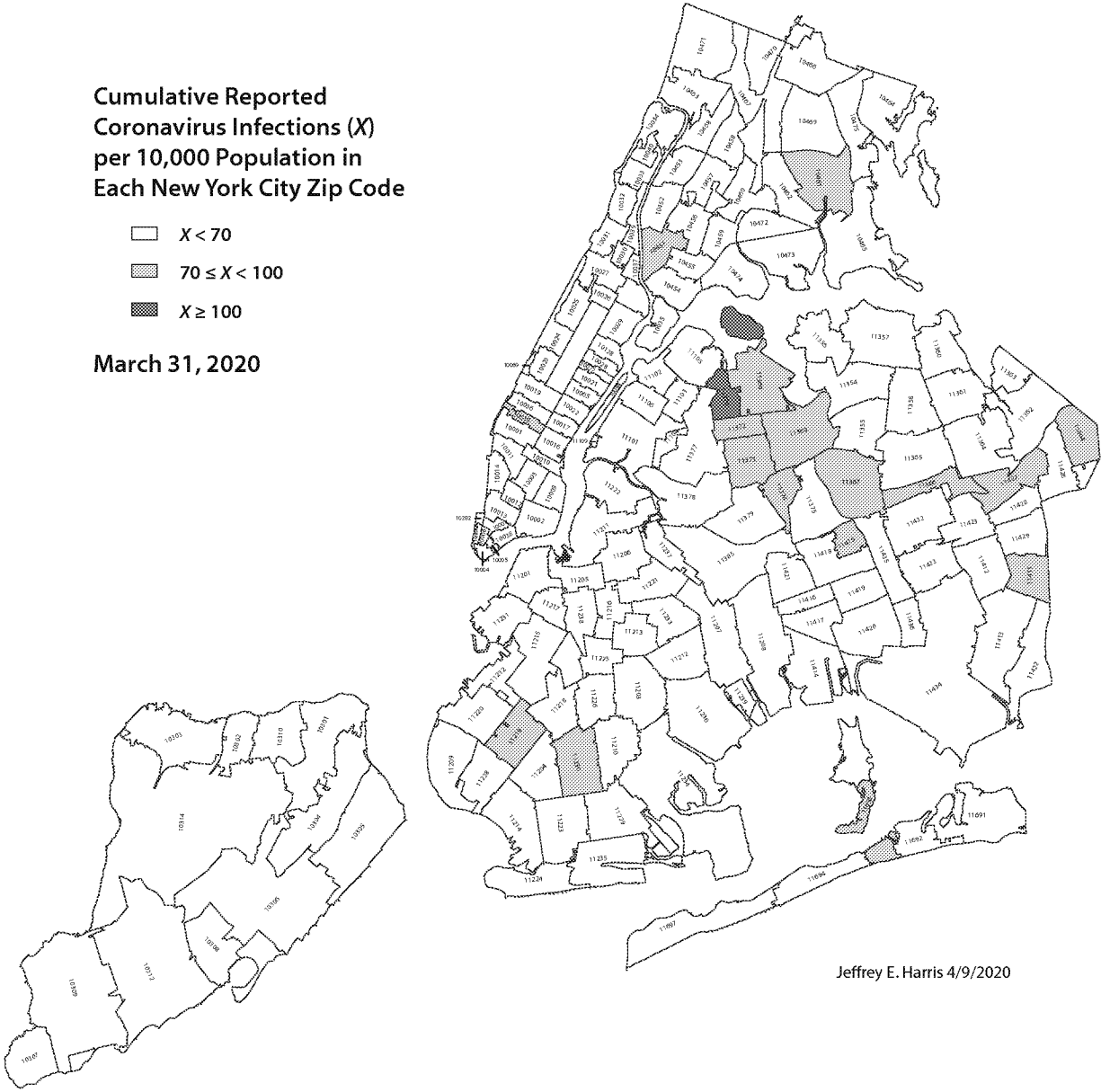


Figure 4. Map of Cumulative Numbers of Coronavirus Infections per 10,000 Population According to Zip Code of Residence, New York City, as of March 31, 2020.

Looking at the data on subway station-specific turnstile entries and zip code-specific infection rates, many economists may see the makings of a *difference-in-differences analysis*. For each station, the idea is first to compute the time trends in turnstile entries and coronavirus incidence, and then assesses whether there is a relation between the two trends across different



subway stations (Fredriksson and Oliviera 2019). Unfortunately, there is a serious problem with this extraordinarily popular method of doing policy analysis (Bertrand, Duflo, and Mullainathan 2004). In particular, there is likely to be significant *serial correlation* in the outcomes among adjacent subway stations situated along the same line.

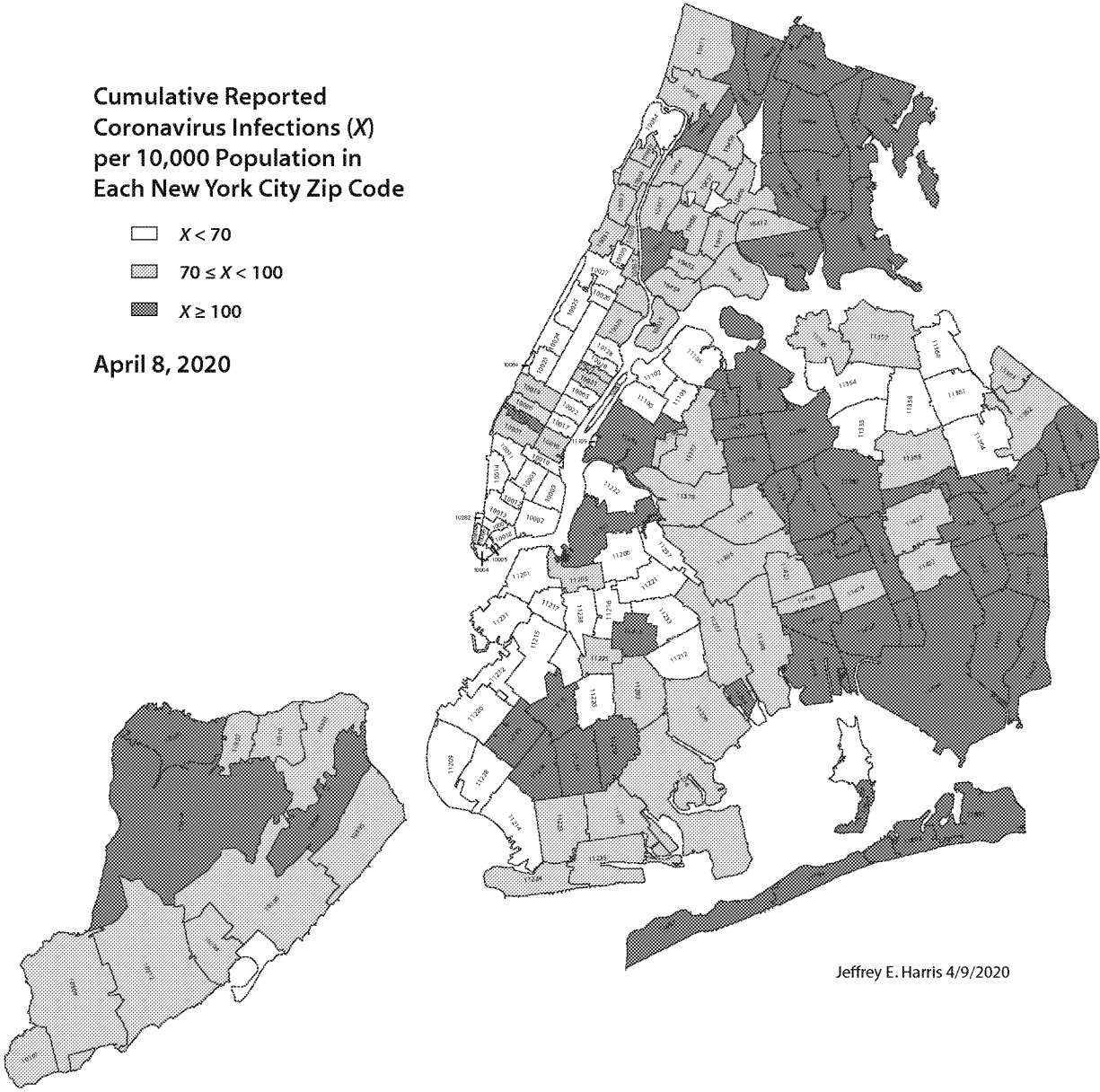


Figure 5. Map of Cumulative Numbers of Coronavirus Infections per 10,000 Population According to Zip Code of Residence, New York City, as of April 8, 2020.

The problem, put differently, is that the individual subway stations are not epidemiologically independent entities. Consider a service worker using public transportation in


New York City, who typically takes more than a half-hour to commute to work (Choi, Velasquez, and Welch 2020). Specifically, she takes the Flushing Local line, entering the turnstile at the Junction Boulevard stop, located within the Corona zip code (11368) in Queens, getting off at the 34<sup>th</sup> Street–11<sup>th</sup> Avenue stop at the end of the line, from which she walks to her work in the Midtown West zip code (10018).

We'll call our commuter Milagros, a name honoring Nuestra Señora de Los Milagros, inasmuch as zip code 11368 is 74% Hispanic-Latino (USZip 2020b). Once Milagros boards the train, the next two stops are 90<sup>th</sup> Street–Elmhurst Avenue and 82<sup>nd</sup> Street–Jackson Heights, smack-dab between zip codes 11372 (Jackson Heights) and Elmhurst (11373), which were already emerging hot spots of infection by March 31. From 82<sup>nd</sup> St.–Jackson Heights, it would take Milagros just five minutes to walk to the Elmhurst Hospital Emergency Department.

Milagros's exposure to coronavirus is not accurately gauged by the number of commuters who passed through the turnstile at her entry point at Junction Boulevard. That's because she'll come into contact with potentially infectious passengers at each of the remaining 17 stops until she gets off at 34<sup>th</sup> Street–11<sup>th</sup> Avenue, which happens to be located in another coronavirus hotspot. On the way back home, she will also be exposed to those passengers staying on the Flushing Local and disembarking after Milagros does – at the 103<sup>rd</sup> St–Corona Plaza, 111<sup>th</sup> Street, and Mets–Willets Point stations likewise located in hotspot zip codes. In view of these independencies between units of observation, the classic technique of difference-in-differences routinely employed in policy evaluation is, as Milagros would put it, *arrojado por la ventana*.

### **Subway Lines Are the Correct Units of Analysis.**

Figure 6 superimposes the stops along the 7 Local Line (historically, the Flushing Local Line) that tens of thousands of passengers like Milagros took every day back and forth between a station at the eastern end of the line in Queens and a station at the western end in downtown Manhattan.

The outer area of each circle  corresponds to the volume of turnstile entries at that station during the first week in March, while the inner area corresponds to the volume during the third week in March. As we would anticipate from the data in Figures 3 and 4, the volume of turnstile entries declined to some extent at all of the station stops along the Flushing Local line. While the percentage decline was considerably greater at the Manhattan stops, the absolute

numbers of entries at Grand Central–42<sup>nd</sup> Street and Times Square–42<sup>nd</sup> Street turnstiles during the third week in March were still comparable to those at the other end of the line.



Figure 6. Stops Along the Flushing Local Line in the New York City Subway System Superimposed on a Section of the Zip Code Map in Figure 5. The outer area of each point corresponds to the volume of turnstile entries during the first week in March 2020, while the inner area corresponds to the volume during the third week of that month.

The data in Figure 6 are compatible with continued but reduced propagation of coronavirus infection along the Flushing Local line during the third week of March. The stations run through the hot spots in the Elmhurst area and terminate at the hotspot zip code in West Midtown Manhattan. The line also runs through Long Island City zip code 11101, another hotspot with a 34.5% Hispanic-Latino, 18.5% African-American and 15.9% Asian demographic profile, where 71.6 percent of workers take public transportation (USZip 2020a).

The data in Figure 6 are further compatible with reciprocal or reverse seeding of the hotspots in Midtown West from the hotspots along the periphery of the Flushing Local Line. While the volume of turnstile entries at the Midtown West stations (especially Times Square and Grand Central) were substantially reduced by the third week in March, the absolute volume still remained elevated as daily workers like Milagros reentered these stations from the periphery.

In the classic, static model of epidemic propagation (Harris 2020, Kermack and McKendrick 1991), susceptible individuals (the  $S$ 's) make contact with infective individuals (the

*I*'s). The incidence of new infections depends on two factors: the frequency of contact between an *S* and an *I*, and the probability that each contact results in transmission of the infection. The model was borrowed from the basic law of mass action in chemistry, where *S* and *I* molecules bombard against each other, bounding around in a gas or a liquid. In an innovative series of papers, Goscé and colleagues generalized this model to consider contagion when the *S*'s and *I*'s move along a corridor (Gosce, Barton, and Johansson 2014, Gosce and Johansson 2018). They applied their framework to the study of the spread of influenza-like illness in the London Underground, a vast network opened just nine years after Dr. John Snow got public officials to disable a pump at Broad (now Broadwick) and Lexington Streets, now about a five-minute walk from the Oxford Circus station.

The Goscé model offers a number of insights that are immediately applicable to the data from the New York City Flushing subway line. The first is that the rate of disease transmission is related to the number of trips and average number of stations per trip along the entire subway line, and not just to the number of entries at any one subway station. Second, passengers entering the subway line even at a remote, less populous station are slowing down the system, thus increasing the transit time that the *S*'s stay in contact with the *I*'s. Third, those uninfected *S*-passengers who cram shoulder-to-shoulder into a particular subway are increasing train-car density and thus raising the average number of other *S*-passengers infected by an *I*-passenger who happens to be standing in the middle of the train. Fourth, *local* trains – like the Flushing local – are more likely to seed epidemic infections than express lines. Finally, an entire subway line, rather than the individual stations or subway cars, is the appropriate unit of analysis.

For 32 subway lines in the MTA's database (Metropolitan Transportation Authority (MTA) 2020b), Figure 7 plots the cumulative per capita incidence of coronavirus infection as of April 3, 2020 against the percentage reduction in turnstile entries between the first and third weeks of March. To compute cumulative per capita incidence, we linked each subway station along each line to its nearest zip code, based on the geocodes of the stations (Metropolitan Transportation Authority (MTA) 2020b) and the centroids of each zip code (Open Data Soft 2020). For each subway line, we then calculated the total number of reported coronavirus cases in all linked zip codes combined and divided that number by the total population of all linked zip codes combined. Thus, each subway line's cumulative incidence was the population-weighted average of cumulative incidence rates among each of its station-linked zip codes.

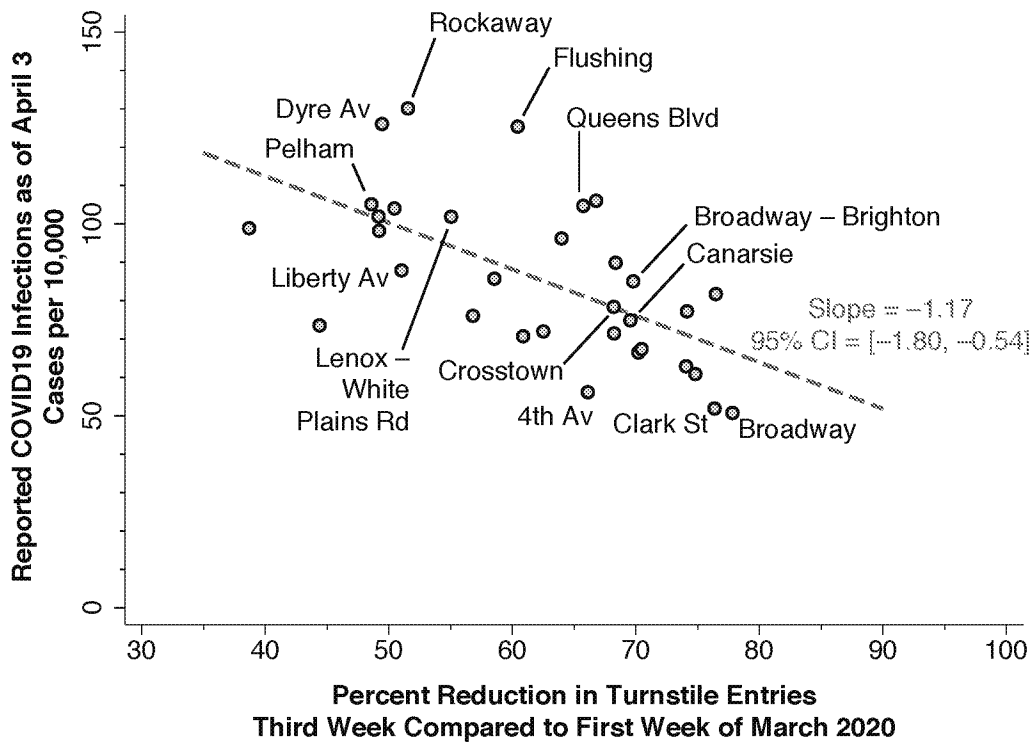


Figure 7. Reported COVID-19 Infections per 10,000 Residents in Each of 32 Subway Lines Relation to the Percentage Reduction in Turnstile Entries from the First to the Third Weeks of March 2020.,

Comparing entire subway lines, Figure 7 thus relates the change in ridership of each line with the overall rate of coronavirus infection in the zip codes traversed by that line. Those lines showing the largest decline in ridership from the first to the third week of March had significantly lower rates of coronavirus infection by the beginning of April. A least squares regression line gives an estimated slope of  $-1.17$  ( $p = 0.001$ ). That is, for every 10-percentage point reduction in subway ridership during the first three weeks of March, the cumulative incidence of infection declined by an estimated 11.7 cases per 10,000. While the Flushing line shows one of the three highest infection rates, the 66-percent decline appears to make it an outlier in the plot. That's because the estimated decline includes the marked reductions in ridership in the two major stops in Midtown Manhattan (Figure 6).

### A Bunch of Garbage

While we've got a few more maps up our sleeve, we're already at a juncture where some readers may react with extreme skepticism. We don't have a cleanly designed *natural experiment* comparable to the removal of the handle on the Broad Street pump in St. James's parish,

advocated by Dr. John Snow, which dramatically shut down a cholera outbreak in mid-nineteenth century London (Snow 1855). Without such evidence, the naysayers will assert that any diffuse, multitentacled network that traverses most of the city could be correlated spatially with the spread of coronavirus infection documented above. To be sure, serious critics won't point to the electromagnetic signals from power lines, but they could argue that the path traced in Figure 6 could just as well represent the stops of sanitation trucks. Put bluntly, the critique goes, the evidence presented thus far would be consistent with contaminated garbage as the vehicle for the massive spread of deadly COVID-19.

Except for one thing – namely, we know that the garbage hypothesis is entirely implausible, while the subway hypothesis is entirely plausible.

We know that coronavirus is transmitted from one person to another by two principal means. First, an infected person exhales moist air containing very small droplets loaded with the virus. A passenger standing two feet away from an infected rider for just 15 minutes would almost certainly inhale virus particles, even if the infected rider never coughed or sneezed (New York City Rapid Transit 1988, Santarpia, Rivera, and Herrera 2020). Second, an infected person constantly sheds virus particles on almost every surface he touches, such as glasses, keys and phones. That would include the vertical metallic poles shared by standing passengers. A crowded subway train is thus an ideal incubator for coronavirus transmission (Qian et al. 2020).


Other places where people congregate might be fairly dense at peak hours, just as restaurants, gyms, retail stores and some workplaces. But the subway system is much more efficient at propagating infection from Midtown to the periphery and back many times in a day.

We know that the flattening of the epidemic curve in Manhattan two weeks after that borough had cut its subway ridership by 65 percent adds tellingly to the circumstantial evidence. So does the finding that those lines with the largest decline in ridership from the first to the third week of March had significantly lower rates of coronavirus infection by the beginning of April.

We know that we can't dismiss out of hand our finding of reciprocal seeding from the periphery of the Flushing local line to Manhattan's only hotspot in Midtown West, and from that central hub back to the periphery. We know that many workers – especially non-White workers – have been trapped by economic necessity into continuing to expose themselves to the bad stuff millions of times daily (Goldbaum and Cook 2020). We know that it would be inappropriate to require the subway hypothesis to explain every aspect of the diffusion of coronavirus, if only

because we have buses and schools, too, if only because Milagros, once she got sick, didn't have her own bedroom and bathroom to isolate herself.

### Overlaying the Other Subway Lines on the Epidemic Map

Figure 8 superimposes comparable data from the 6<sup>th</sup> Avenue Local line (also called the Queens Blvd Local line) to the epidemic map of Figure 6. As in the previous figure, the subway stops of 6<sup>th</sup> Avenue Local run right through the hotspot zip codes. What's more, the inner circles, colored dark blue , show a significantly greater decrease in volume in the Manhattan stops by the third week in March. These additional data in Figure 8 are further compatible with the conclusion that propagation of coronavirus, while reduced in comparison to the first week of March, was continuing to spread along subway lines through at least the third week of March.



*Figure 8. Stops Along the Flushing Local Line and 6<sup>th</sup> Avenue Local Line in the New York City Subway System Superimposed on a Section of the Zip Code Map in Figure 5. The outer area of each point corresponds to the volume of turnstile entries during the first week in March 2020, while the inner area corresponds to the volume during the third week of that month.*

The last station on the 6<sup>th</sup> Avenue Local line is Jamaica – 179<sup>th</sup> Street, a major hub for local bus routes in Queens (Metropolitan Transportation Authority (MTA) 2018). From there, one can take the 43 bus along Hillside Avenue to reach Bellerose Manor (zip code 11426), at the

eastern end of the conglomeration of zip code hotspots within the borough shown in Figure 5. Alternatively, one can take the 111 bus down to Rosedale (zip code 11422) in the southeast corner, where 81 percent of residents are African-American (USZip 2020c).

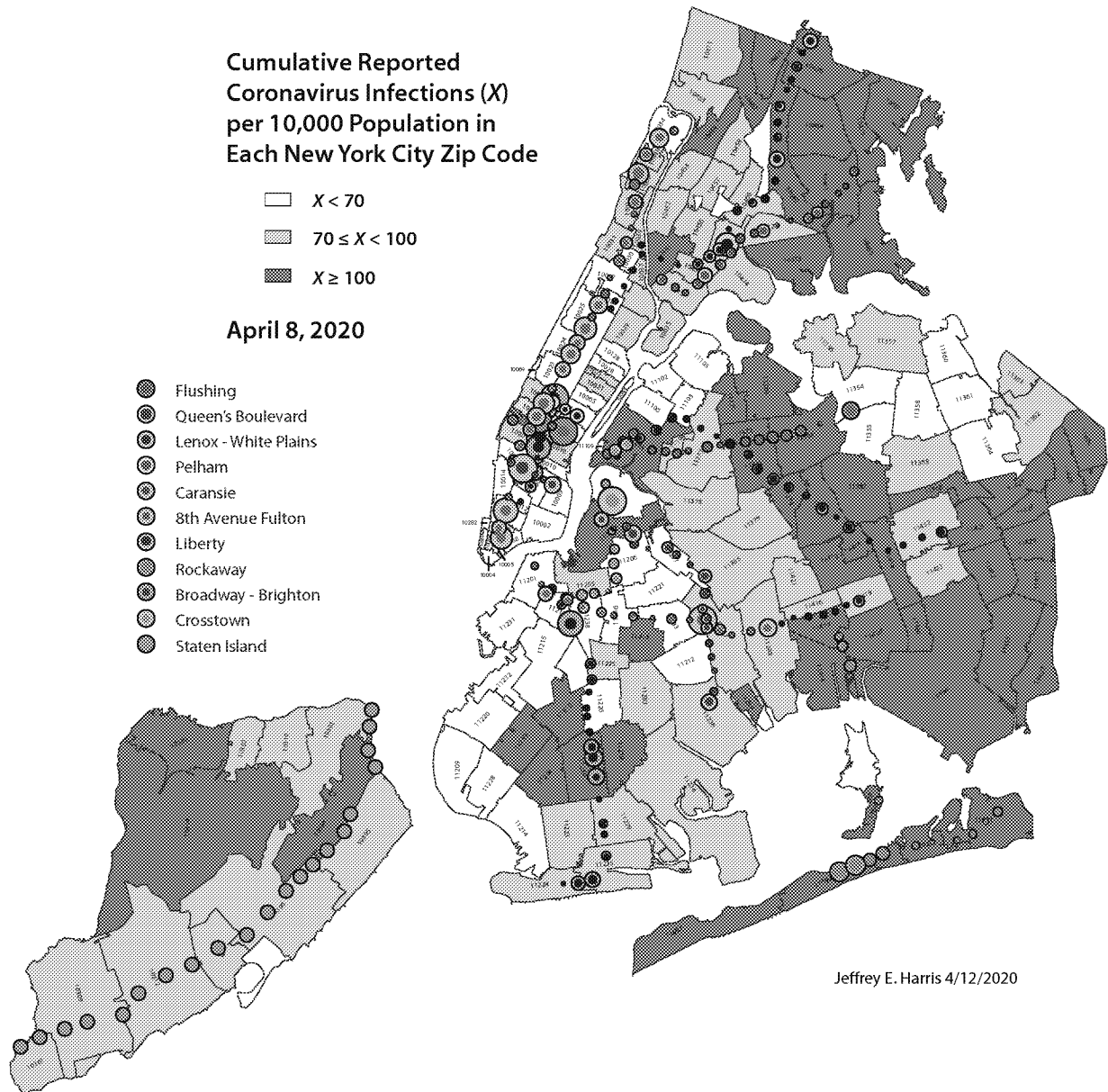


Figure 9. Subway Stops Along Multiple Routes in the Four Principal Boroughs of New York City, Superimposed Upon the Zip Code Map of Figure 5. See text for details.



Following the same conventions as Figure 6 and 8, Figure 9 (displayed above) overlays multiple subway lines on the zip code map of Figure 5. The key shows the historical names of the lines, as reflected in the MTA's geocode database (Metropolitan Transportation Authority (MTA) 2020b). The individual stops for the Staten Island line are included, although the MTA database does not provide sufficient data to show the changes over time within each station. While Figure 9 does not show every subway line in the city, it is intended here to illustrate the breadth and reach of the subway system.

### **Irony Along Eighth Avenue**

The Metropolitan Transit Authority's decision to cut back its train service to accommodate the reduced demand may have indeed helped to shore up the agency's financial position, but it most likely accelerated the spread of coronavirus throughout the city. That's because the resulting reduction in train service tended to maintain passenger density, the key factor driving viral propagation (Goldbaum and Cook 2020). How ironic it is that, from the public health perspective, the optimal policy would have been to double – maybe even triple – the frequency of train service. The agency's decision to convert multiple express lines into local service only enhanced the risk of contagion (Goldbaum 2020). How ironic it is that the preferred policy would have been to run even more express lines. We have not seen any public data on the incremental cost of the agency's decision to begin to disinfect subway cars twice daily. Still, it is natural to inquire why the cars weren't disinfected every time they emptied out of passengers at both ends of the line.

The press has recently reported a significant number of coronavirus infections and deaths among front-line MTA workers. As of April 10, 2020, there were 50 deaths among 1,900 workers who had tested positive (Guse and Rayman 2020). Tragically, the counts of infected and fallen workers have continued to grow. By April 16, the MTA had reported 68 deaths among more than 2,400 subway and bus employees who had tested positive. "Another 4,400 are on home quarantine and thousands more are calling out sick." (Metropolitan Transportation Authority (MTA) 2020a)

Data from TWU Local 100 indicate that the agency has 40,000 front-line transit workers (TWU Local 100 2019). That would imply a cumulative incidence of infection equal to 600 per 10,000, more than three times the rate of 180 per 10,000 reported in East Elmhurst (zip code 11370), the most affected hotspot in Figures 4 and 5 above. While the MTA announced on April

15 that it would begin its own testing of symptomatic employees, the agency's workers had previously been directed to find tests on their own accord. "Nor has [the agency] offered any theories as to why the transit division's workforce is suffering such losses." (Rubinstein 2020)

To be sure, not all MTA workers had direct contact with passengers or subway cars, but once those with direct contact got sick, they gave their infections to their coworkers. What we're seeing now is the second wave of infections among MTA workers, having failed to detect the first wave. It is hard to imagine any plausible explanation for these workers' losses except that their place of work was the principal source of their coronavirus infections. How ironic it is that unfathomable tragedy of these frontline workers turns out to be the clincher that transports us from correlation to causation.

With the incidence of new infections and COVID-19 hospitalizations leveling off (Harris 2020), there will be increasing interest in relaxing social distancing measures. During these renormalization times, the public transportation system will surely require enhanced scrutiny. That means even more attention to staggered work hours, limits on the numbers of passengers per transport unit, refurbished vehicles with enhanced ventilation, subsidies for drivers to transport workers in SUVs, vans and minibuses, new technologies to determine which stations an infected person entered and exited, and redirection of passenger traffic to less dense lines.

This study has touched upon the differential impact of the COVID-19 pandemic on those with the fewest resources. As we put this working paper to press, there have been mounting calls for more data on racial and ethnic minorities. How ironic it is that this point was well aired more than two decades ago (Farmer 1996).

Quite apart from the present study and the above-cited work by Goscé and colleagues (Gosce, Barton, and Johansson 2014, Gosce and Johansson 2018), a few other researchers have attempted to test whether public transport has served as a critical vehicle for the propagation of contagious respiratory diseases (Sun et al. 2013, Troko et al. 2011, Cooley et al. 2011). One distinguishing factor between the present study and prior work is that seasonal influenza has generally had a reproductive number  $R$  in the range of 1.2–1.4, while pandemic influenza has had an  $R$  in the range of 1.4–1.8, with the high end representing the 1918 pandemic (Biggerstaff et al. 2014). By contrast, we have estimated the  $R$  in New York City during the initial surge of infections in early March to be on the order of 3.4 (Harris 2020).

Studies of the role of subways – and public transit generally – in the recent propagation of coronavirus in other major urban centers warrant attention. Urban transport systems are highly heterogeneous with differing design and age. Some systems have many above-ground stations, while others, like New York City, are predominantly below-ground. More modernized signal systems allow higher train frequency and less crowding. Some systems focus on local service, while others, like New York City, serve as effective mixers of traffic, running from the edge to the center, then back out to another part of the periphery. Of particular interest will be forthcoming evaluations of the timing of the closure of the subways on the subsequent path of the COVID-19 outbreak in Wuhan, China (Xu 2020)

In sum, several lines of evidence point to the subway system as a major disseminator – if not the principal transmission vehicle – of coronavirus infection during the initial exponential takeoff of the coronavirus epidemic in New York City during the first two weeks of March 2020. The evidence further supports the conclusion that the ensuing marked decline in subway use was the main vehicle by which the public's growing perception of risk was translated into reduced transmission of the virus. Since the evidence is observational, we can imagine that some scientific reviewers will nonetheless conclude that cause-and-effect remains difficult to prove. Still, we doubt whether any public health practitioner would be reluctant to take action on the basis of the facts we now know.

## References

- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics* 119 (1):249-275.
- Biggerstaff, M., S. Cauchemez, C. Reed, M. Gambhir, and L. Finelli. 2014. "Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature." *BMC Infect Dis* 14:480. doi: 10.1186/1471-2334-14-480.
- Choi, A., J. Velasquez, and W. Welch. 2020. "Queens Neighborhoods Hardest Hit by Virus Home to Many Service Workers." *The City* (April 2, 2020).
- Cooley, P., S. Brown, J. Cajka, B. Chasteen, L. Ganapathi, J. Grefenstette, C. R. Hollingsworth, B. Y. Lee, B. Levine, W. D. Wheaton, and D. K. Wagener. 2011. "The role of subway travel in an influenza epidemic: a New York City simulation." *J Urban Health* 88 (5):982-95. doi: 10.1007/s11524-011-9603-4.

- Farmer, P. 1996. "Social inequalities and emerging infectious diseases." *Emerg Infect Dis* 2 (4):259-69. doi: 10.3201/eid0204.960402.
- Florida, Richard. 2020. *The coronavirus class divide by cities*. City Lab, April 7, 2020: <https://www.citylab.com/equity/2020/04/coronavirus-risk-jobs-essential-workers-data-class-divide/609529/>.
- Fredriksson, A., and G. Oliviera. 2019. "Impact evaluation using Difference-in-Differences." *RAUSP Management Journal* 54 (4):519-532. doi: 10.1108/RAUSP-05-2019-0112.
- Goldbaum, Christina. 2020. "Subway Service Is Cut by a Quarter Because of Coronavirus." *New York Times* (March 24, 2020).
- Goldbaum, Christina, and Lindsay Rogers Cook. 2020. "They Can't Afford to Quarantine. So They Brave the Subway." *New York Times* (March 30, 2020).
- Gosce, L., D. A. Barton, and A. Johansson. 2014. "Analytical modelling of the spread of disease in confined and crowded spaces." *Sci Rep* 4:4856. doi: 10.1038/srep04856.
- Gosce, L., and A. Johansson. 2018. "Analysing the link between public transport use and airborne transmission: mobility and contagion in the London underground." *Environ Health* 17 (1):84. doi: 10.1186/s12940-018-0427-5.
- Guse, Clayton, and Graham Rayman. 2020. *50 MTA workers now dead from coronavirus: officials*. <https://www.nydailynews.com/coronavirus/ny-coronavirus-mta-workers-dead-pat-foye-20200410-f46sug5gf5huhmdii4wxyuebqa-story.html>: New York Daily News, April 10, 2020.
- Harris, J. E. 2020. *The Coronavirus Epidemic Curve Is Already Flattening in New York City*. <https://www.nber.org/papers/w26917>, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3563985](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3563985): National Bureau of Economic Research Working Paper No. 26917, April 3, 2020.
- Kermack, W.O., and A.G. McKendrick. 1991. "A contribution to the mathematical theory of epidemics - I." *Bulletin of Mathematical Biology (reprinted)* 53 (1-2):33-55.
- Metropolitan Transportation Authority (MTA). 2018. *Queens Bus Map*. <http://web.mta.info/nyct/maps/busqns.pdf>: Last Visited April 12, 2020.
- Metropolitan Transportation Authority (MTA). 2020a. *Letter from Chairman Patrick J. Foye to Senator Charles E. Schumer and Congresswoman Nita Lowey*. April 16, 2020.

- Metropolitan Transportation Authority (MTA). 2020b. *Stations.csv (text file, comma-separated)*.  
<http://web.mta.info/developers/data/nyct/subway/Stations.csv>: Last accessed April 4, 2020.
- Metropolitan Transportation Authority (MTA). 2020c. *Turnstile Data*.  
<http://web.mta.info/developers/turnstile.html>: Last accessed April 4, 2020.
- New York City Rapid Transit. 1988. *Rapid Transit Loading Guidelines*.  
NYCTA\_RAPID\_TRANSIT\_LOADING\_GUIDELINES\_FEBRUARY\_1988.pdf.
- New York Department of Health and Mental Hygiene. 2020. *COVID-19: Data*.  
<https://www1.nyc.gov/site/doh/covid/covid-19-data.page>: March 31, 2020.
- Open Data Soft. 2020. *US Zip Code Latitude and Longitude*  
<https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/>: Last accessed April 7, 2020.
- Qian, H., T. Miao, L. Liu, and et al. 2020. *Indoor transmission of SARS-CoV-2*.  
<https://www.medrxiv.org/content/10.1101/2020.04.04.20053058v1.full.pdf>: Last accessed April 23, 2020.
- Reichert, T. A., N. Sugaya, D. S. Fedson, W. P. Glezen, L. Simonsen, and M. Tashiro. 2001.  
"The Japanese experience with vaccinating schoolchildren against influenza." *N Engl J Med* 344 (12):889-96.
- Rubinstein, Dana. 2020. *Subway and bus workers are bearing a disproportionate coronavirus death toll*. <https://www.politico.com/states/new-york/albany/story/2020/04/07/subway-and-bus-workers-are-bearing-a-disproportionate-coronavirus-death-toll-1273457>: Politico, April 7, 2020.
- Santarpia, J.L., D.N. Rivera, and V. Herrera. 2020. *Transmission Potential of SARS-CoV-2 in Viral Shedding Observed at the University of Nebraska Medical Center*.  
<https://www.medrxiv.org/content/10.1101/2020.03.23.20039446v2.full.pdf>: Last accessed April 17, 2020.
- Schneider, Todd W. 2020. "New York City Subway Usage." In.  
<https://toddwtschneider.com/dashboards/nyc-subway-turnstile/>: Last visited April 4, 2020.
- Snow, John. 1855. *On the Mode of Communication of Cholera (Second Edition, Much Enlarged)*. London: John Churchill, New Burlington Street.

- Sun, L., K. W. Axhausen, D. H. Lee, and X. Huang. 2013. "Understanding metropolitan patterns of daily encounters." *Proc Natl Acad Sci U S A* 110 (34):13774-9. doi: 10.1073/pnas.1306440110.
- Troko, J., P. Myles, J. Gibson, A. Hashim, J. Enstone, S. Kingdon, C. Packham, S. Amin, A. Hayward, and J. Nguyen Van-Tam. 2011. "Is public transport a risk factor for acute respiratory infection?" *BMC Infect Dis* 11:16. doi: 10.1186/1471-2334-11-16.
- TWU Local 100. 2019. *Transit Workers Age Demographics*. September 20, 2019: Transit Workers Age Demographics (Demographics09202019.pdf).
- USZip. 2020a. *Zip Code 11101*. <https://www.uszip.com/zip/11101>: Last Accessed April 11, 2020.
- USZip. 2020b. *Zip Code 11368*. <https://www.uszip.com/zip/11368>: Last Accessed April 11, 2020.
- USZip. 2020c. *Zip Code 11422*. <https://www.uszip.com/zip/11101>: Last Accessed April 12, 2020.
- Wellington, Ben. 2020. *I Quant NY: Mapping Friday's 30% Drop in NYC Subway Ridership*. <https://iquantny.tumblr.com/post/612712380924903424/mapping-fridays-30-drop-in-nyc-subway-ridership>: March 15, 2020.
- Whong, Chris. 2020. *Taming the MTA's Unruly Turnstile Data*. <https://medium.com/qri-io/taming-the-mtas-unruly-turnstile-data-c945f5f96ba0>: March 31, 2020.
- Xu, Yuhan. 2020. *Public Transport In Wuhan Suspended Due To Coronavirus Concerns*. <https://www.npr.org/sections/goatsandsoda/2020/01/22/798602296/public-transport-in-wuhan-suspended-due-to-coronavirus-concerns>: NPR, January 22, 2020.