NBER WORKING PAPER SERIES

COVID-19: TESTING INEQUALITY IN NEW YORK CITY

Stephanie Schmitt-Grohé
Ken Teoh
Martín Uribe

Working Paper 27019
http://www.nber.org/papers/w27019

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2020

We thank for comments Dragan Filipovich, Antti Ilmanen, Gabriel Picone, Shehzad Ali, and seminar participants at Columbia University. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Covid-19: Testing Inequality in New York City
Stephanie Schmitt-Grohé, Ken Teoh, and Martín Uribe
NBER Working Paper No. 27019
April 2020
JEL No. I14,R1

ABSTRACT

Motivated by reports in the media suggesting unequal access to Covid-19 testing across incomes, we analyze zip-code level data on the number of Covid-19 tests, test results, and income per capita in New York City. We find that the number of tests administered is evenly distributed across income levels. In particular, the test distribution across income levels is significantly more egalitarian than the distribution of income itself: The ten percent of the city's population living in the richest zip codes received 11 percent of the Covid-19 tests and 29 percent of the city's income. The ten percent of the city's population living in the poorest zip codes received 10 percent of the tests but only 4 percent of the city's income. At the same time, we find significant disparity in the fraction of tests that come back negative for the Covid-19 disease across income levels: moving from the poorest zip codes to the richest zip codes is associated with an increase in the fraction of negative Covid-19 test results from 38 to 65 percent.

Stephanie Schmitt-Grohé
Department of Economics
Columbia University
420 West 118th Street, MC 3308
New York, NY 10027
and NBER
stephanie.schmittgrohe@columbia.edu

Martín Uribe
Department of Economics
Columbia University
International Affairs Building
New York, NY 10027
and NBER
martin.uribe@columbia.edu

Ken Teoh
420 West 118th Street
Suite 1109
Department of Economics
Columbia University
New York, NY 10027
USA
ht2467@columbia.edu

A data appendix is available at http://www.nber.org/data-appendix/w27019

1 Introduction

The United States is the epicenter of the 2020 coronavirus outbreak. An important aspect of this health crisis is how it impacts different income groups. In particular, a much debated issue is how access to health care is linked to peoples' relative position in the income ladder. This paper addresses a specific aspect of this debate, namely, whether higher income gave privileged access to tests for the Covid-19 disease. Several media outlets and observers have suggested that the distribution of tests inherited the well-known inequalities in the distribution of income. For example, on March 18, 2020, the New York Times ran an article with the headline, "Need a Coronavirus Test? Being Rich and Famous May Help."

To analyze the question of inequality in testing for Covid-19 across income groups, we use zip-code-level data from New York City on the number of Covid-19 tests and test outcomes as of April 2 and April 13, 2020. We combine this information with data on income per capita and population by zip code. New York City is a relevant laboratory for evaluating this question for the following reasons: First, it is the city hardest hit by the coronavirus. Second, the residents of all New York city zip codes are subject to the same health policies and regulations at the local, state, and federal levels. Third, income per capita displays significant variation across zip codes. And fourth, data across all zip codes is produced by the same statistical agencies, guaranteeing cross-sectional comparability. Data on Covid-19 tests and test results for the 177 zip codes in New York come from the New York City Department of Health and Mental Hygiene and data on zip-code level income per capita and population is from the 2014-2018 American Community Survey (ACS).

We find that the number of tests administered is evenly distributed across income levels. In particular, the test distribution is significantly more egalitarian than the distribution of income: The 10 percent of the population living in the richest zip codes received 11 percent of the Covid-19 tests and 29 percent of the city's income. The 10 percent of the population living in the poorest zip codes received 9 percent of the tests and only 4 percent of the city's income.

If disease were equally distributed across income levels, then the reported even distribution of tests could be consistent with equal access to testing. However, if the Covid-19 disease were more prevalent at lower incomes, then this finding could reflect unequal access, for in this case, the poorest income groups should account for a disproportionally larger share of the administered tests. Establishing the spread of the coronavirus would require random testing, which is not available. The available data on test results is influenced by patients' self-selection into testing and health-care providers' selection criteria on whom to test. It is nonetheless of interest to investigate whether the combination of incidence, self-selection, and rationing by the health department jointly have an impact on the relation between Covid-19 test results and income. To shed some light on this issue, we analyze the distribution of negative test results across income groups. A negative test result indicates that the patient does not have Covid-19. We find significant disparities across income levels in the fraction of tests for Covid-19 that come back negative: moving from the poorest zip codes to the richest zip codes is associated with an increase in the fraction of negative Covid-19 test results of 27 percentage points, from 38 to 65 percent. Furthermore, controlling for income, the fraction of negative tests is lower in zip codes with larger shares of non-white residents, although the effect is quantitatively small.

Taken together the results on tests and test-outcome distributions by income suggest that an egalitarian distribution of tests may not be tantamount to equal access. The findings are consistent with the non-mutually exclusive hypotheses that Covid-19 is more widespread in poorer zip codes and that the severity of symptoms that triggers a Covid-19 test to be administered is higher for poor New York City residents. The latter hypothesis, in turn, could be driven by both supply and demand factors. For example, a supply side explanation could be that residents of poorer zip codes get a disproportionate share of positive test results because health care providers apply a more stringent symptom threshold to them than to residents of richer zip codes. A demand side explanation could be that poorer residents tend to wait longer before they seek medical assistance.

This paper is related to contemporaneous and independent work by Borjas (2020), who used a similar dataset on Covid-19 tests, test results, population, and income in New York City at the zip code level.[1] The main difference in the empirical strategy is that the present study estimates Lorenz curves and Gini coefficients for testing and negative test results, whereas Borjas focuses on regression analysis. Similar to the present study, Borjas finds that persons who reside in richer neighborhoods were more likely to test negative for Covid-19. However, the methodological differences lead to seemingly different conclusions regarding equity in testing. While the present study finds that the distribution of Covid-19 tests was egalitarian across income, Borjas finds that people residing in poor neighborhoods were less likely to be tested than people residing in rich neighborhoods. These two conclusions are not necessarily inconsistent with each other, for regressions of the odds of Covid-19 tests on household income do not necessarily reflect inequity in the distribution of tests across income levels. The reason is that such regressions do not take into account the weight of the different zip codes in the income distribution, a feature that, by construction, is factored in

[1] There are some differences in the datasets: (1) this paper uses test and test result data for April 2 and 13, whereas Borjas uses data for April 5; (2) this paper uses income and population data from the 2014-2018 ACS, whereas Borjas uses the 2010-2014 ACS; and (3) This paper uses mean per capita income at the zip code level, whereas Borjas uses mean household income at the zip code level.

by Lorenz curves.

The remainder of the paper is in five sections. Section 2 describes the data and their sources. Section 3 documents the distribution of tests for Covid-19 across income levels, and section 4 the distribution of negative test outcomes. Section 5 shows that the findings are robust to extending the sample from April 2 to April 13, 2020, a period in which the number of Covid-19 tests administered in New York City almost tripled. Section 6 concludes with a discussion of the results. An appendix contains the formulas used for the construction of the inequality measures (Lorenz curves and Gini coefficients). The data used in this investigation along with the Matlab code to replicate the tables and figures are available at http://www.columbia.edu/~mu2166/stu_covid19/.

2 Data

The data used in this investigation include the cumulative number of New York City residents who were ever tested for Covid-19 and the number of residents who tested positive as of April 2, 2020 for each of the city's 177 zip codes. The data source is the New York City Department of Health and Mental Hygiene (DOHMH) Incident Command System for Covid-19 Response (https://github.com/nychealth/coronavirus-data#tests-by-zctacsv). In addition, the dataset includes zip-code level data on per capita income in the past 12 months in dollars of 2018 surveyed over the period from 2014-2018, population, and racial composition as of 2018. The source is the American Community Survey, series code B19301_001E and B01003_001E, respectively (https://www.census.gov/data/developers/data-sets/acs-5year.html).

Table 1 displays summary statistics for Covid-19 testing, per capita income, and population across the 177 zip codes of New York City. The number of tests and of negative test results are cumulative as of April 2, 2020. Each zip code received on average 908 Covid-19 tests per 100,000 residents. On average half of the tests came back negative. The sample displays significant cross sectional variation, with standard deviations of 268 and 9 percent for tests and negatives, respectively. The City also displays large cross sectional variation in income per capita with residents of the poorest zip code receiving $13,394 on average per year and residents of the richest $147,547.

3 Covid-19 Test Distribution By Income

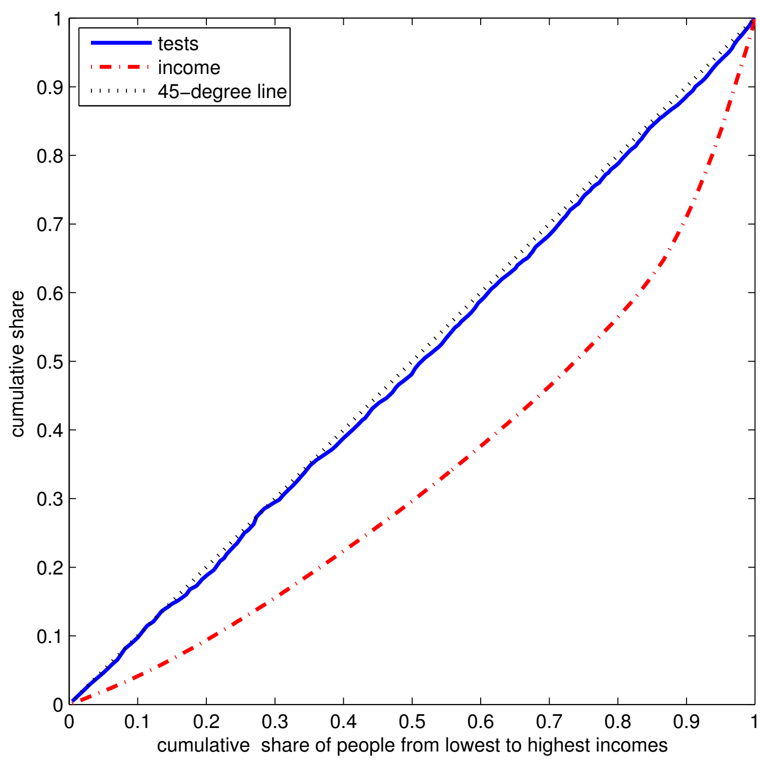
One of the goals of this paper is to document the distribution of tests for Covid-19 across income levels. Figure 1 displays the Lorenz curve of the number of tests for Covid-19 across

Table 1: Summary Statistics

Statistic	Covid-19 Total Tests per 100,000	Covid-19 Share of Negative Tests (%)	Per Capita Income dollars of 2018	Population
Mean	908	49	44287	47645
Median	860	49	31779	42653
Std.Dev.	268	9	31919	26698
Max	2390	75	147547	112425
Min	450	23	13394	3028

Notes. Summary statistics are computed across the 177 New York City zip codes. Total tests and negative tests are cumulative as of April 2, 2020. Replication code `summary_statistics.m` in `stu_covid19.zip`.

Figure 1: Lorenz Curves of Covid-19 Tests and Mean Income Across New York City Zip Codes



Notes. Own calculations based on data from the New York City Department of Health and Mental Hygiene, as of April 2, 2020, and American Community Survey. Replication code `gini_testing.m` in `stu_covid19.zip`.

Table 2: Gini Coefficients

Income	0.32
Covid-19 Tests	0.02
Covid-19 Negative Test Results	0.09

Note. Own calculations based on data from the New York City Department of Health and Mental Hygiene as of April 2, 2020, and American Community Survey. Replication code `gini_testing.m` and `gini_negatives_testing.m` in `stu_covid19.zip`.

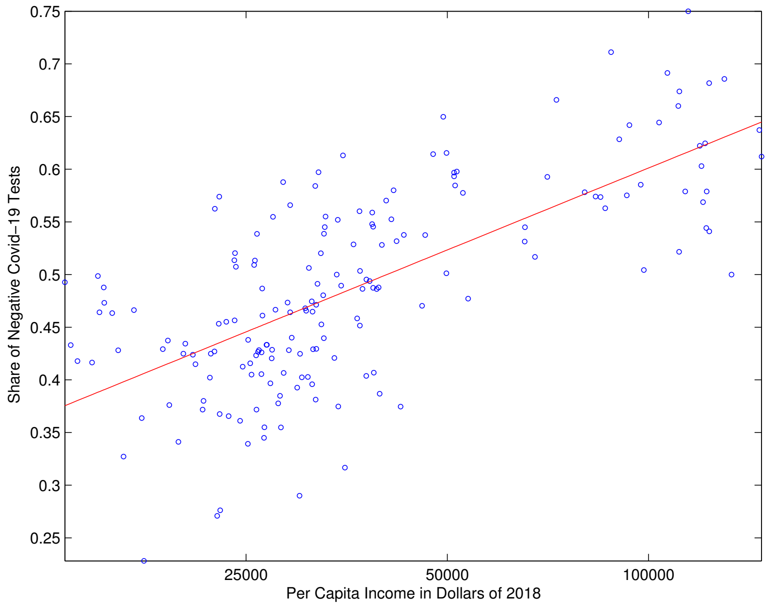
New York City zip codes as of April 2, 2020. For comparison, the figure also displays the Lorenz curve for the distribution of per capita income. The horizontal axis measures cumulative population shares sorted by income. The vertical axis displays with a solid line the cumulative share of Covid-19 tests and with a broken line the cumulative share of income. The appendix presents the formulas used to construct the Lorenz curves and the Gini coefficients discussed below.

The key message of the figure is that Covid-19 tests are almost perfectly distributed across income groups in New York City. Graphically, this is reflected in the fact that the Lorenz curve is nearly equal to the 45-degree line. The 10 percent of the population living in zip codes with the highest income per capita received 11 percent of all Covid-19 tests administered in the City and the 10 percent of the population living in zip codes with the lowest income per capita received 10 percent of the tests.

A more comprehensive and frequently used measure of inequality is the Gini coefficient, which is given by the ratio of the area between the 45-degree line and the Lorenz curve to the triangular area below the 45-degree line. The Gini coefficient associated with the distribution of Covid-19 testing across income levels is equal to 0.02, with a value of 0 representing a perfectly even distribution.

The evenness of the distribution of Covid-19 tests across income levels contrasts with the inequality in the distribution of income per capita across zip codes. This is reflected in the Lorenz curve for the income distribution being significantly below the 45-degree line. The top decile of the population earns 29 percent of total income, whereas the bottom decile earns only 4 percent. The Gini coefficient of income inequality is 0.32, sixteen times higher than the Gini coefficient of testing inequality. Because of the averaging of income per capita within each zip code, the reported Gini coefficient of New York's income distribution, 0.32, is lower than the one that results from using data at the household level, 0.55 for 2018 according to the American Community Survey. However, the Gini coefficient of 0.32 is the

Figure 2: Share of Negative Tests and Mean Income Per Capita Across New York City Zip Codes



Notes. Own calculations based on data from the New York City Department of Health and Mental Hygiene, as of April 2, 2020, and American Community Survey. The negative share is defined as the number of negative test results divided by the total number of tests. The solid line is the OLS regression. Replication code `negatives_vs_income.m` in `stu_covid19.zip`.

relevant one for the purpose of the present analysis because the most disaggregated level at which test for Covid-19 statistics are available is the zip code.

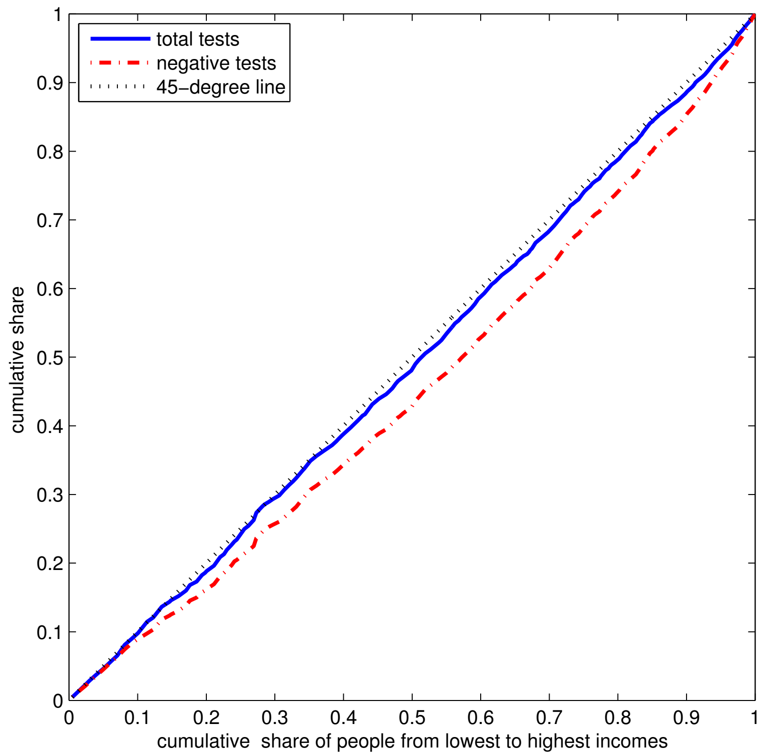
4 Test-Result Inequality

The available data make it possible to address the question of how outcomes of Covid-19 tests vary across income at the zip code level. Figure 2 plots with circles the share of tests that came back negative (the patient is not infected with the coronavirus) against income per capita at the zip code level. Income per capita is plotted on a logarithmic scale. The scatterplot displays a clear positive relationship between income per capita and the corresponding share of negative Covid-19 test results. The solid line is the OLS regression, which is given by

$$s_i^n = -0.69 + 0.11 \ln y_i^c + \epsilon_i, \quad R^2 = 0.48,$$

where s_i^n, y_i^c, and ϵ_i denote, respectively, the share of negatives, income per capita, and the regression residual in zip code $i = 1, \ldots, 177$. The slope coefficient, 0.11, is significant at the

Figure 3: Lorenz Curves of Negative Tests and Total Tests for Covid-19 Across New York City Zip Codes



Notes. Own calculations based on data from the New York City Department of Health and Mental Hygiene, as of April 2, 2020, and American Community Survey. Replication code `gini_negatives_testing.m` in `stu_covid19.zip`.

1 percent confidence level. It implies that moving from the poorest zip codes to the richest zip codes is associated with an increase in the share of negative Covid-19 test results of 27 percentage points, from 38 percent to 65 percent. It follows that unlike the distribution of tests for Covid-19, the distribution of test outcomes across income is significantly regressive. This suggests that the observed egalitarian distribution of tests need not reflect equal access to tests.

Expanding the OLS regression to include the share of black residents, s_i^b, and the share of other racial minorities, s_i^o (neither blacks nor whites), yields $s_i^n = -0.44 + 0.09 \ln y_i^c - 0.08\, s_i^b - 0.05\, s_i^o + \epsilon_i$, with $R^2 = 0.51$. According to this expression, race has a negative but relatively minor effect on the share of negative tests. Controlling for income, a one standard deviation increase in the share of black residents (25 percentage points) is associated with a fall in the share of negative test results of 2 percentage points. The association is even weaker for other racial minorities.

To emphasize the finding that the distribution of negative test results is unequal across income levels, figure 3 displays with a dash-dotted line the Lorenz curve for the distribution of the number of negative test results for Covid-19. For comparison, the figure reproduces with a solid line from figure 1 the Lorenz curve of the distribution of total tests. The Lorenz curve associated with negative test results is farther below the 45-degree line than the one associated with total tests, reflecting more inequality across income levels in test outcomes than in the number of tests administered. The Gini coefficient for the distribution of negatives is 0.09, almost five times larger than the one corresponding to the distribution of total tests.

5 Dynamics

The analysis thus far was conducted on data of cumulated tests and test results as of April 2, 2020. As a robustness check, this section examines data up to April 13. In the intermittent period, the number of administered Covid-19 tests in New York City increased from 73,215 to 182,099 or 150 percent.

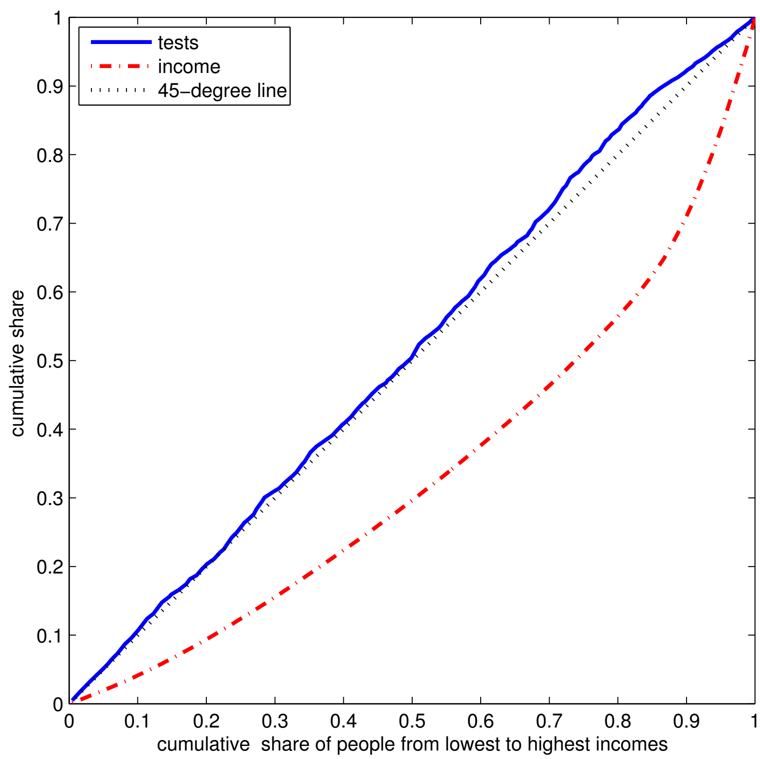
Figure 4 displays the Lorenz curve of the Covid-19 test distribution as of April 13, 2020. For comparison, the figure reproduces from Figure 1 the income distribution. The main message conveyed by the figure is that the distribution of tests continues to be egalitarian (i.e., close to the 45-degree line) after the significant increase in the number of New Yorkers that were tested. If at all, it became slightly more in favor of low income groups: Between April 2 and April 13, the fraction of test going to the bottom decile of the income distribution increased from 10 to 11 percent and that going to the top decile fell from 11 to 8 percent.

Figure 5 displays the Lorenz curve of the distribution of negative test results for Covid-19 in New York City as of April 13, 2020. The figure indicates that it continues to be the case that the distribution of negative test results is more unequal than that of administered tests. The bottom decile of the income distribution received 11 percent of the tests but 8 percent of the negative test results, while the top decile received 8 percent of tests administered, but 11 percent of the negative results.

6 Discussion and Conclusion

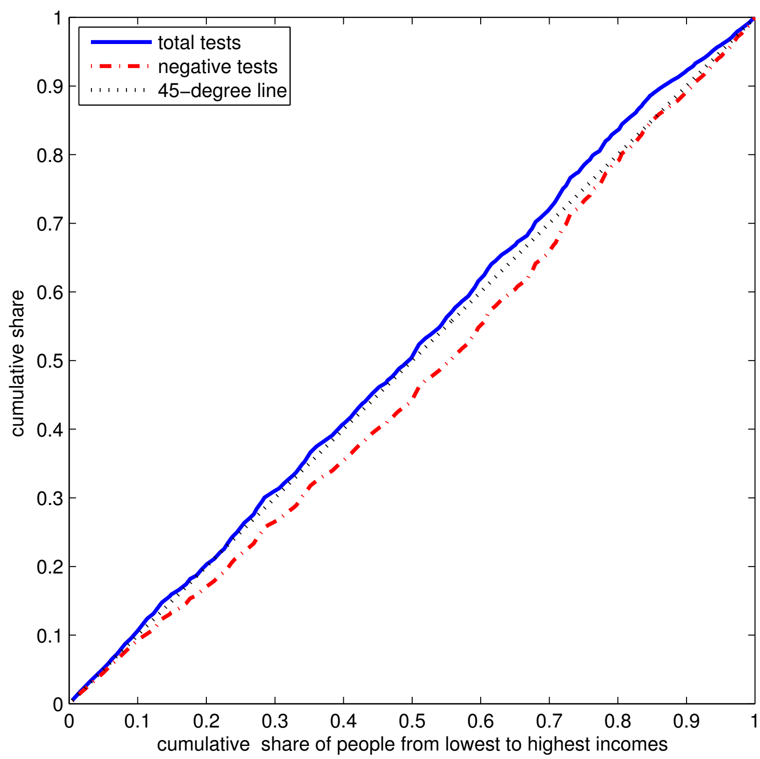
This paper contributes to the economic analysis of pandemics. It documents that in New York City, the most castigated city by the 2020 coronavirus outbreak, the ex-post likelihood of being tested for Covid-19 was evenly distributed across income levels measured at the zip-code unit. The bottom decile of the income distribution received 10 percent of all tests and the top decile 11 percent. The distribution of test outcomes, by contrast, displays significant

Figure 4: Lorenz Curves of Covid-19 Tests and Mean Income Across New York City Zip Codes as of April 13, 2020



Notes. Own calculations based on data from the New York City Department of Health and Mental Hygiene, as of April 13, 2020, and American Community Survey.

Figure 5: Lorenz Curves of Negative Tests and Total Tests for Covid-19 Across New York City Zip Codes, April 13, 2020



Notes. Own calculations based on data from the New York City Department of Health and Mental Hygiene, as of April 13, 2020, and American Community Survey.

inequality across income. The ex-post probability of testing negative for Covid-19 in the zip codes with the lowest per capita incomes was 38 percent compared to 65 percent in the zip codes enjoying the highest per capita incomes.

In light of the reported unequal distribution of test outcomes against lower income groups, it is possible that the observed egalitarian distribution of tests was associated with testing not being proportional to incidence. However, the data analyzed in this paper does not provide sufficient information to establish this conclusion. To see this, apply Bayes law, to obtain

$$P(pos|test)P(test) = P(test|pos)P(pos),$$

where P denotes probability, $test$ denotes being tested, and pos denotes a positive test result. Evaluate this expression at the bottom and top deciles of the income distribution, and take ratios, to obtain

$$\frac{P^{poor}(pos|test)}{P^{rich}(pos|test)} \times \frac{P^{poor}(test)}{P^{rich}(test)} = \frac{P^{poor}(test|pos)}{P^{rich}(test|pos)} \times \frac{P^{poor}(pos)}{P^{rich}(pos)}$$

The estimates obtained in section 3 suggests that $\frac{P^{poor}(pos|test)}{P^{rich}(pos|test)} = 1.5$ and the estimates of section 4 that $\frac{P^{poor}(test)}{P^{rich}(test)} = 1$. Therefore we can write

$$1.5 \times 1 = \frac{P^{poor}(test|pos)}{P^{rich}(test|pos)} \times \frac{P^{poor}(pos)}{P^{rich}(pos)}.$$

If the probability of getting tested for the coronavirus conditional on being infected is the same for the bottom and top deciles, $\frac{P^{poor}(test|pos)}{P^{rich}(test|pos)} = 1$, then it follows that the incidence rate is 50 percent higher in the bottom decile than in the top decile. In this case, the estimated egalitarian distribution of tests would not reflect the relevant incidence of Covid-19 across income groups.

However, testing selection could introduce variation in $P(test|pos)$ across income, which in turn will affect the inference about differences across income levels in incidence of Covid-19, $P(pos)$. In particular, if testing criteria are more stringent in poor neighborhoods than in rich ones, then $P(test|pos)$ could be larger in poor zip codes than in rich ones. To illustrate how testing selection criteria can affect $P(test|pos)$, consider the following example. Suppose there are 100 people in the population, of which 50 have the flu, 30 have corona, and 20 have corona and the flu, so that only 10 have corona but not the flu. People who have the flu have only 1 symptom, say fever. People who have corona have 1 additional symptom, say lack of smell. The health authority has only 5 Covid-19 tests. Suppose initially the testing criterion is just fever. Sixty people therefore meet the criterion, so $P(test|pos) = 1/12$. Now suppose the criterion is fever and lack of smell. In this case only 30 people qualify, namely, the people infested with corona. So $P(test|pos) = 2/12$. This example suggests that if testing selection criteria were more stringent in poor neighborhoods, then, given the results reported in this paper, incidence, $P(pos)$ would be less than fifty percent larger in the poor zip codes than in the rich ones and in principle could even be smaller.

Obtaining reliable measures of incidence would require randomized testing. This would make it possible to design more efficient allocations of tests than relying on what, based on the evidence presented in this paper, appears to be a simple egalitarian rule.

Appendix: Calculation of Gini Coefficients and Lorenz Curves

Let y_i^c denote per capita income in zip code $i = 1, \ldots, 177$. Suppose that y_i^c is sorted in ascending order, so that $y_i^c < y_{i+1}^c$ for any $1 \leq i < 177$ and let p_i be the population of zip code i. Then income in zip code i, denoted y_i is approximated by

$$y_i = y_i^c p_i$$

The share of income of zip code i in total New York City income is defined as

$$s_i^y = \frac{y_i}{\sum_{i=1}^{177} y_i}$$

The cumulative income share up to the ith poorest zip code, denoted S_i^y, is given by

$$S_i^y = \sum_{k=1}^{i} s_k^y$$

Similarly, the population share of the ith poorest zip code, denoted s_i^p is given by

$$s_i^p = \frac{p_i}{\sum_{i=1}^{177} p_i}.$$

And the cumulative population share up to the ith poorest zip code, denoted S_i^p, is given by

$$S_i^p = \sum_{k=1}^{i} s_k^p.$$

Let τ_i denote the number of Covid-19 tests in the ith poorest zip code. Then, the share of tests in zip code i, denoted s_i^τ is given by

$$s_i^\tau = \frac{\tau_i}{\sum_{i=1}^{177} \tau_i}$$

And the corresponding cumulative share up to the ith poorest zip code, denoted S_i^τ, is

$$S_i^\tau = \sum_{k=1}^{i} s_k^\tau.$$

Figure 1 plots the variables S_i^y and S_i^τ (vertical axis) against the variable S_i^p (horizontal axis).

The Gini coefficient of the income distribution across zip codes is measured as

$$\text{Gini coefficient of income distribution} = 1 - \frac{\sum_{i=1}^{177} s_i^p S_i^y}{\sum_{i=1}^{177} s_i^p S_i^p},$$

and the Gini coefficient of the Covid-19 testing distribution across income levels by

$$\text{Gini coefficient of Covid-19 testing across income levels} = 1 - \frac{\sum_{i=1}^{177} s_i^p S_i^\tau}{\sum_{i=1}^{177} s_i^p S_i^p}$$

References

Borjas, George J., "Demographics Determinants of Testing Incidence and Covid-19 Infections in New York City Neighborhoods," NBER working paper 26952, April 2020.