The determinants of the differential exposure to COVID-19 in New York City and their evolution over time*

Milena Almagro† Angelo Orane-Hutchinson‡

June 25, 2020

Updated frequently. Click here for most updated version.

Abstract

We argue that occupations are a key explanatory variable for understanding the early transmission of COVID-19 in New York City, finding that they play a larger role than other key demographics such as race or income. Moreover, we find no evidence that commuting patterns are significant after controlling for occupations. On the other hand, racial disparities still persist for Blacks and Hispanics compared to Whites, although their magnitudes are economically small. We perform a daily analysis over a range of one month to evaluate how different channels interact with the progression of the pandemic and the stay-at-home order. While the coefficient magnitudes of many occupations and demographics decrease, we find evidence consistent with higher intra-household contagion over time. Finally, our results also suggest that crowded spaces play a more important role than population density in the spread of COVID-19.

*We thank Michael Dickstein, Sharon Traiberman, and Daniel Waldinger for their useful comments. Any errors or omissions are our own.

†Federal Reserve of Minneapolis and Chicago Booth. Email: milena.almagro@chicagobooth.edu

‡Department of Economics, New York University. Email: aoh227@nyu.edu

1 Introduction

COVID-19 has affected different locations to very different extents, with much of this variation being explained by characteristics such as the number of international travellers, weather conditions, local policies to control the pandemic, and the timing of those policies. Surprisingly, large differences exist even across smaller geographical units such as neighborhoods *within* a city. For example, Figure 1 shows the differences in the rates of positive tests by zip code of residence in New York City (NYC).

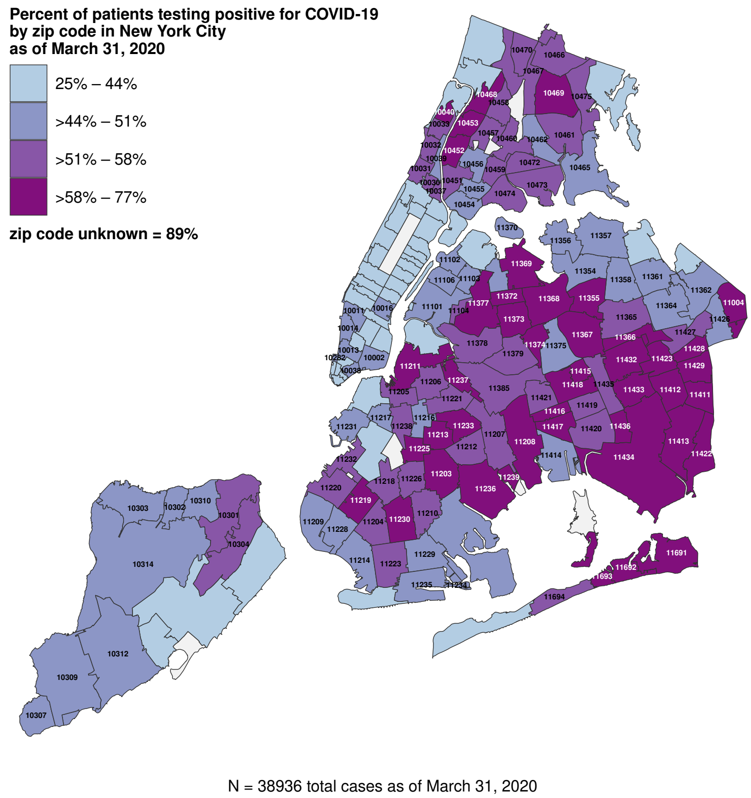

Figure 1: Map of the rate of positives by zip code as of March 31, 2020.

From simple inspection, zip codes with the highest rates are found in the boroughs of Bronx, Brooklyn, and Queens. These boroughs are not only those with the lowest levels of average income but also home to the majority of Blacks and Hispanics living in NYC.[1]

The spatial correlations between the incidence of the pandemic and demographics has garnered the attention of many economists and policy makers. For example, Borjas (2020) and Schmitt-Grohé et al. (2020) show that many of the spatial disparities

[1] These groups compose 29% and 56%, respectively, of all Bronx residents, 31% and 19% for Brooklyn, and 17% and 28% for Queens.

in testing and positive rates across NYC neighborhoods are explained by demographics, and Sá (2020) shows how different socioeconomic variables relate to the number of cases and deaths in the UK. Given that COVID-19 does not intrinsically discriminate across demographic groups, the reason for such disparities still remains an open question. Therefore, the goal of this paper is to assess the importance of a set of observable factors in explaining the existing disparities across NYC neighborhoods, such as population density, commuting patterns, and occupations. We first focus on occupations motivated by the hypothesis that workers in jobs with a higher degree of human exposure are more likely to contract the disease.[2] Because households from a low socioeconomic status are more likely to have jobs with a higher degree of human exposure, the pandemic has an unequal impact across society, with a larger impact on those from a lower socioeconomic status. In light of this mechanism, the pandemic can magnify existing inequalities – a problem that the world has previously witnessed with past pandemics (Furceri et al., 2020). To the best of our knowledge, our analysis is the first to find empirical evidence that occupations play a key role in the different impact of COVID-19 across demographics.[3]

To understand the relevance of different mechanisms, we use data on the number of tests and positives across NYC zip codes provided by Department of Health and Mental Hygiene of New York City (DOH).[4] Because these data have been released (almost) on a daily basis, we are able to keep track of the number of tests and the fraction of those that are positive since April 1. We combine the data on testing with neighborhood and demographic indicators, which are provided by the American Community Survey (ACS). Namely, we use zip code level data on population density, commuting patterns, income, as well as race and age composition. We also include employment data. To analyze the role of occupations, we include the share of workers for 13 categories in each zip code constructed from the ACS according to their degree of human interaction.[5]

Our results show that occupations are a key component in explaining the observed differences across NYC areas at early stages of the pandemic. For example, in our preferred specification, we find that a one-percentage-point increase in the number of workers employed in transportation, an occupation that has been declared essential

[2]Another explanation could be selection of workers along comorbidities across occupations. For example, as mining has been traditionally related to respiratory diseases, miners may show a higher propensity of contracting the disease and of more severe symptoms. However, given as interactive occupations have become more important in larger metros over time (Michaels et al., 2019), we believe that explaining disparities through different degrees of human exposure is a more credible hypothesis for workers in NYC.

[3]Barbieri et al. (2020) provide a descriptive analysis of different occupations in terms of human exposure and point out that this could have played a key role in the early spread of the pandemic in Italy.

[4]Unfortunately, at the time of this analysis, there is no data available with the number of deaths by zip code.

[5]The ACS provides the number of workers employed at different occupations, all at the zip code level.

and has a high degree of exposure to human interaction, increases the share of positive tests by 1% by April 20, six weeks into the pandemic. Moreover, we show that length of commute and the use of public transport are not significant after controlling for occupations.[6] In terms of neighborhood characteristics, we also find that the magnitude of the coefficient for household size is roughly six times larger than the coefficient for neighborhood density for our preferred specification for April 20. This result suggests that crowding of shared spaces plays a more important role than the density of locations.

Several policy implication arise from our analysis. First, policy makers can target specific groups of riskier occupations with the distribution of protective gear, testing, and vaccination when these are scarce. These types of policies can easily complement existing ones and it has a twofold purpose. First, it helps mitigate the unequal effects of COVID-19 by targeting those who are more vulnerable, either because they are more exposed to the virus or because other demographics or comorbidities make them more likely to contract the disease. Second, if this type of policy targets those workers that are more likely to be in contact with the rest of the population, it also mitigates the risk of exposure for everyone in the population at large. Additionally, another implication that arises from our analysis is that, while the subway may have played a role in the spread of COVID-19 at the city level, focusing policy on the subway may not help mitigate contagion at the local level. Finally, observe that our results also suggest that low income households are more vulnerable to infection in part because these households are on average more crowded that high income households. In particular, we expect transmission due to crowded homes to be specially important during lockdowns, when household members tend to spend most of the day together. Thus, our results suggest a potential drawback for such policies and invite policy-makers to rethink the way stay-at-home orders are designed. For example, local governments could offer low income families the option of an alternative shelter during the pandemic, such as vacant hotels. We believe that this type of policy is useful not only for containing the pandemic but also for closing the gap of its impact across demographic groups.

In each of our analyses we use the fraction of positive tests to date across NYC zip codes as our dependent variable.[7] Moreover, in all of our specifications we include day

[6]Harries (2020) argues that the NYC subway was crucial for spreading the pandemic in NYC. More recently, Furth (2020) shows that "local infections are negatively correlated with subway use." A key aspect to bear in mind is that the data on commuting patterns provided by ACS documents average public transport usage and commute times before the pandemic. Hence, the best way to analyze such hypothesis is using daily commuting patterns across geographical units during the time of the pandemic.

[7]We could also focus on the number of positive tests per capita. We refrain from doing so for two reasons. First, random testing has not been possible in NYC due to the limited availability of tests in early stages of the pandemic. Second, Borjas (2020) highlights that the incidence of different variables on positive results per capita is composed of two things: A differential incidence on those who are tested, but also a differential incidence on those with a positive result conditional on being

fixed effects that capture any time trends that are common to all neighborhoods in NYC. We start by including a small set of neighborhood controls, such as commuting patterns, population density, and health insurance controls.[8]

Additionally, our results are robust after including demographics, as well as borough fixed effects.[9] Including demographics leads to several striking patterns. Whereas simple correlations show that wealthier neighborhoods have a lower rate of positives, we show that income is not significant when occupations are included. However, we still see significant and positive effects on positive rates for minorities. These results could be because minorities are less likely to get tested, they have to be in worse conditions than Whites in order to get tested, or that they are more likely to contract the disease due to existing comorbidities.[10] However, whether these racial disparities are economically relevant can be questioned. Moreover, their magnitudes decrease over time as more testing becomes available – with Asians showing no statistical significance at the end of our sample. For example, on April 1, one month after the pandemic started in NYC, we find that a one-percentage point increase in the share of Blacks correlates with an increase of 0.34% in the share of positive tests, for an average number of 51% of positive cases. By April 20th, these numbers are 0.15% and 54%, respectively. For Hispanics, the disparity is larger, where a one-percentage-point increase in their population corresponds to an increase of 0.38% and 0.23% in the rate of positives, for the same two dates.

In all of our specifications, we also include the share of the population being tested, which we call "tests-per-capita." We include this variable for two reasons. First, it allows us to interpret our results as conditional on the number of tests per capita that have been administered. Therefore, it helps to control for selection on testing. Second, the limited availability of tests in NYC forced health authorities to constrain testing to people showing sufficiently acute symptoms or determined to be at high risk of infection at the beginning of the pandemic. Thus, we expect the daily number of tests administered to be close to the population in that segment, which in turn should be roughly proportional to the number of infected people.[11] Therefore, we use the number of tests per capita as a proxy for the overall level of the spread of the pandemic *within* a neighborhood. We find that when the number of tests per

tested. Therefore, we believe that the fraction of positive tests is the variable that maps most closely the actual spread of the disease within a neighborhood throughout our sample.

[8]The data provided by ACS are averages at the neighborhood level. Hence we cannot incorporate neighborhood fixed effects given the lack of temporal variation of our covariates of interest.

[9]We use similar controls to those in Borjas (2020) for comparability purposes.

[10]Some evidence that the first two are plausible mechanism can be found in `www.modernhealthcare.com/safety-quality/long-standing-racial-and-income-disparities-seen-creeping-covid-19-care`. An example for the third channel can be found in `https://www.sciencedaily.com/releases/2020/05/200507121353.htm`. This study found that lower levels of vitamin D, which varies with the levels of melanin in the skin, were positively correlated with higher mortality rates.

[11]As a matter of fact, at earlier dates, tests were performed only on those who required hospitalization.

capita increases, the share of positive tests also increases. This result stems from both variables co-moving with the true number of infected people within a neighborhood. However, we also find that, as testing becomes more widely available and more tests are performed on the asymptomatic population, the magnitude of tests-per-capita decreases over our analyzed time period.

Our daily analysis also reveals that, as the stay-at-home order starts to be effective, the magnitude of many occupations begins to decreases. For example, a one-percentage-point increase in the number of workers employed in transportation decreases its size to 1% as of April 20, almost two months into the pandemic and one month after the stay-at-home order went into effect. On the other hand, we still find a rather stable coefficient of household size over time, which is consistent with the stay-at-home order being more helpful at mitigating contagion at work or in public spaces than within the household.[12]

We conclude that much of the disparity in the rates of positives across demographic groups can be partially explained a heterogeneous distribution of demographics across occupations. In particular, a key channel appears to be the differences in exposure to human contact across jobs. However, our results also suggest that the relevance of these variables decreases over time, and that this change occurs in tandem with an increase in intra-household contagion as days go by. These trends are consistent with the progression of the pandemic and its interaction with the policies set in place. In light of our results, we also propose policies focusing on minorities that can not only help mitigate the effect of the pandemic among those demographic groups, but that may have substantial positive spillovers on the rest of the population.

2 Data description and patterns

Our data on COVID-19 incidence and the number of tests performed is from the NYC DOH. The DOH releases (almost) daily data on the cumulative count of COVID-19 cases and the total number of residents that have been tested, organized by the zip code of residence. We have collected data starting from April 1, with only April 2 and April 6 missing from our sample.[13]

We obtain demographic and occupation data at the zip code level from the ACS. The demographic characteristics we include are zip code median income, average age, racial breakdown, and health insurance status. We also include commuting-related variables: average commute time to work as well as means of transportation. A simple analysis reveals that share of Blacks and Hispanics have a correlation coefficient of 0.426 and 0.312 respectively with a p-value smaller than 0.01 with the share of positive tests. For Asians we observe no significant relationship, with a correlation coefficient

[12] Sá (2020) also finds a positive relationship between the number of household members and number of cases for the UK.

[13] Unfortunately these days have never been made publicly available.

of 0.009 with a p-value of 0.905. Finally, we observe a negative correlation coefficient between log of median income and share of positives, with a correlation coefficient of -0.530 with a p-value smaller than 0.01. To summarize, these correlations show that, a priori, locations populated with more vulnerable groups show higher rates of positive tests.

We also construct the shares of the working-age population employed at different occupation categories. The ACS provides the number of workers employed in each occupation by zip code of residence. We then categorize them according to their essential definition, spatial correlations between them, and similarity in work environments and social exposure.[14] The final occupation groups that we use in our regressions are: (1) Essential - Professional: Management, Business, Finance; (2) Non essential - Professional: Computer and Mathematical, Architecture and Engineering, Sales and Related, Community and Social Services, Education, Training, and Library, Arts, Design, Entertainment, Sports, and Media, Administrative and Office Support; (3) Science fields: Life, Physical, and Social Science; (4) Law and Related: Legal; (5) Health practitioners; (6) Other health: Health technologists, Technicians, and Healthcare Support; (7) Firefighting: Firefighting and prevention; (8) Law enforcement; (9) Essential - Service: Food Preparation and Serving, Buildings and grounds cleaning, and Maintenance; (10) Non essential - Service: Personal care and Service; (11) Industrial, Natural resources, and Construction: Construction and extraction, Material Moving, Farming, Fishing, and Forestry production; (12) Essential - Technical: Installation, Maintenance, and Repair; (13) Transportation. For our occupational regressors, we count the number for workers in each of these occupations and normalize by working-age population, which includes people between the ages of 18 to 65 years old.

[14]Leibovici et al. (2020) rank occupations according to an index of occupational contact-intensity, defined from a survey by O*NET. They use ACS individual-level data at the four digit Standard Occupation Classification (SOC) level and match it to 107 ACS-defined occupations. Unfortunately, we only observe occupations at the SOC first level of aggregation for zip code data and cannot match their classification to our spatial distribution. Nonetheless, our categorization closely follows the intensity index grouping for the more specific group of occupations when aggregated to the first SOC level. More importantly, when defining our 13 categories, we avoid mixing occupations with large differences in their contact-intensity values. For robustness we have also performed our analysis with two alternative classifications for occupations. First, we divided occupations between essential and non-essential as declared by the US government. Second, we used the four categories defined in Kaplan et al. (2020). In both cases, the high level of aggregation lead to non-significant estimates or results that were hard to reconcile with observational evidence.

3 Results

3.1 General Results

In this section, we present the main empirical results for our four different specifications. Our unit of analysis is the zip code, and all models include the cumulative share of positive tests by day as the dependent variable. In all of our specifications we attach weights equal to population size to our geographical units. Additionally, we include tests-per-capita as a proxy for the overall spread of infection within the neighborhoods. The first specification includes some widely discussed potential factors of the spread of COVID-19 in NYC: density and commuting patterns, specifically, log of population density, percentage of workers using public transport, average commute time, and the percentage of the population that is uninsured. We expand our second specification by including our proposed mechanism, namely, the percentage of the working-age population employed in each of the 13 occupation categories defined above. The third specification adds demographic controls related to income, age, gender, household size, race, and borough fixed effects. Exploiting the fact that we have daily data over multiple days, we estimate a separate regression for each day, which allows us to detect any time patterns in the correlations. Therefore, in all of our specifications we run the following regression equation

$$\text{share of positive tests}_{it} = \alpha_t + \beta_t \text{tests per capita}_{it} + \gamma_t X_i + \epsilon_{it},$$

where the set of controls X_i vary for each specification according to the description above.

The first specification - see Table 1 columns 1,4, and 7, which correspond to April 1, April 20, and April 30, respectively - shows the effect of the variables commonly used to explain the incidence of COVID-19 in NYC. Whereas Harries (2020) finds subway use was a major factor of the virus spread, we find it does not have a significant effect. This result could be due to the lack of cross-neighborhood variation to identify this effect, because most New Yorkers use public transportation in their daily commute. Nonetheless, commute time is a significant factor. For example, for April 20 - column 4 of Table 1 - a 10% increase on average which equals to a four-minute increase in commute time, correlates with a 0.02-point increase in the share of positive tests, equivalent to approximately a 4-percentage-point increase in the share of positive tests.[15] We also find that the share of population that is uninsured has a significant positive coefficient for most of our sample. This result could be because uninsured patients are less likely to get tested for fear of medical charges, and therefore only submit a test when experiencing acute symptoms. For example, for April 20 - column 4 of Table 1, we find that a one-percentage-point increase in the share of uninsured population is correlated with a 1.7-percentage-point (0.924*0.01/0.54)

[15]The average rate of positive tests on April 20 was 54%.

increase in the share of positive tests. Although the magnitude of this variable decreases as we include other covariates, its estimated coefficient remains positive and significant.

In specification (2), columns 2, 5, and 8 in Table 1, we test the importance of different occupations. We include the variables defined as the shares of the working-age population employed in these occupations, so the coefficients are relative to the working-age population who are unemployed. The coefficients can be read as the effect of a one-percentage-point increase in the population employed in the particular category on the share of positive tests. We find some occupations explain a significant part of the variation in COVID-19 incidence. On the one hand, an increase in the share of workers employed in non-essential - professional, other health (not health practitioners), and transportation occupations are all associated with a higher percentage of positive tests. On the other hand, higher shares of workers in the science fields category, legal occupations, and law enforcement have a negative correlation with the share of positive tests. These results are discussed further in the time-trends section.

Perhaps surprisingly, under this specification, neither commute time nor the share of the population using public transport have a significant effect. This result suggests commuting patterns are closely related to occupations, and most of the explanatory variation for commuting patterns comes through this channel. This result also implies the existence of within-city location and mobility patterns that are occupation specific.

We include demographic variables in the third model - columns 3, 6, and 9 in Table 1. Notably, the income effect disappears when we control for occupations, suggesting the previous correlation presented in Section 2 is due to income differences across jobs. Still, some demographic effects remain significant, even after including borough fixed effects. For example, on April 20 - column 6 in able 1, a one-percentage-point increase in the share of Blacks and Hispanics leads to a 0.15% and 0.23% increase respectively in the rate of positives, an effect that is economically small. A plausible explanation for these patterns could be driven by a racial bias on the incidence of testing, as pointed out by Borjas (2020). Another explanation is differences in adherence to the shelter-in-place policy, as explored by Coven and Gupta (2020). We also find that household size has a positive correlation with the share of tests that are positive. For example, for the same regression, adding one extra person to the average household, a 37% increase, corresponds with a 6.7% (0.37*0.099/0.54) increase in the percentage of positive tests. On the other hand, for this specification, we do not find a significant effect for neighborhood density. This result suggests that crowding of spaces may be a more important factor in explaining the spread of COVID-19 than density.

The tests-per-capita coefficient is positive and highly significant across all days for specification (3) - columns 3, 6, and 9 in Table 1. Because of the scarcity of tests, testing was only performed on those showing sufficiently severe symptoms or who had a high risk of infection. As argued above, we interpret this variable as a proxy for the rate of infections within the neighborhood especially at early stages of the pandemic.[16] Its magnitude decreases over time as testing becomes more available to the rest of the population.[17]

[16]A concern are potential large differences in the age distribution across NYC zip codes. In the data, we find that the average age ranges from 27.5 to 45.5 across neighborhoods in NYC, with the exception of zip code 11005. It is a fairly small zip code with 1700 residents, an average age of 76, and mainly composed of retired immigrant women. Given such differences, we have excluded it from our analysis.

[17]One should bear in mind that the variable tests-per-capita may be capturing other time-varying unobservables that correlate with the share of positive tests, such as (lagged) daily commuting patterns at the neighborhood level. Thus, the interpretation of its coefficient should not be done lightly. Due to data limitations, we are not able to test for such hypotheses.

Table 1: Dependent variable - share of positive tests (cumulative, up to specified date)

	April 1			April 20			April 30				
	(1) Controls Neighborhood Controls	(2) + Occup.	(3) Borough FE + Dem. & Borough FE	(4) Controls Neighborhood Controls	(5) + Occup.	(6) Borough FE + Dem. & Borough FE	(7) Controls Neighborhood Controls	(8) + Occup.	(9) Borough FE + Dem. & Borough FE		
Tests per capita	9.017***	11.186***	12.050***	0.667	0.262	2.553***	0.723**	0.291	1.437***		
Log Density	0.015	0.022*	0.032***	0.024**	0.015*	0.016***	0.013*	0.005	0.010*		
% Public Transport	-0.015	0.013	-0.059	0.010	-0.001	-0.017	0.006	0.019	0.023		
Log Commuting Time	0.237***	-0.016	-0.054	0.232***	0.001	-0.008	0.196***	0.014	-0.012		
% Uninsured	1.002***	0.662***	0.150	0.924***	0.417**	0.351***	0.831***	0.424***	0.302***		
% Essential - Professional		0.156	0.766***		-0.210	0.235		-0.153	0.039		
% Non ess. - Professional		0.669***	0.544**		0.329**	0.224*		0.192*	-0.008		
% Science fields		-4.703***	-2.965***		-1.931*	-1.609**		-1.000	-1.148*		
% Law and related		-0.410	-1.427**		-0.492	-0.898**		-0.364	-0.876**		
% Health practitioners		-0.432	-0.167		-0.155	-0.206		0.065	-0.130		
% Other health		0.947***	0.346		0.815***	0.365		0.703***	0.307		
% Firefighting		2.743**	1.629*		0.379	-0.156		-0.223	-0.953		
% Law enforcement		-0.301	-0.223		-1.970*	-1.344**		-1.017	-0.729		
% Essential - Service		-0.100	0.245		0.312	0.082		0.236	-0.075		
% Non ess. - Service		0.769	1.154**		-0.046	0.578*		-0.042	0.281		
% Ind. and Construction		1.091**	0.839**		0.271	-0.079		0.254	-0.007		
% Essential - Technical		-2.025*	-0.319		-0.785	-0.908*		-0.159	-0.917**		
% Transportation		1.752***	1.102**		1.253***	0.541*		1.028***	0.229		
Log Income			-0.010			-0.022			0.015		
Share $\geq 20, \leq 40$			-0.357**			-0.208**			-0.028		
Share $\geq 40, \leq 60$			-0.611**			-0.198			-0.067		
Share ≥ 60			-0.347*			0.002			0.276***		
Share Male			-0.146			0.318**			0.218		
Log Household Size			0.037			0.099***			0.090***		
% Black			0.175***			0.081***			0.083***		
% Hispanic			0.194***			0.125***			0.135***		
% Asian			0.141***			0.012			0.019		
Bronx			-0.014			-0.062***			-0.036***		
Brooklyn			0.086***			0.034**			0.036***		
Queens			0.084***			0.023			0.034***		
Staten Island			0.083***			-0.064***			-0.029		
Constant	-0.671**	-0.149	0.196	-0.682***	0.201	0.111	-0.513***	0.190	-0.024		
Observations	174	174	174	174	174	174	174	174	174		
R^2	0.514	0.694	0.839	0.673	0.800	0.920	0.718	0.807	0.911 .763	0.821	0.896

Weighted OLS by population size. Robust standard errors.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, *** $p < 0.01$

3.2 Daily comparison and time trends

In this section we present a time-varying analysis that could provide insights on both the evolution of the pandemic effects as well as the health policies in place. We follow specification (3):

$$\text{share of positive tests}_{it} = \alpha_t + \beta_t \text{tests per capita}_{it} + \gamma_t X_i + \epsilon_{it},$$

where X_i is the vector of neighborhood characteristics including commuting patterns, share of occupations, demographics, and borough fixed effects, and compare the estimated coefficient of the same variable across different dates.

For illustrative purposes we start by plotting the evolution of some specific variables for all days in April. In Figure 2 we plot the estimated coefficients, γ_t, for different variables in X_i, and confidence interval at 95% level from April 1 to April 30. For the top left panel, we plot the coefficients for the share of workers in Transportation and Science. For the top right panel, we plot the coefficients for Density and Household size. Finally, for the bottom panel we plot the coefficients for the share of population who is Black as well as the share of Asians. For the evolution of all variables see Section B of the Online Appendix.

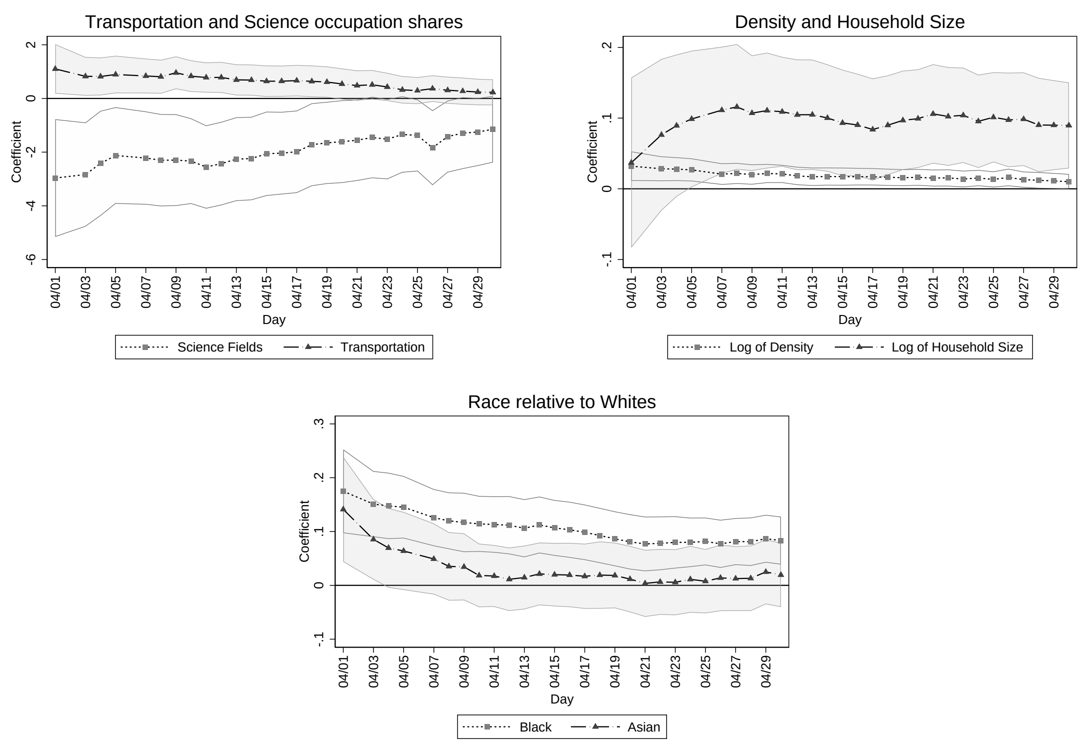

Figure 2: Daily evolution of coefficients

The first pattern the we observe is that the magnitude of the coefficients for Transportation and Science fields decreases over time. This result suggests that the importance of occupations in the spread of the disease decreased with time. This pattern is consistent with the effects associated with the stay-at-home order that was effective on March 22: As people were required to stay in their homes, the transmission through human exposure in work places or in public spaces became less prevalent.

Second, we analyze the time-varying effects of neighborhood density and household size. While both coefficients are positive and significant for most of our sample, we see that the coefficient of household size is orders of magnitude larger than the coefficient for neighborhood density, ranging from roughly 6 to 9 times larger after April 7. The first conclusion that we draw is that intra-household crowding appears to be a more important channel in the transmission of COVID-19 than neighborhood density. Another implication that arises from these first two graphs together is that, while shelter-in-place policies are useful at mitigating contagion in public spaces or in workplace locations, it may not have been as useful to prevent intra-household contagion.

Finally, another outstanding time pattern is that the coefficients on racial composition decrease in magnitude as testing becomes more widely available. This result may suggest a stronger racial-selection component is at play among those in worse condition at earlier dates. For example, an explanation for this pattern could be that Black citizens were less likely to be tested or had to be in worse condition to access testing compared to White citizens.[18] See bottom panel of Figure 2 for the evolution of the coefficients of racial composition.

Now we turn to the analysis of the rest of the variables using 1, by comparing the estimated coefficients across columns 3, 6, and 9.

The result for the tests-per-capita variable is particularly salient; we observe a strong correlation on the share of positive tests that becomes progressively smaller over time. This result could be reconciled with the fact that in the earlier days of the crisis, testing was severely limited. Zip codes with more tests implied a higher share of people at high risk of having the disease. In light of this, a key takeaway from the results of our daily comparison is the importance of widespread testing, because it allows for a more accurate identification of the mechanisms that create demographic and occupational differences in COVID-19 exposure.

Notable time trends exist in the correlations associated with occupations. Higher shares of essential - professional and non-essential - service categories were associated with higher percentage-point increases in the rate of positive tests at earlier dates. On April 1, a one-percentage-point increase implied a 1.5- and a 2.2-percentage point increase in the positive rate of tests. However, they eventually decrease, averaging closer to a 0.3- and a one-percentage-point increase respectively on April 20, with essential - professional not being statistically significant. A plausible explanation is that these professions are either non-essential, or have the highest shares of remote workers. Although they were highly exposed to the virus before the shelter-at-home order, once the workers shelter in place, their correlation with positive tests decreases.

[18] Some evidence that this is a plausible mechanism can be found here.

The opposite happens in science fields and law occupations — they are negatively correlated with COVID-19 incidence pre-shelter order, but the effect trends towards zero.

We find interesting patterns for the essential occupations as well. An additional percentage point in the share of transportation workers is associated with between a 0.5- and a one-percentage-point increase in the rate of positive tests. The effect seems to decay over time, but at a slower rate than other occupations. This result could be due to its essential designation, but also due to the high exposure faced by workers in this occupation. See top left panel of Figure 2 for the evolution of the coefficients of the share of workers in Transportation and Science fields. The share of industrial, natural-resources, and construction occupations begins with a positive correlation with COVID-19 incidence. However, a week after the general stay-at-home order, the Governor of New York determined construction was not essential, and this order could explain the eventual attenuation of the correlation. Law-enforcement-occupation shares have a consistently negative correlation on the share of positive tests, whereas firefighter shares have a declining trajectory toward zero. A plausible explanation for this difference could be the partnership between the NYPD and health care groups to provide free testing to its members.[19] Furthermore, the NYPD provided additional work flexibility for members with pre-existing conditions and extensive sick leave. It's possible that early adoption of these measures protected the most vulnerable workers from infection from the onset.[20] The share of the uninsured population increasingly predicts the variation in positive test results. We find that an additional percentage point in the share of uninsured people predicts an almost 0.3-percentage-point increase in the share of positive tests. Although many health care providers are waiving COVID-19-related out-of-pocket costs, these fees remain very high for the uninsured, and so a higher incidence of COVID-19 in this group could imply a severe financial burden.

4 Conclusions and policy implications

In this paper, we present evidence showing that occupations are an important channel for explaining differences in the rates of COVID-19 across neighborhoods at the early stages of the pandemic. Using data from NYC at the zip code level, we study the relationship between the share of positive tests and the share of workers in different occupations. The DOH provides daily updates of COVID-19 test data, allowing us to study the aforementioned relationships over multiple days and to detect time patterns in their magnitudes.

We begin by showing descriptive evidence of the heterogeneous incidence of positive cases across neighborhoods, income, race, gender, and household size. A zip

[19] See this link for more information.

[20] More information on this can be read here.

code's median income is negatively correlated with its share of positive tests. Furthermore, we find that the shares of Black and Hispanic residents, and average household size, positively correlate with the share of positive tests. Highlighting these differences is important because these observations confirm that the disease has had more harmful effects on vulnerable communities. Finding an occupation mechanism that explains it could guide policy measures intended to alleviate its impact.

We estimate several models to explore the effect of occupations. Our first specification only includes neighborhood characteristics, such as the use of public transportation and the average length of daily commutes. Although commuting patterns have been put forth as a major factor in the spread of the disease in NYC, we show that, after including occupation controls, they fail to significantly explain variation in share of positive tests at the zip code level.

We find the strongest positive correlation on the share of positive tests with the share of workers in Transportation, Industrial, Natural-resources, Construction, and Non essential - Professional, with clear time trends in their estimated coefficients. For example, in the case of Transportation, a one-percentage-point increase in the share of workers in these occupations leads to a one- to two-percentage-point increase in the rates of positive results. Although the other two have a significant effect in positive shares at earlier dates, their magnitude becomes insignificant by the end of our sample period. This trend could be a result of the stay-at-home order. Conversely, higher shares of workers in Science Fields and Law Enforcement reduce the number of positive rates, with Science Fields decreasing in magnitude over time.

When adding demographic controls, we observe that racial patterns do persist, suggesting that the occupation mechanism does not fully explain all of the racial differences. However, their magnitude is small and arguably not economically relevant. Income and most age groups do not contribute to the variation in positive tests, suggesting the occupation mechanism can explain to a greater extent the disparities along those demographics observed in the data.

In all of our regression models we include the number of tests per capita, and find that it is a strong predictor of the share of positive tests. However, its relative importance declines over time, as tests become more widely available. Moreover, as this variable loses relevancy, more of the variation in COVID-19 incidence is explained through the occupation channel.

Our results suggest clear implications for policy. First, they suggest that policymakers can target specific groups in the provision of protective gear, tests, and vaccinations. The purpose of this policy is twofold: while it provides extra protection against the disease for those who are more vulnerable, it also has positive effects which will mitigate the risk of contagion for the rest of the population. For example, a policy that starts vaccinating and/or testing those workers with higher rates of human interaction affects not only those directly targeted by the policy, but also those who are likely to be in contact with them. Our results also suggest that health insurance condition, namely lack of insurance, plays a significant role, and its importance increases over time. Hence, local governments could incentivize the population without medical insurance to get tested, implementing policies such as full coverage of out-of-pocket costs in relation to COVID-19. Finally, we provide suggestive evidence that the stay-at-home order has mitigated contagion rates at work or in public spaces, while it has increased the probability of intra-household infections. This last result suggests the importance of policy or guidance measures to decrease spread within households.

References

Barbieri, T., Basso, G., and Scicchitano, S. (2020). Italian workers at risk during the covid-19 epidemic. *Unpublished Manuscript.*

Borjas, G. J. (2020). Demographic determinants of testing incidence and covid-19 infections in new york city neighbourhoods. *Covid Economics, Vetted and Real-Time Papers.*

Coven, J. and Gupta, A. (2020). Disparities in mobility responses to covid-19. *Working Paper.*

Furceri, D., Loungani, P., Ostry, J., and Pizzuto, P. (2020). Will covid-19 affect inequality? evidence from past pandemics. *Covid Economics*, 12:138–57.

Furth, S. (2020). Automobiles seeded the massive coronavirus epidemic in new york city. Available at `https://marketurbanism.com/2020/04/19/automobiles-seeded-the-massive-coronavirus-epidemic-in-new-york-city/`.

Harries, J. E. (2020). The subways seeded the massive coronavirus epidemic in new york city. *NBER Working Papers.*

Kaplan, G., Moll, B., and G., V. (2020). Pandemics according to hank. Available at `https://benjaminmoll.com/wp-content/uploads/2020/03/HANK_pandemic.pdf`.

Leibovici, F., Santacreu, A. M., and Famiglietti, M. (2020). Social distancing and contact-intensive occupations. Available at `https://www.stlouisfed.org/on-the-economy/2020/march/social-distancing-contact-intensive-occupations`.

Michaels, G., Rauch, F., and Redding, S. J. (2019). Task specialization in us cities from 1880 to 2000. *Journal of the European Economic Association*, 17(3):754–798.

Sá, F. (2020). Socioeconomic determinants of covid-19 infections and mortality: Evidence from england and wales.

Schmitt-Grohé, S., Teoh, K., and Uribe, M. (2020). Covid-19: Testing inequality in new york city. *NBER Working Papers.*