Journal of Urban Economics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Urban Economics

journal homepage: www.elsevier.com/locate/jue

JUE insight: The determinants of the differential exposure to COVID-19 in New York city and their evolution over time[☆]

Milena Almagro[a,*], Angelo Orane-Hutchinson[b]

[a] Federal Reserve Bank of Minneapolis and Booth School of Business, University of Chicago United States
[b] Department of Economics, New York University United States

ARTICLE INFO

Keywords:
Demographic disparities
Income disparities
Racial disparities
COVID-19 exposure
Occupations
Density
Housing density

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 with Whites, although the disparities' magnitudes are economically small. We perform our analysis over a range of several weeks 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.

1. Introduction

COVID-19 has affected different locations to very different extents, with much of this variation explicable 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, Fig. 1 shows the differences in the rates of positive tests by zip code of residence in New York City (NYC).

From simple inspection, zip codes with the highest rates are found in the boroughs of the 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 demographics explain many of the spatial disparities in testing and positive rates across NYC neighborhoods, 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 remains an open question. Therefore, the goal of this paper is to assess the importance of a set of observable factors such as population density, commuting patterns, and occupations in explaining the existing disparities across NYC neighborhoods.

To understand the relevance of different mechanisms, we use data provided by Department of Health and Mental Hygiene of New York City (DOH) on the number of tests and positives across NYC zip codes.[2] Because these data have been released 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, provided by the American Community Survey (ACS). Namely, we use zip code level data on population density, commuting patterns, income, and race and age composition. We also include employment data. To analyze the role of occupations in each zip code, we include the share of workers for 13 categories constructed from the ACS according to their degree of human interaction.[3] In each

☆ We thank Michael Dickstein, Sharon Traiberman, and Daniel Waldinger for their useful comments. Any errors or omissions are our own. The views expressed in articles and other content on this website are those of the authors and don't necessarily reflect the position of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

* Corresponding author.
E-mail addresses: m.almagro@nyu.edu, milena.almagro@chicagobooth.edu (M. Almagro), angelo.orane@nyu.edu (A. Orane-Hutchinson).

[1] These groups compose 29% and 56%, respectively, of all Bronx residents; 31% and 19% for Brooklyn; and 17% and 28% for Queens.
[2] Unfortunately, at the time of this analysis, there are no data available with the number of deaths by zip code.
[3] The ACS provides the number of workers employed at different occupations, all at the zip code level.

https://doi.org/10.1016/j.jue.2020.103293
Received 27 June 2020; Received in revised form 12 October 2020
Available online xxx

Please cite this article as: M. Almagro and A. Orane-Hutchinson, JUE insight: The determinants of the differential exposure to COVID-19 in New York city and their evolution over time, Journal of Urban Economics, https://doi.org/10.1016/j.jue.2020.103293

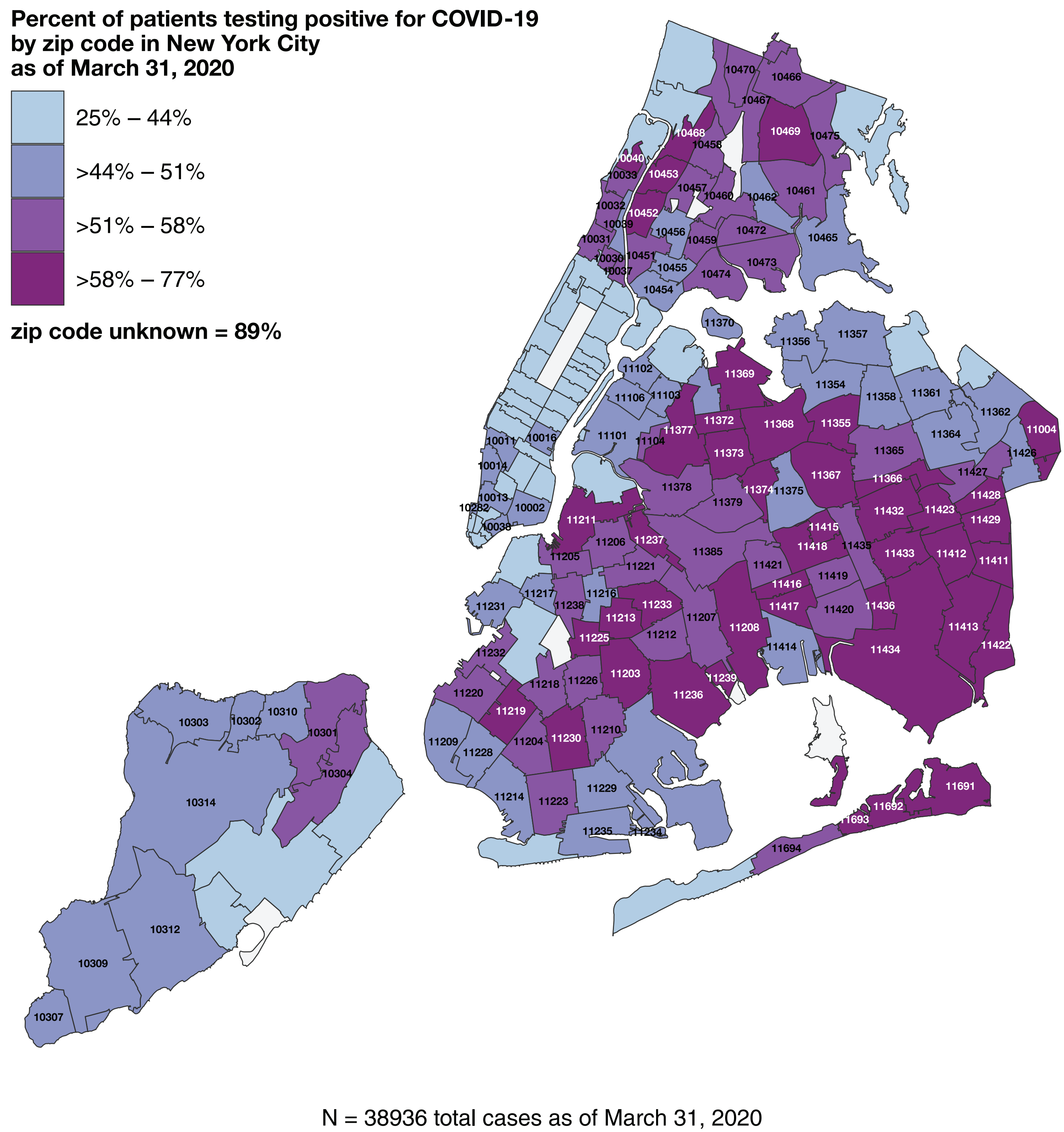

Fig. 1. Rate of Positive Tests ($\frac{\text{Positive tests}}{\text{Total tests}} \cdot 100$) by Zip Code as of March 31, 2020.

of our analyses, we use the fraction of positive tests to date across NYC zip codes as our dependent variable.[4] Moreover, in all of our specifications, we include day 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.[5]

[4] 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 because of 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 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.

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

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.[6] As different types of jobs are more concentrated in certain socioeconomic groups, the pandemic has had a larger impact on those from a lower socioeconomic status because they are more likely to have jobs with a higher degree of human exposure. To the best of our knowledge, our analysis is the first to find empirical evidence that occupations play a key role in the risk of exposure to COVID-19.[7] 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 and has a high degree of exposure to human interaction, increases the share of positive tests by 0.9% by April 30, six weeks into the pandemic.

In light of the importance of the occupational channel, we can draw several general conclusions. First, the combination of sorting of occupations across demographic groups with residential sorting on demographics can reinforce the risk of exposure. This feedback effect can cause large clusters of contagions in specific areas within a city, while keeping the spread contained in other areas. For example, as a result of homophily, low skilled workers, those with a higher risk of exposure, tend to interact more with other low skilled workers outside work. Thus, it is likely that their entire community becomes infected. This argument is similar to that of Azzimonti et al. (2020), who argue that infections occur in clusters defined by social and occupational networks. Second, the pandemic can magnify existing inequalities — a problem that the world has witnessed with past pandemics (Furceri et al., 2020). Third, understanding risk differentials across occupations could help in the optimal allocation of a future vaccine, as shown by Babus et al. (2020). Lastly, occupational composition can be a critical omitted variable beyond the study of the 2020 pandemic. For example, occupations may be an important factor in how social interactions are defined, and if omitted, their effect may be wrongly attributed to other factors such as race, income, or education. Moreover, they can also be a key factor in how people choose where to live, both across and within cities, given the geographical sorting of specific industries or types of jobs. As a result, omitting occupations in a context of residential choice may overstate the preferences of residents for the local demographic composition.

Moving beyond occupations, we also show that length of commute and use of public transport are not significant after controlling for occupations.[8] In terms of neighborhood characteristics, we also find that the effect of a 1% increase in household size is roughly seven times larger than the effect of a 1% increase in neighborhood density for our preferred specification for April 30. This result suggests that crowding of shared spaces plays a more important role than the density of locations.

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. On the other hand, we still see significant and positive effects on positive rates for minorities, but whether these racial disparities are economically relevant can be questioned, with their magnitudes decreasing over time. 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 30th, these numbers are 0.11% and 45%, respectively. For Hispanics, a one-percentage-point increase in their population corresponds to an increase of 0.38% and 0.29% in the rate of positive tests for the same two dates. These results could be because minorities are less likely to get tested, have to be in worse condition than Whites in order to get tested, or are more likely to contract the disease because of existing comorbidities.[10]

After analyzing two isolated dates, we move on to draw a more precise picture of how the role of different characteristics has evolved over time. We do so by running regressions week by week in which we allow coefficients to change over time. This analysis reveals that as the stay-at-home order starts to be effective, the magnitude of most occupations decreases, with only one category of Health workers being fairly stable over time. 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.[11] Combining both trends, our results suggest that transmission at the workplace was key earlier in the pandemic, but intra-household contagion gradually became relatively more important as days went by.

Several policy implications and general lessons arise from our analysis. 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 not only help mitigate the effects of the disease on those who are more vulnerable but also indirectly protect the population at large. Another policy is to mitigate intra-household transmission, a channel that is particularly important during lockdowns, when household members tend to spend most of the day together. Thus, our work highlights how lockdowns may have important backfiring effects, especially in low income areas, where people tend to experience higher housing density.

We conclude that much of the disparity in the rates of positives across demographic groups can be partially explained by 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. 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 not only can 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 are from the NYC DOH. The DOH releases (almost) daily data on the cumulative count of COVID-19 cases and the total number of residents

[6] 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 that 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.

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

[8] Evidence on whether public transport played a key role has been ambiguous. Harries (2020) argues that the NYC subway was crucial for spreading the pandemic in NYC. On the other hand, Furth (2020) shows that “local infections are negatively correlated with subway use.”

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

[10] Some evidence that the first two are plausible mechanisms can be found in https://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] Sá (2020) also finds a positive relationship between the number of household members and number of cases for the UK.

who have been tested, organized by the zip code of residence.[12] 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 from the ACS using the five-year estimates from 2014–2018. We download data at the five digit zip code level to be able to directly merge it with our test data. 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.[14] A simple analysis reveals that shares 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 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 in 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.[15] 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 and 65.

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 as the dependent variable the cumulative share of positive tests by day. 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, to control for the population without access to health care. 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 testing data are released on a daily basis, we estimate separate regressions over multiple days. More concretely, we run the following regression:

$$\text{share of positive tests}_{id} = \alpha_d + \gamma_d X_i + \epsilon_{id}, \tag{1}$$

where the set of controls X_i varies for each specification according to the description above for two different days d, April 1, and April 30, where we allow these two regressions to have different coefficients.[16] The reason to do so is to study how the role of different characteristics has changed over time.[17]

The first specification – see Table 1, columns 1 and 4, which correspond to April 1, and April 30, respectively – shows the effect of the variables commonly used to explain the incidence of COVID-19 in NYC. First, we find that the use of public transport does not have a significant effect across all of our specifications. This result could be due to the lack of cross-neighborhood variation to identify this effect, as most New Yorkers use public transportation in their daily commute, or because the ACS data reflect only pre-pandemic measures.[18] Nonetheless, for this specification, commute time is a significant factor. For example, for April 30, – column 4 of Table 1 – a 10% increase on average commute time, which equals to a four-minute increase, correlates with a 0.013-point increase in the share of positive tests, approximately equivalent to a three-percentage-point increase in the share of positive tests.[19] We also find that the share of the population that is uninsured has a significant positive coefficient for most of our sample. For example, for April 30, – see column 4 of Table 1 – we find that a one-percentage-point increase in the share of the uninsured population is correlated with a

[12] Although tests were conducted in both private clinics and city-run testing sites during our study period, most of the tests were processed by Quest Diagnostics, BioReference, or LabCorp, three of the biggest commercial labs. We have not found any reports of differences in testing procedures across these labs that may affect the probabilities of type I or type II errors. Moreover, clinics in the same network usually send their tests to the same lab, but we do not find any geographical sorting of how tests are distributed across labs. For example, all CityMD locations send their tests to Quest Diagnostics, but CityMD clinics are not concentrated in particular areas of NYC.

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

[14] Unfortunately, data on workplace location by residential zip code are not available from the ACS.

[15] 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 observe only 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 leads to non-significant estimates or results that were hard to reconcile with observational evidence.

[16] In all of our regressions, we compute Conley standard errors allowing for spatial heteroskedasticity in a radius of 2 km computed using code provided by Hsiang (2010).

[17] We focus on April to make sure that the occupational shares provided by the ACS are as precise as possible. For later dates, we expect occupation shares in this dataset to differ dramatically from the actual ones, given the large economic shock that NYC experienced, especially after May. These large changes make results on occupations after April harder to interpret owing to the presence of measurement error, which we believe increases over time. On the other hand, we do not expect such dramatic changes in average demographics across zip codes, as people tend to change their place of residence very infrequently, even during recessions. Despite our concerns, we have extended this analysis to May 14, and May 27, (see Table D1 in the Online Appendix), finding similar trends for most of our covariates. Additionally, we have also plotted all daily coefficients for the entire months of April and May, finding that the most dramatic changes happen during April, while the patterns in May present a more stable profile for most of our controls. These results can be found in Section C of the Online Appendix.

[18] Even though the lockdown was in effect on March 22, there is extensive evidence that people started adjusting their behavior days, and even weeks before that date (Goolsbee and Syverson, 2020; Almagro et al., 2020). This result is in contrast with Glaeser et al. (2020), who find that subway use – captured by daily turnstile data, which we believe is a better measure – corresponds to a higher spread of the virus.

[19] The average rate of positive tests on April 30, was 45%.

Table 1
Regressions of Rate of Positive Tests on Occupations and Demographics.

Dependent Variable:	Daily Cumulative Rate of Positive Tests up to Date											
	April 1						April 30					
	(1) Neighborhood Controls		(2) + Occup.		(3) + Dem. & Borough FE		(4) Neighborhood Controls		(5) + Occup.		(6) + Dem. & Borough FE	
Log Density	-0.008	(0.007)	0.020**	(0.009)	0.029***	(0.009)	-0.011**	(0.005)	0.007	(0.005)	0.006	(0.005)
% Public Transport	-0.001	(0.065)	0.003	(0.049)	0.000	(0.063)	0.032	(0.040)	0.001	(0.028)	0.000	(0.029)
Log Commuting Time	0.137***	(0.017)	0.014	(0.028)	-0.030	(0.037)	0.126***	(0.011)	0.077***	(0.015)	0.008	(0.023)
% Uninsured	0.982***	(0.210)	0.327	(0.240)	-0.056	(0.218)	0.963***	(0.116)	0.417***	(0.109)	0.258***	(0.096)
% Essential - Professional			0.219	(0.183)	0.770***	(0.205)			-0.039	(0.084)	-0.046	(0.120)
% Non ess. - Professional			0.332**	(0.143)	0.299*	(0.159)			0.105	(0.103)	-0.095	(0.095)
% Science fields			-2.932***	(1.116)	-1.751*	(1.009)			-0.689	(0.743)	-0.914	(0.605)
% Law and related			-1.222**	(0.529)	-2.243***	(0.424)			-0.968***	(0.304)	-1.043***	(0.220)
% Health practitioners			-0.063	(0.407)	0.017	(0.343)			0.155	(0.298)	0.058	(0.238)
% Other health			0.872***	(0.288)	0.166	(0.302)			0.489**	(0.207)	0.233	(0.195)
% Firefighting			1.273*	(0.659)	0.910	(0.625)			0.355	(0.437)	-0.142	(0.365)
% Law enforcement			0.019	(0.923)	0.548	(0.914)			-1.733***	(0.623)	-0.687	(0.471)
% Essential - Service			-0.073	(0.396)	0.036	(0.380)			0.231	(0.191)	-0.041	(0.194)
% Non ess. - Service			0.560	(0.626)	1.117**	(0.525)			-0.284	(0.379)	-0.109	(0.301)
% Ind. and Construction			0.796**	(0.387)	0.750**	(0.348)			0.033	(0.216)	-0.118	(0.186)
% Essential - Technical			-1.472	(0.909)	-1.122	(0.891)			0.182	(0.552)	-0.691	(0.497)
% Transportation			1.918***	(0.518)	1.164**	(0.508)			1.042***	(0.296)	0.406*	(0.234)
Log Income					0.010	(0.030)					0.015	(0.020)
Share ≥20, ≤40					-0.414**	(0.187)					0.135	(0.103)
Share ≥40, ≤60					-0.592***	(0.217)					-0.040	(0.118)
Share ≥60					-0.199	(0.177)					0.416***	(0.104)
Share Male					0.357	(0.235)					0.125	(0.125)
Log Household Size					0.048	(0.066)					0.102***	(0.032)
% Black					0.180***	(0.038)					0.060***	(0.019)
% Hispanic					0.164***	(0.057)					0.131***	(0.035)
% Asian					0.100**	(0.049)					-0.002	(0.032)
Bronx					0.009	(0.019)					-0.019*	(0.011)
Brooklyn					0.069***	(0.018)					0.036***	(0.012)
Queens					0.082***	(0.024)					0.035***	(0.013)
Staten Island					0.063**	(0.030)					-0.017	(0.016)
Observations	174		174		174		174		174		174	
R^2	0.520		0.660		0.761		0.711		0.810		0.888	

Spatial HAC (2km) standard errors in parentheses
$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01, ^{***}p < 0.01$

1.8-percentage-point (0.963*0.01/0.45) increase in the share of positive tests.[20]

In specification (2) – columns 2 and 5 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 portion of the working-age population that is 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 next section.

Perhaps surprisingly, under this specification, neither commute time nor the share of the population using public transport has 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 – see columns 3 and 6 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 30, – column 6 in Table 1 – a one-percentage-point increase in the share of Blacks and Hispanics leads respectively to a 0.13% and 0.29% increase in the rate of positives, an effect that is economically small. A plausible explanation is that 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). Observe that in this specification, the coefficient on the share of population that is uninsured remains significant and positive, although smaller compared with the previous results, even after including income as a control. This result could be because uninsured patients are more likely to suffer from comorbidities or are less likely to get tested for fear of medical charges and therefore submit a test only when experiencing acute symptoms. 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

[20] It is worth noting that our coefficients may suffer from an attenuation bias, as the ACS data is constructed with a 5% sample, and thus, may be very noisy. In the univariate regression model, this bias is given by:

$$Bias(\hat{\beta}) = -\frac{\sigma_u^2}{\sigma_x^2 + \sigma_u^2}\beta,$$

where $\sigma_x^2$and $\sigma_u^2$are the variances of the regressor $x$and the measurement error u,respectively. In Section B1 of the Online Appendix, we quantify this bias by estimating $\sigma_u^2$with data on the Margin of Error provided by the ACS. We find that our estimated coefficients have an average downward bias of 8.6% but that the bias-corrected coefficients are not statistically different from the ones reported in the main text.

Table 2
Regressions of Rate of Positive Tests on Occupations and Demographics (Days Pooled in Given Week).

Dependent Variable:	Daily Cumulative Rate of Positive Tests up to Date									
	(1) April 1, - 7		(2) April 8, - 14		(3) April 15, - 21		(4) April 22, - 28		(5) April 29, - May 5,	
Log Density	0.024***	(0.003)	0.016***	(0.002)	0.013***	(0.002)	0.010***	(0.002)	0.005***	(0.002)
% Public Transport	-0.028	(0.028)	-0.017	(0.018)	-0.030*	(0.017)	-0.023	(0.017)	-0.002	(0.015)
Log Commuting Time	-0.028	(0.020)	0.008	(0.012)	0.029**	(0.012)	0.019	(0.013)	0.009	(0.012)
% Uninsured	0.028	(0.082)	0.224***	(0.057)	0.315***	(0.046)	0.295***	(0.050)	0.249***	(0.041)
% Essential - Professional	0.628***	(0.080)	0.387***	(0.053)	0.179***	(0.053)	0.038	(0.053)	-0.086	(0.054)
% Non ess. - Professional	0.265***	(0.087)	0.187***	(0.050)	0.117**	(0.054)	0.034	(0.059)	-0.109**	(0.052)
% Science fields	-1.519***	(0.462)	-1.401***	(0.258)	-1.169***	(0.253)	-0.952***	(0.244)	-0.785***	(0.222)
% Law and related	-2.021***	(0.124)	-1.581***	(0.090)	-1.246***	(0.101)	-1.105***	(0.109)	-1.042***	(0.094)
% Health practitioners	-0.064	(0.138)	-0.016	(0.060)	-0.094	(0.090)	-0.027	(0.116)	0.085	(0.102)
% Other health	0.293**	(0.144)	0.156	(0.099)	0.207**	(0.090)	0.226**	(0.091)	0.261***	(0.077)
% Firefighting	0.836***	(0.195)	0.459***	(0.136)	0.312**	(0.149)	-0.019	(0.185)	-0.140	(0.176)
% Law enforcement	-0.184	(0.316)	-0.899***	(0.172)	-0.950***	(0.203)	-0.712***	(0.241)	-0.681***	(0.209)
% Essential - Service	0.042	(0.045)	0.121**	(0.055)	0.073	(0.071)	-0.033	(0.078)	-0.061	(0.069)
% Non ess. - Service	0.811***	(0.189)	0.460***	(0.159)	0.247*	(0.136)	0.025	(0.132)	-0.173	(0.113)
% Ind. and Construction	0.436***	(0.158)	0.038	(0.079)	-0.110	(0.083)	-0.144	(0.090)	-0.091	(0.085)
% Essential - Technical	-1.464***	(0.385)	-0.749***	(0.204)	-0.597***	(0.219)	-0.712***	(0.262)	-0.698***	(0.265)
% Transportation	0.959***	(0.098)	0.900***	(0.075)	0.722***	(0.084)	0.539***	(0.087)	0.351***	(0.089)
Log Income	-0.008	(0.016)	-0.021***	(0.007)	-0.016**	(0.007)	-0.005	(0.009)	0.019**	(0.008)
Share ≥ 20, ≤ 40	-0.218**	(0.090)	-0.139***	(0.049)	-0.077*	(0.047)	0.054	(0.053)	0.139***	(0.050)
Share ≥ 40, ≤ 60	-0.266**	(0.114)	-0.191***	(0.059)	-0.177***	(0.044)	-0.078*	(0.041)	-0.006	(0.043)
Share ≥ 60	0.012	(0.079)	0.174***	(0.046)	0.211***	(0.043)	0.331***	(0.049)	0.396***	(0.048)
Share Male	0.311***	(0.071)	0.377***	(0.040)	0.358***	(0.046)	0.251***	(0.053)	0.115**	(0.052)
Log Household Size	0.106***	(0.034)	0.119***	(0.019)	0.100***	(0.018)	0.105***	(0.019)	0.090***	(0.019)
% Black	0.152***	(0.020)	0.118***	(0.011)	0.086***	(0.011)	0.067***	(0.011)	0.058***	(0.011)
% Hispanic	0.167***	(0.025)	0.140***	(0.014)	0.129***	(0.016)	0.131***	(0.018)	0.129***	(0.016)
% Asian	0.049**	(0.025)	0.001	(0.011)	-0.010	(0.015)	-0.010	(0.016)	0.003	(0.015)
Bronx	0.006	(0.007)	-0.017***	(0.004)	-0.034***	(0.005)	-0.031***	(0.007)	-0.017***	(0.006)
Brooklyn	0.073***	(0.008)	0.050***	(0.004)	0.035***	(0.006)	0.033***	(0.009)	0.033***	(0.007)
Queens	0.083***	(0.008)	0.058***	(0.004)	0.031***	(0.006)	0.027***	(0.008)	0.035***	(0.006)
Staten Island	0.053***	(0.011)	0.001	(0.007)	-0.031***	(0.009)	-0.032***	(0.010)	-0.016*	(0.009)
Observations	870		1217		1218		1218		1218	
R^2	0.810		0.889		0.885		0.852		0.852	

Spatial HAC (2km) standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$, $^{***}p < 0.01$

37% increase, corresponds with an 8.4% (0.37*0.102/0.45) 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, rather than density, may be a more important factor in explaining the spread of COVID-19.

3.2. Time trends

Motivated by the difference in coefficients between April 1, and April 30, in this section we present a time-varying analysis to provide insights on how different factors interact with the evolution of the pandemic as well as the health policies in place. Because we expect a smooth evolution of coefficients, we pool together all days in a given week to increase our sample size. Our specification includes a day fixed effect, which should control for common factors across all zip codes, and coefficient-specific time trends with coefficients that change week by week. Our main regression equation for this section is

$$\text{share of positive tests}_{iw(t)} = \alpha_t + \gamma_{w(t)} X_i + \epsilon_{it}, \tag{2}$$

where α_t is a day fixed effect, X_i is the vector of neighborhood characteristics including commuting patterns, share of occupations, demographics, and borough fixed effects, t corresponds to date, and coefficients are allowed to vary across different weeks as indexed by $w(t)$.

The results for the first five weeks of our data, covering April 1, to May 5, are given in Table 2.[21] While many more coefficients are significant for this specification, we still find that the coefficients for the use of public transport and commuting time are not statistically significant.

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. However, they eventually decrease, with neither 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. The opposite happens in Science Fields and Law occupations: they are negatively correlated with COVID-19 incidence before the shelter order, but their effect trends toward 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 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.[22] Overall, for all occupations we find coefficients

[21] In Section D of the Online Appendix, we perform a similar analysis for the next five weeks until June 9.

[22] A plausible explanation for this difference could be the partnership between the NYPD and health care groups to provide free testing to its members. See this link for more information. Furthermore, the NYPD provided additional work flexibility for members with

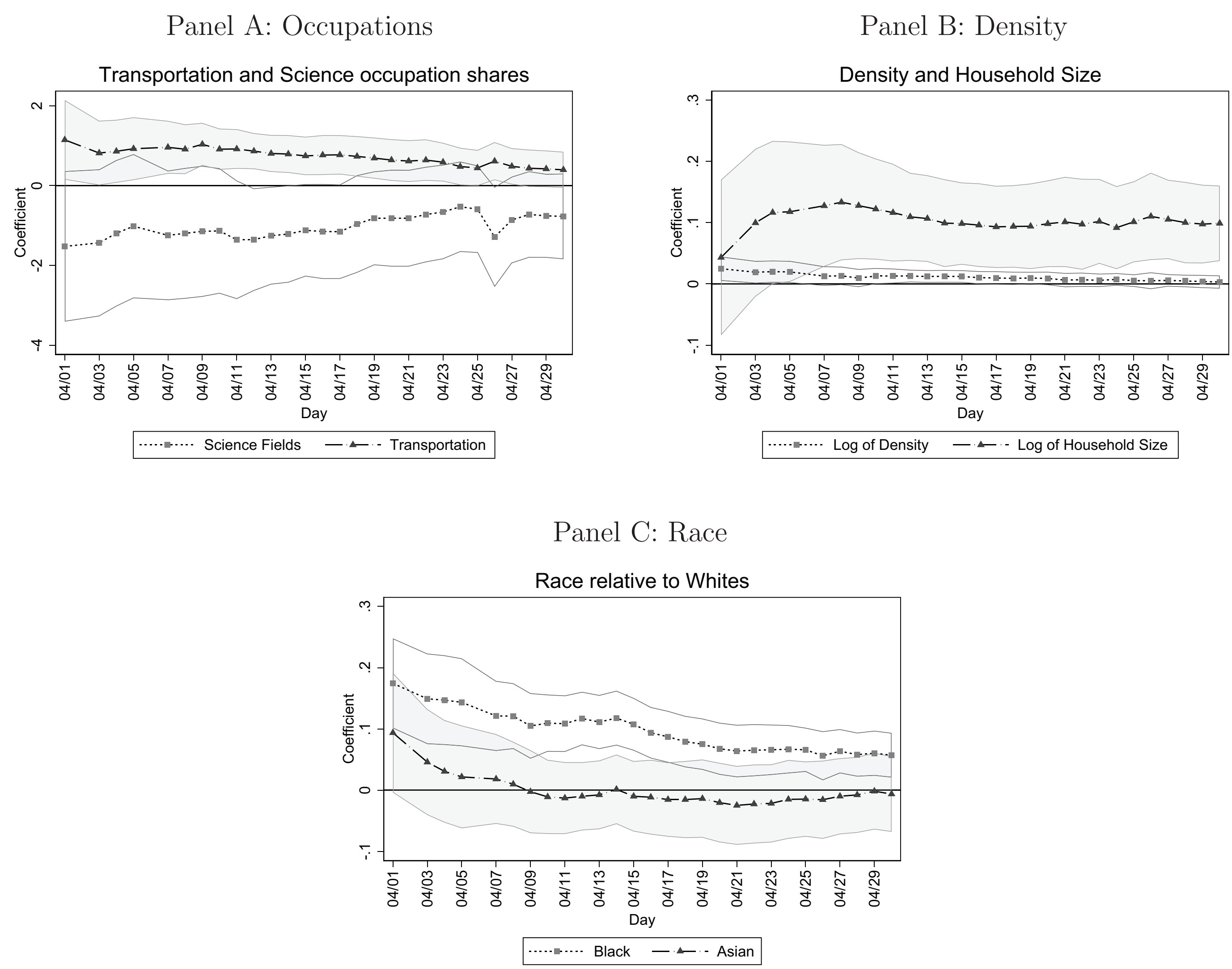

Fig. 2. Daily Evolution of Coefficients.

that decrease in magnitude over time, except for Other Health, which is positive, significant, and fairly stable across weeks. This pattern is consistent with exposure to the disease at the workplace location playing a more important role at the beginning of the pandemic with the introduction of the stay-at-home in late March.

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 out-of-pocket costs related to COVID-19, 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.

One important difference compared with the daily regressions is that Log Income becomes significant for the weeks of April 8, – 15, April 15, – 21, and April 29, – May 5. Its coefficient is negative at the beginning of the pandemic and becomes positive during only this last week.[23] However, these coefficients do not seem economically relevant. For example for the second week in April, a 10% increase in income leads to a 0.0021-point decrease, which corresponds to a decrease of 0.38% in the percentage of positive tests.[24]

For the age composition, we observe a patter similar to the one before: the presence of young residents decreases the rate of positive results, while having older residents has a positive and significant effect with a magnitude that increases over time.

Another time pattern in line with previous findings 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 with White citizens (Borjas, 2020). See bottom panel of Fig. 2 for the evolution of the coefficients of racial composition.

Overall, most of the coefficients decrease their magnitude over time, but they do so at different rates. An interesting result is that we observe only a slight drop for Log of Household Size compared with other demographics. For example, when we compare the first week in column 1 with the last week in column 5 of Table 2, we observe a 63% drop that is statistically significant for the share of males and only a 15% reduction that is not statistically significant for the Log of Household Size.

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. More information on this can be read here.

[23] As a matter of fact, Log Income is never significant for any subsequent week. See Table D2 in the Online Appendix.

[24] For the week of April 8, – 15 the average percentage of positive tests was 55%.

Moreover, these coefficients also imply that the effect associated with a 1% increase in Household Size is of orders of magnitude larger than a 1% increase in neighborhood density. One conclusion that we may draw is that while shelter-in-place policies are useful at mitigating contagion in public spaces or in workplaces, they may not have been as useful for preventing intra-household contagion. Putting all of these trends together, our results suggest that the relative importance of household size increases over time compared with the rest of the other covariates.[25,26]

Finally, to illustrate the gradual evolution of some variables for all days in April, we plot the coefficients for daily regressions:

$$\text{share of positive tests}_{it} = \alpha_t + \gamma_t X_i + \epsilon_{it}.$$

In Fig. 2, we plot the estimated coefficients, γ_t,for different variables in $X_i$and their confidence interval at the 95% level from April 1, to April 30.[27] 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 Log of Household Size. Finally, for the bottom panel, we plot the coefficients for the share of population that is Black as well as the share of Asians. Again, we observe how the importance of most covariates decreases over time, while the magnitude for Log of Household Size stays relatively more stable. For the evolution of all variables, see Section C of the Online Appendix.

4. Conclusions

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 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 includes only 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 find that after including occupation controls, they fail to significantly explain variation in the 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 and 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 and income 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. These results suggest that the occupation mechanism can explain to a greater extent the disparities along those demographics observed in the data.

We draw several policy implications motivated by our results. First, our results suggest that policy makers 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 that 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 related to COVID-19. Finally, we provide evidence suggesting 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.

Finally, we can also draw from our results important lessons that extend beyond the study of COVID-19. Namely, occupations may have been an important factor omitted in other areas of study, such as the nature of social interactions or the geographical sorting of workers. First, occupations may be an important mechanism the shaping of social interactions, therefore their effects can be wrongly attributed to demographic characteristics if occupations are omitted, as certain occupations are concentrated among specific socioeconomic groups. Second, because commuting is costly and industries and jobs tend to be concentrated in specific areas, omitting occupations in studies of residential choice may overstate the preferences of residents for the local demographic composition. Thus, given the sorting of jobs across socioeconomic groups, our study highlights that, in many contexts, omitting occupations may overstate the effects attributed to demographics.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jue.2020.103293.

CRediT authorship contribution statement

Milena Almagro: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Writing - review & editing. **Angelo Orane-Hutchinson:** Conceptualization, Methodology, Software, Data curation, Writing - original draft, Writing - review & editing.

References

Almagro, M., Coven, J., Gupta, A., Orane-Hutchinson, A., 2020. Racial Disparities in Frontline Workers and Housing Crowding during COVID-19: Evidence from Geolocation Data. Working Paper. Federal Reserve Bank of Minneapolis, Opportunity and Inclusive Growth Institute.
Azzimonti, M., Fogli, A., Perri, F., Ponder, M., 2020. Pandemic Control in ECON-EPI Networks. Working Paper. National Bureau of Economic Research.
Babus, A., Das, S., Lee, S., 2020. The optimal allocation of covid-19 vaccines. medRxiv.
Barbieri, T., Basso, G., Scicchitano, S., 2020. Italian workers at risk during the Covid-19 epidemic. Working Paper. SSRN.
Borjas, G.J., 2020. Demographic determinants of testing incidence and covid-19 infections in new york city neighbourhoods. COVID Econ. Vetted Real-Time Pap. (3) 12–39.
Coven, J., Gupta, A., 2020. Disparities in mobility responses to covid-19. Unpublished Manuscript. https://static1.squarespace.com/static/56086d00e4b0fb7874bc2d42/t/5ebf201183c6f016ca3abd91/1589583893816/DemographicCovid.pdf.

[25] An exception is the share of the population above the age of 60.
[26] Similar trends persist in May. See Table D2 in the Online Appendix.
[27] Not surprisingly, we have wider confidence intervals due to the reduction in the number of observations compared with specification 2.

Furceri, D., Loungani, P., Ostry, J., Pizzuto, P., 2020. Will covid-19 affect inequality? evidence from past pandemics. COVID Econ. Vetted Real-Time Pap. (12) 138–157.

Furth, S., 2020. Automobiles Seeded the Massive Coronavirus Epidemic in New York City. Onlinearticle at MarketUrbanism.com. Available at https://marketurbanism.com/2020/04/19/automobiles-seeded-the-massive-coronavirus-epidemic-in-new-york-city/.

Glaeser, E.L., Gorback, C.S., Redding, S.J., 2020. How Much does COVID-19 Increase with Mobility? Evidence from New York and Four Other U.S. Cities. Working Paper. National Bureau of Economic Research doi:10.3386/w27519.

Goolsbee, A., Syverson, C., 2020. Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020. Working Paper. National Bureau of Economic Research doi:10.3386/w27432.

Harries, J.E., 2020. The subways seeded the massive coronavirus epidemic in New York City. Working Paper. National Bureau of Economic Research.

Hsiang, S.M., 2010. Temperatures and cyclones strongly associated with economic production in the caribbean and central america. Proceed. Natl. Acad. Sci. 107 (35), 15367–15372.

Kaplan, G., Moll, B., Violante, G.L., 2020. The Great Lockdown and the Big Stimulus: Tracing the Pandemic Possibility Frontier for the United States. Working Papers. National Bureau of Economic Research.

Leibovici, F., Santacreu, A. M., Famiglietti, M., 2020. Social distancing and contact-intensive occupations. Technical report, Federal Reserve bank of St Louis, On the Economy. Available at https://www.stlouisfed.org/on-the-economy/2020/march/social-distancing-contact-intensive-occupations.

Michaels, G., Rauch, F., Redding, S.J., 2019. Task specialization in us cities from 1880 to 2000. J Eur Econ Assoc 17 (3), 754–798.

Sá, F., 2020. Socioeconomic Determinants of COVID-19 Infections and Mortality: Evidence from England and Wales. Policy Paper. Institute for Labor Economics.

Schmitt-Grohé, S., Teoh, K., Uribe, M., 2020. Covid-19: testing inequality in new york city. Covid Econ. Vetted Real-Time Pap. (27) 27–43.