The differential impact of COVID-19 across demographic groups: Evidence from NYC*

Milena Almagro[†] Angelo Orane-Hutchinson[‡]

April 11, 2020

Updated frequently. Click here for most updated version.

Abstract

In this paper we show that occupations are a key channel for explaining the differential impact of the COVID-19 outbreak across locations and demographic groups within a city. To do so, we estimate several models controlling for demographics, commuting controls, access to healthcare, and finally, occupations. We show that after controlling for occupations, commuting patterns and healthcare controls are not significant. Our results also indicate that other demographics such as gender, income, and age become not significant, suggesting the effects across demographics groups can be partially explained by correlation between demographics and occupations. However, racial effects still persist suggesting that there are other important channels explaining racial disparities. We also show that the effect of occupations on positive tests are heterogeneous and arguably related to their rates of human interaction. We also include the number of tests per capita to control for selection on testing. Our results suggest that selection is significant selection of testing but that it becomes less relevant over time as tests become more accessible. This result emphasizes the importance of widespread testing for identifying the relevant channels of the demographic differences in COVID-19 exposure.

*We thank Fernando Cirelli for useful comments. Any errors or omissions are our own.

[†]Department of Economics, New York University. Email: m.almagro@nyu.edu

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

1 Introduction

A priori, COVID-19 is a disease that does not discriminate across different demographic groups or locations. In practice, not only are there differences across locations but there is extensive evidence that some demographic groups are severely more affected than others. For example, in the US African Americans are being hit the hardest by the pandemic, which creates an additional source of worry to this already vulnerable socio-economic group.[1] Moreover, the total cost of medical expenses to treat COVID-19 can range from 10k to 21k USD, with out-of-pocket costs hovering around 1.4k USD (Rae et al., 2020), a medical expense that not all households may be able to afford. Similarly in relation to gender, Alon et al. (2020) find that the effect of COVID-19 affects male and female workers in different ways. They argue there are two channels for this disparity. First, the loss of jobs and the recession component of the pandemic affect male and female workers differently. Second, through social norms, as childcare is a task predominantly carried out by women within a household.

Moreover, it has been extensively documented that the number of cases and death rates vary greatly across locations, where much of this variation is explained by characteristics of the location, such as population density or weather. However, we still see great disparities across demographic groups even *within the same location*. For example, on April 6 2020, the Department of Health of New York City (DOH) released race-specific data that revealed that compared to Whites, Blacks and Hispanics have a 47% and a 36% higher crude death rate (number of deaths per 100,000 Population), respectively. On the other hand, Asians have a 42% lower crude death rate.[2] These patterns repeat themselves in other areas of the US, however the disparities across racial groups are less pronounced compared to NYC.[3]

Although the fact that more vulnerable groups and minorities are being hit harder by the pandemic has called the attention of many economists and policy makers, it is still unclear which are the channels that lead to such disparities. To shed some light on the mechanisms, we use data on the share of positives across neighborhoods for New York City provided by DOH, which releases (almost) daily updates.[4] We have collected these data updates across different days, which not only allows us to explain cross-sectional results, but also the time-varying importance of different channels. We combine this test data with demographic data provided by the American Community Survey (ACS), also at the zip code level.

[1]For a recent article see `https://nyti.ms/2UUl5je`

[2]For the latest data release visit: `www1.nyc.gov/site/doh/covid/covid-19-data.page`

[3]For a summary see: `https://nyti.ms/2wu9jCK`

[4]Unfortunately DOH has not released data on death cases across neighborhoods.

1.1 Patterns in NYC

The differences metioned above become starker when looking at maps, as shown in Figure 1. From simple inspection, zip codes with the highest rates of positives belong to the boroughs of Bronx, Brooklyn, and Queens. These boroughs are also home of the majority of Blacks and Hispanics living in NYC.[5] Moreover, the zip codes with the lowest income levels are also located in those three boroughs. We plot a simple correlation between the share of positives and demographics. We see that share of Blacks and Hispanics are positively correlated with rate of positives, while a flat relationship for share of Asians as shown in Figure 3. We observe a negative relationship for income, slightly positive for share of males, and positive and significant for household size, as shown in Figure 4. Finally, for age composition, we see an increasing relation with the share of the population below 20, an increasing relationship for those between 20 and 40 and above 60, and mainly flat for those between 40 and 60 as shown in Figure 4. Along the same lines, Borjas (2020) shows that much of the variation of the incidence of testing across neighborhoods in NYC can be explained by demographics.

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

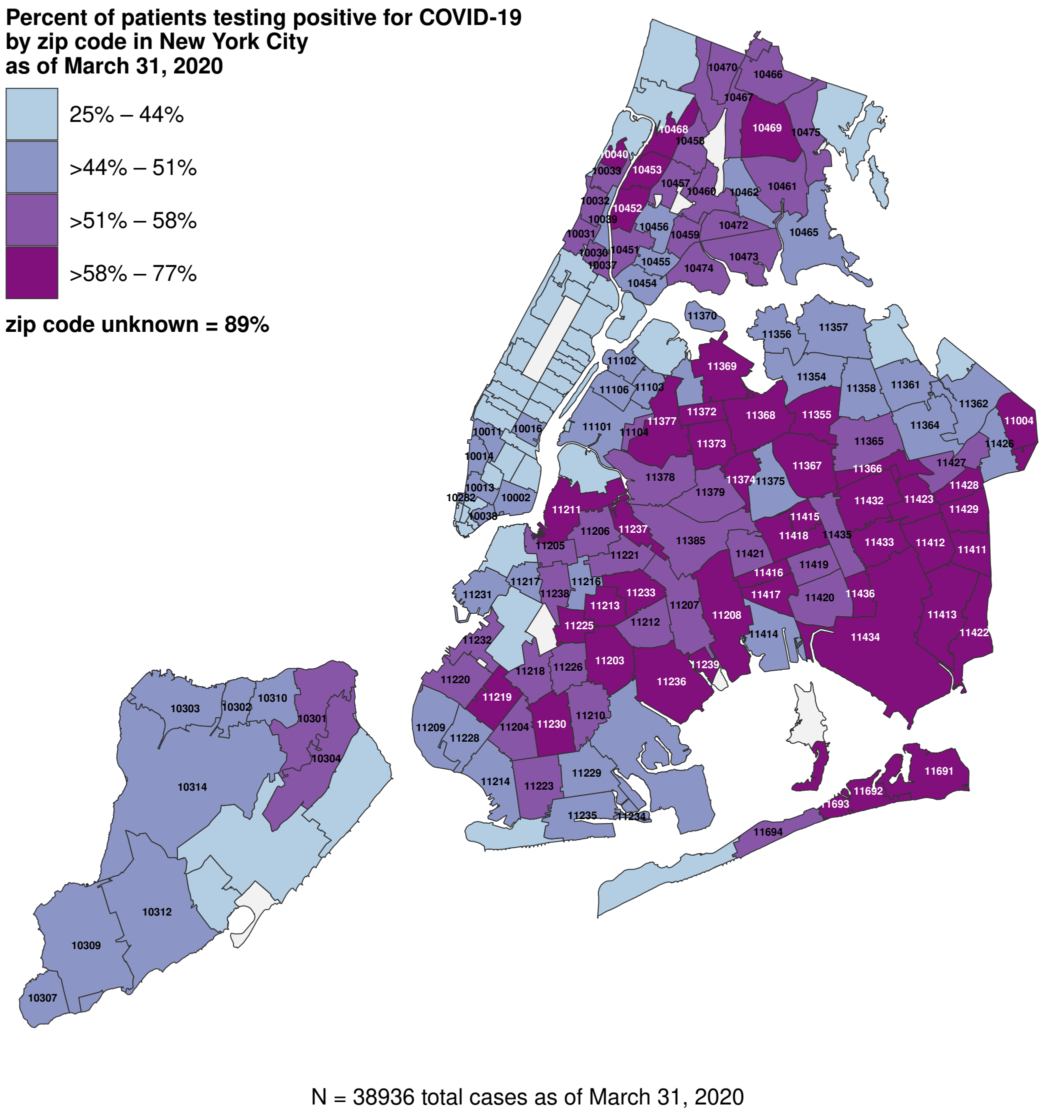

Figure 1: Map of rate of positives by zip code as of March 31, 2020.

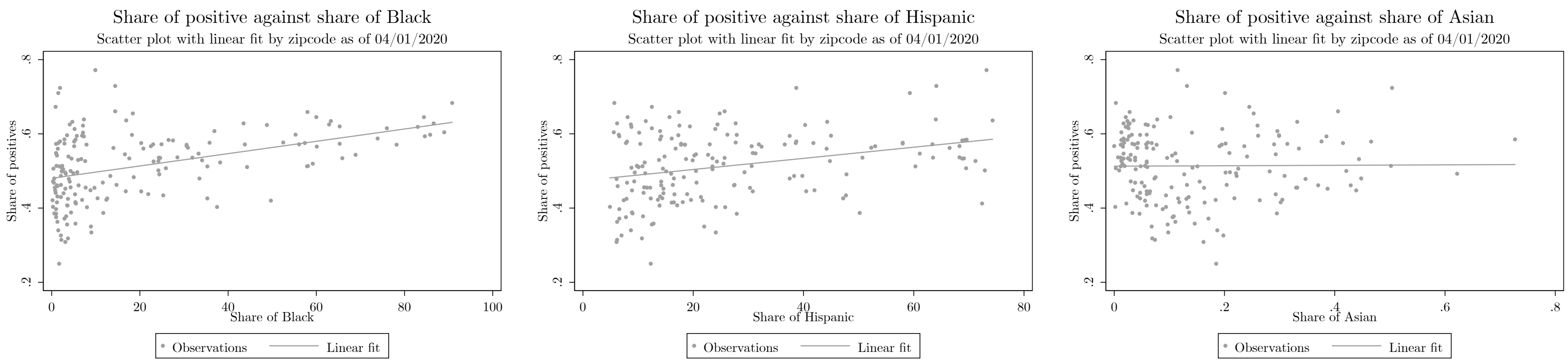

Figure 2: Share of positives against income, share of males, and household size by zipcode.

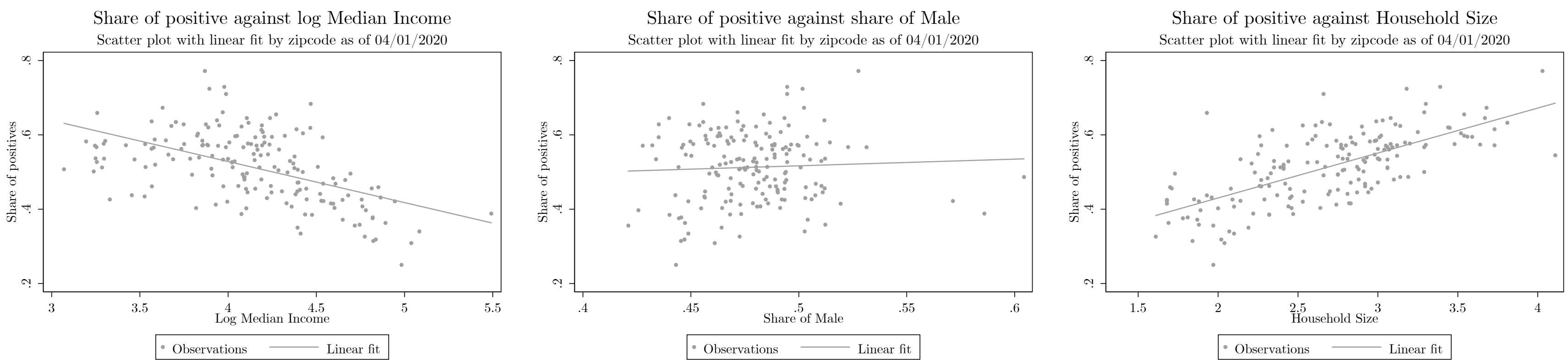

Figure 3: Share of positives against income, share of males, and household size by zipcode.

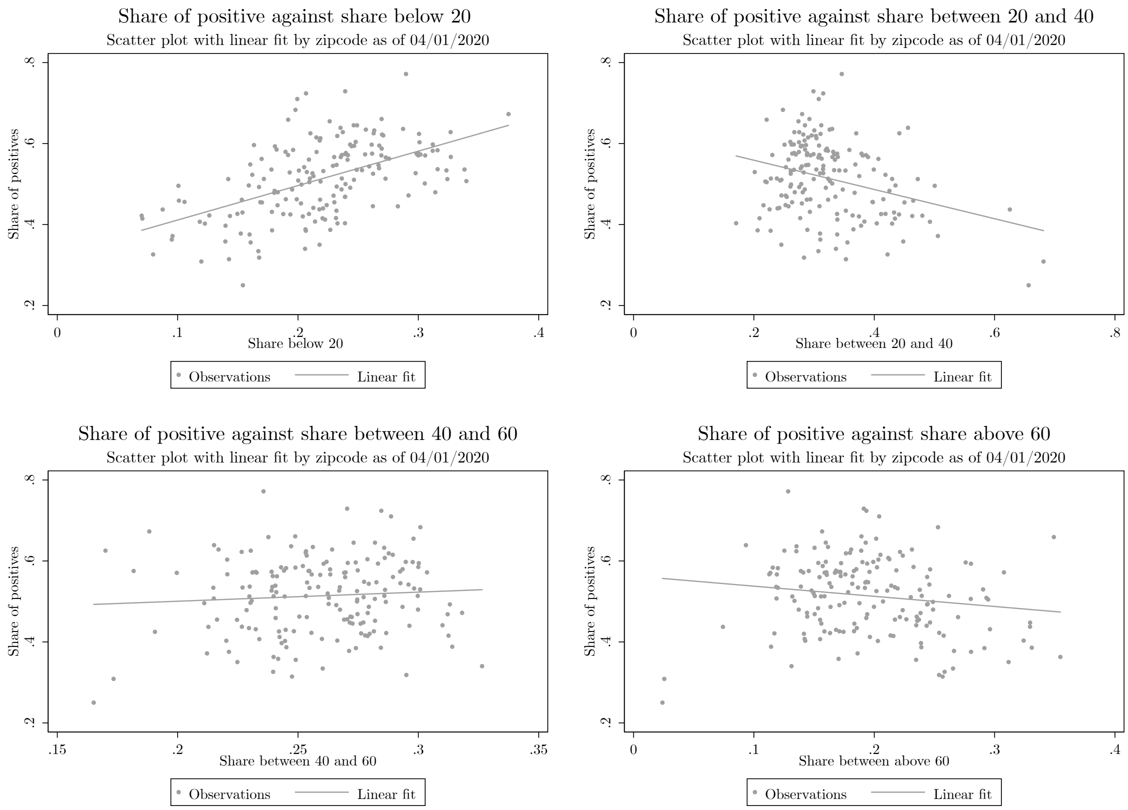

Figure 4: Share of positives against age composition by zipcode.

We proceed to estimate different model specifications to unravel the relationship between demographics and rate of positives described above. In all of our specifications we also include the number of tests per capita to control for a potential selection on testing, which seems to be an important concern as pointed out by Borjas (2020). As the author argues if there is a selection in who us being tested in NYC, we should expect a selection on who has tested positive. We find that when the number of tests per capita increases, the share of positives also increases. This result suggests that people in worse conditions are more likely to be tested, which lines up with how testing has been handled in New York City, where only those who are admitted into a hospital have access to tests. Surprisingly, we find that this selection on testing becomes less important as days go by, which can be explained by the increasing availability of tests in New York City over time.

In our first set of regressions we include controls for age, racial composition and income. The estimation results from this model are similar to what we find in our graphical analysis, with income being negatively correlated with rate of positives and an increasing share of positives with the share of Black, Hispanic, and Asian population. In our second set of results we also include commuting patterns (commuting time, share of population using public transportation, and share of working from home). The commuting choice is in principle an important channel as it can lead to more human contact in small and confined places such as the NYC subway. In a similar spirit, we also include the population density to control for human exposure happening during activities taking place near home. We find that longer commutes as well as denser zip codes correlate with higher share of testing positive. Another important channel could be access to tests through medical insurance. To control for it we also include the share of the population with private insurance as well as the share of the population that are uninsured. We see that areas with a higher share of private coverage also have lower rates of positives and a positive effect for areas with a higher share of uninsured population. Notably, after controlling for insurance, we observe that zip codes with higher incomes also have a higher rate of positive, contrary to our initial findings. However, this result makes sense as income was absorbing the effect of access to medical care in our first two specifications.

Finally, in our last specification (and our most preferred one) we include the shares of different occupations. A priori, different occupations have different rates of exposure to human contact and therefore they could be an important channel.[6] Our results confirm this intuition, as we find that occupations that require a higher human interaction are also related with higher positive rates, such as occupations in services. Similarly, occupations that require less human exposure move in the opposite direction reducing the rate of positives, such as jobs related to science. We also find in our final specification that commuting patterns do not have a significant effect in the rate of positives after controlling for occupations. More importantly, we also find that the differential effects on racial groups become smaller after controlling for occupations.

We conclude that much of the racial and income disparities in the number of cases and deaths can be explained by different demographic groups being more or less representative across different occupations. In particular, one key channel that explain these differences is there are different exposures to human contact across different jobs. If these channels are important, our main takeaway is that the design of policy on testing and vaccination should target these more sensitive groups, not only considering their risk of exposure but also the potential spillovers that it may have on the rest of the population.

2 Data

Our source of incidence rates of COVID-19 and the number of tests performed is the NYC Department of Health and Mental Hygiene (DOH) data release. The DOH releases the cumulative count of COVID-19 cases and the total number of residents that have been tested, divided zip code of residence. This allows us to construct the tests per capita variable we use to control for selection in testing. Demographic

[6] A recent paper by Barbieri et al. (2020) shows evidence of this mechanism for workers in Italy.

Table 1: Occupations classified as essential

Occupation	**Essential**
Business and Financial Operations	✓
Management	
Computer and Mathematical	
Architecture and Engineering	
Life, Physical, and Social Science	
Community and Social Services	
Legal	
Education, Training, and Library	
Arts, Design, Entertainment, Sports, and Media	
Healthcare Practitioners and Technicians	✓
Healthcare Support	✓
Firefighting and prevention	✓
Law Enforcement	✓
Food Preparation and Serving	✓
Building and Grounds Cleaning and Maintenance	✓
Personal Care and Service	
Sales and Related	
Administrative and Office Support	
Farming, Fishing, and Forestry	✓
Construction and Extraction	
Installation, Maintenance, and Repair	✓
Production	
Transportation	✓
Material Moving	✓

and occupation data at the zip code level is obtained from the American Community Survey (ACS). The demographic characteristics we include are zip code median income, average age, racial breakdown, and health insurance condition. We also include commuting related variables: average commuting time to work as well as means of transportation.

We also construct the shares of working age population employed at different occupations. These are categorized according to the groups listed in column 1 of Table 1 and the ACS provides the number of workers in each one of them, by zip code of residence. In column 2 of Table 1 we classify occupations as essential or not essential, based on the general list provided in the 'PAUSE' Executive Order enacted by New York Governor Andrew Cuomo. The executive order lists which businesses are allowed to remain open. Neighborhoods with higher share of the population in essential sectors could be more exposed to COVID-19 if a relatively high share of workers are to remain working during the crisis.

Table 2: Percentage of employed who worked at home on an average day

Occupation	**Worked from Home**
Management, Business and Financial Operations	33.6
Professional and related	32.4
Service	10.0
Sales and Related	27.4
Administrative and Office Support	15.0
Farming, Fishing, and Forestry	-
Construction and Extraction	8.3
Installation, Maintenance, and Repair	-
Production	7.7
Transportation and Material Moving	7.0

Note: Table obtained from the U.S. Bureau of Labor Statistics, 2018 annual average

We argue, however, that there may be other occupation-related factors affecting COVID-19 rates aside from classification alone. Table 2 shows the percentage of employed workers in major occupation categories who worked from home on an average day, according to the Bureau of Labor Statistics. To further explore this channel, we obtain the share of employed workers who work at home for each zip code. For those not working from home, their commuting time and choice of transportation could affect exposure to COVID-19. For example, those with a longer commuting time and using public transportation could be thought to be more at risk.

Taking all of this into account, we divide the occupations into 14 categories for our empirical exercise, according to correlations between them or similarity in work environments. Table 3 shows the occupation groups that will be used in the estimating regressions.

We explore how different demographics correlate to these occupation categories. Different COVID-19 exposure across occupations would translate to a heterogeneous impact across demographics if there is sorting between. We do this by computing the correlation between each occupations share and the racial, age, and male shares by zip code, as well as median income. The results are shown in the graphs of Figure 5.

There are strong positive and negative correlations with income, according to the different occupations. In particular a higher share of workers in any of the first 5 categories (all generally high skilled) is positively correlated with median income, while the opposite happens for the latter occupations with the notable exception

Table 3: Occupation categories

Category	ACS Occupations
(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	Health practitioners
(6) Other health	Health technologists, technicians, and Healthcare Support
(7) Firefighting	Firefighting and prevention
(8) Law enforcement	Law enforcement
(9) Essential - Service	Food Preparation and Serving, Building 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	Transportation

of Law enforcement. This pattern provides additional evidence that the negative correlation between income and share of positive tests could be due to a different COVID-19 exposure that is occupation related.

The next panels plot both the correlation between share of whites and occupation shares, and the correlation between another race group and occupation shares. In our empirical analysis we will measure the effects of racial shares relative to the share of whites, so we plot the racial groups together for an easier comparison. The correlation between share of whites and occupations follows very closely the one with income, an unsurprising result. Also unsurprising is that higher shares of blacks or hispanics are correlated with the lower income occupations. There are, however, some race specific correlations that merit further discussion. Higher share of blacks is highly correlated with occupation (6), workers in the health sector not including health practitioners. Similarly, high share of hispanics is particularly correlated with share of worker in occupation (9), which includes cooks, food servers, and cleaners. It's notable that the share of Asians is not particularly correlated with most occupations, in either direction. This would suggest that the occupation mechanism may not be too determinant in the positive rates in neighborhoods with high share of Asians.

There is evidence of different rates of COVID-19 across ages, and males seem be relative more affected by the disease. The occupation mechanism could be explaining it. The next panels plot the correlation between mean age and occupation shares, and share of males and occupation shares. Zip codes with older populations are strongly correlated with shares of workers in category (2) and category (4), all non essential and professional occupations, as well as category (12), essential - technical occupations. Higher share of males is correlated with more workers in category (11) occupations.

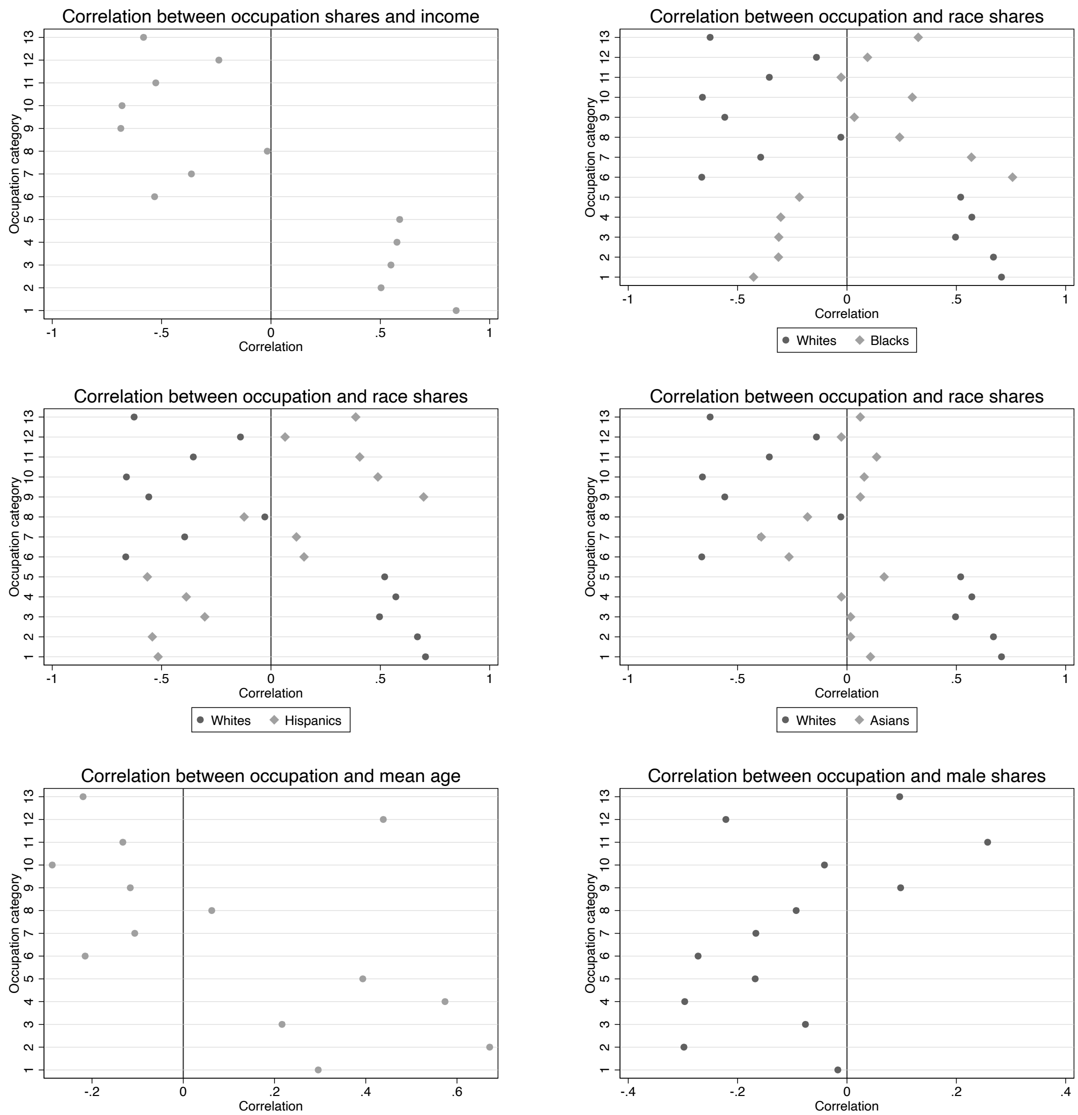

Figure 5: Correlation between shares of demographic characteristics and occupation shares

3 Results

We present the main empirical results in this section, for our 4 different specifications. The baseline includes the share of positive tests as the dependent variable and the main demographic characteristics as the independent variables: log of income, log of household size, share of males age, and race breakdown, similar to the controls included in Borjas (2020). The second model adds commuting and location characteristics by including population density, share of commuters using public transportation, and share of employed who worked from home. The third specification adds health insurance characteristics as controls, including the share of uninsured and of those with private insurance. The last specification incorporates the share of each occupation category, as defined in Table 3 and including the previous variables as controls. Exploiting the fact that we have daily data over multiple days, we estimate a separate regression for each one of them, allowing us to detect any time variation in the effects.

In all specifications we include a variable with tests per capita to control for selection on testing. This selection is important in the context of New York as testing was extremely scarce at earlier dates and only available to those sick enough to require hospitalization. One could expect that those in worse condition and showing more symptoms are also more likely to be tested positive than a random person picked from the street. Therefore those being testes in NYC were selected from a population with a higher probability of testing positive. This reasoning is precisely what the positive coefficient on tests per capita tells us: if only those who almost certainty have the disease are tested, more tests should unambiguously lead to higher rates of positive tests.[7]

For specification (1), we do confirm the observations made in the earlier sections and find a positive relationship between shares of races other than white and positive cases, even after controlling for income. As expected, the coefficient for income is negative. We also see a positive coefficient for household size. These results hold over time and as the volume of tests increases. Part of the income effect is attenuated when including commuting time and location density, as the latter have a positive effect under specification (2). Commuting time and income are negatively correlated, since neighborhoods further away from economic centers in Manhattan or Downtown Brooklyn tend to have cheaper housing and thus, lower income workers have higher commuting times. The share of public transportation users is not significant at earlier dates. A possible explanation is that within NYC roughly 85% of the population uses public transport in their daily commute so there may not be enough variation to identify that coefficient. Similarly, the share of workers working from home is

[7]Another type of selection could be driven by higher income households being able to pay for more testing. In this case, we should expect a negative sign. However, given the way testing has been managed in NYC (only testing after hospitalization at first with the introduction of free testing later on), we believe that this type of selection is unlikely.

significant. This result is explained by the fact commuting time absorbs the effects from the share of workers working from home, as those are likely to input 0 when asked for their daily commuting time. For robustness, we estimate a similar specification but excluding commuting time, obtaining a negative and significant coefficient on the share of workers working remotely.

Specification (3) adds health insurance variables, and we find that a higher share of uninsured correlates with higher rates of positive cases while the share of workers with private insurance has the opposite effect. More notably, including these insurance variables flips the effect we observed on income and now after controlling on these income has a positive effect on rate of positive cases. This implies that another mechanism of the income effect is through differences in health care access and quality. A possible explanation for the positive effect on income is that lower income households tend to spend more of their leisure time watching television, listening to the radio, or playing video-games in a computer and less time socializing as found by the American Time Use Survey.

We can see the effect of occupation categories in our main specification, (4). Note that as the selection of testing decreases, more of these occupation categories become significant. Throughout all our sample we observe that higher shares of workers in three occupation categories are associated with significantly positive results: (2) Non essential - Professional, (11) Industrial, Natural Resources and Construction, and (13) Transportation. We also see that category (10) Non essential Service is significant but only at earlier dates. While not all of these occupations remain active during the Executive Order, it could be argued that these are all relatively high exposure occupations; due to their nature they may require a higher degree of personal interaction. By contrast, we find that the share of workers in occupations related to (3) Science and (4) Law fields have a significantly negative impact on share of positives. However, as days go by and the selection on testing decreases, higher share of workers in (1) Essential - Professional and (7) Firefighting also become positive and significant, while (8) Law enforcement becomes negative and significant.

It should be noted that after including occupation variables, commuting time no longer has an effect on the share of positives. This suggests part of this effect was due to some type of workers being more or less likely to commute depending on their occupations. Something similar occurs for share of the population with private health insurance or uninsured. Also, our results suggest that age composition does not have an effect after controlling for occupations starting from April 4. The share of males and household size also become not significant in most of our specifications and only at the 10% and 5% respectively when controlling for occupations at later dates. This result indicates that the disparities observed in simple correlations can be partially explained by the different representation of demographic across occupations.

Finally, some race breakdown coefficients remain significant an positive throughout all specifications and across time, implying that there are are still some race-related characteristics aside from occupations affecting the rates of COVID-19. This type of result can be driven by a racial bias on the incidence of testing as pointed out by Borjas (2020). Another explanation could be different levels of information or attitudes across demographic groups. Unfortunately, better and richer data is required to assess whether the information and/or attitude channels are important.

Table 4: Dependent variable - Share of Positives as of April 1, 2020

	(1) Demographics		(2) + Commuting		(3) + Insurance		(4) + Occupations	
Tests per capita	4.531	(2.802)	6.685*	(3.447)	8.241***	(3.041)	9.388***	(2.318)
Log Income	-0.049**	(0.022)	-0.016	(0.025)	0.052	(0.040)	0.053	(0.048)
Share $\geq$ 20, $\leq$ 40	0.125	(0.137)	0.021	(0.134)	-0.079	(0.155)	-0.426**	(0.182)
Share $\geq$ 40, $\leq$ 60	-0.215	(0.260)	-0.359	(0.303)	-0.480	(0.306)	-0.633**	(0.246)
Share $\geq$ 60	0.212	(0.147)	0.097	(0.149)	-0.004	(0.150)	-0.356*	(0.189)
Share Male	0.281	(0.281)	0.398	(0.323)	0.223	(0.314)	0.095	(0.245)
Log Household Size	0.250***	(0.038)	0.194***	(0.059)	0.094	(0.068)	0.007	(0.075)
% Black	0.138***	(0.027)	0.120***	(0.028)	0.114***	(0.029)	0.173***	(0.038)
% Hispanic	0.021	(0.048)	0.003	(0.050)	-0.044	(0.045)	0.043	(0.050)
% Asian	0.114***	(0.044)	0.128***	(0.047)	0.075	(0.048)	0.157***	(0.049)
Log Density			0.018*	(0.010)	0.018*	(0.010)	0.025**	(0.010)
% Public Transport			0.064	(0.067)	0.006	(0.064)	0.038	(0.059)
Log Commuting Time			0.116	(0.073)	0.143**	(0.071)	0.059	(0.072)
% Working Home			-0.329	(0.420)	-0.321	(0.406)	-0.271	(0.350)
% Private Insurance					-0.178	(0.131)	-0.125	(0.128)
% Uninsured					0.567**	(0.234)	0.163	(0.241)
% Essential - Professional							0.447	(0.281)
% Non ess. - Professional							0.595***	(0.186)
% Science fields							-2.648***	(1.006)
% Law and related							-1.293*	(0.751)
% Health practitioners							-0.302	(0.484)
% Other health							-0.024	(0.420)
% Firefighting							1.036	(0.837)
% Law enforcement							-0.304	(0.922)
% Essential - Service							-0.033	(0.374)
% Non ess. - Service							0.926*	(0.548)
% Ind. and Construction							1.210***	(0.406)
% Essential - Technical							-0.896	(0.923)
% Transportation							1.371**	(0.534)
Constant	0.222	(0.169)	-0.476	(0.398)	-0.496	(0.382)	-0.274	(0.334)
Observations	174		174		174		174	
R^2	0.604		0.637		0.670		0.766	

Robust standard errors. Standard errors in parentheses

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

Table 5: Dependent variable - Share of Positives as of April 4, 2020

	(1) Demographics		(2) + Commuting		(3) + Insurance		(4) + Occupations	
Tests per capita	3.730**	(1.684)	5.566***	(2.055)	6.644***	(1.795)	7.367***	(1.335)
Log Income	-0.071***	(0.019)	-0.042*	(0.023)	0.026	(0.034)	0.039	(0.038)
Share $\geq$ 20, $\leq$ 40	0.270**	(0.127)	0.155	(0.122)	0.059	(0.136)	-0.241	(0.151)
Share $\geq$ 40, $\leq$ 60	-0.000	(0.238)	-0.106	(0.276)	-0.216	(0.273)	-0.302	(0.217)
Share $\geq$ 60	0.351**	(0.136)	0.255*	(0.135)	0.162	(0.134)	-0.124	(0.149)
Share Male	0.284	(0.265)	0.402	(0.300)	0.235	(0.284)	0.076	(0.200)
Log Household Size	0.284***	(0.037)	0.247***	(0.053)	0.150**	(0.058)	0.076	(0.063)
% Black	0.127***	(0.026)	0.105***	(0.027)	0.099***	(0.027)	0.138***	(0.035)
% Hispanic	0.019	(0.042)	-0.008	(0.043)	-0.052	(0.039)	0.035	(0.043)
% Asian	0.075*	(0.038)	0.085**	(0.041)	0.035	(0.043)	0.097**	(0.044)
Log Density			0.017*	(0.009)	0.017*	(0.009)	0.020**	(0.009)
% Public Transport			0.087	(0.063)	0.034	(0.058)	0.045	(0.052)
Log Commuting Time			0.102	(0.064)	0.123**	(0.062)	0.045	(0.064)
% Working Home			-0.146	(0.407)	-0.140	(0.389)	-0.223	(0.320)
% Private Insurance					-0.182*	(0.109)	-0.158	(0.104)
% Uninsured					0.510***	(0.181)	0.231	(0.198)
% Essential - Professional							0.327	(0.241)
% Non ess. - Professional							0.587***	(0.156)
% Science fields							-2.497***	(0.922)
% Law and related							-1.066*	(0.612)
% Health practitioners							-0.381	(0.459)
% Other health							0.157	(0.361)
% Firefighting							1.069	(0.724)
% Law enforcement							-1.052	(0.809)
% Essential - Service							-0.037	(0.307)
% Non ess. - Service							0.633	(0.504)
% Ind. and Construction							0.891**	(0.356)
% Essential - Technical							-1.279	(0.824)
% Transportation							1.062**	(0.449)
Constant	0.178	(0.149)	-0.487	(0.349)	-0.487	(0.335)	-0.251	(0.286)
Observations	174		174		174		174	
R^2	0.703		0.735		0.763		0.836	

Robust standard errors. Standard errors in parentheses

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

Table 6: Dependent variable - Share of Positives as of April 5, 2020

	(1) Demographics		(2) + Commuting		(3) + Insurance		(4) + Occupations	
Tests per capita	3.399**	(1.560)	5.320***	(1.885)	6.190***	(1.641)	6.919***	(1.267)
Log Income	-0.075***	(0.019)	-0.046**	(0.022)	0.016	(0.033)	0.036	(0.038)
Share $\geq$ 20, $\leq$ 40	0.291**	(0.123)	0.174	(0.121)	0.076	(0.131)	-0.208	(0.148)
Share $\geq$ 40, $\leq$ 60	0.021	(0.231)	-0.079	(0.270)	-0.196	(0.262)	-0.261	(0.213)
Share $\geq$ 60	0.390***	(0.137)	0.291**	(0.136)	0.201	(0.134)	-0.072	(0.146)
Share Male	0.327	(0.266)	0.441	(0.302)	0.293	(0.287)	0.108	(0.200)
Log Household Size	0.283***	(0.036)	0.247***	(0.052)	0.156***	(0.057)	0.082	(0.062)
% Black	0.126***	(0.026)	0.103***	(0.026)	0.098***	(0.026)	0.137***	(0.033)
% Hispanic	0.022	(0.041)	-0.007	(0.041)	-0.049	(0.038)	0.042	(0.042)
% Asian	0.072*	(0.039)	0.082**	(0.041)	0.034	(0.043)	0.090**	(0.044)
Log Density			0.018*	(0.010)	0.018*	(0.010)	0.019**	(0.009)
% Public Transport			0.087	(0.064)	0.036	(0.059)	0.033	(0.051)
Log Commuting Time			0.099	(0.064)	0.119*	(0.063)	0.044	(0.063)
% Working Home			-0.151	(0.404)	-0.138	(0.387)	-0.271	(0.319)
% Private Insurance					-0.160	(0.109)	-0.157	(0.102)
% Uninsured					0.511***	(0.173)	0.272	(0.198)
% Essential - Professional							0.264	(0.239)
% Non ess. - Professional							0.602***	(0.153)
% Science fields							-2.335**	(0.895)
% Law and related							-0.913	(0.601)
% Health practitioners							-0.409	(0.455)
% Other health							0.085	(0.342)
% Firefighting							1.185*	(0.713)
% Law enforcement							-1.571*	(0.840)
% Essential - Service							-0.110	(0.303)
% Non ess. - Service							0.599	(0.506)
% Ind. and Construction							0.796**	(0.352)
% Essential - Technical							-1.471*	(0.815)
% Transportation							1.166**	(0.451)
Constant	0.160	(0.147)	-0.503	(0.351)	-0.504	(0.343)	-0.255	(0.279)
Observations	174		174		174		174	
R^2	0.713		0.746		0.772		0.842	

Robust standard errors. Standard errors in parentheses

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

Table 7: Dependent variable - Share of Positives as of April 7, 2020

	(1) Demographics		(2) + Commuting		(3) + Insurance		(4) + Occupations	
Tests per capita	2.429**	(1.165)	4.133***	(1.499)	4.914***	(1.313)	5.349***	(1.002)
Log Income	-0.071***	(0.016)	-0.047**	(0.020)	0.008	(0.029)	0.021	(0.035)
Share $\geq$ 20, $\leq$ 40	0.293***	(0.112)	0.195*	(0.114)	0.107	(0.118)	-0.166	(0.128)
Share $\geq$ 40, $\leq$ 60	0.044	(0.214)	-0.028	(0.248)	-0.133	(0.240)	-0.202	(0.195)
Share $\geq$ 60	0.474***	(0.128)	0.398***	(0.125)	0.317***	(0.121)	0.051	(0.127)
Share Male	0.388	(0.246)	0.477*	(0.281)	0.344	(0.262)	0.190	(0.184)
Log Household Size	0.293***	(0.033)	0.269***	(0.048)	0.187***	(0.052)	0.125**	(0.057)
% Black	0.123***	(0.025)	0.103***	(0.024)	0.097***	(0.024)	0.139***	(0.029)
% Hispanic	0.030	(0.038)	0.001	(0.037)	-0.038	(0.034)	0.041	(0.039)
% Asian	0.087**	(0.037)	0.098**	(0.039)	0.055	(0.041)	0.097**	(0.041)
Log Density			0.014	(0.009)	0.014	(0.009)	0.015*	(0.008)
% Public Transport			0.090	(0.060)	0.047	(0.055)	0.049	(0.050)
Log Commuting Time			0.076	(0.061)	0.093	(0.059)	0.033	(0.061)
% Working Home			-0.094	(0.365)	-0.072	(0.344)	-0.209	(0.275)
% Private Insurance					-0.141	(0.091)	-0.143	(0.090)
% Uninsured					0.466***	(0.154)	0.222	(0.169)
% Essential - Professional							0.347	(0.222)
% Non ess. - Professional							0.552***	(0.151)
% Science fields							-2.429***	(0.834)
% Law and related							-1.008*	(0.535)
% Health practitioners							-0.237	(0.419)
% Other health							-0.050	(0.306)
% Firefighting							1.059*	(0.583)
% Law enforcement							-1.728**	(0.804)
% Essential - Service							-0.068	(0.273)
% Non ess. - Service							0.578	(0.462)
% Ind. and Construction							0.700**	(0.323)
% Essential - Technical							-0.915	(0.746)
% Transportation							1.154***	(0.408)
Constant	0.087	(0.135)	-0.449	(0.326)	-0.451	(0.316)	-0.240	(0.261)
Observations	174		174		174		174	
R^2	0.752		0.777		0.799		0.863	

Robust standard errors. Standard errors in parentheses

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

Table 8: Dependent variable - Share of Positives as of April 8, 2020

	(1) Demographics		(2) + Commuting		(3) + Insurance		(4) + Occupations	
Tests per capita	0.999	(1.062)	3.202**	(1.361)	3.865***	(1.234)	4.657***	(0.984)
Log Income	-0.069***	(0.017)	-0.045**	(0.021)	0.005	(0.031)	0.012	(0.036)
Share $\geq$ 20, $\leq$ 40	0.319***	(0.118)	0.214*	(0.122)	0.131	(0.127)	-0.149	(0.135)
Share $\geq$ 40, $\leq$ 60	-0.056	(0.224)	-0.105	(0.251)	-0.206	(0.245)	-0.269	(0.204)
Share $\geq$ 60	0.574***	(0.138)	0.500***	(0.138)	0.426***	(0.136)	0.156	(0.129)
Share Male	0.490*	(0.249)	0.584**	(0.282)	0.466*	(0.265)	0.292	(0.184)
Log Household Size	0.297***	(0.035)	0.282***	(0.053)	0.208***	(0.057)	0.156**	(0.061)
% Black	0.143***	(0.027)	0.121***	(0.026)	0.116***	(0.026)	0.166***	(0.032)
% Hispanic	0.045	(0.040)	0.006	(0.038)	-0.030	(0.035)	0.044	(0.041)
% Asian	0.103**	(0.041)	0.118***	(0.043)	0.079*	(0.045)	0.116**	(0.046)
Log Density			0.017*	(0.010)	0.017*	(0.010)	0.017**	(0.008)
% Public Transport			0.109*	(0.064)	0.070	(0.060)	0.071	(0.052)
Log Commuting Time			0.075	(0.066)	0.091	(0.065)	0.037	(0.066)
% Working Home			-0.021	(0.381)	0.005	(0.364)	-0.124	(0.265)
% Private Insurance					-0.125	(0.098)	-0.120	(0.095)
% Uninsured					0.436**	(0.168)	0.174	(0.185)
% Essential - Professional							0.434*	(0.223)
% Non ess. - Professional							0.551***	(0.160)
% Science fields							-2.552***	(0.869)
% Law and related							-1.044**	(0.519)
% Health practitioners							-0.254	(0.438)
% Other health							-0.175	(0.316)
% Firefighting							1.154*	(0.618)
% Law enforcement							-2.111**	(0.868)
% Essential - Service							-0.040	(0.296)
% Non ess. - Service							0.715	(0.500)
% Ind. and Construction							0.762**	(0.360)
% Essential - Technical							-1.067	(0.758)
% Transportation							1.278***	(0.442)
Constant	0.029	(0.135)	-0.568*	(0.340)	-0.574*	(0.334)	-0.376	(0.272)
Observations	174		174		174		174	
R^2	0.735		0.768		0.785		0.855	

Robust standard errors. Standard errors in parentheses

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

Table 9: Dependent variable - Share of Positives as of April 9, 2020

	(1)		(2)		(3)		(4)	
	Demographics		+ Commuting		+ Insurance		+ Occupations	
Tests per capita	0.651	(0.870)	2.334**	(1.107)	2.425**	(1.069)	3.497***	(0.865)
Log Income	-0.077***	(0.016)	-0.049**	(0.021)	0.007	(0.032)	0.002	(0.036)
Share $\geq$ 20, $\leq$ 40	0.283**	(0.114)	0.171	(0.126)	0.136	(0.132)	-0.197	(0.135)
Share $\geq$ 40, $\leq$ 60	0.005	(0.212)	-0.098	(0.243)	-0.230	(0.248)	-0.298	(0.195)
Share $\geq$ 60	0.548***	(0.134)	0.485***	(0.136)	0.460***	(0.142)	0.122	(0.119)
Share Male	0.473*	(0.245)	0.563**	(0.277)	0.508*	(0.261)	0.318*	(0.179)
Log Household Size	0.294***	(0.033)	0.274***	(0.048)	0.213***	(0.059)	0.146**	(0.057)
% Black	0.130***	(0.023)	0.111***	(0.024)	0.122***	(0.026)	0.147***	(0.029)
% Hispanic	0.053	(0.037)	0.019	(0.036)	-0.023	(0.036)	0.039	(0.038)
% Asian	0.083**	(0.041)	0.099**	(0.041)	0.080*	(0.047)	0.091**	(0.043)
Log Density			0.011	(0.009)	0.016	(0.010)	0.013*	(0.007)
% Public Transport			0.125**	(0.062)	0.063	(0.061)	0.089*	(0.050)
Log Commuting Time			0.059	(0.061)	0.086	(0.067)	0.033	(0.061)
% Working Home			-0.195	(0.348)	-0.032	(0.373)	-0.225	(0.253)
% Private Insurance					-0.128	(0.101)	-0.094	(0.091)
% Uninsured					0.388**	(0.175)	0.257	(0.171)
% Essential - Professional							0.417*	(0.223)
% Non ess. - Professional							0.501***	(0.162)
% Science fields							-2.290**	(0.915)
% Law and related							-1.115**	(0.504)
% Health practitioners							-0.154	(0.435)
% Other health							-0.181	(0.292)
% Firefighting							0.956*	(0.576)
% Law enforcement							-1.748**	(0.849)
% Essential - Service							0.039	(0.278)
% Non ess. - Service							0.589	(0.465)
% Ind. and Construction							0.547*	(0.318)
% Essential - Technical							-0.735	(0.694)
% Transportation							1.349***	(0.411)
Constant	0.098	(0.126)	-0.379	(0.319)	-0.557	(0.343)	-0.227	(0.249)
Observations	174		174		174		174	
R^2	0.772		0.798		0.777		0.874	

Robust standard errors. Standard errors in parentheses

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

3.1 Daily comparison

The previous discussion focused on the effects that remain mostly constant over time. However, we find some time-variant results that could provide insight on both the evolution of the pandemic effects as well as the health policies set in place. The result for the tests per capita variable is particularly salient; we observe a very positive of effect on the share of positive tests. However, its effect becomes progressively smaller each passing day. This result could be reconciled with the fact that in the earlier days of the crisis, testing was severely limited. Only those who had traveled to one of the early affected countries or who had been in direct contact with a positive case were recommended for testing. Over the next few days, testing was performed to those sick enough to need hospitalization. Naturally, zip codes with more tests implicitly meant a higher share of people with higher risk of having the disease. As testing becomes more available, we see that the tests per capita lose relative importance and, hence, widespread testing in the coming weeks should solve this.

After adding the rest of control variables, the income effect is positive but not significant. This indicates that the variation that led to a negative correlation between income and positive tests is now explained by these other variables. Furthermore, the income effects seems to become less positive over time, providing more evidence that selection into testing was generating part of the correlation.

There are many notable time trends related to occupations. First, we observe that the share of workers in law enforcement (8) eventually evolves to a negative effect on share of positives. Second, the share of firefighters (7) tend to have the opposite effect, becoming significantly positive later on. While these occupations seem to be similar in certain aspects, data shows that there is some location sorting between them. In particular, the share of law enforcement is relatively high in Staten Island, a low to medium COVID-19 exposure borough, while the share of firefighters is high in the Bronx and eastern Brooklyn, areas with high shares of positive tests. This would suggest some underlying characteristics that we do not observe common to both occupations is getting absorbed once selection in testing is alleviated. Third, the share of workers in industrial, natural resources, and construction (11) becomes less significant over time. This pattern may be caused by the stay-at-home order issued for the State of New York in March 22.

Finally, another outstanding time pattern is that the coefficients on racial composition decrease in magnitude as the selection of testing decreases. This result may suggest that there could be racial selection component among those in worse conditions 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 conditions to access testing compared to whites.[8]

One key takeaway from the results of our daily comparison is that they emphasize

[8]Some evidence that this is plausible mechanism can be found in www.modernhealthcare.com/safety-quality/long-standing-racial-and-income-disparities-seen-creeping-covid-19-care

the importance of widespread testing, as only with reliable testing we can identify the mechanisms that explain demographic differences in COVID-19 exposure.

Table 10: Dependent variable - Share of Positives as of April 8, 2020

	(1) April 1		(2) April 5		(3) April 9	
Tests per capita	9.388***	(2.318)	6.919***	(1.267)	3.497***	(0.865)
Log Income	0.053	(0.048)	0.036	(0.038)	0.002	(0.036)
Share $\geq$ 20, $\leq$ 40	-0.426**	(0.182)	-0.208	(0.148)	-0.197	(0.135)
Share $\geq$ 40, $\leq$ 60	-0.633**	(0.246)	-0.261	(0.213)	-0.298	(0.195)
Share $\geq$ 60	-0.356*	(0.189)	-0.072	(0.146)	0.122	(0.119)
Share Male	0.095	(0.245)	0.108	(0.200)	0.318*	(0.179)
Log Household Size	0.007	(0.075)	0.082	(0.062)	0.146**	(0.057)
% Black	0.173***	(0.038)	0.137***	(0.033)	0.147***	(0.029)
% Hispanic	0.043	(0.050)	0.042	(0.042)	0.039	(0.038)
% Asian	0.157***	(0.049)	0.090**	(0.044)	0.091**	(0.043)
Log Density	0.025**	(0.010)	0.019**	(0.009)	0.013*	(0.007)
% Public Transport	0.038	(0.059)	0.033	(0.051)	0.089*	(0.050)
Log Commuting Time	0.059	(0.072)	0.044	(0.063)	0.033	(0.061)
% Working Home	-0.271	(0.350)	-0.271	(0.319)	-0.225	(0.253)
% Uninsured	0.163	(0.241)	0.272	(0.198)	0.257	(0.171)
% Private Insurance	-0.125	(0.128)	-0.157	(0.102)	-0.094	(0.091)
% Essential - Professional	0.447	(0.281)	0.264	(0.239)	0.417*	(0.223)
% Non ess. - Professional	0.595***	(0.186)	0.602***	(0.153)	0.501***	(0.162)
% Science fields	-2.648***	(1.006)	-2.335**	(0.895)	-2.290**	(0.915)
% Law and related	-1.293*	(0.751)	-0.913	(0.601)	-1.115**	(0.504)
% Health practitioners	-0.302	(0.484)	-0.409	(0.455)	-0.154	(0.435)
% Other health	-0.024	(0.420)	0.085	(0.342)	-0.181	(0.292)
% Firefighting	1.036	(0.837)	1.185*	(0.713)	0.956*	(0.576)
% Law enforcement	-0.304	(0.922)	-1.571*	(0.840)	-1.748**	(0.849)
% Essential - Service	-0.033	(0.374)	-0.110	(0.303)	0.039	(0.278)
% Non ess. - Service	0.926*	(0.548)	0.599	(0.506)	0.589	(0.465)
% Ind. and Construction	1.210***	(0.406)	0.796**	(0.352)	0.547*	(0.318)
% Essential - Technical	-0.896	(0.923)	-1.471*	(0.815)	-0.735	(0.694)
% Transportation	1.371**	(0.534)	1.166**	(0.451)	1.349***	(0.411)
Constant	-0.274	(0.334)	-0.255	(0.279)	-0.227	(0.249)
Observations	174		174		174	
R^2	0.766		0.842		0.874	

Robust standard errors. Standard errors in parentheses

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

4 Implications and policy recommendations

There are clear implications for policy on testing and vaccinations all over the world, but in particular in NYC. First and foremost, as mentioned earlier, our results highlight the need of mass testing to be able to cleanly identify what are the relevant demographic groups that are at greater risk of exposure. Second, in light of our results, a clear policy implication is to target specific workers when tests, protective gear, and vaccines become available but there are still scarce. We see that the purpose of this policy is twofold. On the one hand, it provides a form of insurance to those who are more vulnerable against contracting the disease. On the other hand, targeting specific workers or demographic groups may have spillovers on the rest of the population. For example, a policy that starts vaccinating and/or testing those workers with higher rates of human interactions does not only have effects on those directly targeted by the policy but also on those who are likely to be in contact with them. We believe, given the high contagion rates of COVID-19, that the spillovers from this policy may be substantial.

5 Conclusions

In this paper, we present evidence of an occupation mechanism in explaining different rates of COVID-19 incidence across income, race, age, and other socio-demographic groups. We argue that part of these demographic disparities can be explained due to differences in occupations and their exposure to the illness. 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, controlling for demographic and location characteristics. The DOH provides daily updates of COVID-19 test data, allowing us to estimate the aforementioned effects over multiple days. Test availability was a major issue in the early days of the crisis, so measuring time-varying effects allows us to detect evidence and potential effects of selection in testing.

We begin by showing descriptive evidence of heterogeneous incidence of positive cases across income, race, age, gender, and household size. A zip code's median income is negatively correlated with its share of positives. Conversely, We find that a positive correlation between shares of black, Hispanic, residents under 20, as well as average household size with the share of positive tests. To justify our proposed occupation mechanism, we study the correlation of the shares of the demographic groups with the share of workers in different occupation categories. We divide the occupations defined in the ACS in 13 more general categories and show how the share of workers in each one of them correlates with median income. Following that, we do a similar exercise between occupations and demographics groups and find that generally the groups with higher rates of positive shares are more likely to be employed in occupations with higher rates of human interaction and these occupations tend to have lower incomes.

To formalize these observations and the effects on share of positive tests, we estimate several models. We start with a simple specification, only including demographic characteristics. We progressively augment the model by including commuting and location density, health insurance indicators, and finally occupation shares. We find the strongest positive effect on share of positive tests in the share of workers on

Transportation, Industrial, Natural resources, Construction, and Non essential - Professional. Conversely, higher shares of workers in science fields and law enforcement have a negative effect on positive tests. These results align with the fact that the previous occupations tend to involve more human interaction. Racial effects do persist after adding the occupation variables for blacks and Asians, suggesting that the occupation mechanism does not explain the entire racial disparity in positive cases and there could a racial bias on the incidence of testing (Borjas, 2020). On the other hand, when adding occupation variables, age and income do not longer contribute to explain the variation in positive tests, suggesting that disparities along those demographics are explained by selection across occupations.

The coefficient for the tests per capita is a strong predictor of share of positive tests. However, its relative importance declines each passing day. This confirms the selection in testing argument we made earlier. As tests become more available, this issue loses relevancy and more of the variation in COVID-19 incidence can be explained through the occupation channel. Having a clear understanding of this mechanism will be vital when assessing the policies to alleviate the health and economic effects that the pandemic is inflicting, particularly to the more vulnerable groups.

References

Alon, T., Doepke, M., Olmstead-Rumsey, J., and Michèle, T. (2020). The impact of covid-19 on gender equality. *Unpublished Manuscript.*

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.*

Rae, M., Claxton, G., Kurani, N., McDermott, D., and Cox, C. (2020). Potential costs of coronavirus treatment for people with employer coverage. Available at `https://www.healthsystemtracker.org/brief/potential-costs-of-coronavirus-treatment-for-people-with-employer-coverage/`.