Racial Disparities in Frontline Workers and Housing Crowding during COVID-19: Evidence from Geolocation Data

Milena Almagro* Joshua Coven† Arpit Gupta‡
Angelo Orane-Hutchinson§

September 18, 2020

Abstract

We document that racial disparities in COVID-19 in New York City stem from patterns of commuting and housing crowding. During the initial wave of the pandemic, we find that out-of-home activity related to commuting is strongly associated with COVID-19 cases at the ZIP Code level and hospitalization at an individual level. After layoffs of essential workers decreased commuting, we find case growth continued through household crowding. A larger share of individuals in crowded housing or commuting to essential work are Black, Hispanic, and lower-income. As a result, structural inequalities, rather than population density, play a role in determining the cross-section of COVID-19 risk exposure in urban areas.

*Federal Reserve Bank of Minneapolis and Booth School of Business, University of Chicago. Email: milena.almagro@chicagobooth.edu

†Department of Finance, Stern School of Business, New York University. Email: jcoven@stern.nyu.edu

‡Department of Finance, Stern School of Business, New York University. Email: arpit.gupta@stern.nyu.edu

§Department of Economics, New York University. Email: angelo.orane@nyu.edu

I Introduction

The novel coronavirus disease 2019 (COVID-19) has disproportionately and negatively impacted disadvantaged populations. The hardest-hit regions of New York City include parts of the Bronx, Brooklyn, and Queens with high fractions of Black, Hispanic, and low-income populations as has been noted by Borjas (2020) and Schmitt-Grohé et al. (2020).[1] Nationwide, infections are three times as likely among Latinos and African-Americans compared to infections among whites.[2] While racial disparities in COVID-19 disease burdens have been widely recognized, the ultimate drivers of these inequities remain unclear.

This paper connects racial disparities in COVID-19 cases with mobility patterns for vulnerable populations by using anonymized mobile phone Global Positioning System (GPS) data. We focus on New York City, the global epicenter for the pandemic in Spring 2020. Our work splits the COVID-19 pandemic in New York City into two periods. In the initial stage of the crisis—lasting from March until early April—we document that the commuting behavior of essential workers placed them at greater risk of infection. We find these commuting patterns changed after early April, when many of these workers were laid off. At this point in the crisis, disease spread continued through a household crowding channel. The relative importance of household crowding compared to mobility patterns doubled during April. We find that racial minorities are over-represented in both mobility-based risk factors—occupational specialization in essential work professions and household overcrowding.

Our analysis is conducted at both the neighborhood and individual levels, allowing us to finely measure the nature of mobility responses in the wake of this pandemic. To do so, we link individual mobile phone data with ZIP Code-level data on daily COVID-19 infection and hospitalization rates, as well as Census data on occupation and household occupancy. A key advantage of our identification approach is that we are able to measure hospitalizations at aggregate levels as well as the *individual* level by isolating mobile phone users who appear in local hospitals. Doing so allows us to control for important local characteristics and use within-tract variation in commuting and housing characteristics.

[1] We document the demographic associations of COVID-19 in Section III.A. New York City official data suggest that African-Americans were 59% more likely to be diagnosed with COVID-19 relative to whites, while Hispanics were 64% more likely. See: https://www1.nyc.gov/site/doh/covid/covid-19-data.page.

[2] See data from the C.D.C. https://www.nytimes.com/interactive/2020/07/05/us/coronavirus-latinos-african-americans-cdc-data.html.

Our results suggest sizeable effects of mobility on disease exposure: increasing time outside of an individual's home census tract from the 10^{th} to the 90^{th} percentile is associated with a 30% increase in the hazard rate of hospitalizations. Similarly, individuals at the 90^{th} percentile of housing crowding are 7% more likely to be hospitalized than those at the 10^{th} percentile. While both measures suggest large and statistically significant impacts of our mobility measures on hospitalization outcomes, we also find that including the housing crowding measure lowers the measured impact of the commuting measure to 25%. This suggests a substantial correlation between mobility patterns and housing density, and it points to possible implications of policies that target these measures separately. For example, shutting down workplaces through lockdowns may lower infectious spread through a commuting channel but may instead result in individuals interacting more in crowded home settings.

Our analysis has implications for ongoing debates on the role of density and urban form on disease exposure. In contrast to research which emphasizes the role of static characteristics of urban design such as density (Duranton and Puga, 2020; Carozzi et al., 2020) or subways (Harris, 2020), we highlight the dynamic responses of individuals and groups which depend on access to preexisting resources. Notably, Manhattan—the densest and wealthiest borough—saw many fewer infections than the other boroughs.

Instead, our results suggest that the types of density that matter the most are the experienced density of front-line workers exposed to contact through direct physical proximity, as well as the crowding of individuals in multi-family households. We document that these mobility-induced densities are disproportionately experienced by vulnerable populations. In turn, structural inequalities lead disadvantaged groups to disproportionately live in crowded housing and specialize in jobs that require physical presence. These underlying inequalities in job presence and housing create temporary pockets of density through which SARS-CoV-2 virus propagates.

We contribute to a growing literature on COVID-19 by providing direct evidence on the role of specific mobility factors in contributing to the spread of the disease, as well as on the racial dimensions of these factors. Many papers have used geolocation data in the context of COVID-19 (Chen et al., 2020; Couture et al., 2020; García-Lopez and Puga, 2020). Our work is most closely related to Glaeser et al. (2020). We differ in four key ways. First, we consider both aggregated data and individual-level mobility data, allowing us to highlight individual risk factors for hospitalization. Second, our central focus is examining racial disparities. Third, we consider an additional housing crowding dimension which was crucial at the stage in the pandemic when many workers stopped commuting due to unemployment. Finally, we complement the approach in Glaeser et al. (2020) by proposing a new identification strategy that leverages the granularity of our data. To do so, we construct a panel of buildings where our main outcome variable is the hospitalization of a building's resident. Our main identification assumption is that daily unobservables are common across all individuals who live in buildings in the same census tract. Chiou and Tucker (2020) also examine income mobility responses and focus on variation in the ability to work from home and internet access. We emphasize both income and racial disparities within urban areas and highlight both commuting- and housing-related disparities.

We build on methods used in prior works such as Athey et al. (2019), Chen et al. (2019), and Chen and Rohla (2018), which used mobile phone geolocation data to examine segregation, racial disparities in voting waiting times, and partisanship.

A growing literature also examines racial disparities specifically in the context of COVID-19 (Borjas, 2020; McLaren, 2020; McCormack et al., 2020; Almagro and Orane-Hutchinson, 2020; Sá, 2020; Karaca-Mandic et al., 2020). Our work adds to this literature by linking important mobility components of this racial disparity, and further directly connecting them to case exposure. Notably, the emerging medical literature, such as Rentsch et al. (2020) and Price-Haywood et al. (2020), finds evidence of racial disparities in exposure to COVID-19—but does not find evidence of racial differences in mortality. This highlights the importance of understanding why different racial groups are potentially exposed to infection at different rates, which our analysis does through considering commuting and housing crowding channels. Finally, our findings complement other well-established health disparities which may impact the severity of COVID-19 for different populations, such as those presented in Wong et al. (2002) and Trivedi et al. (2005).

II Data

II.A Geolocation Data

Mobile location data were sourced from VenPath, a holistic global provider of compliant smartphone data. Our data provider aggregates information from approximately 120 million smart phone users across the United States. GPS data were combined across applications for a given user to produce "pings" corresponding to time stamp–location pairs.

The provider anonymizes information on individual users. Ping data include both background pings (location data provided while the application is running in the background) and foreground pings (activated while users are actively using the application). Our sample period covers the period February 1$^{\text{st}}$–July 12$^{\text{th}}$, 2020.

II.B Estimating Time Outside of Home Tract

To isolate the mobility behavior of New York City residents, we employ multiple screens to filter out commuters, visitors, and those who leave the city either temporarily or permanently.

First we separate those who spend the night in New York City from those who commute into or visit New York City during the day in order to measure the behavioral responses of local residents. We select from the anonymous users those who have the majority of their pings between 6pm and 8am (night hours) in New York City (as opposed to any non-New York City county in the US) on at least three different days in a specific month. We then enforce a minimum required data density and keep only those with at least three pings on at least five nights in the data in New York City, with the same requirements during work hours.

Second, we filter out those who commute to, visit, or leave NYC by joining our remaining users' pings to census tracts from New York City Open Data. On every date, we identify each user's modal tract during night hours. If they ping in the modal tract at least twice a night, on at least five different nights in a month, we assign their most frequent remaining modal night tract as their "home census tract" (HCT). If a user spends an equal amount of nights in more than one HCT, we choose the one that they ping in the most throughout the data.

We repeat this process each month from February to June and exclude those who have been identified as residents in previous months. We use only one month of data at a time to identify residents' home tracts. We then analyze their data in the months after the month that was used to identify their home locations. This gives us a sample population of 647,068 unique users for our base analysis.[3]

Next, we exclude those who have left New York City from our analysis by requiring the county of their HCT to match each day with the county they spent the night in. The resulting dataset allows us to measure the mobility responses among NYC residents.

[3]We find that our estimated mobile phone population correlates with Census population at 0.89, suggesting representative sample coverage.

Finally, to measure out-of home-behavior, we estimate the fraction of pings that occur each hour within that user's home tract, and we then estimate the number of hours a user spends entirely outside of their home tract during the range of 8am to 10pm (the "number of hours entirely outside").

II.C Household Crowding

To construct our housing crowding metrics, we connect the ping data with the geographic data for all building footprints in NYC, which are created by Microsoft using satellite images.[4] We join these building shapes to land use data from the NYC Department of Planning at the lot level.[5] Multiple lots correspond to each Microsoft building. We do a geospatial join of the building lots to the Microsoft building shapes and then aggregate to arrive at the total number of residential units and residential square footage for each building.

To define a metric of housing crowding, we identify each user's modal home building each night and then their modal building across nights. We aggregate to the building level and count unique users for whom that building is their home building. For each building, we calculate the people per housing unit on each date. While our measures of mobility are measured at the individual level, our housing crowding measure is estimated at the building level. We then take the average of buildings for each tract and ZIP Code to get our "mean people per unit" measure of housing crowding. Appendix Figure B2 shows the spatial variation of the housing crowding measure, averaged across our sample. We tend to observe greater housing crowding in the outer boroughs of the city.

II.D Measuring Hospitalizations

To construct our hospitalization measure, we first determine which of the Microsoft building shapes correspond to hospitals by joining them to latitudes and longitudes of hospitals provided by HIFLD Open Data.[6] We include only hospitals within NYC that are not long-term care facilities or psychiatric hospitals. We can then see which pings correspond to which designated hospitals within NYC.

[4]This dataset can be found at: https://github.com/microsoft/USBuildingFootprints.

[5]See: https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page.

[6]See: https://hifld-geoplatform.opendata.arcgis.com/datasets/6ac5e325468c4cb9b905f1728d6fbf0f_0.

We identify whether a user pings within a hospital from our hospital building shapefiles. We observe the first date when a user pings within a hospital as their hospitalization date. To verify that our measured hospitalizations line up with other data sources, we compare against actual hospitalization data for COVID-19 from the NYC Department of Health in Appendix Figure A2. Across the period from March 25th–April 22nd, we find a correlation of 0.78 between our mobility measure of hospitalization and actual hospitalizations, suggesting that we are able to accurately estimate individual hospitalized COVID-19 cases.

II.E NYC COVID-19 Data

Our source of incidence rates of COVID-19 and number of tests performed is the NYC Department of Health (DOH) data release.[7] The DOH releases (almost) daily data on the cumulative count of COVID-19 cases and the total number of residents who have been tested, divided by the ZIP code of residence. We have collected data covering the months of April and May.[8] In our analysis, we drop the first week of April due to these missing dates, and also because the first few days in our sample appear very noisy.

II.F Census and Occupation Shares

We obtain demographic and occupation data at the ZIP code and Census tract level from the American Community Survey (ACS). The demographic characteristics we include are ZIP Code median income, average age, racial breakdown, and health insurance status. We also include commuting-related variables: average commute time to work as well as means of transportation.

We also construct the shares of the working-age population employed in different occupation categories, similar to Almagro and Orane-Hutchinson (2020). The ACS provides the number of workers employed in each occupation by ZIP code of residence. We first divide occupation between flexible and non-flexible occupations. Then, we categorize non-flexible occupations according to their essential definition and similarity in work environments and social exposure. For our occupational variables, we count the number for workers in each of these occupations and normalize by working-age population, which includes everyone with age ranging from 18 to 65 years. The summary statistics of demographics and occupations can be found in Appendix A, which also breaks out the occupational groups in the non-flexible category separately.

[7] See: `https://github.com/nychealth/coronavirus-data.`

[8] Unfortunately April 2nd and April 6th are missing from our sample as these data have never been made publicly available.

II.G Aggregating Mobility and Crowding Measures

To aggregate our mobility and crowding metrics to the ZIP Code level, we use the geospatial shapes of NYC's census tracts provided by NYC Open Data, and we link to ZIP Codes using a crosswalk provided by the Department of Housing and Urban Development.[9] We select the ZIP and tract mapping that has the highest number of residents residing in the ZIP for a given tract to get a 1:1 mapping of tracts to ZIP Codes.

We can aggregate housing crowding data and individual mobility data to the building level. Panel A of Appendix Table A1 presents the summary statistics of the building-level dataset used for our housing crowding analysis. We identified 333,132 buildings in NYC with residential space, from March 1st to July 12th. We were able to link 538 hospitalizations to residents of these buildings, where the hospitalizations occurred between March 18th and April 22nd.

For the individual analysis, we aggregate these variables to the ZIP and date level. We winsorize these variables at the 1% level. All of the daily values are constructed using a seven-day moving average to reduce the noise of daily raw values and to account for weekly seasonality. Panel B of Appendix Table A1 presents the summary statistics of the individual-level dataset used for our survival analysis. We identified 286,367 individual residents in NYC from March 1st to July 12th. We were able to link 597 hospitalizations to these residents, where the hospitalizations occurred between March 18th and April 22nd.

III RESULTS

III.A Descriptive Analysis

We begin with a descriptive analysis of our sample to highlight the key features of the COVID-19 pandemic in New York City. Appendix Figure A1 shows how our mobility metrics, housing crowding, and the daily share of positive tests evolve over time. We observe that the share of positive tests steadily decreases over time. We also observe a large decrease in mobility in early March that started prior to the stay-at-home order issued by Governor Cuomo on March 20. Our finding that mobility responds primarily to the pandemic, rather than the state-imposed order, is consistent with similar nationwide findings in Goolsbee and Syverson (2020). Mobility in our sample hits a low in

[9]See: https://data.cityofnewyork.us/City-Government/2010-Census-Tracts/fxpq-c8ku for the list of tracts and https://www.huduser.gov/portal/datasets/usps_crosswalk.html for the crosswalk.

early April before steadily recovering towards pre-pandemic levels later in our sample. On the other hand, we do not see any stark trend for the average number of people per housing unit.

We also contrast the time series of mobility measures in Appendix Figure B1 across different boroughs of NYC. In the early period of our sample, we observe the greatest sheltering responses in Queens and Staten Island (Richmond County). However, after May, we observe the highest sheltering responses in Manhattan. These differential patterns across boroughs may reflect the ability of different populations to shelter effectively given the tendency for these jobs to be precarious and non-local. We compare across both the time series and the cross-section in Appendix Figure B3. In the key months of the pandemic, in March and April, measured mobility patterns shows sheltering in certain high-income neighborhoods of Manhattan and Brooklyn—while residents in other boroughs were much more likely to spend time outside of their home tract.

We then plot some basic correlations between our mobility and housing density measures with demographics and occupations. Panel A of Figure 1 shows correlations of mobility with certain neighborhood demographics: the fraction of tract residents who are Black, Hispanic, and/or low-income.

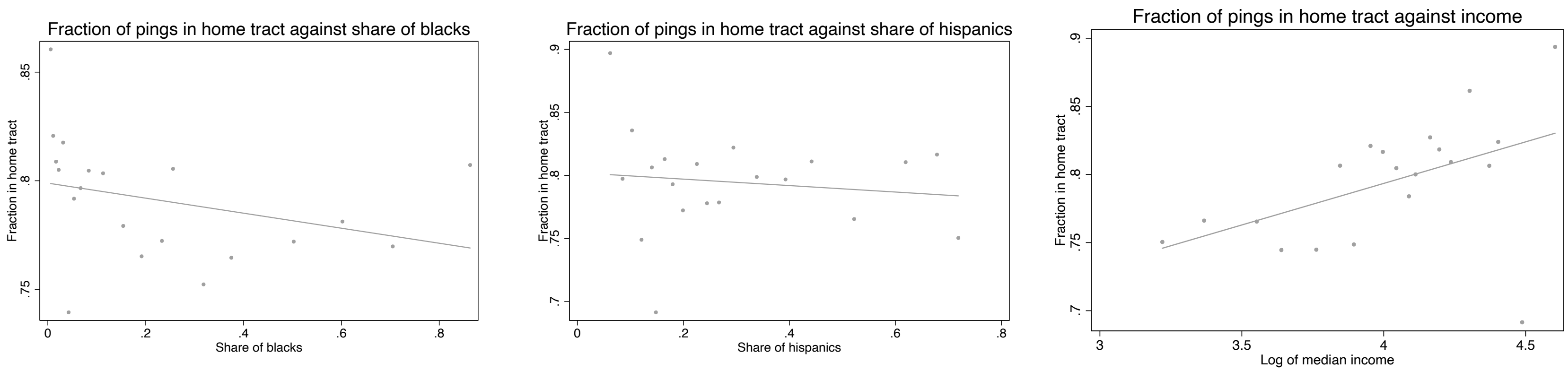
We find substantial positive correlations of increased out-of-home mobility in areas with more low-income, Black, and Hispanic individuals. We also observe a positive correlation between crowded spaces and neighborhoods with a higher share of minorities and lower average income in Panel B of Figure 1.

Given that many workers were laid off during April and May due to the pandemic, we turn to a more dynamic analysis to illustrate correlations between our two mobility measures (commuting and housing crowding) across occupations. In Panel C of Figure 1, we show the coefficients obtained when we regress daily mobility patterns and housing crowding on the share of flexible occupations after controlling for time trends. We find that ZIP Codes with higher shares of flexible occupations are consistently positively correlated with less mobility throughout our sample, as shown by the positive daily coefficients plotted on the left graph. This suggests that a key driver of our mobility results is related to commuting to essential work.[10] We conduct the same exercise for our measure of housing crowding and find a negative, although not significant, coefficient for most of our sample.

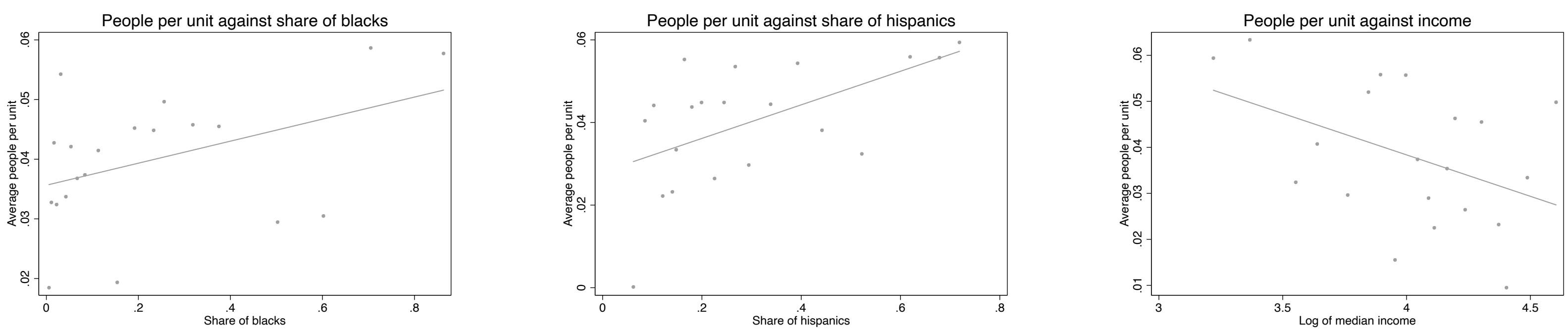
[10]The lack of significance in most of these regressions is mainly driven by our having a small number of observations for each daily regression.

Figure 1: Demography and Mobility Measures

Panel A: Binscatter plots of share of mobile phone pings in home tract and demographics



Panel B: Binscatter plots of crowded spaces and demographics



Panel C: Daily correlations of mobility patterns and housing crowding, and share of flexible occupations

III.B ZIP Code-Level Analysis

Having established our basic variables, we turn next to a deeper analysis of the relationship between structural inequalities and the incidence of COVID-19. We start by constructing a panel of the daily share of positive tests across NYC ZIP codes from April 8$^{\text{th}}$ to May 26$^{\text{th}}$. For this specification, we estimate the following equation:

$$\text{share of positive tests}_{jt} = \beta_1 \text{mobility}_{jt} + \beta_2 \text{housing density}_{jt} + \gamma X_j + \mu_t + \varepsilon_{jt}$$

where X_j contains demographic and occupational characteristics at the ZIP Code level and μ_t is a day fixed effect that controls for the aggregate evolution of the pandemic in NYC.

Table 1 shows the estimation results of regressing the daily share of positive tests across ZIP Codes on mobility and housing crowding measures for several specifications that vary in their set of neighborhood controls. The first specification, column (1), includes only basic demographics such as race and income, while column (2) includes only our mobility and housing density measures. Column (3) includes all of these covariates together. A comparison of column (1) with column (3) shows that the initial racial disparities are partially explained by differences in mobility patterns and housing density, as all coefficients for racial groups shrink towards zero. Similarly, a comparison of column (2) with column (3) shows that much of the correlation in mobility and housing density is mediated through race and income differences: their coefficient sizes decrease by 10% and 52%, respectively, when including both set of controls.

For this basic specification, column (3), which presents the interpretation of the magnitudes for our mobility measure, is as follows: if the fraction of pings inside a home census tract (HCT) increases by 10%, an average level increase of 0.06, the daily share of positives decreases by 0.00396 points, corresponding to a 1.8% lower share of daily positive tests. On the other hand, a 10% increase in the number of people per unit corresponds to a 0.0019-point increase, or equivalently a 0.8% increase in the share of positive tests.[11] When we include other controls, we find that the size of the coefficient on the out-of-tract mobility decreases, becoming non-significant with the inclusion of the occupation controls. This result indicates that most of the housing density variation can be captured by the occupational composition of neighborhoods.

[11] The average rate of daily tests that came out positive between April 8$^{\text{th}}$ and May 26$^{\text{th}}$ is 22%.

Examining only demographic variables shows evidence of strong racial disparities in test positivity rates. Incorporating mobility, demographic, and occupational controls lowers the coefficient on fraction Black by 45%, and lowers the coefficient on the fraction of population members who are Asian to zero. This suggests that racial disparities in test positivity rates, at least for these groups, can be accounted for by variation in background variables related to mobility and occupation—though we observe larger residual disparities for Hispanic populations.

III.B.1 Weekly Analysis

In principle, the effects of mobility measures may vary over time given the dynamic behavior of other factors such as the natural evolution of the pandemic or its effects on the economy. Motivated by this hypothesis, we estimate the following equation:

$$\text{daily share of positives}_{jt} = \alpha_{1,t}\text{mobility}_{jt} + \alpha_{2,t}\text{housing density}_{jt} + \gamma_t X_j + \mu_t + \varepsilon_{jt}.$$

Panels A and B of Appendix Figure A3 plot the evolution of the daily coefficients for our mobility measure as well as for housing density. These patterns suggest that mobility had a larger impact for the first three weeks of April, while the magnitude of the coefficient of housing crowding was larger for the last week of April and the first week of May.

As pointed out by Glaeser et al. (2020), the estimates of these regressions are likely masking correlations with unobservables that bias the coefficients towards zero. For example, a change in behaviour following the evolution of the pandemic correlates with mobility measures as well as with the number of new infections. For this reason, we turn to our next analysis, where we leverage the granularity of our data to overcome these endogeneity concerns.

Table 1: Neighborhood Associations of Positive Tests

Dependent Variable:	Daily Share of Tests that are Positive									
	(1) Race & Income		(2) Mobility		(3) Mobility & Race, Income		(4) Mobility & Demographics		(5) Mobility, Dem. & Occupations	
Fraction of Pings in Home Tract			-0.076***	(0.024)	-0.066**	(0.026)	-0.144***	(0.026)	-0.065***	(0.025)
Log People per Unit			0.044***	(0.002)	0.019***	(0.002)	0.011***	(0.003)	0.000	(0.003)
Log Income	-0.003	(0.003)			-0.001	(0.003)	-0.015**	(0.006)	-0.000	(0.011)
% Black	0.113***	(0.004)			0.095***	(0.005)	0.061***	(0.008)	0.051***	(0.008)
% Hispanic	0.154***	(0.006)			0.145***	(0.006)	0.125***	(0.008)	0.157***	(0.010)
% Asian	0.170***	(0.007)			0.152***	(0.008)	0.006	(0.013)	0.002	(0.013)
Share $\geq 20, \leq 40$							0.066	(0.059)	0.040	(0.054)
Share $\geq 40, \leq 60$							0.016	(0.090)	0.027	(0.108)
Share ≥ 60							0.456***	(0.058)	0.371***	(0.072)
Share Male							0.192***	(0.050)	0.362***	(0.055)
Log Household Size							0.124***	(0.020)	0.086***	(0.025)
Log Density							0.007***	(0.002)	0.004**	(0.002)
% Public Transport							0.020	(0.014)	-0.016	(0.018)
Log Commute Time							0.016	(0.011)	0.017	(0.012)
% Uninsured							0.203***	(0.038)	0.351***	(0.047)
Bronx							-0.010	(0.007)	-0.017**	(0.007)
Brooklyn							0.026***	(0.008)	0.028***	(0.007)
Queens							0.022***	(0.007)	0.030***	(0.008)
Staten Island							-0.019**	(0.008)	-0.003	(0.009)
% Flexible Occupations									-0.034	(0.058)
% Health Practitioners									-0.121	(0.138)
% Other Health									0.618***	(0.107)
% Firefighting									-0.323	(0.305)
% Law Enforcement									-1.705***	(0.397)
% Essential: Service									-0.287***	(0.068)
% Non Ess.: Service									-0.017	(0.144)
% Ind. and Construction									-0.240**	(0.096)
% Essential: Technical									-0.683**	(0.294)
% Transportation									0.312***	(0.106)
Constant	0.280***	(0.014)	0.570***	(0.019)	0.395***	(0.023)	0.014	(0.059)	-0.105	(0.066)
Day FE	✓		✓		✓		✓		✓	
N	6401		6401		6401		6401		6401	
adj. R^2	0.89		0.88		0.89		0.91		0.91	

Robust standard errors in parentheses

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

III.C Building-Level Analysis

In this section we exploit the panel structure of our sample including both mobility and housing crowding measures in order to better address identification concerns. Unfortunately, test data at more granular geographical levels are unavailable. To overcome this challenge, we instead measure hospitalizations at the individual level by measuring direct individual visits to hospitals. We classify individuals as being hospitalized if they spend more than 24 hours at a hospital. We focus on the first month after the issuance of the stay-at-home order to maximize the probability that new hospitalizations that we observe in our data are due to COVID-19 and not due to something else.[12] Our main outcome variable is defined as the event of a building's resident being hospitalized per building occupancy. Because our unit of observation is the crowding at the building level, we analyze as our key outcome the hospitalization rate per building.

Table 2 shows the result of an analysis in which our dependent variable is the event of a new hospitalization within a building and different specifications include different sets of neighborhood controls. Our main regression equation is

$$\text{hospitalization}_{bt} = \alpha_1 \text{mobility}_{bt} + \alpha_2 \text{housing density}_{bt} + \gamma X_{j(b)} + \mu_t + \varepsilon_{bt},$$

where mobility_{bt} and $\text{housing density}_{bt}$ are respectively the average mobility and housing density measures for date t, $X_{j(b)}$ are demographic and occupational controls for the Census tract where the building b is located, and μ_t are date fixed effects.

Specifications (1)–(4) are similar to those of Table 1 but are done at a different level of aggregation. Given that we have variation across buildings within the same census tract for any given date, we can include a fixed effect at the census tract and date level, $\delta_{j(b)t}$, to control for daily factors common to all individuals within a census tract. That is, our regression equation in this case is:

$$\text{hospitalization}_{bt} = \alpha_1 \text{mobility}_{bt} + \alpha_2 \text{housing density}_{bt} + \delta_{j(b)t} + \varepsilon_{bt}.$$

Our identifying assumption for this specification is based on the hypothesis that unobservables with temporal variation that correlate with mobility and housing density measures are common to all buildings inside the same census tract.

[12]Using data available at `https://github.com/thecityny/covid-19-nyc-data` that constructed total hospitalizations from reports of Governor Cuomo's office, we observe that more than 50% of all hospitalizations for the first three weeks of April were related to COVID-19. This measure includes new hospitalizations as well as as patients with more long-term diseases or patients in palliative care. For the time-series correlation of our measure of hospitalizations and the official numbers, see Appendix A2.

Table 2: Impact of Mobility on Hospitalizations

Dependent Variable	Hospitalizations per Occupancy				
	(1) Mobility	(2) Mobility & Race, Income	(3) Mobility & Demographics	(4) Mobility & Dem., Occ.	(5) Census Tract × Day
Fraction of Pings in Home Tract	-1.815e-04***	-1.780e-04***	-1.766e-04***	-1.770e-04***	-1.837e-04***
	(2.624e-05)	(2.618e-05)	(2.620e-05)	(2.626e-05)	(2.766e-05)
Log People per Unit	1.226e-05***	1.217e-05***	1.283e-05***	1.281e-05***	1.347e-05***
	(2.313e-06)	(2.340e-06)	(2.329e-06)	(2.333e-06)	(2.482e-06)
Constant	3.135e-04***	3.496e-04*	2.832e-04	1.373e-04	3.191e-04***
	(2.873e-05)	(1.425e-04)	(3.389e-04)	(3.547e-04)	(2.979e-05)
Day FE	✓	✓	✓	✓	
Demographic Controls			✓	✓	
Occupation Controls				✓	
Census Tract × Day FE					✓
N	1,355,459	1,354,504	1,354,504	1,354,504	1,355,459

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The interpretation of the coefficients is as follows. For our preferred specification, column (5), if a building's residents increase their fraction of pings within their HCT by 10%, the number of hospitalizations per occupant in that building decreases by 7.8%.[13] Similarly, if a building's number of people per housing unit increases by 10%, we expect to see hospitalizations per occupant increase by 0.96%. These results are similar to those presented in Table 1 where the effect sizes of mobility patterns are larger compared to housing crowding.

Also, we observe that coefficients on mobility patterns are very stable across specifications, showing small differences that are no larger than 5%. This result implies that for this particular specification, unobservables are not producing meaningful biases in the coefficients. On the other hand, we see more pronounced changes in housing density with a bias of 10% when we compare the results in column (1) to those in column (5). In both cases, we see a bias that shrinks the coefficients towards zero, which is consistent with the fact that people adjust their behaviour counter-cyclically with the evolution of the pandemic.

[13]The average probability of hospitalization per occupant is 1.396e-4, and the average number of pings per building is 0.6.

III.C.1 Building-level weekly analysis

Similarly, motivated by the dynamic evolution of different channels of transmission, we estimate the following equation:

$$\text{hospitalization}_{bt} = \alpha_{1,w(t)}\text{mobility}_{bt} + \alpha_{2,w(t)}\text{housing density}_{bt} + \gamma_{w(t)}X_{j(b)} + \mu_t + \varepsilon_{bt},$$

where coefficients are allowed to change week by week as denoted by subindex $w(t)$.

Appendix Table C2 shows that the coefficients for mobility patterns had a larger impact at early stages of the pandemic and that they decrease in magnitude over time, similar to Glaeser et al. (2020). We also find a similar pattern for housing density. Between week 1 (March 25th to March 31st) and week 4 (April 15th to April 21st), the coefficient for mobility patterns declines by 48% while decreasing by 41% for housing density—suggesting that housing density gained more importance over time with the progression of the pandemic, the issuance of the state-at-home order, and the large economic shock that led to high unemployment levels.

III.D Individual-Level Analysis

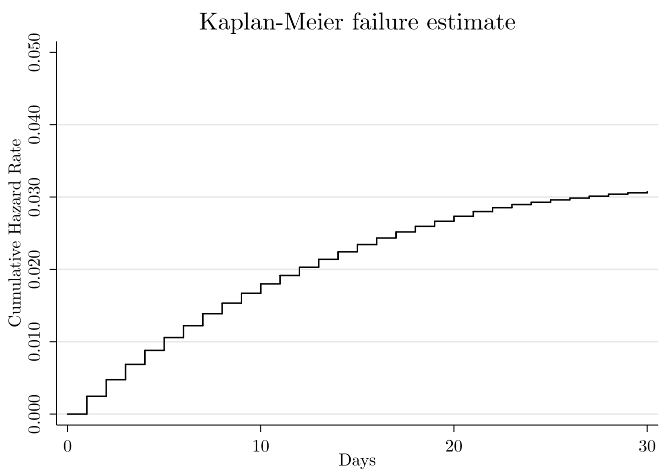
In this section we present results obtained using anonymized individual-level data. Given that we can track individual pings over time, we associate mobility measures to individual phones. While we cannot see whether an individual has been tested, we can observe whether an individual pings inside a hospital (in which case we assign them as a hospitalized individual). Our measurement of individual hospitalizations is an important contribution to the literature, which has generally focused on cases measured at more aggregate levels—and hence has been unable to control for possibly important local covariates. However, we face an important challenge of censoring. The event of being hospitalized due to COVID-19 generally happens only once with a probability that increases over time. For this reason, we borrow tools from the survival analysis literature to appropriately account for this censoring issue as well as for the fact that the probability of this event is not independent of what happened in the past. In doing so, we construct a panel of individuals with the hazard rate of being hospitalized as the outcome variable whose covariates are seven-day moving averages of mobility measures and housing density lagged by two weeks.

We start by plotting Kaplan-Meier graphs with the cumulative probability of failure on a daily time scale. We observe that the probability of being hospitalized increases over time. For the following graphs, we have divided the population into two bins corresponding to above and below median. We do so for two variables: share of pings in HCT and average number of people in the same housing unit.

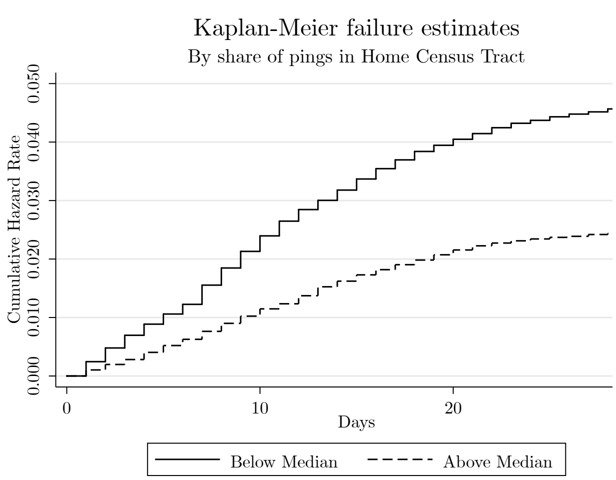
Panel B of Figure 2 shows that spending more time outside of the HCT is associated with a greater cumulative probability of being hospitalized. Similarly, individuals with more people in their housing unit also experience a higher cumulative hazard rate of being hospitalized, as shown in Panel C of Figure 2.

Figure 2: Kaplan-Meier graphs of survival probability

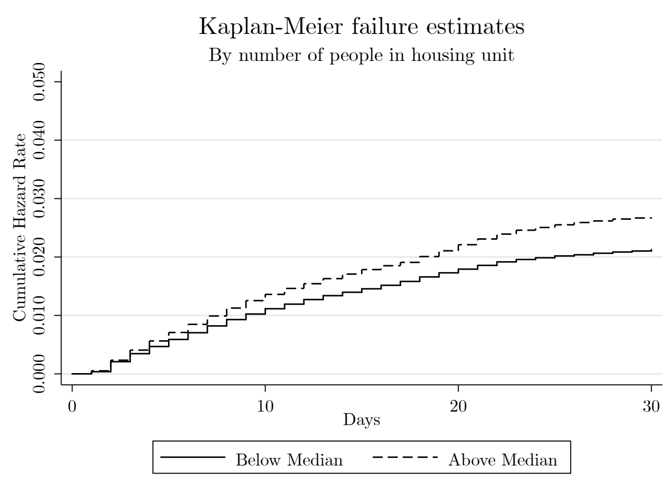
Panel A: Whole Sample



Panel B: By Share of Pings Outside HCT



Panel C: By Average People per Housing Unit



III.E Survival Analysis Regressions

In this section we present estimation results obtained from semiparametric Cox regressions where a failure in our sample is the event of being hospitalized in a panel of individuals. This type of estimation constructs hazard rates of being hospitalized nonparametrically and then uses such hazard rates as the outcome variable in a regression where covariates can be similarly defined as in any standard linear regression.

We first start by looking at all individuals.[14] To understand how mobility and housing crowding measures correlate with each other, we perform our analysis in two steps: (1) including only mobility measures, and (2) including mobility as well as housing crowing measures.

Our estimation results from the Cox regression are presented in Table 3, and they highlight the central role of out-of-home mobility and housing crowding in determining individual hospitalization rates. First, we observe a similar pattern as in Table 2 for mobility patterns in both Panel A and B of Table 3: including demographics and occupational controls decreases the magnitude of the coefficient, suggesting that part of the mobility patterns can actually be mediated by occupations and demographics.

Moreover, we can employ a similar identification strategy as in our aggregate analysis at the Census tract level. Unfortunately, Cox regressions do not allow fixed effects at the same level as the temporal unit level, which in this case is days. Hence, we cannot include day fixed effects. To overcome this problem, we pool days in the same week. We then interact week with ZIP Code to construct fixed effects that control for common factors in a given week and inside a given ZIP Code. The identifying variation comes from differences in individual patterns for those who live in the same ZIP Code in any given week. Our exclusion restriction is that unobservables that correlate with mobility and housing crowding patterns are common to all individuals living in the same ZIP Code in a given week. Reassuringly, we obtain similar results as in Table 2, where we find that the coefficient on mobility increases in magnitude after controlling for such types of unobservables, which possibly indicates that individuals countercyclically adjust their mobility patterns with the evolution of the pandemic.[15]

[14] For 34% of individuals in our sample, we cannot identify a modal building and thus we cannot construct a housing density measure for them.

[15] For full results, see Appendix D.

Table 3: Cox Regression of Mobility on Hospitalization

Panel A: Personal Mobility

Dependent Variable:	Hazard Rate of Hospitalization				
	(1) Baseline	(2) Demographics	(3) Dem. & Occupations	(4) ZIP Code Fixed Effect	(5) ZIP $\times$ Week Fixed Effect
Share of Pings in HCT	-0.224***	-0.216***	-0.215***	-0.211***	-0.284***
	(0.060)	(0.060)	(0.060)	(0.060)	(0.063)
Demographic Controls		✓	✓		
Occupation Controls			✓		
ZIP Fixed Effects				✓	
ZIP $\times$ Week Fixed Effects					✓
Number of Observations	1,798,442	1,795,810	1,795,810	1,798,442	1,795,810

Standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Panel B: Personal Mobility and Housing Crowding

Dependent Variable:	Hazard Rate of Hospitalization				
	(1) Baseline	(2) Demographics	(3) Dem. & Occupations	(4) ZIP Code Fixed Effect	(5) ZIP $\times$ Week Fixed Effect
Share of Pings in HCT	-0.196***	-0.186***	-0.186***	-0.187***	-0.243***
	(0.064)	(0.064)	(0.064)	(0.065)	(0.067)
Log People per Unit	0.023***	0.022***	0.021***	0.025***	0.015***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Demographic controls		✓	✓		
Occupation controls			✓		
ZIP Fixed Effects				✓	
ZIP $\times$ Week Fixed Effects					✓
Number of Observations	1,677,868	1,675,466	1,675,466	1,677,868	1,675,466

Standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The magnitude of mobility patterns decreases when including housing crowding measures when we compare Panel A to Panel B for all specifications. For example, for column (5), the magnitude of mobility patterns decreases by 14%. This pattern indicates that mobility patterns and housing crowding measures share a component that affects the individual level of exposure to the pandemic: if we ignore the housing crowding channel we will overestimate the effect of mobility patterns. However, it is important to disentangle these two channels in order to understand different effects of containment policies. For example, the most common non-pharmaceutical intervention, that is a quarantine on the population, can mitigate contagion by minimizing mobility but will undoubtedly fuel exposure within the household.

In our preferred specification, column (5) of Panel B, a 10% increase in the share of pings inside HCT translates into a hazard rate of being hospitalized that is 1.46%. Similarly, a 10% increase in the number of people per unit leads to a hazard rate of being hospitalized that is 0.15% higher. These estimates are quite large, and they point to economically and statistically important impacts of our estimated risk factors on hospitalization risk.

Following the same specification, column (5) of Panel A in Table 3, we find that individuals at the 10^{th} percentile of share of pings inside HCT have a hazard rate of being hospitalized that is 30% larger than individuals at the 90^{th} percentile when only mobility measures are taken into account. This decreases to 25% when we control for the housing crowding channel, in Panel B . When we do the same comparison for the housing crowding distribution we find a hazard rate of being hospitalized that is 7% higher for individuals at the 90^{th} percentile compared to individuals at the 10^{th} percentile.

IV Conclusion

We document that important inequities in occupations and housing lead to racially disparate outcomes in exposure to COVID-19. We focus on the epicenter of the global pandemic in New York City, showing that infections spread in two waves. First, infections spread through essential workers, who continued to commute to establishments. Next, even after many of these essential workers were laid off, infections continued to spread within more crowded households, and the relative importance of this channel grew.

Racial disparities in infections reflect inequalities in access to both jobs and housing. Black, Hispanic, and low-income workers are more likely to be employed in an essential work occupation and hence exhibit mobility patterns which put them at greater risk of infection in the initial phase of the pandemic. We use novel data drawn from cell phones to measure these mobility patterns, which we use to establish a direct link between outside mobility at both neighborhood and individual levels. Our individual-level analysis advances on prior research using geolocation data by directly linking greater mobility for individual workers and presence in hospitals, controlling for other unobserved local factors.

We also connect both cell phone mobility and Census data on housing occupancy, and we see case increases in the second phase of the pandemic. We find that housing overcrowding predicts a greater caseload, and we also document more Black, Hispanic, and low-income households present in overcrowded buildings. As a result, vulnerable populations are disproportionately burdened by disease exposure through this housing crowding channel.

Our results present a stark contrast to some existing work on the COVID-19 pandemic which highlights the role of static factors such as population density or public transportation. We find that population density per se is not the dominant factor in explaining the cross-section in infections seen throughout this crisis: the densest borough, Manhattan, was relatively less affected. Instead, we find that underlying inequalities in access to jobs and housing explain the racial disparities in outcomes. Crowding and exposure at work, rather than density, best explains the pattern of exposure through the pandemic.

References

Almagro, Milena and Angelo Orane-Hutchinson, "The Determinants of the Differential Exposure to COVID-19 in New York City and Their Evolution Over Time," *Covid Economics, Vetted and Real-Time Papers*, 2020.

Athey, Susan, Billy Ferguson, Matthew Gentzkow, and Tobias Schmidt, "Experienced Segregation," Working Paper 2019.

Borjas, George J., "Demographic determinants of testing incidence and Covid-19 infections in New York City neighbourhoods," *Covid Economics, Vetted and Real-Time Papers*, 2020.

Carozzi, Felipe, Sandro Provenzano, and Sefi Roth, "Urban Density and Covid-19," *CEP Discussion Paper No 1711*, 2020.

Chen, M. Keith and Ryne Rohla, "The effect of partisanship and political advertising on close family ties," *Science*, 2018, *360* (6392), 1020–1024.

_, Judith A. Chevalier, and Elisa F. Long, "Nursing Home Staff Networks and COVID-19," Working Paper 27608, National Bureau of Economic Research July 2020.

_, Kareem Haggag, Devin G. Pope, and Ryne Rohla, "Racial Disparities in Voting Wait Times: Evidence from Smartphone Data," Working Paper 26487, National Bureau of Economic Research November 2019.

Chiou, Lesley and Catherine Tucker, "Social Distancing, Internet Access and Inequality," Working Paper 26982, National Bureau of Economic Research April 2020.

Couture, Victor, Jonathan Dingel, Allison Green, Jessie Handbury, and Kevin Williams, "Measuring movement and social contact with smartphone data: A real-time application to COVID-19," 2020.

Duranton, Gilles and Diego Puga, "The economics of urban density," Technical Report 2020.

García-Lopez, Miquel-Angel and Diego Puga, "Cities reshaped: Mobility patterns and COVID-19's impact on cities," 2020.

Glaeser, Edward L., Caitlin S. Gorback, and Stephen J. Redding, "How Much does COVID-19 Increase with Mobility? Evidence from New York and Four Other U.S. Cities," Working Paper 27519, National Bureau of Economic Research July 2020.

Goolsbee, Austan and Chad Syverson, "Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020," Working Paper 27432, National Bureau of Economic Research June 2020.

Harris, Jeffrey E., "The Subways Seeded the Massive Coronavirus Epidemic in New York City," Working Paper 27021, National Bureau of Economic Research April 2020.

Karaca-Mandic, Pinar, Archelle Georgiou, and Soumya Sen, "Assessment of COVID-19 Hospitalizations by Race/Ethnicity in 12 States," *JAMA Internal Medicine*, 08 2020.

McCormack, Grace, Christopher Avery, Ariella Kahn-Lang Spitzer, and Amitabh Chandra, "Economic Vulnerability of Households With Essential Workers," *JAMA*, 07 2020, *324* (4), 388–390.

McLaren, John, "Racial Disparity in COVID-19 Deaths: Seeking Economic Roots with Census data," Working Paper 27407, National Bureau of Economic Research June 2020.

Price-Haywood, Eboni G., Jeffrey Burton, Daniel Fort, and Leonardo Seoane, "Hospitalization and mortality among black patients and white patients with Covid-19," *New England Journal of Medicine*, 2020.

Rentsch, Christopher T., Farah Kidwai-Khan, Janet P. Tate, Lesley S. Park, Joseph T. King Jr., Melissa Skanderson, Ronald G. Hauser, Anna Schultze, Christopher I. Jarvis, Mark Holodniy et al., "Covid-19 by Race and Ethnicity: A National Cohort Study of 6 Million United States Veterans," *medRxiv*, 2020.

Sá, Filipa, "Socioeconomic Determinants of Covid-19 Infections and Mortality: Evidence from England and Wales," *Covid Economics, Vetted and Real-Time Papers*, 2020.

Schmitt-Grohé, Stephanie, Ken Teoh, and Martín Uribe, "COVID-19: Testing inequality in New York City," Working Paper 27019, National Bureau of Economic Research 2020.

Trivedi, Amal N., Alan M. Zaslavsky, Eric C. Schneider, and John Z. Ayanian, "Trends in the quality of care and racial disparities in Medicare managed care," *New England Journal of Medicine*, 2005, *353* (7), 692–700.

Wong, Mitchell D., Martin F. Shapiro, W. John Boscardin, and Susan L. Ettner, "Contribution of major diseases to disparities in mortality," *New England Journal of Medicine*, 2002, *347* (20), 1585–1592.

Online Appendix

A Data Appendix

Table A1: Summary Statistics

Variable	**Mean**	**Std. Dev.**	**p10**	**Median**	**p90**
Panel A: Housing Crowding Measures					
People per Unit	0.043	0.228	0.000	0.000	0.031
Residential Units per Building	8.306	4.420	1.000	2.000	9.000
Residential Area (sqft)	8071	38,034	1140	2160	9011
Panel B: Individual Level Mobility					
Average Number of Pings in Home Tract	0.62	0.41	0.00	0.80	1.00
Average Distance From Home Tract (km)	2.42	4.50	0.00	0.22	8.13
Number of Hours Entirely Outside of Home Tract	1.46	2.37	0.00	0.00	5.00
Panel C: Other Variables and Local Controls					
Share of Positive Tests	0.563	0.085	0.438	0.583	0.645
Tests per Capita	0.018	0.006	0.012	0.017	0.026
Median Income (in $1000ss)	68.604	31.878	34.122	62.202	115.084
Share $\geq$ 20, $\leq$ 40	0.323	0.084	0.246	0.308	0.433
Share $\geq$ 40, $\leq$ 60	0.258	0.033	0.220	0.261	0.296
Share $\geq$ 60	0.200	0.079	0.132	0.190	0.276
Share Male	0.477	0.029	0.446	0.479	0.508
Household Size	2.683	0.537	1.930	2.750	3.300
% Black	0.200	0.240	0.010	0.076	0.600
% Hispanic	0.263	0.195	0.078	0.189	0.634
% Asian	0.144	0.139	0.017	0.094	0.335
Density (in 1000s of people per unit)	43.380	31.045	10.784	36.639	90.075
% Public Transport	0.532	0.150	0.312	0.543	0.712
Commuting Time (in mins)	40.647	7.054	27.200	42.100	48.100
% Uninsured	0.089	0.043	0.042	0.084	0.143
% Essential: Professional	0.126	0.089	0.046	0.092	0.285
% Essential: Service	0.065	0.033	0.035	0.060	0.107
% Essential: Technical	0.014	0.009	0.004	0.013	0.022
Non-Flexible Occupations:					
- % Health Practitioners	0.029	0.018	0.009	0.026	0.050
- % Other Health	0.038	0.024	0.010	0.035	0.073
- % Firefighting	0.012	0.009	0.003	0.012	0.023
- % Law Enforcement	0.007	0.007	0.001	0.006	0.014
- % Ind. and Construction	0.054	0.027	0.014	0.056	0.090
- % Transportation	0.029	0.016	0.004	0.032	0.048
- % Non Ess.: Professional	0.279	0.075	0.195	0.271	0.359
- % Science Fields	0.006	0.007	0.001	0.004	0.015
- % Law and Related	0.018	0.026	0.003	0.008	0.049
- % Non Ess.: Service	0.032	0.013	0.016	0.032	0.047

Figure A1: Time series of share of positive tests, mobility patterns, and housing density

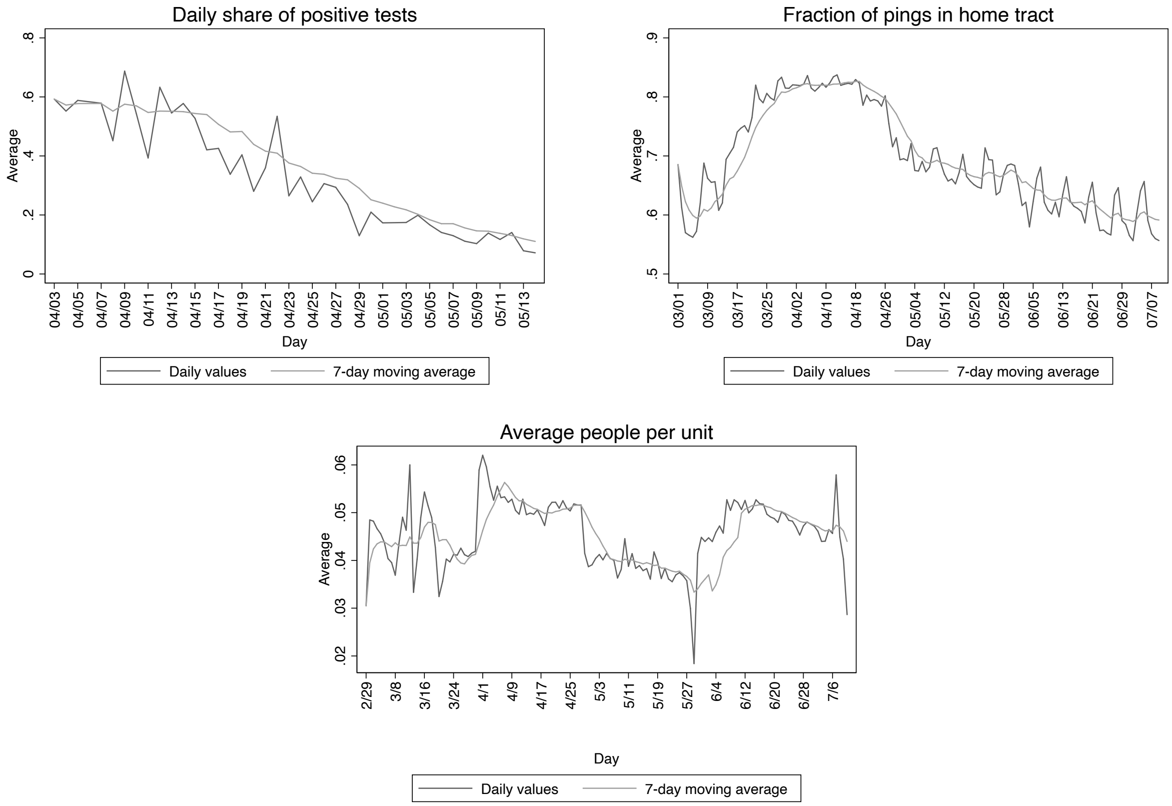


Figure A2: Comparing Hospitalizations in Mobile Phone Sample

These graphs plots the time series for our individual-level mobility-derived measure of hospitalization in comparison with the official figures provided by the DOH. The correlation between the two is 0.79

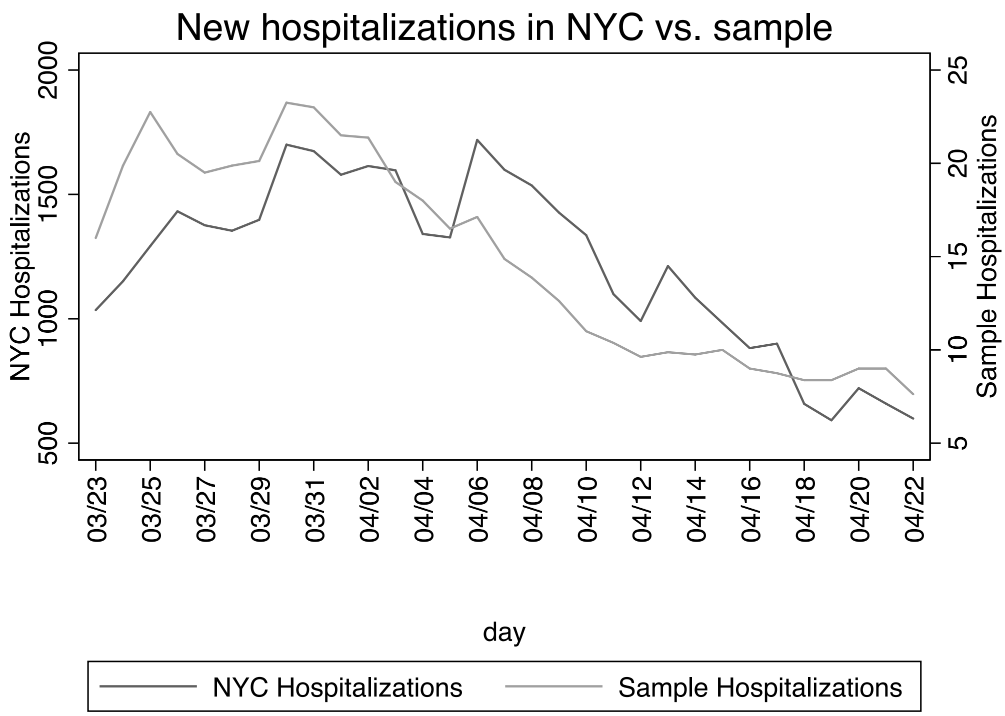
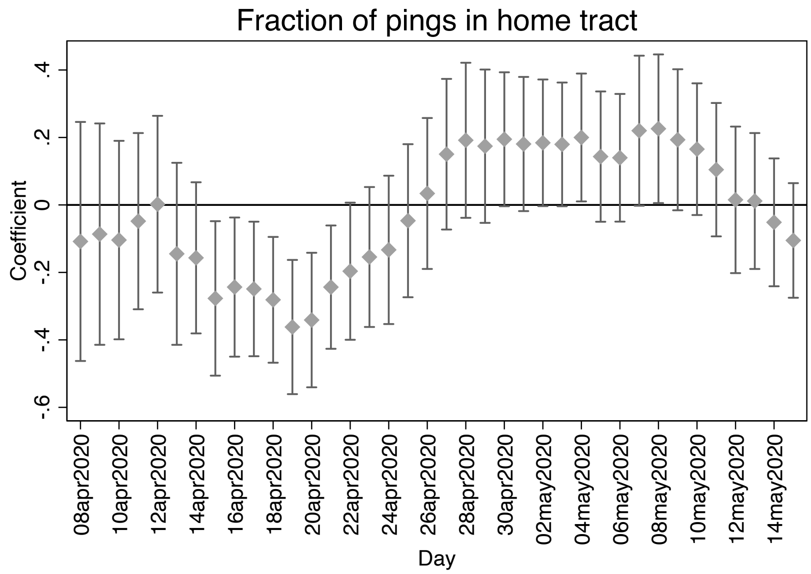
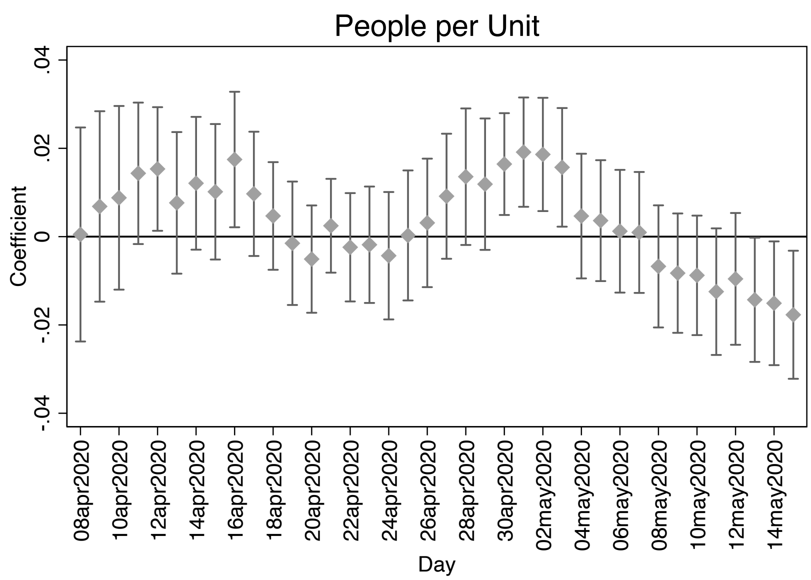


Figure A3: Time series of mobility and crowding coefficients

Panel A: Coefficient of share of pings in HCT on daily positive share



Panel B: Coefficient of log of people per unit on daily positive share



B Time Series and Cross-Section of Mobility measures

Figure B1: Time Series of Mobility Measures

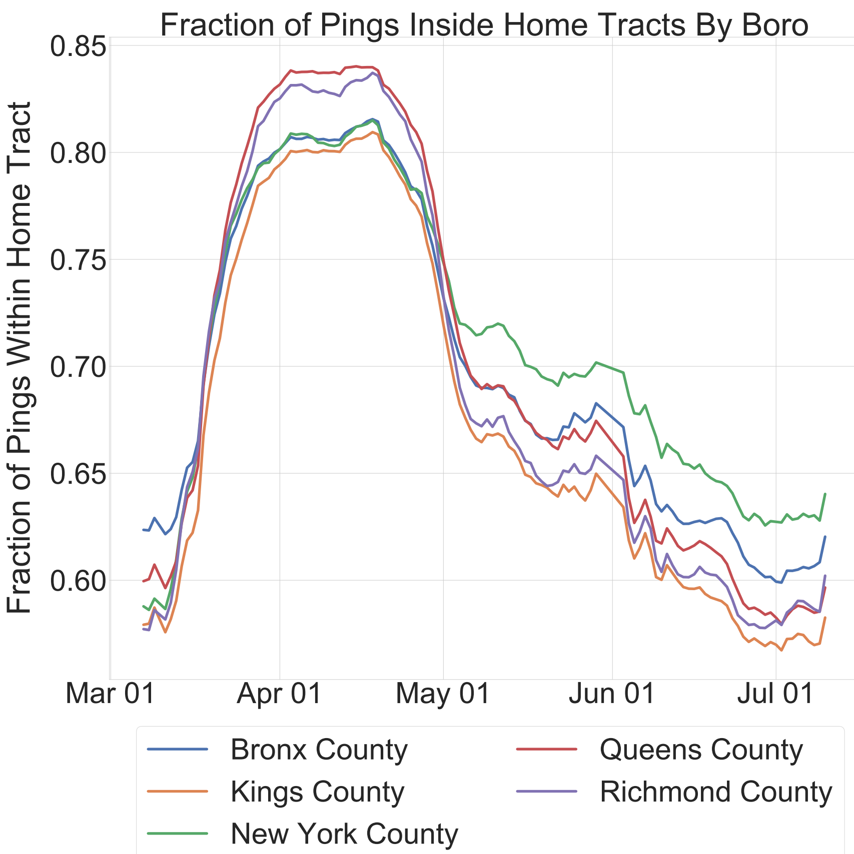


Figure B2: Cross-Section of Housing Crowding Measure

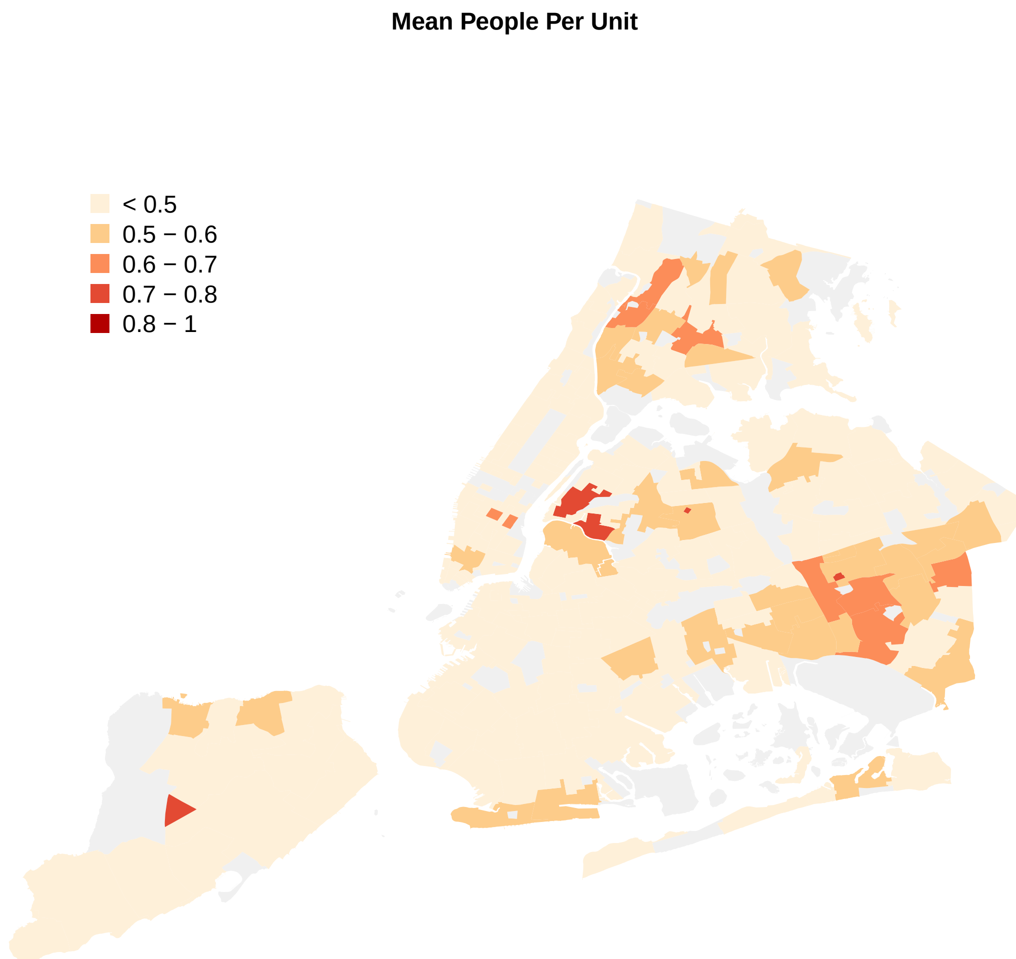
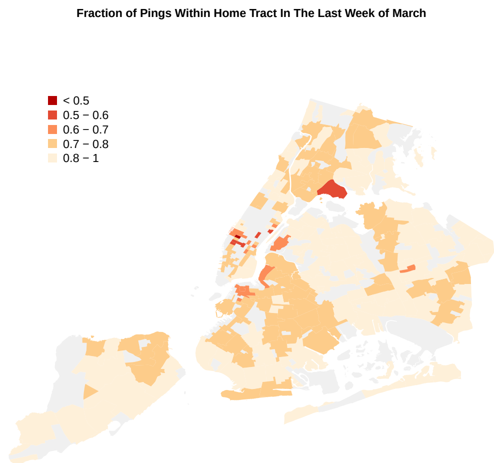
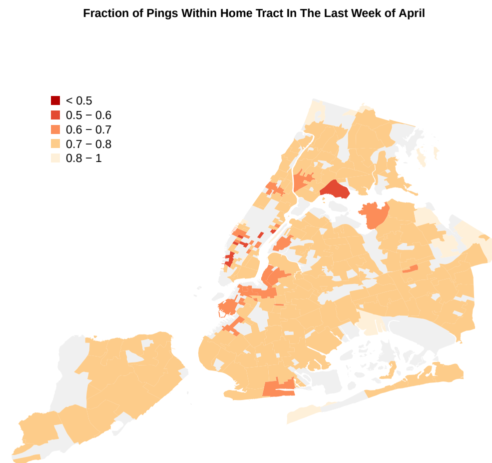


Figure B3: Cross-Section of Outside of HCT Mobility Measure

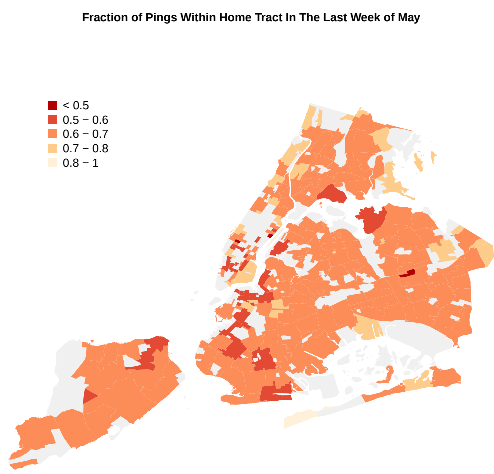
Panel A: March



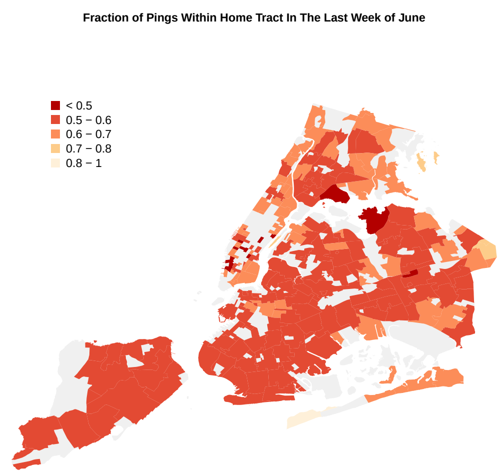
Panel B: April



Panel C: May



Panel D: June



C Building Level Analysis

Table C1: Dependent Variable: Hospitalizations per Occupancy

	(1) Mobility	(2) Demographics	(3) Demographics & Occupations	(4) Mobility & Demographics	(5) Mobility, Dem. & Occupations	(6) Mobility
Fraction of Pings in Home Tract	-1.815e-04***			-1.766e-04***	-1.770e-04***	-1.837e-04***
	(2.624e-05)			(2.620e-05)	(2.626e-05)	(2.766e-05)
Log People per Unit	1.226e-05***			1.283e-05***	1.281e-05***	1.347e-05***
	(2.313e-06)			(2.329e-06)	(2.333e-06)	(2.482e-06)
Log Income		1.669e-05	3.709e-05	1.977e-05	3.564e-05	
		(3.674e-05)	(4.768e-05)	(3.673e-05)	(4.768e-05)	
Share $\geq 20, \leq 40$		8.318e-05	1.696e-04	7.534e-05	1.611e-04	
		(2.365e-04)	(2.459e-04)	(2.365e-04)	(2.458e-04)	
Share $\geq 40, \leq 60$		2.038e-04	3.018e-04	1.846e-04	2.891e-04	
		(2.898e-04)	(2.924e-04)	(2.894e-04)	(2.922e-04)	
Share ≥ 60		2.264e-05	2.130e-04	2.546e-05	2.091e-04	
		(2.284e-04)	(2.477e-04)	(2.284e-04)	(2.476e-04)	
Share Male		-4.539e-04	-5.037e-04	-4.554e-04	-5.023e-04	
		(2.839e-04)	(2.915e-04)	(2.841e-04)	(2.915e-04)	
Log Household Size		-1.076e-06	1.277e-05	-1.182e-05	7.429e-06	
		(7.604e-05)	(8.391e-05)	(7.582e-05)	(8.380e-05)	
% Black		1.478e-05	2.735e-05	3.124e-06	1.856e-05	
		(5.735e-05)	(6.633e-05)	(5.751e-05)	(6.643e-05)	
% Hispanic		1.374e-05	5.078e-07	1.355e-05	2.703e-07	
		(8.876e-05)	(9.829e-05)	(8.877e-05)	(9.829e-05)	
% Asian		-1.067e-04	-1.388e-04	-1.018e-04	-1.345e-04	
		(8.472e-05)	(8.754e-05)	(8.462e-05)	(8.746e-05)	
% Public Transport		-7.439e-05	-5.389e-05	-6.890e-05	-4.948e-05	
		(1.031e-04)	(1.074e-04)	(1.030e-04)	(1.073e-04)	
Log Commute Time		3.133e-05	7.468e-05	3.125e-05	7.561e-05	
		(7.924e-05)	(8.154e-05)	(7.970e-05)	(8.192e-05)	
% Uninsured		7.045e-04*	7.817e-04*	7.054e-04*	7.798e-04*	
		(2.889e-04)	(3.044e-04)	(2.889e-04)	(3.044e-04)	
Bronx		-4.522e-05	-4.126e-05	-4.450e-05	-4.026e-05	
		(5.039e-05)	(5.096e-05)	(5.021e-05)	(5.078e-05)	
Brooklyn		7.926e-06	5.342e-06	9.103e-06	6.997e-06	
		(4.639e-05)	(4.764e-05)	(4.622e-05)	(4.750e-05)	
Queens		-5.847e-05	-6.588e-05	-5.814e-05	-6.561e-05	
		(4.749e-05)	(4.954e-05)	(4.735e-05)	(4.941e-05)	
Staten Island		-6.500e-05	-8.138e-05	-6.649e-05	-8.162e-05	
		(5.649e-05)	(5.927e-05)	(5.641e-05)	(5.919e-05)	
% Flexible Occupations			-2.138e-04		-1.953e-04	
			(1.781e-04)		(1.777e-04)	
% Health Practitioners			-3.211e-06		3.794e-05	
			(4.518e-04)		(4.522e-04)	
% Other Health			-6.136e-04		-6.307e-04	
			(3.278e-04)		(3.279e-04)	
% Firefighting			-1.482e-03*		-1.517e-03*	
			(7.495e-04)		(7.502e-04)	
% Law Enforcement			9.929e-04		1.048e-03	
			(9.474e-04)		(9.484e-04)	
% Essential: Service			-2.094e-04		-1.834e-04	
			(3.328e-04)		(3.330e-04)	
% Non Ess.: Service			-3.793e-04		-3.278e-04	
			(5.994e-04)		(5.995e-04)	
% Ind. and Construction			-3.336e-04		-3.357e-04	
			(3.358e-04)		(3.359e-04)	
% Essential: Technical			-1.397e-03		-1.365e-03	
			(7.253e-04)		(7.255e-04)	
% Transportation			4.253e-04		4.248e-04	
			(4.789e-04)		(4.791e-04)	
Constant	3.135e-04***	1.054e-04	-3.692e-05	2.832e-04	1.373e-04	3.191e-04***
	(2.873e-05)	(3.367e-04)	(3.527e-04)	(3.389e-04)	(3.547e-04)	(2.979e-05)
Day FE	✓	✓	✓	✓	✓	
Demographic Controls		✓	✓	✓	✓	
Occupation Controls			✓		✓	
Census Tract × Day FE						✓
N	1354631	1354631	1354631	1354631	1354631	1354631

Table C2: Weekly Analysis of Mobility Exposures and Hospitalization

Dependent Variable:	Hospitalizations per Occupancy			
	(1) Mar 25–31	(2) Apr 1–7	(3) Apr 8–14	(4) Apr 15–21
Fraction of Pings in Home Tract	-1.962e-04***	-2.715e-04***	-1.890e-04***	-1.020e-04**
	(6.609e-0.5)	(6.700-06)	(5.070e-0.5)	(4.990-0.5)
Log People per Unit	1.220e-05***	2.370e-05***	1.320e-05***	7.210e-06***
	(5.860e-06)	(5.460e-06)	(3.760e-06)	(4.810e-06)
Census Tract $\times$ Day FE	✓	✓	✓	✓
N	336,455	312,529	288,467	280,810

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D Survival Analysis

Table D1: Cox Regression of Mobility on Hospitalization

Dependent Variable:	Hazard Rate of Hospitalization									
	(1) Baseline		(2) + Demographics		(3) + Demographics & Occupations		(4) ZIP Code Fixed Effects		(5) ZIP Code $\times$ week Fixed Effects	
Fraction of Pings in Home Tract	-0.224***	(0.060)	-0.215***	(0.060)	-0.216***	(0.060)	-0.211***	(0.060)	-0.284***	(0.063)
Log Income			0.396***	(0.079)	0.661***	(0.104)				
Share $\geq 20, \leq 40$			-1.247*	(0.562)	-1.058	(0.571)				
Share $\geq 40, \leq 60$			-3.949***	(0.703)	-4.027***	(0.716)				
Share ≥ 60			-1.123	(0.580)	-0.501	(0.623)				
Share Male			-3.166***	(0.642)	-3.370***	(0.666)				
Log Household Size			1.113***	(0.200)	0.852***	(0.215)				
% Black			0.695***	(0.122)	0.450**	(0.141)				
% Hispanic			-0.689***	(0.201)	-1.052***	(0.217)				
% Asian			0.535*	(0.212)	0.299	(0.219)				
% Public Transport			0.051	(0.215)	0.260	(0.236)				
Log Commute Time			-0.092	(0.183)	-0.299	(0.198)				
% Uninsured			2.903***	(0.527)	2.196***	(0.600)				
Bronx			-0.192*	(0.084)	-0.124	(0.087)				
Brooklyn			-0.460***	(0.074)	-0.422***	(0.077)				
Queens			-0.807***	(0.081)	-0.762***	(0.083)				
% Flexible Occupations					-1.457***	(0.385)				
% Health Practitioners					-0.440	(1.128)				
% Other Health					-0.180	(0.786)				
% Firefighting					2.174	(1.712)				
% Law Enforcement					2.775	(2.333)				
% Essential: Service					1.587*	(0.742)				
% Non Ess.: Service					-1.167	(1.279)				
% Ind. and Construction					-0.057	(0.843)				
% Essential: Technical					-3.300	(1.962)				
% Transportation					0.168	(1.050)				
Demographic Controls			✓		✓					
Occupation Controls					✓					
ZIP Code FE							✓			
ZIP Code $\times$ week FE									✓	
Number of Observations	1798442		1795810		1795810		1798442		1795810	

Standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D2: Cox Regression of Mobility and Housing Crowding on Hospitalization

Dependent Variable:	Hazard Rate of Hospitalization									
	(1) Baseline		(2) + Demographics		(3) + Demographics & Occupations		(4) ZIP Code Fixed Effects		(5) ZIP Code × week Fixed Effects	
Fraction of Pings in Home Tract	-0.196**	(0.064)	-0.186**	(0.064)	-0.186**	(0.064)	-0.187**	(0.065)	-0.243***	(0.067)
Log People per Unit	0.023***	(0.006)	0.022***	(0.006)	0.021***	(0.006)	0.025***	(0.006)	0.015*	(0.006)
Log Income			0.450***	(0.084)	0.668***	(0.110)				
Share $\geq 20, \leq 40$			-1.592**	(0.595)	-1.337*	(0.605)				
Share $\geq 40, \leq 60$			-4.615***	(0.745)	-4.671***	(0.760)				
Share ≥ 60			-1.515*	(0.615)	-0.909	(0.662)				
Share Male			-2.978***	(0.679)	-3.188***	(0.705)				
Log Household Size			0.996***	(0.212)	0.767***	(0.228)				
% Black			0.748***	(0.130)	0.544***	(0.151)				
% Hispanic			-0.564**	(0.213)	-0.895***	(0.230)				
% Asian			0.672**	(0.225)	0.441	(0.234)				
% Public Transport			0.084	(0.228)	0.275	(0.250)				
Log Commute Time			0.043	(0.195)	-0.130	(0.211)				
% Uninsured			2.587***	(0.562)	1.830**	(0.640)				
Bronx			-0.266**	(0.089)	-0.199*	(0.092)				
Brooklyn			-0.492***	(0.079)	-0.457***	(0.081)				
Queens			-0.831***	(0.085)	-0.786***	(0.088)				
% Flexible Occupations					-1.323**	(0.408)				
% Health Practitioners					0.468	(1.179)				
% Other Health					-0.526	(0.840)				
% Firefighting					2.436	(1.819)				
% Law Enforcement					1.719	(2.497)				
% Essential: Service					1.289	(0.788)				
% Non Ess.: Service					-0.867	(1.359)				
% Ind. and Construction					0.450	(0.895)				
% Essential: Technical					-4.681*	(2.099)				
% Transportation					0.685	(1.115)				
Demographic Controls			✓		✓					
Occupation Controls					✓					
ZIP Code FE							✓			
ZIP Code × week FE									✓	
Number of Observations	1677868		1675466		1675466		1677868		1675466	

Standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$