Demons of Density:
Do Higher-Density Environments Put People at Greater Risk of Contagious Disease?

Ingrid Gould Ellen
NYU Wagner

Renata Howland
NYU Wagner

Sherry Glied
NYU Wagner

December 10, 2022

Abstract

We study the relationship between density and COVID during three distinct waves of the pandemic in New York City. Unlike prior work, our analysis uses individual Medicaid claims records, which include a rich array of demographic characteristics and pre-existing medical conditions and cover a near universe of low-income New Yorkers. In brief, our results suggest that living in higher density neighborhoods did not heighten the risk of COVID hospitalization. The size of a multifamily building made little difference either, and people living in public housing developments, which are typically highly dense environments, were *less* likely to be hospitalized for COVID. However, while neighborhood and building density do not seem to matter, we find significant, positive relationships between COVID hospitalization rates and household size. Specifically, we see that people living in large households or in neighborhoods with high levels of crowding were more likely to be hospitalized for COVID. In other words, our results suggest that crowded living quarters – which can occur at any level of population density – and not density itself, increase the risk of COVID hospitalization. We also see a strong correlation between being unstably housed or living in institutional settings and COVID hospitalizations.

Acknowledgments
We are grateful to the Russell Sage Foundation for their support of this research. The views and opinions expressed in this article are those of the author(s) and do not necessarily reflect the official policy or position of the New York State Department of Health.

Urban economists have provided ample evidence about the agglomeration economies that come from higher density environments. In areas with a greater concentration of economic activity, firms and workers are more productive, and consumers enjoy a richer and more diverse set of consumer amenities. Critically, density also provides environmental benefits as people living in denser cities drive less and are more likely to live in multifamily buildings (Duranton & Puga, 2020). But density may also bring "demons," such as high housing costs, traffic, crime and most of all, contagious disease (Glaeser, 2011; Glaeser and Cutler, 2021).

In the age of COVID-19, many commentators have asserted that high density living environments put people at greater risk of contracting COVID-19 (Olsen, 2020). The relationship seems intuitive. The more people you live near, the more likely you are to be exposed to the virus. The calamitous initial COVID outbreak in New York City buttressed this view. Later in the pandemic, however, severe outbreaks in much less densely populated areas, such as South Dakota, cast doubt on the initial assumptions. A growing literature examines the relationship between density and COVID (and other infectious disease) transmission, severity, and mortality. Most existing studies, however, use aggregated population data to measure outcomes. This limits their ability to control for differences in the underlying risks of people living in different areas and assess how different components of density affect COVID outcomes. Early studies also treat the pandemic as a single event. It is plausible, however, that the relationship between density and disease is modified by public health control measures.

We extend earlier work on the relationship between COVID and density using Medicaid claims records, which cover a near-universe of poor and near-poor residents of New York City. Unlike most prior work that has relied on aggregated data, these individual-level data allow us to control for demographics and pre-existing medical conditions. Moreover, by using Medicaid data, we limit our analysis to a group that is relatively economically homogenous -- lower-income individuals eligible for Medicaid – which helps in addressing unobserved differences between those living in higher and lower density environments. Second, in contrast to earlier work, we focus on hospitalizations rather than infection rates to capture more serious cases, which helps to address any concerns about differences in testing prevalence across places, and specifically, the likelihood that people living in higher density areas will have easier access to testing sites. Third, we have access to precise residential addresses and so can observe both the size of an individual's building and its subsidy status (public housing, privately owned subsidized, or unsubsidized), which helps to control for micro-neighborhood density as well as unobserved individual characteristics. Fourth, as explained below, we separately examine three distinct waves of the pandemic to capture potential variation over time. Fifth, we consider density at the neighborhood level rather than at the level of a county, as most previous studies have done, which takes advantage of the large variation in density within jurisdictions, particularly in a place like New York City, and more accurately targets the density that people actually experience on a daily basis. Finally, we consider different types of density. While Duranton and Puga (2020) write that "high density is synonymous with crowding," we distinguish between household size, building size and overall population density.

We focus on the three distinct waves of COVID that occurred in 2020, before the introduction of COVID vaccines in late December 2020. In this way, we separately examine whether density was correlated with COVID hospitalizations before most of the protective measures were put in place and people had much chance to adjust their behavior versus after these measures were put into place. We do not look beyond 2020, because the relationship between density and COVID in the post-vaccine era would likely be confounded by the effects of density on the availability of vaccines (likely positive) and by the political polarization of vaccine uptake. It is possible that later mutations of the virus would respond differently to density than did those studied in this analysis.

In brief, across three different COVID waves, we see no evidence that people living in higher density neighborhoods are at greater risk of hospitalization. Further, while people living in single-family homes are less likely to be hospitalized for COVID, there is no difference in hospitalization rates among people living in small or large multifamily buildings as compared to those living in 2-4 family homes. And people living in public housing developments, which are highly dense environments with large concentrations of elevator buildings in close proximity, were somewhat *less* likely to be hospitalized for COVID.

However, while neighborhood and building density do not seem to matter, we do find persistent and positive relationships between COVID hospitalization rates and household size, as proxied by the number of people sharing an individual's same Medicaid case number and using a neighborhood-level crowding measure. Specifically, we see that people living in large Medicaid households or in neighborhoods with high levels of crowding were more likely to be hospitalized for COVID. In other words, our results suggest that crowded living quarters – which can occur at any level of population density – and not density itself, increase the risk of COVID hospitalization. We also see a strong correlation between being unstably housed or living in institutional settings and COVID hospitalizations.

Unpacking Density in the Context of Infectious Disease

Researchers commonly capture density by simply dividing the population of a city by its land area. This measure, however, fails to capture the environments that people actually experience and that may contribute to the spread of infectious disease. For one thing, researchers often capture density at the level of the city or county, while people typically inhabit only a small part of their city on a day-to-day basis. Beyond this aggregation problem, density is likely to affect outcomes differently if it occurs within a housing unit, a building or a neighborhood. Very different building configurations can lead to the same overall city density, depending on design and the composition of households. Simplistically, neighborhood population density can be decomposed into three factors:

Neighborhood density =

People/housing unit * Housing units/building * Buildings/neighborhood land area

All three of these components of density could elevate contagion risk through increasing the number of interactions with other people, whether on sidewalks or in subways in dense neighborhoods, in hallways or elevators in multifamily buildings, or around kitchen tables in large households. But the nature of these interactions is quite different, and thus the experience of density may be quite different in two neighborhoods with identical population densities if one is high density because people live in crowded homes, while the other is high density because the neighborhood contains multifamily buildings. Similarly, the experience of density may differ in a neighborhood with many row homes versus one with the same overall density but with a few large multifamily buildings surrounded by a lot of open space.

These differences may matter for public health too. Growing evidence suggests that much of the spread of COVID occurs through extended interactions within small groups and households, such as in circumstances where people are talking to one another (Hobbs et al., 2020; Whaley, Cantor, Pera, & Jena, 2021). In densely populated neighborhoods, people interact frequently but usually without much communication on streets and in stores. They are also more likely to rely on public transit for commuting, but such commutes generate multiple brief, usually silent interactions with strangers (we include the proportion of residents in the ZIP Code using different commuting modes in analyses reported in the appendix). In multifamily buildings, people are exposed to neighbors in elevators and laundry rooms, but these interactions are again, often brief and silent. In large and crowded households, interactions are more frequent, longer, and more voluble.

Behavioral responses to contagious disease may also differ in higher and lower density environments. For example, density may facilitate the spread of information about virus threats and strategies for protection (Tan et al, 2020). Density may increase the salience of disease, prompting prevalence-elastic responses (Philipson, 2000). Further, people living in higher density areas could have better access to protective equipment, like masks, and to testing sites both of which can help to slow the spread of a virus (Callaghan et al, 2021; Tan et al, 2020). Finally, because of greater perceived risks, policymakers in higher density areas might enact more stringent shut-down laws and mask mandates (or enforce the laws enacted), or people living in higher density areas and in multifamily buildings could simply adopt behavioral responses that shield themselves from exposure. These behavioral responses have limits, however, especially in one's home. It is far easier to shield yourself from exposure to other people outside of the home than it is inside the home. Inside the home, you need to share food, sleeping spaces and bathrooms. Outside the home, you can more easily wear masks and take other precautions. Thus, there are good reasons to think that household size or crowding are more important predictors of COVID hospitalizations than either the number of units in a building or the overall density of a neighborhood. Yet the research on density and disease has failed to separately consider these different components of density. Instead, studies have simply included measures of overall population density, generally at the county level (e.g., Carozzi, 2020; Chen & Li, 2020; Hamidi, Sabouri, & Ewing, 2020; Sy, White, & Nichols, 2021).

A few prior studies have found a positive association between county density and infection rates (Sy et al., 2021). Using county-level estimates of new deaths per population, Chen et al. (2020) find that the spread rate of SARS-CoV-2 was 8 times higher in high density compared to low density counties in the early part of the pandemic. Wong et al., (2020) also find that population density alone accounted for 57 percent of the variation in aspatial models of county-level weekly rates of cumulative infection, and 75 percent in models that include spatial error correlations. But other studies suggest density is not the culprit. Carozzi et al (2020) conclude that density levels have predicted the timing of outbreaks, but not overall infection and death rates. Specifically, higher density areas saw earlier outbreaks, but over time, any differences between high- and low-density areas converged, perhaps because residents of higher density areas took greater precautions, or perhaps because overall population density is simply not that correlated with the kind of indoor interaction that seems to fuel transmission. Hamidi et al (2020) find that case rates are associated with metropolitan area size (which they take as a proxy for connectivity or rates of population flows in and out of an area) but not population density. Their work focused on the early part of the COVID pandemic, however, so it's possible that connectivity invites earlier outbreaks but does not necessarily shape overall infection or mortality rates.

Finally, Hamidi and Hamidi (2021) specifically study ZIP Code-level infection rates in New York City during the first half of 2021, and similar to our analysis, distinguish between household crowding and population density. They find that crowding (rather than population density) was associated with ZIP code level cumulative infection rates. Average household size and racial/socioeconomic compositions were also significant predictors of spatial variation in infections rates, whereas public transit use and nursing home bed rates were not. But they rely on aggregate infection rates, so they are not able to control for individual characteristics or the availability of testing sites. No prior study that we know of considers the association between COVID hospitalizations and household size, building size, as well as overall neighborhood density.

We add to the existing literature by examining how an individual's likelihood of being hospitalized from COVID changes with population density, the number of units in a building, building subsidy status, and Medicaid case size (a proxy for household size). Using Medicaid claims data, we are able to control for age, race/ethnicity, pre-existing health conditions, building age, and neighborhood demographics. We also look separately at COVID hospitalizations during the first wave of COVID in New York City, when knowledge of risk was very limited and virtually no public health policies had yet to be put in place. If density were to matter, we might see it in this early period, before people living in higher density areas did much to adapt their behavior and take additional precautions. We then look at two later waves of COVID, after public health measures had been implemented. Specifically, the New York State on PAUSE (or Policy to Assure the Uniform Safety of Everyone) Act was enacted March 22, 2020. This 10-point policy recommended sanitary practices, restricted social gathering, closed non-essential businesses for 90 days. In New York City, schools were closed for in-person instruction from March 16, 2020 through September 2020. Restaurants, theaters, and similar venues were shut down on March 17, 2020. The state's mask mandate was announced on April 15th and went into effect on April 17th. Following prior research, which describes a median time between infection and hospitalization as 12 days, we classify the second wave of the pandemic as beginning May 1 (Gupta, Archelle, Soumya, Kosali, & Pinar, 2021).

Data

In order to analyze the relationship between population density and COVID-19 hospitalizations in a low-income population, we use Medicaid claims from New York State, which include the universe of encounter and fee-for-service claims from outpatient and inpatient settings. We obtained these data from the Health Evaluation and Analytic Lab (HEAL) at New York University (NYU), which processes the raw Medicaid claims from the New York State Medicaid Data Warehouse into a series of analytical files. Although all addresses in Medicaid are geocoded by the state with latitude and longitude, HEAL joins these values with the TIGER shape files from the US Census Bureau in order to tag the data with other geographic indicators, such as census tract, ZIP Code, and county. In addition, we use the NYC Geosupport software (https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-gde-home.page) to assign a parcel number (Borough-Block-Lot, or BBL) to each address to link to building-level information.

Using eligibility information from Medicaid, we create a sample of 1,865,729 adults (18 years or older) living in New York City and continuously enrolled in both 2019 and 2020 (meaning they had 10 or more months of enrollment in each calendar year). We use the assigned geographic indicators to match claims with other sources of neighborhood and building level data described below. To allow for comparison across relevant indicators, we omitted individuals without a reported Medicaid case number (n=16,049), those living in census tracts with fewer than 200 people (n=6004), and those with P.O. Boxes or nonsensical (primarily numeric) addresses (n=9468).

We identify inpatient hospitalizations with COVID-19 using a newly established International Classification for Disease Version 10 (ICD-10-CM) code U07.1, which was recommend for use by the Centers for Disease Control as of March 18, 2020. This code is designed to capture laboratory-confirmed cases of COVID-19 (separate codes are used for the screening of those potentially exposed) and a validation study in inpatient records suggests that it has high sensitivity and specificity compared to microbiology reports (Kadri et al., 2020).

We define the three waves of COVID as January 1 to April 30, May 1 to August 31, and September 1 to December 31, 2020. As noted above, we divided the year into thirds, with first wave roughly aligned with the peak of cases and hospitalizations and the period before the public health measures could have had much effect. The second wave corresponds to the time period after public health measures had been put in place in late March and April, and when hospitalizations fell significantly. The third wave aligns with the increases in infection and hospitalization rates that occurred late in the year.

We create individual-level controls from Medicaid eligibility files, including age as of January 1, 2020 (categorized as 18-39, 40-49, 50-59, 60-69, 70-79, and 80+), race/ethnicity (White, non-Hispanic Black, Hispanic, Asian and Pacific Islander (API), Other and unknown), and dual Medicaid/Medicare eligibility. We also used one year of pre-COVID claims (2019) to create binary flags for eight relevant pre-existing conditions (heart disease, asthma, diabetes, cancer, kidney disease, liver disease, dementia, and arthritis). We created these definitions using the clinical classification codes created by Healthcare Cost and Utilization Project, which groups related ICD-10-CM diagnosis into clinically-meaningfully categories. Individuals with one or more codes for a given a condition in 2019 were categorized as having that condition. Since historical diagnoses may be carried forward in time, these codes do not represent incident cases and may overestimate the actual burden of disease in the population compared to clinical estimates.

Our key independent variables include population density, building size and subsidy status, and our proxy for household size. We measure overall neighborhood population density as the total number of residents per square mile, using population estimates from the 5-year 2011-2015 American Community Survey (ACS) and land area at the census tract level from Census Shape files (2010). We create neighborhood composition variables from ACS data, including the poverty rate, and the percentages of the census tract population that is poor, college educated, employed, and foreign born. In analyses reported in the appendix, we also include ACS estimates of primary commuting modality shares in the ZIP Code tabulation area. For building size, age, and subsidy status, we draw on the Primary Land Use Tax Lot Output (PLUTO) dataset maintained by NYC agencies, and building-level rent subsidy information from the NYU Furman Center. Specifically, we create a set of indicator variables for building size based on the number of residential units (categorized as 1 unit, 2-4 units, 5-49 units, and 50+ units). We also identify buildings by subsidy type (Public Housing, privately owned subsidized, or no subsidy) and the year built (before 1940, 1940-2000, or after 2000).

Finally, we estimate household size at the individual level and household crowding at the neighborhood level. At the individual-level, we use the Medicaid case number (which is the same for groups of individuals applying for Medicaid together) as a proxy for number of people in a given household. That is, for a given individual in the sample, we count the total number of individuals in the Medicaid database that had the same case number as of January 1, 2020. We then create a set of indicators variables capturing how many Medicaid recipients are on the same case number (1, 2, 3, and 4+ people), which we refer to throughout the paper as the Medicaid case size. Everyone on a case number at the same time has the same reported addresses in Medicaid, so people with the same case number are clearly in the same household. That said, the case size is likely to be an undercount of household members because not all members of a household members will be covered by Medicaid or covered under the same Medicaid case number at the same time; we do not believe that the omission of some household members introduces any systematic bias.

For the neighborhood measure of crowding incidence, we use the ACS data to create a measure of neighborhoods that have 30% of more of residents reporting overcrowding, or more than 1 person per room.[1] Finally, using the street level addresses we also create measures for individuals who are unstably housed s (e.g., 'General Delivery' or 'Undomiciled'), those in congregate living situations (e.g., jails, psychiatric institute, rehabilitation center offices), and those with other administrative addresses (e.g., buildings occupied by city agencies, like the Human Resources Administration or Administrative for Child Services). Our listing of congregate situations captures most settings in which large numbers of Medicaid beneficiaries live, but may not be comprehensive of all congregate situations.

Table 1 displays summary statistics. It shows that 0.82 percent (n=14,990) of our sample was hospitalized for COVID during 2020. More than half (n=8,425) of all hospitalizations occurred during the first third of 2020, 18 percent were hospitalized during the middle third, and 26 percent were hospitalized in the last third of the year. About 39 percent of the population is between the age of 18 and 39, and about 30 percent is age 60 or older. (Note that about a quarter

[1] The Census reports on crowding incidence at the census tract level in the five-year ACS, drawing on individual responses about number of rooms in the housing unit and number of people in the household.

participate in both Medicaid and Medicare, suggesting that they are older adults.) The sample is 57 percent female. Nineteen percent of the individuals in the sample are Black, 19 percent are Hispanic, 17 percent are Asian and Pacific Islander, 16 percent non-Hispanic white, 7 percent are other races, and 21 percent have unknown race or ethnicity. When we imputed race, using surname, census tract, age, and sex, we found that those with missing race were disproportionately Hispanic[2].

The most common prior diagnosis among the eight we examined is diabetes, with a quarter of the sample having a diabetes code recorded in 2019. In addition, 10 percent of the sample had a diagnosis code for asthma in 2019, 9 percent for heart disease, and 7 percent for liver disease.

Our sample mostly lived in multifamily buildings, with 24 percent in 2-4 family homes, 26 percent in small apartment buildings between 5 and 49 units, and 31 percent in large buildings with at least 50 units. About 13 percent live in buildings with missing or zero residential units. About half of these are addresses that are missing parcel numbers (see below), but the other half are addresses for hotels and other commercial buildings. Nearly 20 percent lived in public housing, and another 8 percent lived in privately owned subsidized housing. Most of the sample (60%), not surprisingly given the age of New York's housing stock, were living in buildings constructed prior to 1940 and only a tiny fraction were living in buildings constructed since the year 2000. Note that we are missing building age for 6.2 percent of the sample. Most of these individuals (93 percent) are living at addresses that are missing parcel (borough-block-lot)

[2] We used the Bayesian Improved Surname and Geography (BISG) algorithm, which is an indirect imputation method that uses racial/ethnic probabilities from the census to assign a predicted probability of being Black, White, Hispanic, Asian/Pacific Islander, or other races. We implemented this imputation method using the 'Who are you' package in R, which links the data with census tract probabilities using surname, census tract, age, and sex. We selected the highest predicted category as the imputed race/ethnicity for those with missing information.

identifiers. These buildings are disproportionately located in Queens, given the complicated structure of addresses in the borough and difficulty geocoding them.

The majority of our sample adults were living in households with no other Medicaid recipients on the same Medicaid case number, but 37 percent were living in households with at least one other recipient. As noted above, given that not all members of a household members will be covered by Medicaid or covered under the same Medicaid case number, these estimates are likely to be undercounts of actual household size.

Finally, the adults in our sample were living in census tracts with an average poverty rate of 26 percent, a percent college educated of 18 percent, an employment rate of 89 percent, and a foreign-born share of 40 percent. They lived in census tracts with an average population density of 70,000 people per square mile of land area.[3] Just over 21 percent lived in a neighborhood characterized by high levels of crowding.

Empirical Strategy:

People are not randomly assigned to neighborhoods, buildings, and homes. Unobserved characteristics may shape the homes and neighborhoods where people live, and as a result, we are unable to draw clear causal inferences from our analysis. While we cannot say with certainty that living in higher density buildings or neighborhoods causes/does not cause people to be hospitalized from COVID, our use of individual-level Medicaid data helps us control more precisely for many typically unobserved factors, including pre-existing health conditions, receipt of public assistance, and building characteristics, such as age, conditions, subsidy status, and size. This helps us to isolate the risks associated with living in higher density environments

[3] Note that to make coefficients more readable, we specify as millions of people per square mile in the regressions.

more precisely than existing papers, which simply rely on aggregated data. This gives us more confidence that the relationships between housing characteristics and COVID that we estimate represent causal links.

Specifically, we estimate linear probability models to assess the relationship between population density and the risk of hospitalization with COVID-19. (Note results are qualitatively the same when estimating logit models.) Our core regression for each individual (*i*) in neighborhood j is as follows:

$$Y_{ij} = \beta_0 + \beta_1 population\ density_j + \beta_2 building\ size_i + \beta_3 household\ crowding_i + \omega Nj + \delta B_i + \rho X_i + \varepsilon$$

Where N is a vector of neighborhood composition variables, B is a vector of building attributes, and X is a vector of individual characteristics. All models include robust standard errors clustered at the census tract level.

<u>Results</u>

Figure 1 shows a scatterplot of COVID hospitalization rates at the census tract level and tract population density. There is little visible correlation, and the limited correlation is negative: the census tracts with highest COVID rates have relatively low population densities, and the tracts with the highest population density have relatively low COVID hospitalization rates.

Table 2 shows our regression results for the full 2020 year. The coefficients on the control variables are generally as expected. Not surprisingly, the probability of being hospitalized with COVID increases monotonically with age. Hospitalization risk was also significantly lower for women than for men. Perhaps surprisingly, once we focus on the sample of adults receiving Medicaid, we do not see the same racial and ethnic disparities that so many others have highlighted, perhaps because our sample is far more economically homogenous than others and we are able to control for pre-existing conditions. Other research also shows that among Medicaid recipients in New York City, Black and Hispanic adults were no more likely to be hospitalized with COVID than non-Hispanic white adults (Howland, Wang, Ellen & Glied, 2022). We do, however, find that Asian adults were significantly *less* likely to be hospitalized than other groups. The magnitude of the coefficient estimate suggests that Asian adults were a full half-percentage point less likely to be hospitalized for COVID. Adults with unknown race/ethnicity, who, as described above, appear to be disproportionately Hispanic, were also less likely to be hospitalized. Once we control for age, respondents that are also enrolled in Medicare were less likely to be hospitalized, but this likely reflects that hospitalizations among those enrolled in both Medicare and Medicaid are less likely to be billed and recorded on Medicaid claims than those of beneficiaries who are not dually-enrolled.

Pre-existing conditions are strongly predictive of hospitalizations. Prior history of heart disease, asthma, diabetes, cancer, kidney disease, liver disease, and dementia are all positively associated with the probability of hospitalization. Arthritis is the one health condition that is not significantly associated with hospitalization.

Adults in neighborhoods with more college-educated adults were less likely to be hospitalized for COVID, perhaps a reflection of their greater ability to work at home or their higher compliance with mask mandates. The coefficient on neighborhood employment rate is positive, suggesting that neighborhoods with more workers are riskier because more people may be interacting outside of their homes. Finally, the coefficient on percent foreign-born is negative. It is not clear why this relationship exists, but it's possible that people living in neighborhoods with more foreign-born residents were aware of the risks of COVID earlier or may be more likely to comply with public health guidance such as mask mandates.

Building age appears to be unrelated to COVID risk, but people in buildings of unknown age are somewhat less likely to be hospitalized. These buildings are disproportionately located in Queens, as noted above, and also tend to be in neighborhoods with lower poverty rates and higher foreign-born shares, so this variable could be picking up unobserved protective individual and neighborhood features. Living in public housing is associated with a *reduced* risk of being hospitalized with COVID relative to Medicaid residents living in market-rate housing. The magnitude of the coefficient suggests that the risk of hospitalization for adults living in public housing was 0.1 percentage points lower, which is notable given that the overall hospitalization rate was just over 0.8 percent. This negative relationship runs counter to conventional wisdom and raw correlations. Numbers released by the City's Department of Health show that one year into the pandemic, NYCHA residents accounted for 7 percent of COVID-19 deaths citywide, but only 4 percent of the population (Gonzalez, 2021), but in these comparisons, the full population includes those with much higher incomes (and does not control for differences in age and pre-existing conditions), while our comparison is only among low income people and controls for these risk factors. The numbers reported by the Department of Health may also overstate NYCHA rates because official estimates of people living in public housing significantly undercount the actual population by 20 percent or even more (Blumgart, 2016).

Turning to our core variables, the coefficient on population density is negative, though not statistically significant. We estimated several alternative regressions with non-linear specifications of density (not shown), and we found no evidence that population density heightened the risk of COVID hospitalization. If anything, the estimated coefficients suggest that hospitalization rates were lower in the census tracts in the top third of the population density distribution compared to those in lower density categories. Building size does not appear to matter much either, though people living in single-family homes were at lower risk of hospitalization compared to those living in 2-4 unit buildings. There is no difference in risk between people living in 2-4 unit buildings, 5-49 unit buildings, and larger buildings. Note that people living in buildings with missing unit numbers also saw elevated risk, perhaps because they are likely to be unstably housed and/or living in hotels or other non-residential buildings, suggesting temporary placements.

By contrast, the number of people on the same Medicaid case number (our proxy for household size) is significantly related to hospitalization rates. Interestingly, people that were the only person on their case number were at higher risk than those with a 2-person Medicaid case number. Perhaps people living alone are more likely to interact with others outside their household; it is also possible that very low-income single adults without children have unmeasured risk factors not captured here. People living in households with three or more Medicaid recipients were significantly more likely to be hospitalized than those living in households with just two Medicaid recipients. Compared to adults living in households with two Medicaid recipients, adults living in households with three Medicaid recipients had a .09 percentage point higher hospitalization rate and those living in households with 4+ Medicaid recipients had a 0.15 percentage point higher hospitalization rate. Again, given the overall hospitalization rate of just over 0.8 percent for this population, these are sizable magnitudes. And adults who were living in congregate settings had a 0.5 percentage point higher risk of being hospitalized. Finally, we find that people living in neighborhoods characterized by high levels of household crowding were at elevated risk of hospitalization.

Table 3 shows separate results for the three stages of COVID that we track. The results are generally consistent across the three time periods. One key difference is that the coefficient on population density for the last wave is actually negative and statistically significant. Since we do not find that density is positively correlated with hospitalization in earlier waves, this finding is unlikely to reflect acquired population immunity. Rather, it may be that dense neighborhoods were better able to accommodate adaptations in behavior during this latter period. Interestingly, people who were unstably housed were also less likely to be at risk during this final period, perhaps as a result of greater immunity from earlier infections. One other notable difference is that people with large Medicaid case sizes were at no greater risk of hospitalization than people living alone during the first phase of COVID, perhaps suggesting that more transmission was occurring outside the home during this first phase, before public health measures were put into place.

Sensitivity Analyses

One argument for the relationship between density and infectious disease is that residents of denser neighborhoods are more likely to use public transit, and that transit is itself is a risk factor for disease. In sensitivity analyses reported in the appendix, we control for the share of residents in the ZIP Code using different commuting modes: the proportion of the ZIP code's population whose primary commuting mode is driving a car, biking or walking, carpooling or taxi, or working from home. We omit the proportion of the ZIP Code's population whose primary commuting mode is public transportation, which is the most common mode citywide. Note that the distribution of modality of commuting is likely linked to where commuters travel to – in particular, residents of Queens, the Bronx, and Staten Island who commute in their own car or by car share may be commuting out of the NYC metropolitan area, rather than into it.

Including measures of commuting modality doesn't measurably change our estimates of the relationship between density and COVID, and the association remain negative and statistically insignificant.

As for commuting mode, we find no relationship between the share of residents in a ZIP Code commuting by private car and COVID hospitalizations, though an increase in the share of ZIP Code residents getting to work through a car-share (a very uncommon commuting mode) is associated with lower hospitalization rates (see Appendix Table 1). The relationship between commuting modality distribution and COVID hospitalization, however, varies by borough, consistent with commuting modality reflecting commuting destinations. In Brooklyn, Manhattan, and Queens, residents of ZIP Codes with more private car commuters had *higher* rates of COVID hospitalizations than residents of ZIP Codes with higher public transit shares, while in Staten Island, the results are reversed (see Appendix Table 2).

Conclusion

In summary, our results suggest that among the population of low-income New Yorkers, living in higher density neighborhoods did not heighten the risk of COVID hospitalization. Indeed, in the final four months of 2020, we find that Medicaid recipients living in higher density neighborhoods were at *lower* risk of hospitalization. The size of a multifamily building made little difference either, and people living in public housing developments, which are typically highly dense environments, were *less* likely to be hospitalized for COVID. In other words, it's not so clear that density is a demon, at least with respect to COVID.

However, our results suggest that large households and crowded living quarters – which can occur at any level of population density – increase the risk of COVID hospitalization. We also see a strong correlation between being unstably housed or living in institutional settings and COVID hospitalizations. So housing is consistently correlated with hospitalization, and policymakers may want to target prevention tools to those living unstably or in large households. Further research should explore our intriguing finding that people living in public housing were actually at lower risk of COVID hospitalizations compared to comparable adults living in market-rate buildings.

Importantly, our research focuses on a low-income population, and our results may not be generalizable to the broader population. That said, we believe our focus on Medicaid recipients is a strength of our analysis, as it allows us to control more fully for socioeconomic status. Further, our use of Medicaid claims data allows us to include a much richer set of pre-existing conditions than other work, bolstering our confidence that our results may be capturing the actual effects of density.

References

Blumgart, J. (2016). The Ghost Tenants of New York City. *Slate Magazine*. Retrieved from https://slate.com/business/2016/03/new-york-city-public-housing-could-have-more-than-100000-ghost-tenants-living-off-the-books-heres-why.html

Callaghan T, Lueck JA, Trujillo KL, Ferdinand AO. Rural and Urban Differences in COVID-19 Prevention Behaviors. J Rural Health. 2021 Mar;37(2):287-295. doi: 10.1111/jrh.12556. Epub 2021 Feb 22. PMID: 33619836; PMCID: PMC8013340.

Carozzi, F. (2020). *Urban Density and Covid-19*. Retrieved from London School of Economics & Political Science (LSE) website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3643204

Chen, K., & Li, Z. (2020). The spread rate of SARS-CoV-2 is strongly associated with population density. *Journal of Travel Medicine*, *27*(8), 1–2. https://doi.org/10.1093/JTM/TAAA186

Duranton, G., & Puga, D. (2020). The Economics of Urban Density. *Journal of Economic Perspectives*, *34*(3), 3–26. https://doi.org/10.1257/jep.34.3.3

Glaeser, E. L. (2011). *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*. New York: Penguin Press.

Glaeser, E.L., and David Cutler. (2021). Survival of the City: Living and Thriving in an Age of Isolation. New York: Penguin Press.

Gonzalez, A. (2021). More NYCHA residents are dying of COVID-19 than rest of NYC. Retrieved February 27, 2022, from Spectrum News NY 1 website: https://www.ny1.com/nyc/all-boroughs/news/2021/09/16/more-nycha-residents-are-dying-of-covid-19-than-rest-of-nyc

Gupta, S., Archelle, G., Soumya, S., Kosali, S., & Pinar, K.-M. (2021). US trends in COVID-19–associated hospitalization and mortality rates before and after reopening economies. *JAMA Health Forum, American Medical Association*, *2*(6), e211262–e211262.

Hamidi, S., & Hamidi, I. (2021). Subway Ridership, Crowding, or Population Density: Determinants of COVID-19 Infection Rates in New York City. *American Journal of Preventive Medicine*, *60*(5), 614–620. https://doi.org/10.1016/j.amepre.2020.11.016

Hamidi, S., Sabouri, S., & Ewing, R. (2020). Does Density Aggravate the COVID-19 Pandemic? *Journal of the American Planning Association*, *86*(4), 495–509. https://doi.org/10.1080/01944363.2020.1777891

Hobbs, C. V, Martin, L. M., Kim, S. S., Kirmse, B. M., Haynie, L., Mcgraw, S., … Stephenson, M. E. (2020). Factors Associated with Positive SARS-CoV-2 Test Results in Outpatient Health Facilities and Emergency Departments Among Children and Adolescents Aged <18 Years — Mississippi, September–November 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*(50), 1925–1929. https://doi.org/10.15585/mmwr.mm6950e3

Howland, R.H., Wang S., Ellen, I.G., Glied S. (2022). Not a New Story: Place- and Race-Based Disparities in COVID-19 and Influenza Hospitalizations among Medicaid-Insured Adults in New York City. *J Urban Health, 99*(2):345-358. doi: 10.1007/s11524-022-00612-y. Epub 2022 Feb 22.

Kadri, S. S., Gundrum, J., Warner, S., Cao, Z., Babiker, A., Klompas, M., & Rosenthal, N. (2020). Uptake and Accuracy of the Diagnosis Code for COVID-19 Among US Hospitalizations. *JAMA*, *324*(24), 2553. https://doi.org/10.1001/jama.2020.20323

Olsen, H. (2020). Opinion: The United States might have a secret weapon against coronavirus. *The Washington Post*. Retrieved from https://www.washingtonpost.com/opinions/2020/03/19/united-states-might-have-secret-weapon-against-coronavirus/

Sy, K. T. L., White, L. F., & Nichols, B. E. (2021). Population density and basic reproductive number of COVID-19 across United States counties. *PLOS ONE*, *16*(4), e0249271. https://doi.org/10.1371/journal.pone.0249271

Tina Q Tan, Ravina Kullar, Talia H Swartz, Trini A Mathew, Damani A Piggott, Vladimir Berthaud, Location Matters: Geographic Disparities and Impact of Coronavirus Disease 2019, *The Journal of Infectious Diseases*, Volume 222, Issue 12, 15 December 2020, Pages 1951–1954, https://doi.org/10.1093/infdis/jiaa583

Whaley, C. M., Cantor, J., Pera, M., & Jena, A. B. (2021). Assessing the Association Between Social Gatherings and COVID-19 Risk Using Birthdays. *JAMA Internal Medicine*, *181*(8), 1090. https://doi.org/10.1001/jamainternmed.2021.2915

Wong, D. W. S., & Li, Y. (2020). Spreading of COVID-19: Density matters. *PloS One*, *15*(12). https://doi.org/10.1371/JOURNAL.PONE.0242398

Figure 1
Census tract COVID-19 hospitalization rates per 100,000 Medicaid-insured adults and population density in New York City

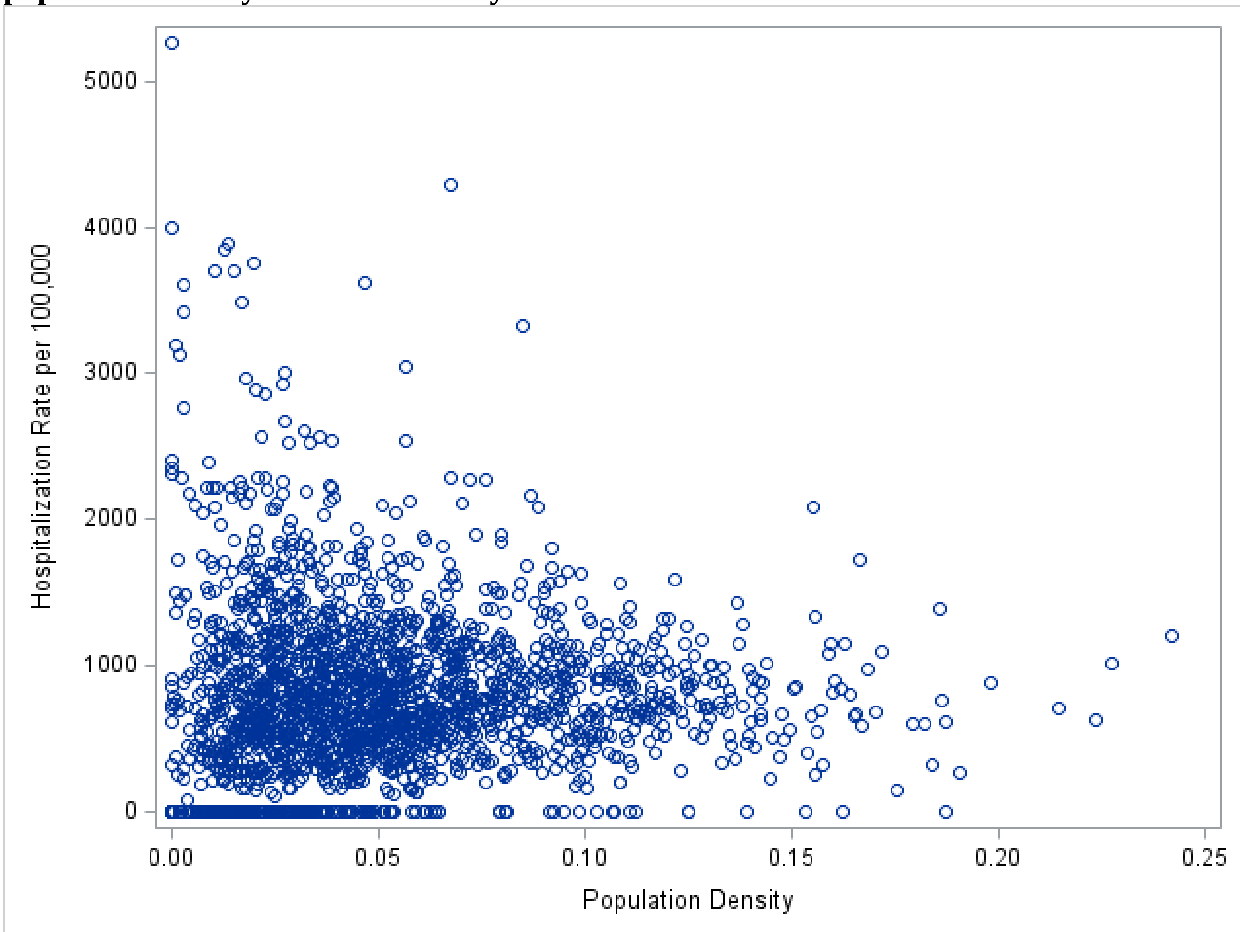

Table 1
Summary statistics

Variable	N	Percent or mean (standard deviation)
Total Medicaid-insured adults	1,834,208	
COVID-19 hospitalizations (total)	14,990	0.82
First wave (January – April 2020)	8,425	0.56
Second wave (May – August 2020)	2,711	0.18
Third wave (September – December 2020)	3,854	0.26
Population Density (millions/sq mile)		0.07 (0.04)
Residential Units		
Missing	251,132	13.7
1 unit	102,494	5.6
2-4 units	448,037	24.4
5-49 units	468,511	25.5
50+ units	564,034	30.8
Unstably housed		
No	1,803,812	98.3
Yes	30,396	1.7
Congregate living		
No	1,797,960	98.0
Yes	36,248	2.0
Other administrative address		
No	1,827,780	99.7
Yes	6,428	0.4
Medicaid case size		
1 person	1,153,778	62.9
2 people	277464	15.1
3 people	171860	9.4
4+ people	231106	12.6
Age category		
18-39	717,590	39.1
40-49	260,693	14.2
50-59	296,678	16.2
60-69	274,453	15.0
70-79	172,408	9.4
80+	112,386	6.1
Sex		
Male/Unknown	786,178	42.9
Female	1,048,030	57.1
Race/ethnicity		
Unknown	381,770	20.8
White	291,957	15.9
Black	354,604	19.3

Asian and Pacific Islander	318,434	17.4
Other	132,143	7.2
Hispanic	355,300	19.4
Dual Medicare/Medicaid		
No	1,369,406	74.7
Yes	464,802	25.3
Pre-existing conditions		
Heart disease	164,658	9.0
Asthma	183,036	10.0
Diabetes	459,372	25.0
Cancer	102,686	5.6
Kidney	77,813	4.2
Liver	130,493	7.1
Dementia	54,341	3.0
Arthritis	25,604	1.4
Subsidy Type		
None/missing	1,345,261	73.3
Other subsidy	149,098	8.1
NYCHA	339,849	18.5
Year Built		
Missing	113,417	6.2
Before 1940	1,082,742	59.0
1940-2000	555,263	30.3
After 2000	82,786	4.5
Neighborhood household crowding rate > 30%		
No	1,437,374	78.4
Yes	396,834	21.6
Neighborhood composition		
Percent poverty		25.7 (13.2)
Percent college education		17.9 (10.1)
Percent employed		89.2 (5.5)
Percent foreign born		40.4 (15.8)
Share of residents in ZIP Code using commuting mode		
Public transportation		59.3 (12.0)
Car (alone)		18.7 (12.6)
Bike/walk		11.8 (8.3)
Car share (car pool + taxi)		5.9 (2.7)
Work from home		3.9 (1.9)

Table 2
Regression Results

Variable	β	SE
Population Density	-0.0046	0.0029
Residential Units		
Missing	0.0040***	0.0006
1 unit	-0.0011***	0.0003
5-49 units	0.0000	0.0003
50+ units	-0.0002	0.0003
2-4 units	REF	
Unstably housed		
Yes	0.0002	0.0011
No	REF	
Congregate living		
No	REF	
Yes	0.0050***	0.0015
Other administrative address		
No	REF	
Yes	0.0060**	0.0030
Medicaid case size		
1 person	0.0008***	0.0002
3 people	0.0009***	0.0002
4+ people	0.0015***	0.0002
2 people	REF	
Age category		
40-49	0.0009***	0.0002
50-59	0.0028***	0.0002
60-69	0.0038***	0.0003
70-79	0.0048***	0.0004
80+	0.0060***	0.0005
18-39	REF	
Sex		
Female	-0.0019***	0.0002
Male/Unknown	REF	
Race/ethnicity		
Unknown	-0.0008***	0.0003
Black	-0.0004	0.0003
API	-0.0055***	0.0003
Other	-0.0009**	0.0004
Hispanic	-0.0001	0.0003
White	REF	
Dual Medicare/Medicaid		
Yes	-0.0019***	0.0003
No	REF	
Pre-existing conditions		
Heart disease	0.0101***	0.0004

Asthma	0.0025***	0.0003
Diabetes	0.0047***	0.0002
Cancer	0.0016***	0.0004
Kidney	0.0209***	0.0007
Liver	0.0054***	0.0004
Dementia	0.0149***	0.0008
Arthritis	0.0010	0.0008
Neighborhood Characteristics		
Poverty	0.0021*	0.0012
College Education	-0.0051**	0.0016
Employed	0.0032	0.0025
Foreign born	-0.0017**	0.0007
Building subsidy		
NYCHA	-0.0011***	0.0004
Other subsidy	-0.0003	0.0002
None	REF	
Year Built		
Missing	-0.0029***	0.0007
1940-2000	0.0002	0.0002
After 2000	-0.0001	0.0003
Before 1940	REF	
Neighborhood household crowding rate >30%		
Yes	0.0014***	0.0003
No	REF	
N	1,834,208	

Standard errors cluster by census tract
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

Table 3.
Regression Results by wave

	First wave Jan-Apr 2020		Second wave May-Aug 2020		Third wave Sept-Dec 2020	
Variable	**β**	**SE**	**β**	**SE**	**β**	**SE**
Population Density	-0.0018	0.0021	0.0005	0.0010	-0.0033***	0.0010
Residential Units						
Missing	0.0035***	0.0005	0.0007***	0.0002	-0.0001	0.0002
1 unit	-0.0005**	0.0002	-0.0002*	0.0001	-0.0004**	0.0002
5-49 units	0.0001	0.0002	0.0000	0.0001	-0.0001	0.0001
50+ units	-0.0001	0.0002	-0.0002**	0.0001	0.0001	0.0001
2-4 units	REF		REF		REF	
Unstably housed						
Yes	-0.0006	0.0009	0.0006*	0.0004	0.0002	0.0003
No	REF		REF		REF	
Congregate living						
No	REF					
Yes	0.0036***	0.0012	0.0015***	0.0004	-0.0001	0.0003
Other administrative address						
No	REF					
Yes	0.0019	0.0021	0.0020*	0.0011	0.0021***	0.0007
Medicaid case size						
1 person	0.0007***	0.0001	0.0001	0.0001	0.0000	0.0001
3 people	0.0007***	0.0002	0.0003***	0.0001	-0.0001	0.0001
4+ person	0.0007***	0.0002	0.0005***	0.0003	0.0003**	0.0001
2 people	REF		REF		REF	
Age category						
40-49	0.0011***	0.0001	-0.0004***	0.0001	0.0002*	0.0001
50-59	0.0023***	0.0002	-0.0002***	0.0001	0.0007***	0.0001
60-69	0.0030***	0.0002	-0.0004***	0.0001	0.0011***	0.0001
70-79	0.0029***	0.0003	-0.0002	0.0002	0.0021***	0.0002
80+	0.0021***	0.0004	0.0001	0.0002	0.0038***	0.0003
18-39	REF		REF		REF	
Sex						
Female	-0.0015***	0.0001	0.0001	0.0001	-0.0005***	0.0001
Male/Unknown	REF		REF		REF	
Race/ethnicity						
Unknown	0.0000	0.0002	-0.0002*	0.0001	-0.0006***	0.0001
Black	0.0005**	0.0002	-0.0001	0.0001	-0.0009***	0.0001
API	-0.0032***	0.0002	-0.0010***	0.0001	-0.0013***	0.0001
Other	-0.0001	0.0003	-0.0002	0.0001	-0.0005***	0.0002
Hispanic	0.0003	0.0002	-0.0001	0.0001	-0.0003*	0.0002
White	REF		REF		REF	
Dual Medicare/Medicaid						

Yes	-0.0012***	0.0002	0.0001	0.0001	-0.0007***	0.0001
No	REF		REF		REF	
Pre-existing conditions						
Heart disease	0.0055***	0.0003	0.0021***	0.0002	0.0025***	0.0002
Asthma	0.0010***	0.0002	0.0005***	0.0001	0.0010***	0.0001
Diabetes	0.0029***	0.0002	0.0006***	0.0001	0.0011***	0.0001
Cancer	0.0003	0.0003	0.0003*	0.0002	0.0010***	0.0002
Kidney	0.0140***	0.0015	0.0032***	0.0003	0.0037***	0.0003
Liver	0.0033***	0.0003	0.0012***	0.0002	0.0009***	0.0002
Dementia	0.0100***	0.0017	0.0032***	0.0004	0.0018***	0.0004
Arthritis	0.0000	0.0015	0.0007**	0.0003	0.0003	0.0004
Neighborhood Characteristics						
Poverty	-0.0009	0.0009	0.0009**	0.0004	0.0021***	0.0004
College Education	-0.0053***	0.0012	-0.0004	0.0005	0.0006	0.0006
Employed	0.0007	0.0021	0.0011	0.0008	0.0014*	0.0008
Foreign born	-0.0012**	0.0005	-0.0006**	0.0002	0.0001	0.0003
Building subsidy						
NYCHA	-0.0006**	0.0003	-0.0002	0.0001	-0.0003*	0.0002
Other subsidy	-0.0003	0.0002	0.0000	0.0001	0.0000	0.0001
None	REF		REF		REF	
Year Built						
Missing	-0.0026***	0.0005	-0.0005**	0.0002	0.0002	0.0002
1940-2000	0.0003	0.0002	0.0000	0.0001	0.0000	0.0001
After 2000	0.0001	0.0003	0.0000	0.0001	-0.0002	0.0002
Before 1940	REF		REF		REF	
Neighborhood household crowding rate >30%						
Yes	0.0009***	0.0002	0.0003***	0.0001	0.0001	0.0001
No	REF		REF		REF	
N	1,834,208		1,834,208		1,834,208	
Total COVID-19 hospitalizations	8,425		2,711		3,854	

Standard errors cluster by census tract
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

Appendix Table 1.
Regression Results including Commuting Mode Variables

Variables	β	SE
Population Density	-0.0024	0.0033
Residential Units		
Missing	0.0043***	0.0006
1 unit	-0.0012***	0.0003
5-49 units	0.0001	0.0002
50+ units	-0.0002	0.0003
2-4 units	REF	
Unstably housed		
Yes	0.0005	0.0011
No	REF	
Congregate living		
No	REF	
Yes	0.0051***	0.0015
Other administrative address		
No	REF	
Yes	0.0060**	0.0030
Medicaid case size		
1 person	0.0008***	0.0002
3 people	0.0008***	0.0002
4+ people	0.0015***	0.0002
2 people	REF	
Age category		
40-49	0.0009***	0.0002
50-59	0.0028***	0.0002
60-69	0.0038***	0.0003
70-79	0.0048***	0.0004
80+	0.0061***	0.0005
18-39	REF	
Sex		
Female	-0.0019***	0.0002
Male/Unknown	REF	
Race/ethnicity		
Unknown	-0.0009***	0.0003
Black	-0.0006**	0.0003
API	-0.0053***	0.0003
Other	-0.0010***	0.0004
Hispanic	-0.0002	0.0003

White	REF	
Dual Medicare/Medicaid		
Yes	-0.0018***	0.0003
No	REF	
Pre-existing conditions		
Heart disease	0.0101***	0.0004
Asthma	0.0025***	0.0003
Diabetes	0.0047***	0.0002
Cancer	0.0017***	0.0004
Kidney	0.0208***	0.0007
Liver	0.0054***	0.0004
Dementia	0.0149***	0.0008
Arthritis	0.0009***	0.0008
Neighborhood Characteristics		
Poverty	0.0028**	0.0013
College Education	-0.0036**	0.0017
Employed	0.0053**	0.0027
Foreign born	-0.0008	0.0008
Building subsidy		
NYCHA	-0.0011**	0.0004
Other subsidy	-0.0003	0.0002
None	REF	
Year Built		
Missing	-0.0032***	0.0007
1940-2000	0.0003	0.0002
After 2000	0.0000	0.0003
Before 1940	REF	
Neighborhood household crowding rate >30%		
Yes	0.0015***	0.0003
No	REF	
Share of residents in ZIP Code using commuting mode		
Car	0.0019	0.0013
Bike/Walk	-0.0025	0.0019
Car Share/Taxi	-0.0226***	0.0046
Work from home	-0.0065	0.0092
Public transportation	REF	
N	1,834,208)	

Standard errors cluster by census tract
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$

Appendix Table 2.
Regression Overall including Commuting Mode Variables, separated by borough

Variables	β	SE
Bronx (005)		
Share of residents in ZIP Code using commuting mode		
Car	-0.0010	0.0034
Bike/Walk	-0.0036	0.0135
Car Share/Taxi	0.0291	0.0185
Work from home	-0.0200	0.0190
Public transportation	REF	
N	363,539	
Brooklyn (047)		
Share of residents in ZIP Code using commuting mode		
Car	0.0105***	0.0031
Bike/Walk	0.0092**	0.0042
Car Share/Taxi	-0.0374***	0.0124
Work from home	0.0075	0.0137
Public transportation	REF	
N	615,599	
Manhattan (061)		
Share of residents in ZIP Code using commuting mode		
Car	0.0186*	0.0110
Bike/Walk	0.0004	0.0030
Car Share/Taxi	-0.0635***	0.0205
Work from home	0.0189	0.0217
Public transportation	REF	
N	413,296	
Queens (081)		
Share of residents in ZIP Code using commuting mode		
Car	0.0085**	0.0034
Bike/Walk	0.0165	0.0104
Car Share/Taxi	-0.0413***	0.0099
Work from home	-0.0076	0.0241
Public transportation	REF	
N	396,735	
Staten Island (085)		
Share of residents in ZIP Code using commuting mode		
Car	-0.0217**	0.0085
Bike/Walk	-0.0009	0.0346

Car Share/Taxi	0.0506	0.0346
Work from home	-0.0397	0.0484
Public transportation	REF	
N	45,039	

Models include all other variable. Standard errors cluster by census tract

*** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$