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

March 1, 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.

Disclaimer: 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. Examples of analysis performed within this article are only examples. They should not be utilized in real-world analytic products

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 brings 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).

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, 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). 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 crowding, 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 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 and 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. As for building size, 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. Further, people living in public housing developments, which are typically highly dense environments, 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. Specifically, we see that people living in large households or in a neighborhood with a high level 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

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 measures of commuting at the zip code level 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. 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. 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 could simply adopt behavioral responses that shield themselves from exposure. These behavioral responses have limits, 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 metropolitan area size (which they take as a proxy for connectivity or rates of population flows in and out of an area) is associated with case rates but not population density. Their work focused on the early part of the COVID pandemic, however, so it's possible that connectivity also affects the timing of outbreaks but not necessarily 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 household size. Using Medicaid claims data, we are able to control for age, race/ethnicity, pre-existing health conditions, building age, building code violations, 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 Department of Health GBAT software (REF) 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 the year 2020 (meaning they had 10 or more months of enrollment during that year). We use the assigned geographic indicators to match claims with other sources of neighborhood and building level data described below.

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 diagnosis of a condition in 2019 were categorized as having that condition.

Our key independent variables include population density, building size and subsidy status, and household crowding. 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 commuting modality at 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-level crowding at both the individual and neighborhood level. At the individual-level, we use the Medicaid case number (which is the same across 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 in the same household (1, 2, 3-4, and 4+ people in a given household). Note that this 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, but we don't expect any systematic bias. There are also a small proportion of people (<1%) for whom we could not assign a case number, likely due to administrative issues in the recording of IDs over time. 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. Finally, we also create a measure for individuals who are unstably housed based on street level addresses that indicate homeless (e.g., 'General Delivery' or 'Undomiciled') or matched with known institutional address for jails, psychiatric institute, rehabilitation centers, or other administrative offices.

Figure 1 shows summary statistics. It shows that 0.83% (n=15,261) of our sample was hospitalized for COVID during 2020. More than half (n=8,583) of all hospitalizations occurred during the first third of 2020, 18.1% percent were hospitalized during the middle third, and

25.6% 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 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 over 20 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.

As for the eight relevant pre-existing conditions, the most common prior diagnosis is diabetes, with a quarter of the sample receiving a diabetes diagnosis in 2019. In addition, 10 percent of the sample was diagnosed with asthma in 2019, 9 percent with heart disease, and 7 percent with liver disease.

Our sample mostly lived in multifamily buildings, with 24 percent in 2-4 family homes, 25 percent in small apartment buildings between 5 and 49 units, and 30 percent in large buildings with at least 50 units. Nearly 20 percent lived in public housing, and another 8 percent lived in privately owned subsidized housing. Most of the sample, 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 parcel (borough-block-lot) identifiers for about 6.3% percent of the sample. This group may be more likely to be living in institutions or unstably housed. They are also disproportionately likely to be located in Queens, given the complicated structure of addresses in the borough and difficulty geocoding them.

The majority of our sample were living in households with no other Medicaid recipients, but over 20 percent were living in households with at least two other recipients. 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 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.[1] 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 more precisely than existing papers, which simply rely on aggregated data. This gives us more

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

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 hospitalization with COVID-19. (Note results are qualitatively the same when estimating logit models.) Our core regression for each individual (i) is as follows:

$$Y_i = \beta_0 + \beta_1 population\ density + \beta_2 building\ size + \beta_3 household\ crowding + \omega\ neighborhood\ composition + \delta\ other\ building\ measures + \rho\ individual\ controls + \varepsilon$$

All models include robust standard errors clustered at the census tract level.

Results

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 is 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 due to the fact that our sample is more economically homogenous than others. Black and Hispanic adults are no more likely to be hospitalized with COVID than non-Hispanic white adults. We do, however, find that Asian adults are

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 also look to be at somewhat lower risk. Our imputation exercise suggests that this group is disproportionately Hispanic. Once we control for age, respondents that are also enrolled in Medicare are less likely to be hospitalized, but this likely reflects that the hospitalization is less likely to be billed and recorded on Medicaid claims.

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 associated with hospitalization.

Adults in neighborhoods with more college-educated adults are 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 living in public housing reduces the risk of being hospitalized with COVID. The magnitude of the coefficient suggests that living in public housing reduces the risk of hospitalization by 0.1 percentage points, 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 these numbers may 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). Plus, these raw population rate comparisons fail to control for risk factors like age and pre-existing conditions. The negative coefficient we find may be due to lower employment rates in these communities, or greater investment in sharing information with residents about COVID risk and ways to minimize it.

Turning to our core variables, the coefficient on population density is negative, though not statistically significant. We estimated several alternative specifications (not shown), and we found no evidence that population density heightened the risk of COVID hospitalization. Building size does not appear to matter much either, though people living in single-family homes are at lower risk of hospitalization. There is no difference in risk between people living in 2-4 unit buildings, 5-49 unit buildings, and larger buildings.

By contrast, household size does make a difference. Interestingly, people living in 2-person households are at lower risk than those living alone. Perhaps people living alone are more likely to interact with others outside their household. People living in households with at least three other Medicaid recipients were significantly more likely to be hospitalized. Specifically, adults living in households with at least three other Medicaid recipients had a 0.1 percentage point higher hospitalization rate than those living in households with no other Medicaid recipients. Again, given the overall hospitalization rate of just over 0.8 percent for this population, this is a sizable effect. And adults who were unstably housed, or living in

institutions (e.g., jails, psychiatric institutions) had a 0.3 percentage point higher risk of being hospitalized.

Table 3 shows separate results for the three stages of COVID that we track. The results are generally fairly 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 living in large households 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.

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 commuting modality patterns at the zip code level. 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 by own car or car share may be commuting out of the NYC metropolitan area, rather than into it. Including measures of commuting modality slightly reduces our estimates of the relationship between density and COVID, though they remain negative. At the City-wide level, we find no effect of zip code area public transit propensity on

COVID hospitalization risk -- residents of zip codes with a high share of commuters who drive their own cars have slightly (and statistically insignificantly) higher COVID hospitalization rates than those with a high share of commuters who take transit (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 a higher share of commuters who drive their own cars have higher rates of COVID hospitalizations than zip codes with a higher share of public transit commuters, 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 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 matters, 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

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*. The 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

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

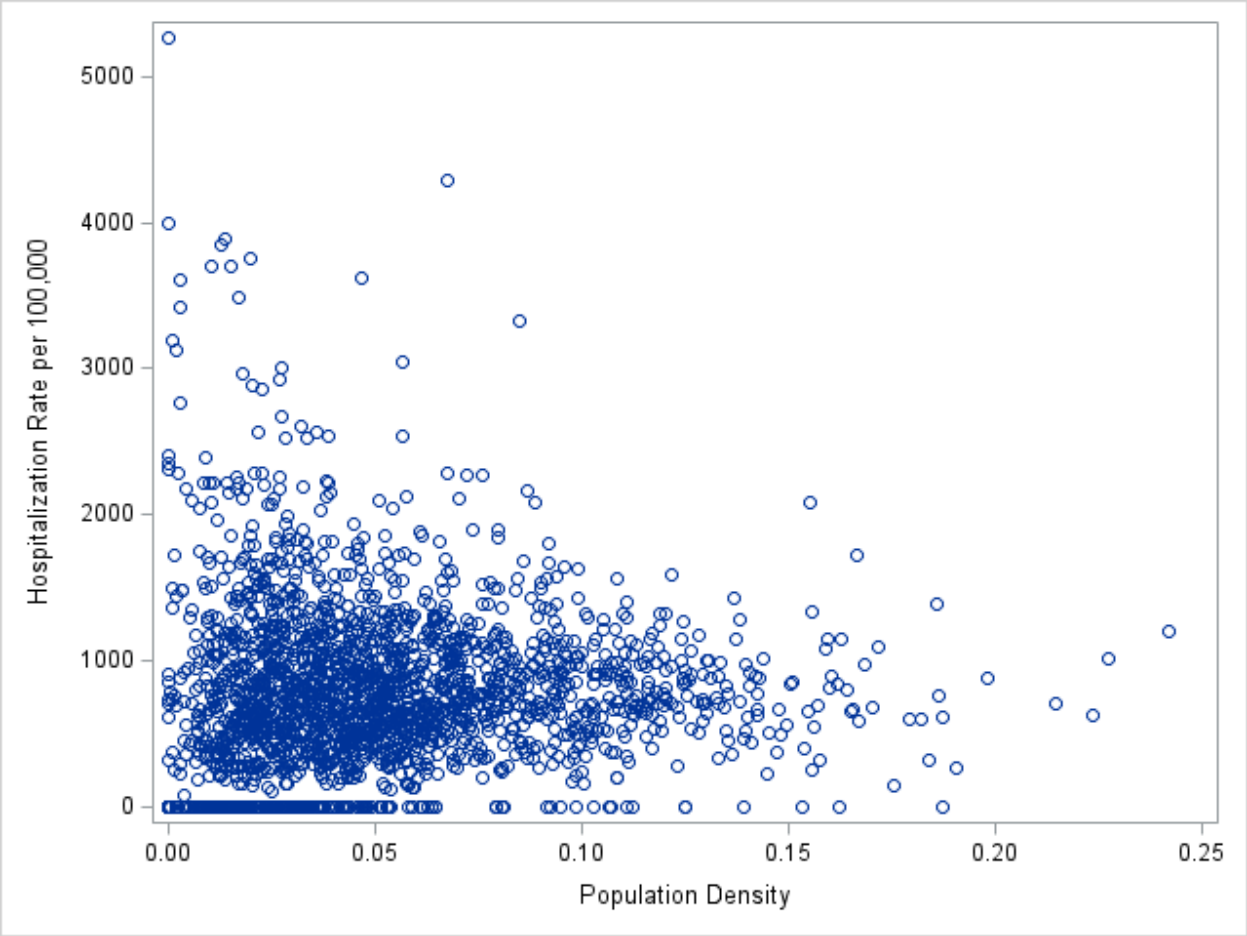
Olsen, H. (2020). Opinion: The United States might have a secret weapon against coronavirus. *The Washington Post*. Retrieved from https://www.washingtonpost.com/opinions/2020/03/19/united-states-might-have-secret-weapon-against-coronavirus/

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

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



Note: To improve readability two census tracts with extreme values were removed from the figure

Table 1
Summary statistics

Variable	N	Percent or mean (standard deviation)
Total Medicaid-insured adults	1,865,729	
COVID-19 hospitalizations (total)	15,261	0.82
First wave (January – April 2020)	8,583	0.46
Second wave (May – August 2020)	2,768	0.15
Third wave (September – December 2020)	3,910	0.21
Population Density (millions/sq mile)		0.07 (0.04)
Residential Units		
0 units	142,711	7.7
1 unit	103,406	5.5
2-4 units	452,533	24.3
5-49 units	473,277	25.4
50+ units	570,662	30.6
Missing	123,140	6.6
Unstably housed		
No	1,792,206	96.1
Yes	73,523	3.9
Household size		
Missing	15,713	0.8
1 person	1,165,245	62.5
2 people	279,246	15.0
3-4 people	292,309	15.7
4+ people	113,216	6.1
Age category		
18-39	731,792	39.2
40-49	265,421	14.2
50-59	301,794	16.2
60-69	278,645	14.9
70-79	174,549	9.4
80+	113,528	6.1
Sex		
Male/Unknown	800,458	42.9
Female	1,065,271	57.1
Race/ethnicity		
Unknown	389,912	20.9
White	296,498	15.9
Black	362,425	19.4
Asian and Pacific Islander	322,077	17.3
Other	134,400	7.2
Hispanic	360,417	19.3
Dual Medicare/Medicaid		
No	1,394,504	74.7

Yes	471,225	25.3
Pre-existing conditions		
Heart disease	166,728	8.9
Asthma	186,271	10.0
Diabetes	465,740	25.0
Cancer	104,178	5.6
Kidney	78,815	4.2
Liver	132,395	7.1
Dementia	55,019	3.0
Arthritis	25,950	1.4
Subsidy Type		
None/missing	1,371,336	73.5
Other subsidy	150,962	8.1
NYCHA	343,431	18.4
Year Built		
Missing	123,110	6.6
Before 1940	1,097,085	58.8
1940-2000	561,791	30.1
After 2000	83,743	4.5
Neighborhood household crowding		
No	1,463,258	78.4
Yes	402,471	21.6
Neighborhood composition		
Percent poverty		25.6 (13.3)
Percent college education		17.9 (10.2)
Percent employed		89.1 (6.6)
Percent foreign born		40.3 (15.8)
Missing BBL		
No	1,748,622	93.7
Yes	117,107	6.3
Primary means of commute to work		
Public transportation		59.2 (12.2)
Car (alone)		18.7 (12.6)
Bike/walk		11.8 (8.4)
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	-0.0035	0.0027
Residential Units		
1 unit	-0.0011***	0.0003
5-49 units	0.0000	0.0002
50+ units	-0.0002	0.0003
0 or missing units	0.0042***	0.0006
2-4 units	REF	
Unstably housing		
Yes	0.0030***	0.0008
No	REF	
Household size		
Missing	0.0011*	0.0007
2 people	-0.0009***	0.0002
3-4 people	0.0001	0.0002
4+ people	0.0011***	0.0002
1 person	REF	
Age category		
40-49	0.0010***	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	-0.0020***	0.0002
Male/Unknown	REF	
Race/ethnicity		
Unknown	-0.0007***	0.0003
Black	-0.0005	0.0003
API	-0.0055***	0.0003
Other	-0.0008**	0.0004
Hispanic	-0.0001	0.0003
White	REF	
Dual Medicare/Medicaid		
Yes	-0.0018***	0.0003
No	REF	
Pre-existing conditions		
Heart disease	0.0101***	0.0004
Asthma	0.0026***	0.0003
Diabetes	0.0046***	0.0002
Cancer	0.0016***	0.0004
Kidney	0.0209***	0.0007
Liver	0.0054***	0.0004
Dementia	0.0151***	0.0008
Arthritis	0.0011	0.0008

Neighborhood Characteristics		
Poverty	0.0015	0.0011
College Education	-0.0054***	0.0015
Employed	0.0036**	0.0017
Foreign born	-0.0018**	0.0007
Building subsidy		
NYCHA	-0.0011***	0.0004
Other subsidy	-0.0003	0.0002
None	REF	
Year Built		
Missing	-0.0058***	0.0010
1940-2000	0.0003	0.0002
After 2000	0.0000	0.0003
Before 1940	REF	
Crowded Housing (neighborhood)		
Yes	0.0014***	0.0003
No	REF	
Missing BBL		
Yes	0.0029***	0.0010
No	REF	

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	-0.0007	0.002	0.0005	0.001	-0.0032***	0.0010
Residential Units						
1 unit	-0.0005**	0.0002	-0.0002*	0.0001	-0.0004**	0.0002
5-49 units	0.0001	0.0002	-0.0001	0.0001	-0.0001	0.0001
50+ units	-0.0001	0.0002	-0.0002**	0.0001	0.0001	0.0001
0 or missing units	0.0035***	0.0005	0.0007***	0.0002	-0.0001	0.0002
2-4 units	REF		REF		REF	
Unstably housing						
Yes	0.0016**	0.0007	0.0012***	0.0003	0.0002	0.0002
No	REF		REF		REF	
Household size						
Missing	0.0013**	0.0006	0.0001	0.0003	-0.0004	0.0003
2 people	-0.0007***	0.0001	-0.0001*	0.0001	0.0000	0.0001
3-4 people	0.0000	0.0001	0.0002**	0.0001	0.0000	0.0001
4+ people	0.0000	0.0002	0.0006***	0.0001	0.0005***	0.0001
1 person	REF		REF		REF	
Age category						
40-49	0.0012***	0.0001	-0.0003***	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	0.0030***	0.0003	-0.0003	0.0002	0.0021***	0.0002
80+	0.0022***	0.0004	0.0001	0.0002	0.0037***	0.0003
18-39	REF		REF		REF	
Sex						
Female	-0.0016***	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	-0.0002	0.0003	-0.0002	0.0001	-0.0005***	0.0002
Hispanic	0.0002	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***	0.0005	0.0033***	0.0003	0.0037***	0.0003
Liver	0.0033***	0.0003	0.0012***	0.0002	0.0009***	0.0002
Dementia	0.0101***	0.0007	0.0033***	0.0004	0.0017***	0.0004
Arthritis	0.0000	0.0005	0.0007**	0.0003	0.0003	0.0004
Neighborhood Characteristics						
Poverty	-0.0013*	0.0008	0.0010**	0.0004	0.0019***	0.0004
College Education	-0.0059***	0.0011	-0.0002	0.0005	0.0007	0.0005
Employed	0.0016	0.0014	0.0010*	0.0006	0.0010	0.0007
Foreign born	-0.0014***	0.0005	-0.0005**	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	-0.0001	0.0001
None	REF		REF		REF	
Year Built						
Missing	-0.0043***	0.0007	-0.0009	0.0007	-0.0006	0.0005
1940-2000	0.0003*	0.0002	0.0000	0.0001	-0.0001	0.0001
After 2000	0.0002	0.0003	0.0000	0.0001	-0.0002	0.0002
Before 1940	REF		REF		REF	
Crowded Housing (neighborhood)						
Yes	0.001***	0.0002	0.0003***	0.0001	0.0001	0.0001
No	REF		REF		REF	
Missing BBL						
Yes	0.0017***	0.0006	0.0004	0.0007	0.0008*	0.0005
No	REF		REF		REF	

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

Appendix Table 1.
Regression Results including Commute Variables

Variables	β	SE
Population Density	-0.0016	0.0031
Residential Units		
1 unit	-0.0011***	0.0003
5-49 units	0.0000	0.0002
50+ units	-0.0002	0.0003
0 or missing units	0.0044***	0.0006
2-4 units	REF	
Unstably housing		
Yes	0.0032***	0.0008
No	REF	
Household size		
Missing	0.0011	0.0007
2 people	-0.0008***	0.0002
3-4 people	0.0001	0.0002
4+ people	0.0011***	0.0002
1 person	REF	
Age category		
40-49	0.0010***	0.0002
50-59	0.0029***	0.0002
60-69	0.0038***	0.0003
70-79	0.0048***	0.0004
80+	0.0061***	0.0005
18-39	REF	
Sex		
Female	-0.002***	0.0002
Male/Unknown	REF	
Race/ethnicity		
Unknown	-0.0008***	0.0003
Black	-0.0007**	0.0003
API	-0.0053***	0.0003
Other	-0.0009***	0.0004
Hispanic	-0.0003	0.0003
White	REF	
Dual Medicare/Medicaid		
Yes	-0.0018***	0.0003
No	REF	
Pre-existing conditions		

Heart disease	0.0101***	0.0004
Asthma	0.0026***	0.0003
Diabetes	0.0046***	0.0002
Cancer	0.0017***	0.0004
Kidney	0.0209***	0.0007
Liver	0.0055***	0.0004
Dementia	0.0150***	0.0008
Arthritis	0.0010	0.0008
Neighborhood Characteristics		
Poverty	0.0021*	0.0011
College Education	-0.0041**	0.0017
Employed	0.0045**	0.0018
Foreign born	-0.0009	0.0007
Building subsidy		
NYCHA	-0.0010***	0.0004
Other subsidy	-0.0003	0.0002
None	REF	
Year Built		
Missing	-0.0056***	0.001
1940-2000	0.0003	0.0002
After 2000	0.0001	0.0003
Before 1940	REF	
Crowded housing (neighborhood)		
Yes	0.0015***	0.0003
No	REF	
Missing BBL		
Yes	0.0023**	0.001
No	REF	
Primary commute (neighborhood)		
Car	0.0020	0.0013
Bike/Walk	-0.0024	0.0017
Car Share/Taxi	-0.0211***	0.0043
Work from home	-0.0041	0.0087
Public transportation	REF	

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

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

Variables	β	SE
Bronx (005)		
Primary commute (neighborhood)		
Car	0.0002	0.0031
Bike/Walk	-0.0006	0.0113
Car Share/Taxi	0.0315*	0.0165
Work from home	-0.0119	0.0182
Public transportation	REF	
Brooklyn (047)		
Primary commute (neighborhood)		
Car	0.0097***	0.0031
Bike/Walk	0.0083**	0.0042
Car Share/Taxi	-0.0374***	0.0124
Work from home	0.0031	0.0134
Public transportation	REF	
Manhattan (061)		
Primary commute (neighborhood)		
Car	0.0229**	0.0100
Bike/Walk	0.0010	0.0026
Car Share/Taxi	-0.0655***	0.0213
Work from home	0.0184	0.0203
Public transportation	REF	
Queens (081)		
Primary commute (neighborhood)		
Car	0.0083***	0.0030
Bike/Walk	0.0190**	0.0096
Car Share/Taxi	-0.0396***	0.0093
Work from home	-0.0019	0.0237
Public transportation	REF	
Staten Island (085)		
Primary commute (neighborhood)		
Car	-0.0143**	0.0070
Bike/Walk	0.0211	0.0291
Car Share/Taxi	0.0499	0.0340
Work from home	-0.0227	0.0495
Public transportation	REF	

Models include all other variable. Standard errors cluster by census tract
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.1$