



# Exposure density and neighborhood disparities in COVID-19 infection risk

Boyeong Hong<sup>a</sup>, Bartosz J. Bonczak<sup>a</sup>, Arpit Gupta<sup>b</sup>, Lorna E. Thorpe<sup>c</sup>, and Constantine E. Kontokosta<sup>a,d,1</sup>

<sup>a</sup>Marron Institute of Urban Management, New York University, New York, NY 10011; <sup>b</sup>Stern School of Business, New York University, New York, NY 10012; <sup>c</sup>Department of Population Health, New York University School of Medicine, New York, NY 10016; and <sup>d</sup>Center for Urban Science and Progress, New York University, Brooklyn, NY 11201

Edited by Douglas S. Massey, Princeton University, Princeton, NJ, and approved February 8, 2021 (received for review October 10, 2020)

Although there is increasing awareness of disparities in COVID-19 infection risk among vulnerable communities, the effect of behavioral interventions at the scale of individual neighborhoods has not been fully studied. We develop a method to quantify neighborhood activity behaviors at high spatial and temporal resolutions and test whether, and to what extent, behavioral responses to social-distancing policies vary with socioeconomic and demographic characteristics. We define *exposure density* ( $E_{\rho}$ ) as a measure of both the localized volume of activity in a defined area and the proportion of activity occurring in distinct land-use types. Using detailed neighborhood data for New York City, we quantify neighborhood exposure density using anonymized smartphone geolocation data over a 3-mo period covering more than 12 million unique devices and rasterize granular land-use information to contextualize observed activity. Next, we analyze disparities in community social distancing by estimating variations in neighborhood activity by land-use type before and after a mandated stay-at-home order. Finally, we evaluate the effects of localized demographic, socioeconomic, and built-environment density characteristics on infection rates and deaths in order to identify disparities in health outcomes related to exposure risk. Our findings demonstrate distinct behavioral patterns across neighborhoods after the stay-at-home order and that these variations in exposure density had a direct and measurable impact on the risk of infection. Notably, we find that an additional 10% reduction in exposure density city-wide could have saved between 1,849 and 4,068 lives during the study period, predominantly in lower-income and minority communities.

mobility behavior | neighborhood disparities | COVID-19 | computational modeling | geolocation data

As of December 17, 2020, there have been 73 million cases of COVID-19 in more than 200 countries, and 1.6 million people have lost their lives to the disease (1). The COVID-19 pandemic is considered the most severe public health crisis since the 1918 flu pandemic due to its transmission and infection characteristics (2–5). Social distancing (also referred to as physical distancing) has been shown to be an effective behavioral nonpharmaceutical intervention to reduce the transmission rate of COVID-19 (3–7). Social distancing reduces the probability of contacts between individuals who might be infected, resulting in reduced exposure risk (7, 8). Governments have implemented a range of social-distancing policies, including travel bans, restrictions on gatherings, school closures, nonessential business closures, and restaurant restrictions. In particularly hard-hit locations, mandatory “stay-at-home” orders have been issued to limit or avoid unnecessary close contacts outside of the home (7–9).

Studies have found that social-distancing measures help to prevent transmission of the virus and reduce the reproduction ( $R_0$ ) number (5–7, 10–14). These practices help to avoid overwhelming hospital intensive care units and healthcare systems, control doubling time of infections, and ultimately save lives (5, 8, 14, 15). Although not without potentially significant hardship

to individuals and communities, social distancing is an important public health tool to flatten the epidemic curve and support longer-term economic and public health benefits (3, 15–17).

However, the impact of, and response to, stay-at-home orders and social-distancing guidelines is not uniform across neighborhoods and communities (18, 19). In order to maximize the positive effects of social distancing, individuals need to change their typical behavior, often dramatically (3, 20). Despite government-mandated social-distancing policies (such as New York State’s PAUSE order), socio-behavioral responses vary across neighborhoods, further contributing to disparities in risk of infection (4, 7, 21). Disparities in social-distancing practices—namely, geographic or population subgroup differences in adopting behavior changes in response to the same policy context—may stem from varying levels of awareness, perception, or belief in the severity of the virus threat; differences in social and cultural norms; or the ability of households and communities to alter normal activity patterns given economic constraints or other existing responsibilities (7, 20–23). For example, lower-income households typically do not have the option to work from home, and going to a place of work (often in essential services) is unavoidable, meaning higher risk of exposure to COVID-19 for themselves, as well

## Significance

We present a computational approach to measure exposure density at high spatial and temporal resolution to understand neighborhood disparities in transmission risk of COVID-19. By integrating geolocation data and granular land-use information, we are able to establish both the extent of activity in a particular neighborhood and the nature of that activity across residential, nonresidential, and outdoor activities. We then analyze the differential behavioral response to social-distancing policies based on local risk factors, built-environment characteristics, and socioeconomic inequality. Our results highlight the significant disparities in health outcomes for racial and ethnic minorities and lower-income households. Exposure density provides an additional metric to further explain and understand the disparate impact of COVID-19 on vulnerable communities.

Author contributions: B.H., B.J.B., A.G., L.E.T., and C.E.K. designed research; B.H., B.J.B., and C.E.K. performed research; B.H., B.J.B., and C.E.K. analyzed data; B.H., B.J.B., and C.E.K. wrote the paper; B.H. designed the study and methods and analyzed data; B.J.B. contributed to the study design and methods and processed and analyzed data; A.G. provided data and reviewed the study design; and L.E.T. reviewed the study design and analysis; and C.E.K. conceived the research idea, designed and structured the research study and methods, and contributed to data processing and review of results.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

<sup>1</sup>To whom correspondence may be addressed. Email: ckontokosta@nyu.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2021258118/-DCSupplemental>.

Published March 16, 2021.

as their families and communities (7, 24). Within specific neighborhoods, norms can also be reinforcing; if large numbers of residents are essential workers and not socially distancing, other residents may have similar behavioral responses (20).

A growing number of outbreaks are occurring in densely populated areas (25), with disproportionate impacts on lower-income and predominantly minority communities (18, 26–28). Measuring and understanding social distancing and behavior change across neighborhoods can provide critical insight into the design and implementation of more effective—and equitable—public health policy. Given the potential heterogeneity in localized responses to social-distancing recommendations, quantifying local patterns of activity represents an emerging tool to understand and eventually reduce local exposure risk and limit community outbreaks (7, 29, 30). Although there has been increasing awareness of the troubling disparities in infection rates and outcomes in vulnerable communities, the effectiveness of behavioral interventions at the scale of individual neighborhoods has not been fully studied. Often, studies that do attempt to observe effects at higher spatial resolutions rely on simulations or are limited to relatively coarse areal units (e.g., county or state) due to data availability and computational constraints (31–34). Absent a more complete understanding of neighborhood activity patterns in response to nonpharmaceutical interventions, disaggregating built-environment, behavioral, and social determinants of health in the context of COVID-19 remains a challenge.

We develop a method to quantify neighborhood activity at high spatial and temporal resolutions to test whether—and to what extent—behavioral responses to social-distancing policies vary with socioeconomic, demographic, and built-environment characteristics. We define *exposure density* ( $E_x\rho$ ) as a measure of both the localized volume of activity in a defined area and the proportion of activity occurring in nonresidential and outdoor land uses, areas that can be associated with an increased risk of exposure to others that may be infected. We utilize this approach to capture community inflows/outflows of people as a result of the pandemic and changes in mobility behavior for those that remain.

Our focus is on New York City (NYC), the first epicenter of the pandemic in the United States, where a statewide stay-at-home order (NY on PAUSE) was introduced on March 22, 2020. By June 30, 2020, NYC had more than 212,000 confirmed cases of COVID-19, accounting for 8% of the nationwide total, resulting in at least 18,492 confirmed deaths and 4,604 probable deaths (35). Our methodology proceeds in three steps. First, we develop a generalizable method for assessing neighborhood activity levels using smartphone geolocation data over a 3-mo period (February, March, and April) covering more than 12 million unique devices within the Greater New York area, together with land-use classifications at 1-m grid resolution. Second, we measure and analyze disparities in community social distancing by estimating variations in neighborhood activity and associated patterns in community characteristics before and after the stay-at-home order. Finally, we evaluate the effect of exposure density on COVID-19 infection rates associated with localized demographic, socioeconomic, and built-environment characteristics in order to identify disparities in health outcomes related to mobility behavior. Our findings provide insight into the timely evaluation of the effectiveness of social distancing at the scale of individual neighborhoods and support a more equitable allocation of resources to vulnerable and at-risk communities.

### Measuring Exposure Density by Neighborhood over Time

We explore three hypotheses. First, large-scale mobility data can represent neighborhood activity levels over time, and neighborhood social distancing can be measured by changes in

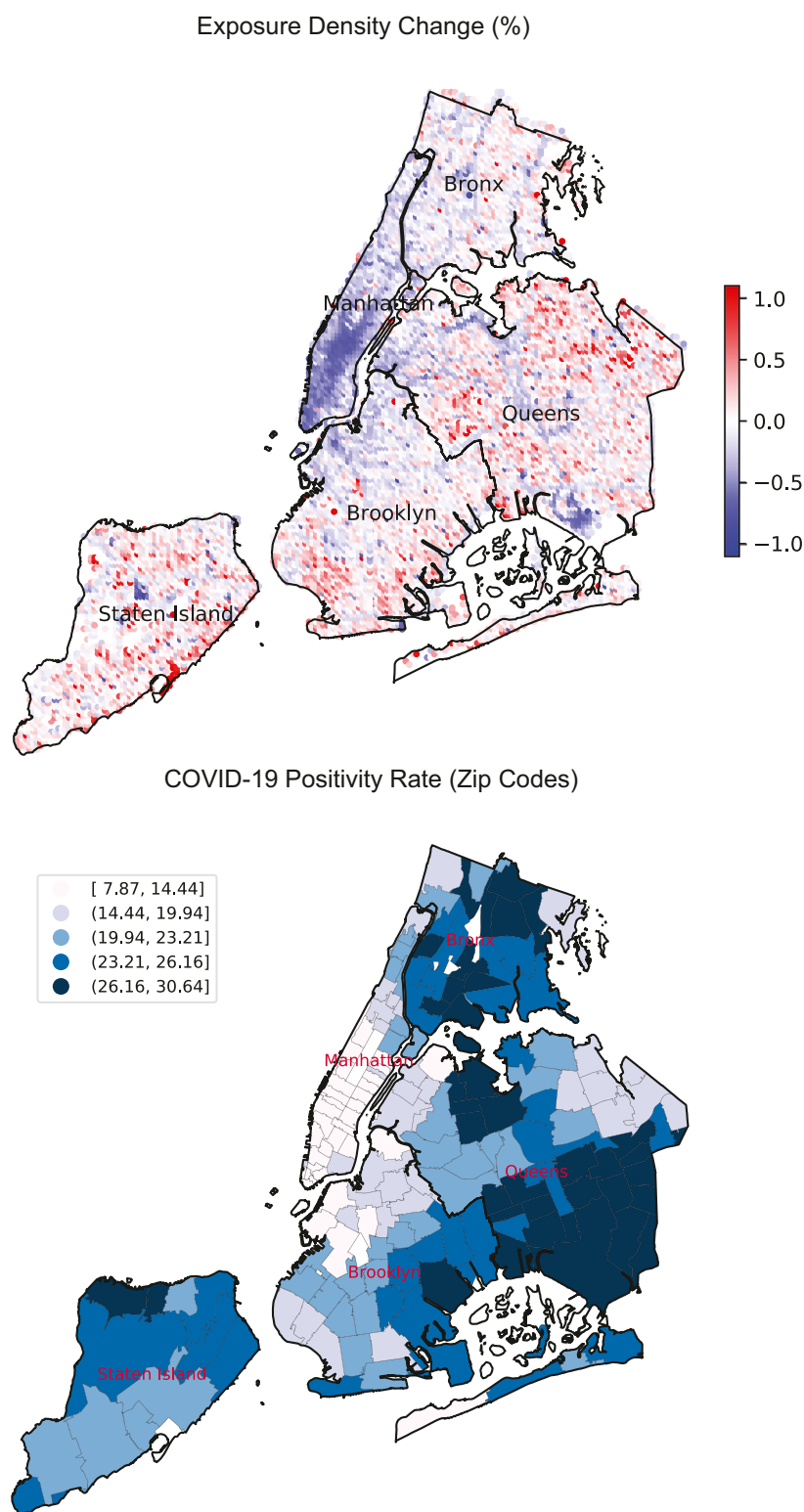
this observed activity. Second, disparities in community activity changes before and after a stay-at-home order are associated with neighborhood socioeconomic, demographic, and built-environment characteristics. Third, variations in neighborhood social distancing result in disparities in COVID-19 infections and outcomes, controlling for differences in population health risk.

To examine these questions, we introduce exposure density ( $E_x\rho$ ) as a high-spatiotemporal-resolution social-distancing metric using large-scale mobility data without tracking individual devices. The goal of social distancing is to reduce the probability of contact between potentially infected and noninfected individuals; therefore, it can be defined mathematically as the inverse proportion of human activity density, represented by the number of people in a given area at a given time. Naively, a lower activity volume, holding spatial area constant, results in a lower dynamic population density, thus decreasing the probability of close contacts. However, this metric needs to account for both the volume of activity in an area and the type of land use where activities occur. For example, activities in residential buildings can be a measure of people staying at home, while activities outside of residential buildings, depending on the specific nature of those activities, are more likely to increase exposure risk by raising the likelihood of contact with those outside of the family or household unit. As transmission risk increases with a greater probability of close contacts outside of the household or family unit, we quantify  $E_x\rho$  based on activities in nonresidential buildings (e.g., office buildings, hotels, and retail stores) and outdoor areas (e.g., parks, sidewalks, and open spaces). We measure the average number of hourly users per grid cell (250 m × 250 m) outside of residential buildings for 177 zip code tabulation areas during the pre-COVID period and after the stay-at-home order.

The average change in neighborhood exposure density before and after the New York stay-at-home order (by grid cell) and COVID-19 infection positivity rates (by zip code) are presented in Fig. 1. The positivity rate is a measure of the prevalence of disease infection, represented by the percentage of COVID-positive tests out of all tests conducted in a given area using a PCR test (*SI Appendix, Table S3*). The citywide overall activity volume decreased approximately 20% after the stay-at-home order when compared to the pre-COVID baseline (*SI Appendix, Fig. S2*). However, there are significant disparities in neighborhood exposure density levels across the city, as shown in Fig. 1, *Upper*. A majority of neighborhoods in Manhattan, and several in Brooklyn, experienced large reductions in exposure density, a result, in part, of a decrease in overall population as many residents left the city, and a shift in activities from nonresidential and outdoor areas to residential buildings for those that remained. On the other hand, neighborhoods in South Brooklyn, East Queens, and Staten Island showed an increase in exposure density, despite having relatively lower urban densities, as more residents stayed within their local communities. The measured change in exposure density corresponds with higher positivity rates, as illustrated in Fig. 1, *Lower*. Overall, this visual representation suggests that areas with lower median incomes and lower housing density had greater infection risk during the study period.

### Community Disparities in Behavioral Responses to Social Distancing

Neighborhoods are classified into groups based on changes in community exposure density before and after the stay-at-home order by using a hierarchical agglomerative clustering algorithm (see *Materials and Methods* for a detailed description). Fig. 2 visualizes the spatial patterns of the clustering output with associated time series of neighborhood activity and where (by land-use type) that activity is occurring. In order to contextualize neighborhood activity patterns, we collect and integrate

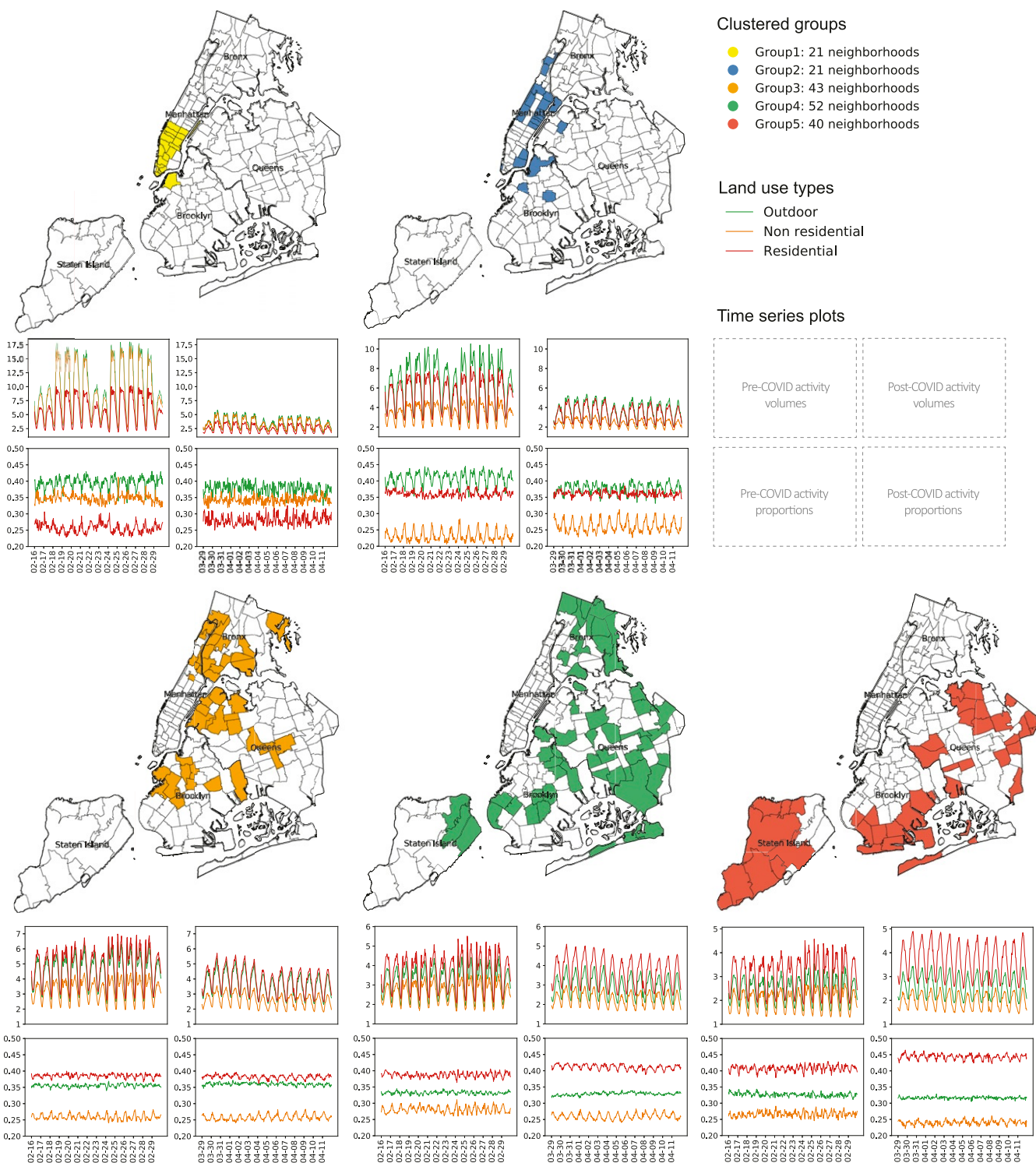


**Fig. 1.** Neighborhood exposure density change by 250-m  $\times$  250-m grid cell (*Upper*) and COVID-19 positivity rate by zip code (*Lower*).

a range of demographic, socioeconomic, housing, and public-health-related variables retrieved from multiple data sources (*Materials and Methods* and *SI Appendix, Table S1*). Descriptive statistics of input variables and neighborhood features for each group, shown in Table 1, reveal distinct neighborhood profiles based on changes in  $E_x\rho$  over time.

We identify five neighborhood clusters based on this analysis. Group 1 (21 zip codes) and group 2 (21 zip codes), which we label “outflow” neighborhoods based on observed activity patterns, are primarily located in Manhattan and downtown Brooklyn and represent substantial changes in  $E_x\rho$  after the stay-at-home order. As shown in Table 1, the average activity





**Fig. 2.** Agglomerative clustering results and associated neighborhood activity change. (Upper) Activity volume by land use. (Lower) Activity proportion by land use.

volume change for group 1 and group 2 is  $-56.5\%$  and  $-33.5\%$ , respectively, meaning that these two neighborhood groups experienced nontrivial declines in normal activity levels—across all land-use types—during the pandemic. Most neighborhoods in group 1 and group 2 have a higher percentage of younger, non-Hispanic White residents, relatively smaller average household size, and higher incomes and educational attainment. This

indicates that residents in these clusters are among the least vulnerable population groups. As such, they may be more likely to have the opportunity to leave their home neighborhoods (or stay at home) by shifting to remote working environments to avoid exposure risk, resulting in reduced exposure density. Even though these two clusters present similar outflow patterns with respect to neighborhood activity volume, the activity-proportion

**Table 1. Descriptive statistics of neighborhood clusters**

Feature	Group 1: Outflow-mixed use (yellow)	Group 2: Outflow-residential (blue)	Group 3: Stable-outflow (orange)	Group 4: Stable-stable (green)	Group 5: Shelter-in-place (red)
Clustering input variables					
Residential volume change, %	−0.52	−0.37	−0.20	−0.01	0.20
Residential proportion change, %	0.12	0.01	−0.01	0.07	0.09
Nonresidential volume change, %	−0.60	−0.28	−0.19	−0.13	−0.00
Nonresidential proportion change, %	−0.01	−0.14	0.00	−0.07	−0.09
Outdoor volume change, %	−0.61	−0.42	−0.18	−0.07	0.07
Outdoor proportion change, %	−0.04	−0.07	0.02	−0.01	−0.03
Exposure density change					
Neighborhood activity change, %	−0.63	−0.40	−0.20	−0.11	0.03
Demographic and socioeconomic features					
Age group 25–34, %	0.28 (0.08)	0.22 (0.05)	0.19 (0.04)	0.16 (0.06)	0.13 (0.02)
Age group over 65, %	0.12 (0.07)	0.15 (0.06)	0.12 (0.04)	0.14 (0.05)	0.17 (0.07)
Black, %	0.05 (0.04)	0.14 (0.17)	0.27 (0.22)	0.31 (0.28)	0.16 (0.25)
Non-Hispanic, %	0.90 (0.05)	0.77 (0.19)	0.60 (0.21)	0.72 (0.07)	0.82 (0.11)
Foreign-born, %	0.16 (0.08)	0.14 (0.05)	0.18 (0.08)	0.15 (0.07)	0.13 (0.08)
Avg. household size	1.92 (0.26)	2.21 (0.35)	2.61 (0.31)	2.90 (0.45)	2.91 (0.37)
College degree, %	0.40 (0.07)	0.31 (0.09)	0.20 (0.08)	0.19 (0.07)	0.20 (0.04)
Unemployment rate	0.04 (0.01)	0.05 (0.03)	0.08 (0.03)	0.08 (0.04)	0.06 (0.02)
Healthcare support workers, %	0.01 (0.01)	0.03 (0.02)	0.06 (0.04)	0.07 (0.04)	0.05 (0.03)
Retail service workers, %	0.03 (0.01)	0.04 (0.02)	0.06 (0.01)	0.05 (0.02)	0.05 (0.02)
Median income, \$	133,000	90,000	54,000	62,000	72,000
Avg. commute time, minutes	27.05 (3.00)	33.83 (4.15)	41.86 (3.23)	44.7 (3.87)	45.30 (3.73)
No health insurance, %	0.04 (0.02)	0.06 (0.03)	0.09 (0.03)	0.09 (0.04)	0.07 (0.04)
Owner-occupied units, %	0.26 (0.12)	0.23 (0.12)	0.22 (0.14)	0.41 (0.21)	0.59 (0.20)
Urban form features					
Residential area, %	0.30 (0.20)	0.71 (0.13)	0.69 (0.14)	0.69 (0.14)	0.71 (0.18)
Office area, %	0.43 (0.24)	0.05 (0.06)	0.05 (0.03)	0.04 (0.03)	0.03 (0.02)
Commercial area, %	0.57 (0.22)	0.24 (0.10)	0.25 (0.12)	0.25 (0.13)	0.21 (0.13)
One or two family units, %	0.00 (0.00)	0.03 (0.05)	0.15 (0.15)	0.41 (0.27)	0.64 (0.26)
Population	959,780	988,652	2,489,946	2,970,495	985,480
COVID-19 features					
Case counts	12,740	20,735	62,151	79,755	25,715
Deaths counts	1,131	2,061	5,397	6,906	1,909
Case rate	1,166.60 (431.88)	1,570.96 (621.38)	2,475.90 (786.84)	2,790.36 (777.17)	2,534.96 (630.57)
Death rate	91.12 (76.79)	150.63 (84.10)	219.87 (83.11)	224.46 (97.73)	195.78 (116.87)
Positivity rate	0.11 (0.03)	0.15 (0.05)	0.22 (0.05)	0.24 (0.04)	0.23 (0.04)

Statistically significant differences between groups are based on one-way ANOVA and Tukey's multicomparison method. Mean values are shown with SD in parentheses. COVID-19 features are based on data provided by the NYCDOH through June 4, 2020.

changes exhibit some notable differences. While the proportion of residential activities in group 1 increased by 12% without any significant changes in nonresidential and outdoor activities, group 2 showed a 14% increase in nonresidential activity, a function of the pre-COVID resident population size. Therefore, we refine the labels for group 1 and group 2 as “outflow-mixed use” and “outflow-residential,” respectively.

Group 3 (43 zip codes) neighborhoods exhibit a 19% decrease, on average, in exposure density. Although we see a marked outflow of residents, these neighborhoods maintain a stable proportion of activity between the different land uses, indicating that residents who remained in these communities largely maintained their regular behavior patterns. When compared to the outflow groups (groups 1 and 2), these “stable-outflow” communities have higher proportions of racial and ethnic minorities, foreign-born residents, and lower median incomes, as well as significantly higher proportions of renter households and those without health insurance. Additionally, a greater percentage of employees in these neighborhoods work in retail services and healthcare support occupations, essential businesses that were not required to close during the outbreak. Like group 3 neigh-

borhoods, communities in the group 4 cluster have stable activity patterns over time; however, these neighborhoods did not see a significant out-mover population. These communities, which we label “stable-stable,” comprise socioeconomically vulnerable households and a high proportion of racial minorities (accounting for approximately 75% of the population), coupled with the second lowest median income, large average household size, high unemployment rate, lower educational attainment, and a large share of healthcare support workers. Such socially and economically vulnerable neighborhoods are less likely to be able to work from home, as the nature of the predominant occupations in these communities often requires physical presence at the workplace, leading to fewer opportunities to reduce exposure to others. We also find that the relatively modest change in exposure density in these “stable” groups (18% and 10% decrease in nonresidential activity density for group 3 and group 4, respectively) is associated with significantly higher infection rates. Particularly, the stable-stable neighborhood group shows the highest case rate (2,790 cases per 100,000 population), death rate (224 deaths per 100,000 population), and positivity rate (24%) in the city.

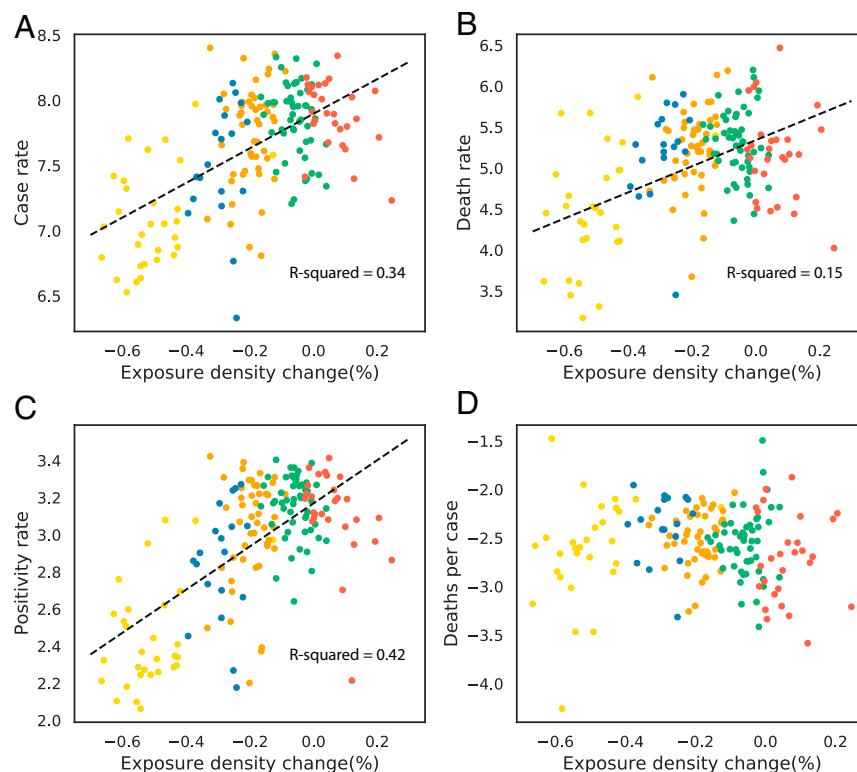
In comparison to other clusters, group 5 (“shelter-in-place”) neighborhoods demonstrate a 20% increase in local activity volume for residential activities and a 7% increase in outdoor activities. In addition to increasing overall neighborhood activity volume, residents staying in these neighborhoods are found to shift activity to residential buildings (by 10%) and away from nonresidential and outdoor activities (by 6%). While nonresidential activities are found to decrease as a proportion of the three activity types, the increase in the overall volume of activity leads to a net increase in exposure density. This group has the highest proportion of elderly population, the largest household size, moderate incomes, a relatively lower percentage of racial and ethnic minorities, and a significantly higher homeownership rate. This indicates that activity in these neighborhoods, where housing density is the lowest in the city, became more localized. As a result, group 5 experienced the second-highest infection rate (2,534 case rate), despite the relatively low built-environment density compared to other neighborhoods.

### Socioeconomic Disparities in Neighborhood Health Outcomes and the Effects of Exposure Density

The results of the bivariate regression model are shown in Fig. 3. Exposure density is found to be correlated with case rate ( $R^2 = 0.34$ ), death rate ( $R^2 = 0.15$ ), and positivity rate ( $R^2 = 0.42$ ), while, as expected, not being a statistically significant determinant of fatality rate (deaths per case). Based on these simple relationships, a 1-percentage-point decrease in exposure density is associated with a 1.33% reduction in case rate, a 1.59% reduction in death rate, and a 1.16% decrease in positivity rate in NYC. By extension, if all neighborhoods reduced exposure density by 10% as compared to normal activity levels prior to the stay-at-home order, approximately 28,960 COVID-19 cases [95% CI of 23,320 to 33,920] could have been avoided, and 2,940 [1,849; 4,068] lives saved through the end of June 2020.

The results of our multivariate regression models, which control for neighborhood socioeconomic, demographic, and built-environment covariates, are described in Table 2. We combine both outflow cluster groups (groups 1 and 2) and use the stable-stable neighborhood group (group 4) as the reference case. As a robustness check, we also specify these models using exposure density as a continuous variable, replacing the neighborhood-cluster dummy variables. After accounting for neighborhood covariates, we continue to observe statistically significant coefficients for the exposure-density variables. The positivity-rate model (model 3) shows the most substantial effects of behavior change on measured health incomes. Neighborhoods (those in groups 1 and 2) that reduced exposure density, largely through outmigration of local population, are shown to have a 44.3% lower positivity rate compared to the reference group. For outflow neighborhoods that maintain the distribution of activities across land-use types (classified as the stable-outflow group), the output shows a 23% lower positivity rate. A similar pattern is also found in the case-rate model (model 1), and the direction and significance of the coefficients are similar in the model specifications using the continuous exposure density variable. These findings provide additional empirical evidence for the effectiveness of social distancing as a nonpharmaceutical intervention strategy to reduce COVID-19 spread, reinforcing that proactive neighborhood behavior change can help to prevent transmission of the virus (5–7, 10–13).

As importantly, race and ethnicity, age group, and socioeconomic status are found to have statistically significant effects on neighborhood infection rates and disease outcomes. Communities with larger proportions of minority and lower-income populations are more likely to be at risk for virus transmission. For example, for every 10% increase of Hispanic residents in a community, the positivity rate increases by 5%, the case rate increases by 9%, and the death rate increases by 6%. This



**Fig. 3.** Scatter plot of exposure density versus the log-transformed cumulative number of COVID-19 cases through June 4, 2020, with linear best-fit lines for significant correlations. (A) Case rate. (B) Death rate. (C) Positivity rate. (D) Deaths per case. Colors represent individual clusters.

**Table 2. Multivariate regression model results**

Feature	Model 1: Case rate	Model 2: Death rate	Model 3: Positivity rate	Model 4: Deaths per case
Num of obs.	177	177	177	177
F-stats.	35.69	16.59	53.26	10.90
Prob > F	0	0	0	0
R <sup>2</sup>	0.77	0.61	0.83	0.50
Intercept	7.040 (0.171)***	2.862 (0.403)***	2.359 (0.116)***	−3.848 (0.249)***
Group outflow-mixed and outflow-residential	−0.632 (0.135)***	−0.716 (0.318)***	−0.443 (0.091)***	0.128 (0.197)
Group stable-outflow	−0.436 (0.142)***	−0.003 (0.335)	−0.228 (0.096)**	0.426 (0.207)**
Group shelter-in-place	−0.051 (0.115)	−0.010 (0.273)	−0.130 (0.078)*	0.050 (0.169)
% Black	0.005 (0.001)***	0.007 (0.002)***	0.004 (0.001)***	0.002 (0.001)*
% Hispanic	0.009 (0.001)***	0.003 (0.003)***	0.005 (0.001)***	0.006 (0.002)*
% Units occupied by owner	0.002 (0.001)*	−0.005 (0.003)	0.003 (0.001)***	−0.007 (0.002)***
% Household with kids	0.012 (0.003)***	0.028 (0.008)***	0.013 (0.002)***	0.014 (0.005)***
% Employees working from home	−0.018 (0.008)**	0.010 (0.019)	−0.016 (0.005)***	0.015 (0.011)
Num of occupied nursing home beds per 100 people	0.036 (0.010)***	0.086 (0.024)***	0.008 (0.007)	0.059 (0.015)***
% Household without health insurance	−0.018 (0.011)*	0.046 (0.025)*	−0.003 (0.007)	0.056 (0.015)***
Insurance × group effect 1	0.062 (0.017)***	0.088 (0.041)***	0.046 (0.012)***	0.001 (0.025)*
Insurance × group effect 2	0.042 (0.014)***	0.010 (0.033)	0.021 (0.010)**	−0.031 (0.021)
Insurance × group effect 3	0.001 (0.013)	−0.008 (0.031)	0.018 (0.009)**	−0.008 (0.019)
Age group over 65	0.014 (0.005)***	0.069 (0.011)***	0.008 (0.003)***	0.043 (0.007)***
% Public housing area	−0.005 (0.003)*	0.006 (0.007)	−0.002 (0.002)	0.009 (0.004)**

SEs are in parentheses. F-stats., F-statistics; Num of obs., number of observations; Prob, probability.

\* $P < 0.10$ ; \*\* $P < 0.05$ ; \*\*\* $P < 0.01$ .

finding holds after accounting for changes in exposure density. As expected, exposure density is not shown to be a statistically significant feature in the death-rate and the deaths-per-case models, while the variables related to the presence of vulnerable populations have significant negative impact on survival probability. We find that the proportion of residents over the age of 65, without health insurance coverage, or living in public housing have positive and statistically significant associations with death rates across the city. Thus, the mortality risk of the virus is higher in socially vulnerable neighborhoods than in other communities, exacerbated by pre-existing health conditions and lack of adequate access to healthcare. This also helps to explain, in part, why the stable-outflow group, which includes neighborhoods with the highest proportion of lower-income residents without health insurance, experienced an approximately 43% higher fatality rate compared to the reference group, despite observed lower infection rates.

## Discussion and Conclusion

We present a computational approach to measure exposure density at high spatial and temporal resolution to understand localized disparities in transmission risk of COVID-19. By integrating geolocation data and granular land-use classifications, we are able to establish both the extent of activity in a particular area and the nature of that activity across residential, nonresidential, and outdoor spaces. This approach is scalable to any areal unit of interest: Here, we utilize a 250-m grid and aggregate to the zip code level to match the geography of reported health data. However, it is possible to apply the same methodology to point locations or grids of any size and then aggregate the units to other common administrative or political boundaries, such as census tracts, counties, and metropolitan areas. We normalize our data to enable comparative studies between regions and to scale the analysis to other cities with similar land-use data resources.

Our findings demonstrate distinct patterns of activity before and after the stay-at-home order across neighborhoods in NYC. These neighborhood patterns are clustered into five distinct groups, each exhibiting statistically significant differences in

socioeconomic, demographic, and built-environment characteristics. In wealthier neighborhoods of Manhattan and Brooklyn, we observe an exodus of residents leaving for other areas around NYC or regions further afield. Presumably, these residents have the means to relocate to second homes or rental homes that provide a greater degree of (perceived) safety from the virus. In addition, residents in these neighborhoods were more likely to work from home before the pandemic, suggesting that these residents had similar opportunities to work remotely after the stay-at-home order, thus reducing the transaction costs of leaving their primary residence. Conversely, we observe clusters of lower-income neighborhoods and areas of minority concentration that faced greater infection risk. While some residents in neighborhoods in the stable groups did relocate, the large majority stayed in their communities and continued on with their typical (pre-COVID) routines. As a result, we find that exposure density in these neighborhoods remained relatively constant over the study period, reflecting the continued need to commute to work and other places of responsibility, especially given that many of those employed worked in occupations deemed essential services (for instance, 12% of employed residents work in retail or healthcare support services) (7). Finally, we find a cluster of neighborhoods that increased their exposure density due to an increase in localized activity. These neighborhoods, characterized by lower-density, single-family homes in areas further from the Manhattan central business district, are found to have both a greater volume of activity and more activity taking place in nonresidential and outdoor areas than normal. The effect of this local activity was an increase, compared to pre-COVID levels, in the probability of coming in contact with others outside of the household or family unit.

The variation in exposure density has a direct and measurable impact on the risk of infection. In neighborhoods where exposure density decreased the most, we find lower rates of infection, positivity rates, and death rates per capita, controlling for other covariates associated with social determinants of health. The communities hardest hit by the virus were in the stable-stable neighborhoods, where residents faced multiple challenges and



risk factors. In addition to continuing their normal activity patterns, and thus exposing themselves to greater risk of infection while commuting and in their place of work, these communities have the largest proportion of racial minorities, among the lowest median incomes, and the lowest rate of health insurance coverage. These compound risks resulted in these vulnerable communities facing the burden of the highest rate of infection, death rate, and positivity rate in the city during the study period. Notably, if these neighborhoods were able to reduce their exposure density by as much as the wealthiest neighborhoods, more than 1,300 lives could have been saved through the end of June 2020.

We note several potential limitations to this work. These include the data availability and coverage, the spatial accuracy of the geolocation data used to assign land-use classifications, and the use of zip codes as an areal unit of analysis. We acknowledge and account for these constraints, as described in *Materials and Methods*.

Nonetheless, our study highlights the importance of understanding neighborhood activity patterns in evaluating the determinants of health outcomes and risk factors for future infection outbreaks. By measuring exposure density at the community scale, we are able to determine the differential behavioral response to social-distancing policies based on local risk factors and socioeconomic inequality. Our results expose the significant disparities in health outcomes for racial and ethnic minorities and lower-income households. Exposure density provides an additional metric to further explain and understand the disparate impact of COVID-19 on vulnerable communities and a tool for the design and evaluation of equitable, targeted public health interventions.

## Materials and Methods

**Data.** Our primary data are anonymized smartphone geolocations collected by VenPath, Inc.—a data-marketplace company providing mobile-application data and business-analytics consulting based on more than 200 smartphone applications across the United States. The approximately 5-TB dataset covers the period from February through April, 2020, and contains more than 127 billion geotagged data points associated with 120 million unique devices every month. Due to the level of granularity and potential reverse-identification risk, a dedicated data-management plan detailing the protocols for access, use, and security of these data was developed, and data were stored in a secured and access-controlled database environment maintained by New York University's (NYU's) High Performance Computing infrastructure. Both the data-processing methods and data-management plan were approved by NYU's Institutional Review Board (approval no. IRB-FY2018-1645), with input from, and review by, NYU Data Services. Furthermore, we developed our methodology so as to avoid tracking of individual devices and, instead, focused on spatial and temporal aggregation of device counts. For the purpose of this study, the data were processed and spatially aggregated to counts at the 250-m grid cell level, which preserves the anonymity of users, especially in a densely populated region such as NYC. For this study, we extracted a subset of data falling within the Greater New York area bounding box extent (40°29'46.0"N 74°15'20.1"W: 40°54'55.9"N 73°42'00.0"W) and adjusted timestamps to the Eastern Standard Time zone, resulting in 12,858,781 unique devices over the study period. After filtering for devices active for at least 14 d over the study period, the processed dataset includes 744,147 unique devices, representing approximately 8.9% of the NYC population. To complement our mobility data, we used a range of ancillary data for analysis and modeling (*SI Appendix, Table S1*). NYC Primary Land Use Tax Lot Output (PLUTO) data were used to obtain land-use and building-type information for every property in the city (36). The building-footprint shapefile was used to identify the exact perimeter of individual buildings (37). NYC LION data—a single line street base map—were used to extract street-segment geometries (38). We used daily NYC COVID-19 information by zip code, which includes confirmed cases, deaths, and positive test rates, provided by the NYC Department of Health and Mental Hygiene (NYCDOH) (35). In order to contextualize neighborhood demographic, socioeconomic, housing, and public-health-related characteristics, we used American Community Survey data from the US Census Bureau, NYC hospital locations from NYC OpenData, and nursing-home data provided by the US Centers for Disease Control and Prevention (39–41).

With the exception of the smartphone geolocation data, all data are publicly available and extracted from NYC or federal open-data platforms.

**Building the Exposure-Density Metric.** Here,  $E_{x\rho}$  is measured as the number of unique devices in a given geographical and temporal unit by land-use type, specified as:

$$E_{x\rho} = f(A_{g,t,L}), \quad [1]$$

where  $g$  is a given geographical unit (e.g., grid cell or census block group),  $t$  is a given temporal unit (e.g., hourly or daily), and  $L$  is the land-use class.

In order to maintain a scalable and uniform areal unit that can be applied across different cities and regions, we divided the NYC study area into a 250-m grid (187 × 186 cells), which we used for aggregation of the mobility data. To integrate the mobility data with land-use information, we created a 1-m resolution raster with the extents and the coordinate system matching the aforementioned 250-m grid. The land-use raster combines the geographical city limits and land-use classification derived from PLUTO data, together with street and sidewalk boundaries and building footprints for more than 1 million buildings. Each category of land cover was then classified by an integer (e.g., 10 for residential property, 50 for outdoor open space, and so on). Each 1-m cell was then identified by its index, location, and associated land-use category. This allowed us to assign each geolocation data point from the mobility dataset to a specific land-use cell.

One limitation to this method is the horizontal accuracy of the mobility data, which can add nontrivial uncertainty to the reported ping location. The geolocation error is a function of the source of the data (application type) and the technology it relies on. Mobile-device locations can be retrieved by using Global Positioning Systems with an estimated accuracy ranging from 1 to 20 m, depending on the area; local Wi-Fi network signals with accuracy up to several hundred meters; cell triangulation providing location at the neighborhood level; and network internet-protocol address location or user-registration information yielding a static location associated with the network hardware (42–44). VenPath data collect geolocation information from a variety of applications that can utilize one or multiple of these technologies, resulting in varying CIs for the geolocation coordinates of a device at a particular time. The calculated average horizontal accuracy obtained directly from the dataset is 52.6 m, and the median accuracy is 16.0 m. Given this uncertainty, and the inability to validate it based on the data provided, we used the reported geolocation coordinates as the device location for the purposes of land-use classification, but also conducted a robustness check using 20-m-grid and 50-m-grid land-use rasters (see *SI Appendix* for results).

To estimate dynamic population density, we counted the hourly number of unique devices in each 250-m grid cell and the corresponding land-use category based on the raster cell. We classified land-use types into three groups to account for mobility behavior and varying infection risk associated with certain places and activities (45, 46). Our data-processing workflow is visualized in *SI Appendix, Fig. S1*. The rasterization process was implemented in Python and deployed on NYU Center for Urban Science and Progress' (CUSP's) Research Computing Facility (RCF), and the activity computation was performed with PySpark on a Hadoop distributed computing cluster using NYU's High Performance Computing platform.

Our 250-m grid-cell-level measurement can be aggregated into larger geospatial units in order to estimate neighborhood activities at different scales. In this work, we used zip code aggregation to align with the spatial resolution of COVID-19 infection data provided by the NYCDOH. The zip code aggregated  $E_{x\rho}$  is defined as:

$$A_{z,g,t,L} = \frac{1}{N_z} \sum_{i=1}^{N_z} A_{g,t,L}, \quad [2]$$

where  $A_{z,g,t,L}$  is the average number of hourly unique devices in a 250-m × 250-m grid cell by land-use type  $L$  in a given zip code  $z$ , and  $N_z$  is the number of grid cells in zip code  $z$ . The various spatial aggregation levels used in our study can introduce the potential risk of bias caused by the modifiable areal unit problem. Our data-integration methods are designed to minimize bias while accounting for privacy concerns and the spatial resolution of available data, particularly the zip-code-level COVID-19 infection data from the NYCDOH. We acknowledge that zip codes are not necessarily socioeconomically or demographically homogeneous and provide only approximations of neighborhood boundaries. However, given the density of zip code areas in NYC, the geographic boundaries provide reasonable proxies for distinct communities. Furthermore, the modified zip code tabulation areas provided by the NYCDOH combine zip code areas with smaller populations to create more stable estimates to reduce statistical uncertainty.



(35). Based on our social-distancing metric, changes in mobility activity by residential, nonresidential, and outdoor land uses in a neighborhood over the study time period were examined. We filtered out activities from major roads used exclusively by motor vehicles (those without sidewalks or pedestrian access) to remove vehicular activity within a given neighborhood. A descriptive summary of citywide hourly average activity volumes and proportions in each land-use category before and after the stay-at-home order can be found in *SI Appendix, Fig. S3 and Table S2*.

**Analyzing Disparities in Exposure Density across the City.** To understand disparities in exposure density and behavioral responses to social-distancing mandates across neighborhoods, we applied an unsupervised machine-learning clustering algorithm based on a *pre/post* comparative analysis. We extracted  $E_{x\rho}$  subsets for two 2-wk periods, defined as the *preimpact* period (February 16 through February 29, 2020) and the *postimpact* period (March 29 through April 11, 2020), to measure changes in  $E_{x\rho}$  before and after the state-mandated stay-at-home order on March 22, 2020. In order to take into account both the absolute change in activity volume and the change in the proportion of activity type, we created six input variables for the zip code clustering analysis, specified as:

$$\begin{aligned} A_{\text{change}, z, L} &= \frac{A_{\text{post}, z, L} - A_{\text{pre}, z, L}}{A_{\text{pre}, z, L}} \\ P_{\text{change}, z, L} &= \frac{P_{\text{post}, z, L} - P_{\text{pre}, z, L}}{P_{\text{pre}, z, L}} \end{aligned} \quad [3]$$

( $L \in$  residential, non-residential, and outdoor)

where  $A_{\text{change}, z, L}$  is average hourly activity volume change for residential, nonresidential, and outdoor land uses in zip code  $z$  based on the preimpact period activity level ( $A_{\text{pre}, z, L}$ ), and the impact period level ( $A_{\text{post}, z, L}$ ).  $P_{\text{change}, z, L}$  is the average hourly change in activity based on the proportion of those activities occurring in different land-use types. Neighborhood activity by land-use classification is defined as the proportion of activity in a given land-use (residential, nonresidential, and outdoor) grid cell.

To identify similarities in the change in  $E_{x\rho}$  across neighborhoods, we applied a hierarchical agglomerative clustering algorithm. Initially, each data point was considered an individual cluster. At each iteration, the closest two clusters merge with one another based on the proximity matrix measured by Euclidean distance until all data points form a single cluster (47). Input data are in the form of a  $177 \times 6$  vector—177 zip code neighborhoods and 6 features—and the optimized number of clusters was determined by the corresponding dendrogram (hierarchical tree diagram) based on the similarities and dissimilarities of the objects. We ran different agglomerative clustering models using complete, average, and Ward's linkage methods, and the resultant dendrograms are included in *SI Appendix, Fig. S4*. We selected the Ward's linkage method in order to minimize within-group variance while maximizing efficiency and variance among groups, instead of comparing the direct sample distances, as explained by smaller merging cost. This clustering process is specified as:

$$\begin{aligned} \Delta(C_i, C_j) &= \sum_{k \in C_i \cup C_j} \|\bar{x}_k - \bar{m}_{C_i \cup C_j}\|^2 - \\ &\sum_{k \in C_i} \|\bar{x}_k - \bar{m}_{C_i}\|^2 - \sum_{k \in C_j} \|\bar{x}_k - \bar{m}_{C_j}\|^2 \end{aligned} \quad [4]$$

where  $\Delta(C_i, C_j)$  is a merging cost of combining clusters  $C_i$  and  $C_j$  (distance between clusters),  $\bar{m}_C$  is the centroid of cluster  $C$ , and  $\bar{x}_k$  is an individual element within a cluster. The initial number of optimized clusters suggested by the hierarchical tree diagram (*SI Appendix, Fig. S4*) is two ( $n = 2$ ), which maximizes between-group variance. When using  $n = 2$ , the clustering result is significantly influenced by activity volume change, rather than proportion change, features due to the larger variable scale, resulting in an imbalanced cluster size that divides neighborhoods into either Manhattan or non-Manhattan groups. In order to take into account neighborhood activity proportion changes and to balance cluster-group size, we selected five cluster groups ( $n = 5$ ), keeping within-group variance small and between-group variance large, while satisfying the balancing condition. The resultant clustered neighborhood groups were then integrated with demographic and socioeconomic characteristics; housing and urban form features; and COVID-19 infection and outcome data. By using a one-way ANOVA test and a Tukey's test for posthoc analysis, we identified statistically significant differences in neighborhood characteristics between classified groups.

**Identifying the Impact of Exposure Density and Neighborhood Behavior Change on Infection Risk.** In order to evaluate the effect of neighborhood behavior change on COVID-19 infection rates for the 177 zip code neighborhoods included in the study, we first estimated Pearson correlation coefficients for observed community-activity changes before and after the stay-at-home order and disease-infection case rates—daily new confirmed cases per 100,000 people and cumulative cases per 100,000 people—while accounting for an incubation period. We observed statistically significant positive correlations between exposure density and infection rates ( $r = 0.52$  and  $r = 0.47$ , respectively).

Then, we developed bivariate and multivariate log-transformed regression models to identify any statistically significant effects of  $E_{x\rho}$  on infection risk, controlling for neighborhood characteristics. Four ordinary least squares models were specified, each with a dependent variable representing one of four measures of COVID-19 infection risk (*SI Appendix, Table S3*), including case rate, death rate, positivity rate, and deaths per case. One limitation of COVID-19 per capita infection-rate measures is that they are based on annual census population estimates as of July 1, 2019 (35). These rates, therefore, do not account for dynamic changes in localized resident population, such as those caused by out-movers in response to the pandemic. Therefore, we focused on positivity rates in our model and confirmed using ANOVA that there were no statistically significant differences in testing rates across neighborhoods. In order to account for this limitation for death rates, we created a deaths-per-case variable based on the World Health Organization's case-fatality ratio (48). *SI Appendix, Table S4* provides descriptive statistics for the included independent variables. The bivariate models take  $E_{x\rho}$  change (as a percent) as a continuous variable to measure the marginal effects of activity change on infection rates. The multivariate models use dummy variables for each clustered neighborhood group to evaluate disparities between groups and are respecified to include a continuous exposure density variable as a robustness check. The linear models are specified as:

$$\begin{aligned} y &= \beta_0 + \beta_1 X_1 + \varepsilon \quad (\text{Bivariate model}) \\ y &= \beta_0 + \sum_{i=1}^n \beta_i X_i + \varepsilon \quad (\text{if } n > 1, \text{ multivariate model}) \end{aligned} \quad [5]$$

where  $y$  is the logarithmic transformed zip-code-level COVID-19 outcome variable, cumulative COVID-19 case data from March 1 through June 4, 2020;  $X_1$  for the bivariate model is  $E_{x\rho}$  change;  $X_i (i > 1)$  for the multivariate model includes the cluster group dummy variables and the set of neighborhood demographic, socioeconomic, and built-environment features; and  $\varepsilon$  is the error term. We also considered interaction terms between neighborhood groups and other social determinants of health. Spatial dependence of COVID infection risk was a consideration. The benefit of using the neighborhood cluster dummies is the geographic proximity of the grouped neighborhoods. As shown in Fig. 2, the clusters reflect the socioeconomic and demographic landscape of NYC, which also accounts for variations in infection prevalence across zip code boundaries. Thus, we are capturing potential spatial spillover effects by using the cluster dummies in the regression model. In order to test the spatial dependency of COVID-19 infections more fully, we respecified our multivariate regression models by including a spatial dummy variable to account for adjacency to neighborhoods with high disease burden and ran a spatial lag model using the  $k$ -nearest-neighbor method to create spatial weights. The results of both modeling approaches reinforce the results as presented and do not substantially change the magnitude or direction of the exposure density coefficients. Finally, we used correlation tests and Variance Inflation Factors analysis to identify multicollinearity as part of the feature-selection process. The coefficients  $\beta_i$  quantify the effects of neighborhood  $E_{x\rho}$ .

**Data Availability.** The primary mobility data that support the findings of this study are available from VenPath, Inc., but restrictions apply to the availability of these data, which were used under data sharing agreement. The processed aggregate mobility data that are directly used for clustering and regression analyses are available, upon reasonable request, from the corresponding author, subject to any restrictions related to the NYU Institutional Review Board approval and with permission of the data provider.

Additional data needed to evaluate the analyses in the paper are described in *SI Appendix, Table S1*. Land use and building type information data is available through the New York City Department of City Planning (NYCDP) website ([https://www1.nyc.gov/site/planning/data-maps/open-data/bytes-archive.page?sorts\[year\]=0](https://www1.nyc.gov/site/planning/data-maps/open-data/bytes-archive.page?sorts[year]=0)). Data on building footprints is available from the New York City Department of Information Technology &

Telecommunications (DoITT) (<https://www1.nyc.gov/site/doitt/residents/gis-2d-data.page>), and road network information from NYCDP (<https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-lion.page>). The COVID-19 case, death, and test information was obtained from the New York City Department of Health and Mental Hygiene (NYCDOH) GitHub repository (<https://github.com/nychealth/coronavirus-data>), the hospital location information through the NYC OpenData platform (<https://data.cityofnewyork.us/Health/NYC-Health-Hospitals-Facilities-2011/yhmhw-9cz9>), and the nursing home locations through the Centers for Disease Control's (CDC) National Healthcare Safety Network System (<https://data.cms.gov/stories/s/bkwz-xpvg>). All demographic and socioeconomic data were retrieved from the US American Community Survey (ACS) administrated by Census Bureau (<https://www.census.gov/data/developers/data-sets/acs-5year.2018.html>). All ancillary data related to the current study may be requested from the corresponding author upon reasonable

request and with permission of the data provider if data are not publicly available.

The code used to process data and perform the analysis for this paper, as well as resulting models and figures, is openly available in the publicly accessible repository GitHub (<https://github.com/UrbanIntelligenceLab/Exposure-Density-and-Neighborhood-Disparities-in-COVID-19-Infection-Risk>) under MIT License.

**ACKNOWLEDGMENTS.** We thank the NYU High Performance Computing team and the NYU CUSP RCF for providing the computing infrastructure necessary for this work; and VenPath, Inc. for providing geolocation data. This work was supported by NSF Grant 2028687; and by NYU C2SMART, a US Department of Transportation Tier 1 University Transportation Center. Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of any supporting institution. All errors remain our own.

- World Health Organization, WHO coronavirus disease (COVID-19) dashboard (2020). <https://covid19.who.int/>. Accessed 17 December 2020.
- World Health Organization, "Coronavirus disease 2019 (COVID-19): Situation report" (Situation Report 72, World Health Organization, Geneva, Switzerland, 2020).
- B. Sen-Crowe, M. McKenney, A. Elkbali, Social distancing during the COVID-19 pandemic: Staying home save lives. *Am. J. Emerg. Med.* **38**, 1519–1520 (2020).
- C. Courtemanche, J. Garuccio, A. Le, J. Pinkston, A. Yelowitz, Strong social distancing measures in the United States reduced the COVID-19 growth rate: Study evaluates the impact of social distancing measures on the growth rate of confirmed COVID-19 cases across the United States. *Health Aff.* **39**, 1237–1246 (2020).
- S. Gao et al., Mobile phone location data reveal the effect and geographic variation of social distancing on the spread of the COVID-19 epidemic. *JAMA Netw. Open* **3**, e2020485 (2020).
- A. Chudik, M. H. Pesaran, A. Rebucci, "Voluntary and mandatory social distancing: Evidence on COVID-19 and exposure rates from Chinese provinces and selected countries" (NBER Working Paper 27039, National Bureau of Economic Research, Cambridge, MA, 2020). [https://cepr.org/active/publications/discussion\\_papers/dp.php?dpno=14646](https://cepr.org/active/publications/discussion_papers/dp.php?dpno=14646). Accessed 1 July 2020.
- J. Jay et al., Neighbourhood income and physical distancing during the COVID-19 pandemic in the United States. *Nature Hum. Behav.* **4**, 1294–1302 (2020).
- C. Adolph, K. Amano, B. Bang-Jensen, N. Fullman, J. Wilkerson, Pandemic politics: Timing state-level social distancing responses to COVID-19. *Science*, 10.1215/03616878-8802162 (2020).
- M. Chinazzi et al., The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
- R. Abouk, B. Heydari, The immediate effect of COVID-19 policies on social-distancing behavior in the United States. *Public Health Rep.* **136**, 245–252 (2021).
- R. J. Hatchett, C. E. Mecher, M. Lipsitch, Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7582–7587 (2007).
- M. Farboodi, G. Jarosch, R. Shimer, "Internal and external effects of social distancing in a pandemic" (NBER Working Paper 27059, National Bureau of Economic Research, Cambridge, MA, 2020) [https://www.nber.org/system/files/working\\_papers/w27059/w27059.pdf](https://www.nber.org/system/files/working_papers/w27059/w27059.pdf). Accessed 1 July 2020.
- L. Matrajt, T. Leung, Evaluating the effectiveness of social distancing interventions to delay or flatten the epidemic curve of coronavirus disease. *Emerg. Infect. Dis.* **26**, 1740–1748 (2020).
- D. Duque et al., Timing social distancing to avert unmanageable COVID-19 hospital surges. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 19873–19878 (2020).
- Greenstone M., Nigam V. "Does social distancing matter?" (Becker Friedman Institute for Economics Working Paper 2020-26, University of Chicago, Chicago, IL, 2020). <http://dx.doi.org/10.2139/ssrn.3561244>. Accessed 1 July 2020.
- L. Thunstrom, S. Newbold, D. Finnoff, M. Ashworth, J. Shogren, The benefits and costs of using social distancing to flatten the curve for COVID-19. *J. Benefit-Cost Anal.* **11**, 179–195 (2020).
- A. Sheridan, A. L. Andersen, E. T. Hansen, N. Johannesen, Social distancing laws cause only small losses of economic activity during the COVID-19 pandemic in Scandinavia. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20468–20473 (2020).
- M. W. Hooper, A. M. Nápoles, E. J. Pérez-Stable, COVID-19 and racial/ethnic disparities. *JAMA* **323**, 2466–2467 (2020).
- G. A. Millett et al., Assessing differential impacts of COVID-19 on black communities. *Ann. Epidemiol.* **47**, 37–44 (2020).
- J. J. Van Bavel et al., Using social and behavioural science to support COVID-19 pandemic response. *Nat. Hum. Behav.* **4**, 460–471 (2020).
- J. A. Weill, M. Stigler, O. Deschenes, M. R. Springborn, Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 19658–19660 (2020).
- T. Wise et al., Changes in risk perception and protective behavior during the first week of the COVID-19 pandemic in the United States. *R. Soc. Open Sci.* **7**, 200742 (2020).
- P. Caley, D. J. Philp, K. McCracken, Quantifying social distancing arising from pandemic influenza. *J. R. Soc. Interface* **5**, 631–639 (2008).
- C. J. Atchison et al., Early perceptions and behavioural responses of the general public during the COVID-19 pandemic: A cross-sectional survey of UK adults. *BMJ Open* **11**, e043577 (2021).
- D. Desai, "Urban densities and the Covid-19 pandemic: Upending the sustainability myth of global megacities." ORF Occasional Paper 244, Observer Research Foundation, New Delhi, India, 2020.
- M. S. Bryan et al., Coronavirus disease 2019 (COVID-19) mortality and neighborhood characteristics in Chicago. *Ann. Epidemiol.*, 10.1016/j.annepidem.2020.10.011 (2020).
- T. Kirby, Evidence mounts on the disproportionate effect of COVID-19 on ethnic minorities. *Lancet. Resp. Med.* **8**, 547–548 (2020).
- E. G. Price-Haywood, J. Burton, D. Fort, L. Seoane, Hospitalization and mortality among black patients and white patients with covid-19. *N. Engl. J. Med.* **382**, 2534–2543 (2020).
- S. Adhikari et al., Assessment of community-level disparities in coronavirus disease 2019 (COVID-19) infections and deaths in large US metropolitan areas. *JAMA Netw. Open* **3**, e2016938 (2020).
- L. J. Thomas et al., Spatial heterogeneity can lead to substantial local variations in COVID-19 timing and severity. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 24180–24187 (2020).
- G. A. Wellenius et al., Impacts of state-level policies on social distancing in the United States using aggregated mobility data during the COVID-19 pandemic. arXiv [Preprint] (2020). arXiv:2004.10172 (Accessed 1 July 2020).
- A. Aktay et al., Google COVID-19 community mobility reports: Anonymization process description (version 1.0). arXiv [Preprint] (2020). arXiv:2004.04145 (Accessed 1 July 2020).
- J. Bayham et al., Colorado COVID-19 Modeling Group, Colorado Mobility Patterns During the COVID-19 Response. (Denver, CO, 2020) <http://www.ucdenver.edu/academics/colleges/PublicHealth/coronavirus/Documents/Mobility%20Report.final.pdf>. Accessed 1 July 2020.
- K. H. Grantz et al., The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nat. Commun.* **11**, 4961 (2020).
- New York City Department of Health and Mental Hygiene, NYC coronavirus disease 2019 (COVID-19) data (2020) <https://github.com/nychealth/coronavirus-data>. Accessed 30 June 2020.
- New York City Department of City Planning, Primary Land Use Tax Lot Output (2020). [https://www1.nyc.gov/site/planning/data-maps/open-data/bytes-archive.page?sorts\[year\]=0](https://www1.nyc.gov/site/planning/data-maps/open-data/bytes-archive.page?sorts[year]=0). Accessed 30 June 2020.
- New York City Department of Information Technology & Telecommunication, Building Footprint Data (2020). <https://www1.nyc.gov/site/doitt/residents/gis-2d-data.page>. Accessed 30 June 2020.
- New York City Department of City Planning, Road Network Data (LION). <https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-lion.page>. Accessed June 30 2020.
- Census Bureau, American Community Survey 2018 5-Year Estimate. <https://www.census.gov/data/developers/data-sets/acs-5year.2018.html>. Accessed June 1 2020.
- New York City OpenData, NYC Health + Hospitals Facilities. <https://data.cityofnewyork.us/Health/NYC-Health-Hospitals-Facilities-2011/yhmhw-9cz9>. Accessed June 1 2020.
- Centers for Disease Control, National Healthcare Safety Network System, Nursing Home Data. <https://data.cms.gov/stories/s/bkwz-xpvg>. Accessed June 30 2020.
- Mobile Marketing Association et al., Demystifying location data accuracy (2017). <https://www.mmaglobal.com/files/documents/location-data-accuracy-v3.pdf>. Accessed 15 November 2020.
- C. Wu, Y. Zhao, Y. Liu, Y. Xu, Human mobility enhances global positioning accuracy for mobile phone localization. *IEEE Trans. Parallel Distr. Syst.* **26**, 131–141 (2015).
- J. Wang, X. Kong, F. Xia, L. Sun, Urban human mobility: Data-driven modeling and prediction. *SIGKDD Explor. Newsl.* **21**, 1–19 (2019).
- S. G. Benzell, A. Collis, C. Nicolaidis, Rationing social contact during the COVID-19 pandemic: Transmission risk and social benefits of us locations. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 14642–14644 (2020).
- R. K. Bhagat, M. D. Wykes, S. B. Dalziel, P. Linden, Effects of ventilation on the indoor spread of COVID-19. *J. Fluid Mech.* **903** (2020).
- J. Friedman, T. Hastie, R. Tibshirani, *The Elements of Statistical Learning* (Springer Series in Statistics, Springer, New York, NY, 2001), vol. 1.
- World Health Organization, Estimating mortality from COVID-19 (Tech. Rep., World Health Organization, Geneva, Switzerland, 2020; <https://www.who.int/news-room/commentaries/detail/estimating-mortality-from-covid-19>).