Article

Socio-spatial inequality and the effects of density on COVID-19 transmission in US cities

Constantine E. Kontokosta [1,2,3] ✉, Boyeong Hong[1] & Bartosz J. Bonczak [1,2,3]

Cities are often associated with the rapid spread of infectious diseases, driven by the perceived risks of urban density and overcrowding. However, transmission risk can vary considerably within urbanized areas as a function of socio-spatial disparities and the adoption of mitigating behaviors across communities. Here we examine the effect of density on coronavirus disease 2019 (COVID-19) infection rates at the neighborhood scale, within and across US cities. We integrate high-spatial resolution measures of land use and residential population density, mobility, infection rates and social determinants of health to evaluate the impact of neighborhood context on infection risk, while controlling for the potential mitigating effects of social distancing behavior. We are particularly focused on disparities among marginalized and vulnerable neighborhoods, and the generalizability of the results across political, socioeconomic, regional and built environment contexts. Our findings demonstrate a nonlinear relationship between urban density and infection rates, with higher-density neighborhoods more likely to adopt mitigating behaviors to reduce transmission. However, low-income and minority communities, facing cascading health challenges, are found to be least able to modify mobility behavior and therefore experienced a disproportionate burden of COVID-19 infection risk during the first wave of the pandemic.

Urban density is often considered a contributing factor in the transmission of infectious diseases. Although urban agglomerations have been connected to the risk of disease throughout history, the rapid population growth in cities during the industrialization period of the nineteenth century, and the rise of epidemiological methods, more directly linked cities to disease spread[1]. As increasing demand outpaced infrastructure capacity, the resultant overcrowded housing, congestion and insufficient sanitary and public health services led to widespread disease outbreaks, particularly among disadvantaged populations. Since higher urban densities are associated with a greater probability of close contacts, it follows that the rate of infectious disease transmission should increase with population density, represented mathematically by a scaling function[2,3]. In response, residents living in higher-density communities may acknowledge this risk, and modify their behavior to reduce the potential for exposure to infected individuals[4].

Emerging technologies, such as food home-delivery apps, and changing labor market structures, which enable new opportunities for remote work for some occupations, contribute to the viability of social distancing behaviors for certain groups[5,6]. Studies on density and disease spread have been constrained by the confounding effects of the positive correlation between population density and contact probability, and the potential mitigating influence of behavior change among urban residents[7,8]. The ability of a particular community to shift mobility behavior through social distancing, however, is mediated by racial and ethnic disparities, social and cultural norms, political ideologies, education, occupation and income, among other household and neighborhood attributes[9–11]. Beyond this polarity, the influence of density on local health risk is further shaped by a range of social determinants, including access to health care and housing conditions[12]. Thus, transmission rates can vary considerably across different communities in the same city[13].

[1]Marron Institute of Urban Management, New York University, New York, NY, USA. [2]Center for Urban Science and Progress, New York University, Brooklyn, NY, USA. [3]Tandon School of Engineering, New York University, Brooklyn, NY, USA. ✉e-mail: ckontokosta@nyu.edu

Taken together, it is important to consider the role of urban population density in disease transmission at the scale of the neighborhood, where variations in built environment and socioeconomic characteristics can be observed and analyzed[14–16]. During what is considered the first wave of the coronavirus disease 2019 (COVID-19) pandemic in the United States, many cities and states issued 'stay-at-home' or 'shelter-in-place' orders to encourage, or mandate, social distancing and the reduction of close contacts outside of the home[17]. This behavioral intervention was intended to reduce exposure density and thus minimize the spread of the disease. The response to these orders, however, was not uniform[9,18]. Such interventions have been found to have had a range of negative effects on individuals and communities, from mental health and emotional well-being, to educational outcomes and financial burdens, reinforcing the need to understand the relationship between stay-at-home orders, neighborhood context and transmission risk[19–21]. Measuring and understanding contact heterogeneity resulting from social distancing practices within neighborhoods is also a necessary step in evaluating the effect of urban density on infection rates, as well as the role social and economic disparities play in increasing risk for vulnerable communities.

There has been growing interest in social distancing behavior during the COVID-19 pandemic, as measured using various data sources, such as aggregated mobility data derived from smartphone geolocation information and geotagged social media activity[22–24]. However, the value of analyzing social distancing independent of the built environment context, and vice versa, is limited. For instance, a household in a rural community sheltering-in-place would reduce their potential risk of exposure to an infected individual less than a similar household in an urban environment simply as a function of the lower likelihood of close contacts in a less-dense neighborhood. Thus, we must consider both the baseline risk of transmission in a particular place, and the impact of social distancing behavior on an adjusted risk basis.

Several recent studies have specifically examined the effect of urban density on the transmission of COVID-19 in different global contexts[25–31]. The results from these papers show inconsistent results on the link between population density and COVID-19 transmission rates[31,32]. Most are limited by the geographic scale of the analysis, the granularity of COVID-19 outcome data and the extent to which social distancing behavior is appropriately represented[27,30,33–38]. First, density and disease transmission are mediated by the likelihood of close contacts in urbanized areas. Effectively accounting for mobility behaviors when examining differential socio-spatial impacts of density creates nontrivial data and computational challenges[39]. Second, many studies analyze the effects of density at the county or regional level, which can obscure localized variations in urban density, particularly at the neighborhood scale[36,40], or focus on a single city, which can constrain the generalizability of findings[30]. Finally, it is important to account for local health policy context during the study period, particularly as it relates to stay-at-home orders. Many cities and states adopted these orders in the early stages of the pandemic to limit disease transmission. However, the timing of these orders, and the extent of the associated mandates, varied considerably across the United States, given local and regional differences in disease prevalence and political ideology[41,42].

In this Article, we analyze the effects of urban residential population density on COVID-19 infection rates at the neighborhood (census tract) level, while controlling for changes in mobility behavior and other social determinants of health. We model both residential population density and exposure density across neighborhoods in 15 US cities to identify and evaluate how racial and income disparities influence the effects of density and the adoption of mitigating behaviors in response to COVID-19 transmission risk. We address three questions related to local public health policy and the impact of density and land use on infection rates during the first wave of the COVID-19 pandemic: (1) how did neighborhoods in different urban and socioeconomic contexts respond to social distancing policies during the early stage of the pandemic?; (2) what are the effects of land use and density on social distancing behaviors and COVID-19 infection rates? and (3) how do racial and income disparities influence mitigating behaviors in higher-density neighborhoods? To examine heterogeneity in neighborhood effects and capture regional variations, we focus our study on a diversity of cities in the United States for the time periods before, during and after stay-at-home orders were issued in each respective city.

Our approach is as follows: we first employ an indicator of exposure density as a measure of social distancing behavior to quantify changes in neighborhood activity associated with stay-at-home orders, using mobility data provided by VenPath, Inc. This indicator captures changes in human mobility behavior by both the volume of activity in a given place and the nature of that activity across land-use types. Exposure density provides a measure of local adoption of social distancing behaviors accounting for both migration patterns and changes in typical routines. We then analyze the effects of density and land use on COVID-19 infection rates by neighborhood, controlling for measured changes in exposure density. We define density here as the residential population per unit area of residential land use in a given census tract. Finally, we identify neighborhood racial and income disparities in observed COVID-19 outcomes by analyzing socioeconomic and demographic correlates of infection rates across neighborhoods with similar densities and land-use typologies. Throughout our analysis, we account for appropriate covariates that influence health outcomes. Our findings provide insight into the role of urban density on disease transmission and the effect of mitigating behaviors on neighborhood health risk, supporting targeted and equity-driven policy decisions.

Results

Neighborhood disparities in social distancing behavior

We calculate the average exposure density (as defined in Methods) for each census tract in the 15 cities (Supplementary Table 1) included in the study before the COVID-19 pandemic period (prepandemic or 'typical' period, covering the 2 weeks from 16 February to 29 February 2020), in the 2-week period after the stay-at-home order was issued for each respective city, and the 2-week period following the lifting of the stay-at-home order, also known as the 'phase 1' reopening (Supplementary Table 1). The neighborhood exposure density for each city is presented in Supplementary Fig. 2.

Figure 1 (selected cities) and Supplementary Fig. 3 (all cities) visualize the spatial patterns of neighborhood exposure density change over the study period (for additional details, see Supplementary Information). In a majority of cities, the overall activity volume in downtown areas decreased after the stay-at-home order compared with pre-COVID-19 levels, while activities in the peripheral areas remained relatively constant or increased, a result of a shift to more localized activity around residential areas. The nature of activity changed over this time period, with a substantial increase in the proportion of activity in residential land uses and commensurate decreases in nonresidential (office, retail, school and so on) and outdoor (park, sidewalks and so on) areas. Varying levels of social distancing behavior adoption across the country are observed. In Miami, Florida, for example, activity proportions by land use type are found to be relatively unchanged before, during and after the state-wide stay-at-home order, indicating a relatively stable mobility pattern. On the other hand, nonresidential and outdoor activities in New York City decreased approximately 20% after the stay-at-home order when compared with prepandemic levels. Independent of the total volume of activity in New York City, and the migration of population out of the urban core, those residents that remained changed their mobility behavior to avoid activities outside of the home.

Figure 2 shows the neighborhood distribution of exposure density change across the studied cities. Neighborhoods in New York City, on average, experienced a decrease in exposure density of approximately 49%, and almost all neighborhoods adopted social distancing behaviors
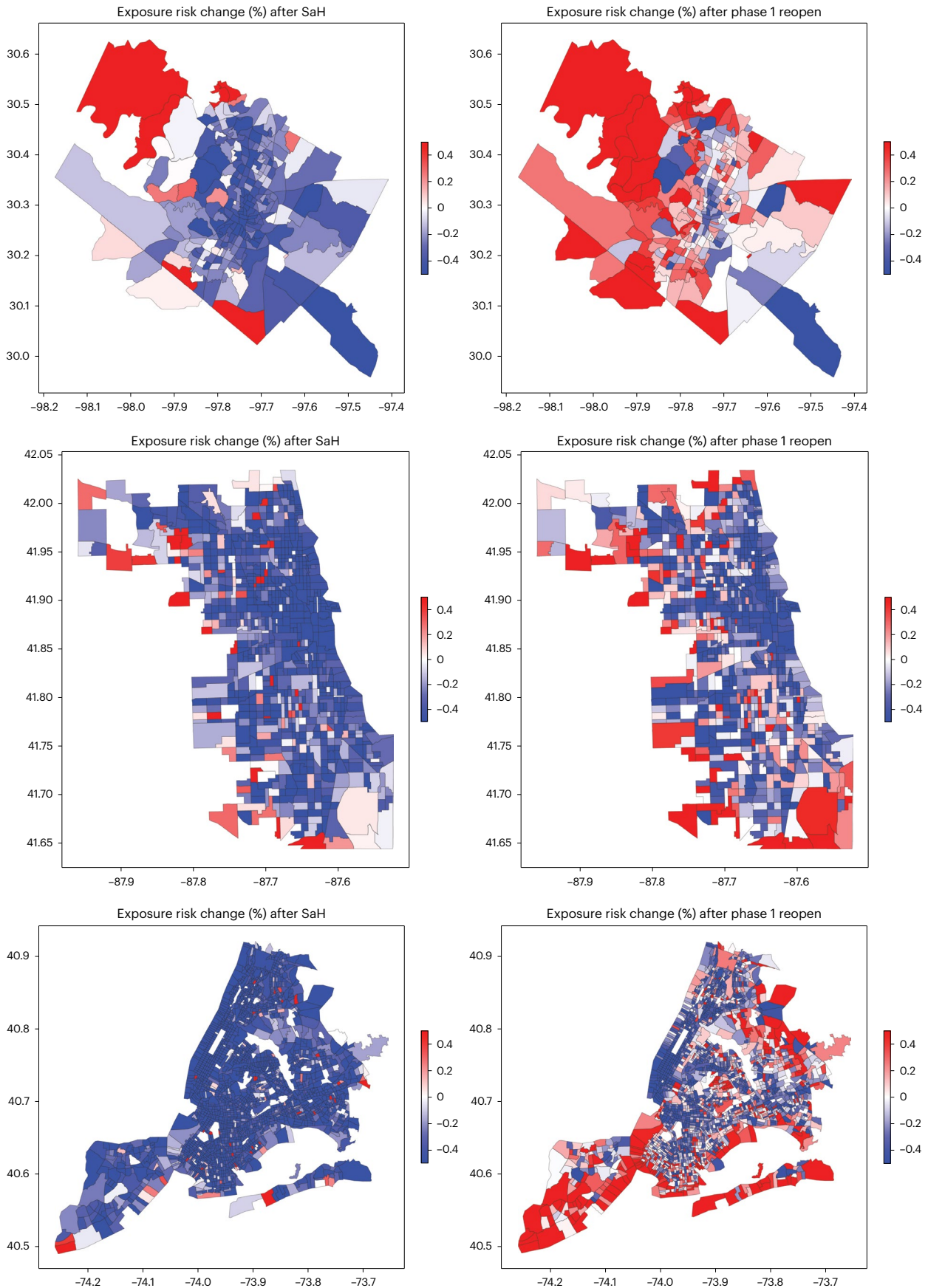
Fig. 1 | Census tract-level exposure density changes between the prepandemic period and the period after the stay-at-home order (left), and between the stay-at-home period and the period after phase 1 reopening (right). Austin, TX (top), Chicago, IL (middle) and New York City, NY (bottom) are shown here as examples. Maps for all cities are found in Supplementary Fig. 3. The minimum and maximum values of −0.4 and 0.4 are threshold levels for the lowest and highest bins, respectively, for visual representation.
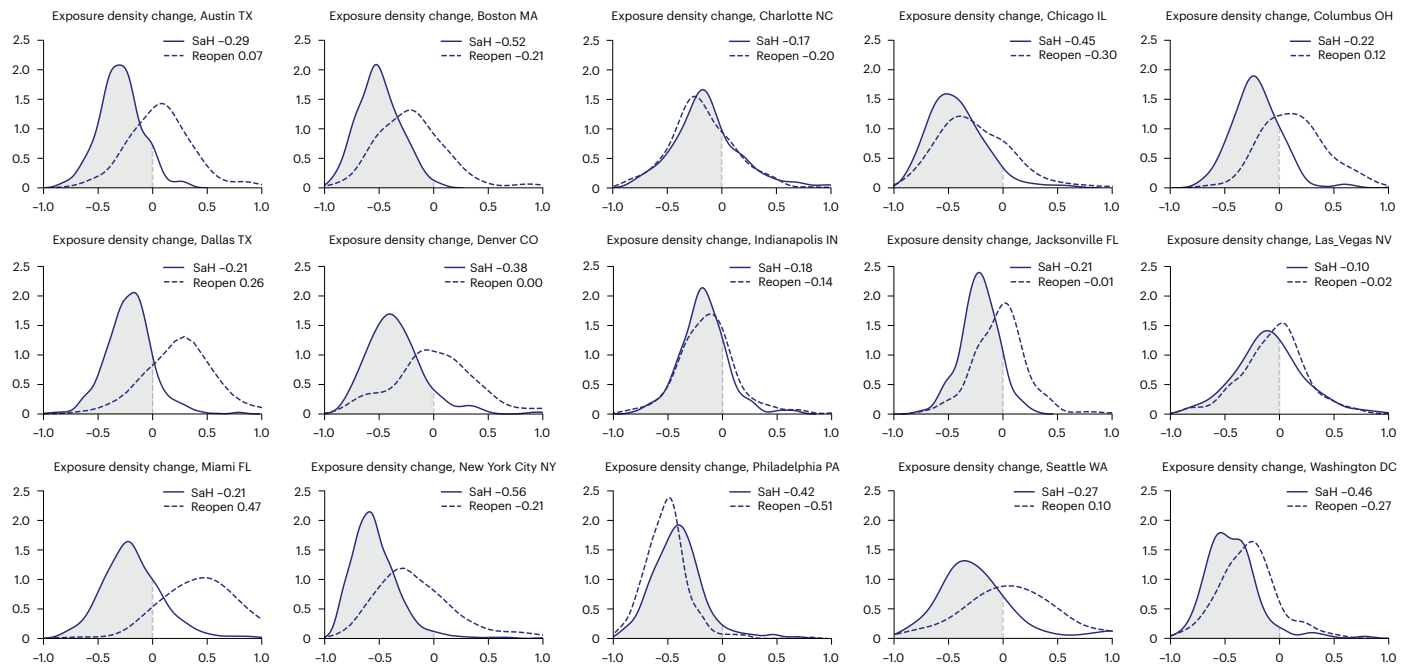
Fig. 2 | Neighborhood distribution of exposure density change across the studied cities. The solid lines represent exposure density change after the stay-at-home order and dotted curves show the change after the lifting of the order.

after the stay-at-home order (represented as the gray-colored area under the curve). In the case of neighborhoods in Las Vegas, Nevada, on the other hand, average exposure density decreased by only 6%, and approximately half of all neighborhoods did not adjust their typical mobility behaviors. Overall, neighborhoods in east coast cities, including Boston, New York City, Philadelphia and Washington DC, exhibit larger decreases in exposure density when compared with the other cities in the study group, as represented by positively skewed exposure density distributions. The distributions of the neighborhood response to the lifting of the stay-at-home order are shown as dotted curves. For example, the exposure density curve for New York City after its phase 1 reopening demonstrates a clear positive shift from the stay-at-home period. The mean value for the exposure density change curve indicates a return to near-normal levels, on average, of neighborhood activity. As the figures represent, exposure density change varied considerably across cities and within different neighborhoods in each city. Despite similar social distancing policies, there are clear disparities in community social distancing behaviors.

Effects of density on social distancing and infection rates

After observing community differences in mobility behaviors during the first wave of the COVID-19 pandemic, we focus on the effect of neighborhood residential population density on mitigating behaviors and COVID-19 infection rates. Specifically, we hypothesize that neighborhoods with similar density and land-use characteristics will have similar social distancing behavioral responses and comparable infection rates, after controlling for other social determinants of health. Residents in higher-density neighborhoods may perceive a higher risk of infection due to the greater probability of close contacts with potentially infected individuals in their everyday lives. As a risk-mitigation measure, this awareness could lead to mobility behavior changes to reduce the likelihood of interaction with others outside of the household or family unit. Conversely, residents in lower-density neighborhoods may not perceive an increased risk of transmission in their communities, where the number of random social interactions may be relatively limited. This could result, then, in more modest mobility behavior changes in response to social distancing restrictions. Other factors, such as

political ideology, demographic and socioeconomic characteristics, and risk tolerance are also considered.

To test this hypothesis, we first group neighborhoods in all 15 cities based on density and land-use characteristics using k-means clustering, an unsupervised machine learning method. The algorithm is applied to a total of 6,216 census tract neighborhoods after removing outliers based on population size (for descriptions of the methodology and land use data, see Methods). The clustering output identifies five distinct neighborhood groups with similar built environment characteristics across the cities in the study group. Neighborhoods assigned to the same cluster group, although they may not be in the same city, share similar land use profiles and population densities. Figure 3 visualizes the spatial pattern of the clusters for selected cities (for all cities, see Supplementary Information) and Supplementary Table 2 presents descriptive statistics for the built environment features for each cluster group.

The five neighborhood groups have distinct patterns of exposure density change, as shown in Supplementary Table 2. Group 1 (defined as 'ex-urban residential' given the relatively high proportions of residential and open space land uses in these neighborhoods) and group 2 ('institutional and industrial') neighborhoods decreased overall activity 26% and 27%, respectively, which represents the lowest rate of exposure density change across the five groups. Residents in these neighborhood groups may be more likely to maintain their behaviors and activity outside of the home because of a lower perceived risk of close contacts. 'Low-density residential' (group 3, characterized by one- and two-family housing) and 'mixed-use' (group 4) neighborhoods show similar exposure density change rates, decreasing overall by 34% and 36%, respectively, with a noted shift away from activities in nonresidential and outdoor areas. Activity levels in 'high-density' neighborhoods (group 5), which includes urban core neighborhoods, dropped dramatically, approximately 61% when compared with baseline (prepandemic) levels. Residents in this cluster group appear to adjust their normal mobility behaviors substantially in response to social distancing mandates. Additionally, neighborhoods in group 5 see many visitors for various purposes (for example, commuting, tourism, education, work and so on), which may influence risk perception and
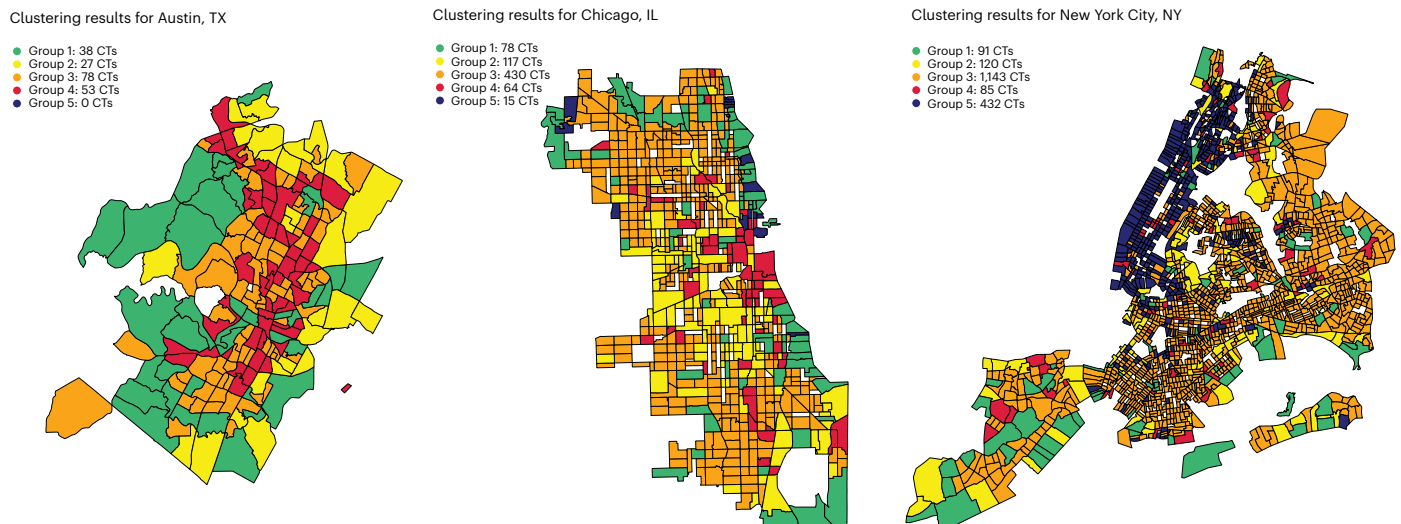
Fig. 3 | The results of the neighborhood clustering algorithm based on density and land use, with Austin, TX (left), Chicago, IL (middle) and New York City, NY (right) as examples. The number of census tract neighborhoods assigned to each clustered group are shown in the legend. The results for all cities can be found in Supplementary Fig. 4. CT, census tract.

thus behavior change. Furthermore, there were considerable decreases in activity volume in group 5 neighborhoods resulting from population out-migration, thus accounting for the large decrease in exposure density. The relationship between neighborhood density and social distancing behavior is more clearly demonstrated in the scatter plots of population density and exposure density change for each neighborhood group. As Fig. 4 shows, there is a statistically significant negative relationship between residential population density and exposure density change within each neighborhood group. This suggests that neighborhoods with higher residential population density reduced their exposure density more than lower-density neighborhoods.

The magnitude of the density effect on infection rates can be estimated using a spatial contact model[2,43,44]. Airborne transmission of respiratory viruses generally follows a density-dependent curve within a relatively lower-density setting that saturates in a higher-density setting based on a frequency-dependent constraint (Fig. 5a and ref. 2). From empirical data as shown in Fig. 5b, however, we observe a quadratic function as the best-fit relationship between residential population density and neighborhood infection rates (here, we use cumulative COVID-19 case rates through the second week of August 2020). This suggests that within higher-density communities, there are mitigating behaviors that reduce transmission rates from what would be expected based on density alone.

We assume that mitigation effects begin at the inflection point of the curve, and identify communities with densities greater than the x value of the inflection point and case rates lower than its associated y value (the lower-right shaded quadrant). Neighborhoods in this quadrant are assumed to exhibit mitigating behaviors (for example, mobility changes to reduce exposure density) to lower case rates below what would be expected on the basis of residential density alone. The magnitude of the mitigation effect can be measured by the relative vertical distance between the inflection point in the curve and a given neighborhood's case rate (for a detailed description, see Methods). Supplementary Table 3 represents the summary of mitigation effects for each clustered group (city-specific mitigation effects are summarized in Supplementary Table 4 and Supplementary Fig. 5).

To understand differences between higher-density neighborhoods with and without mitigating behaviors, we examine exposure density change, total residential population change, activity proportion change and socioeconomic characteristics of neighborhoods above and below the estimated inflection point of case rates for each cluster. We specifically examine the role of racial composition and income on disparities in infection rates and exposure density change, controlling for neighborhood density and land-use characteristics. We analyze statistically significant differences within and across cluster groups using t-tests and a logistic regression model (for detailed descriptions of the data and methodology, see Methods).

The t-test results are presented in Table 1, and the results of the logistic regression model are presented in Supplementary Table 5. Race and ethnicity are found to be significant factors in mitigating behavior for higher-density neighborhoods. For example, the percentage of non-Hispanic White population is positively associated with mitigating behavior (making up 41.65% of the population of neighborhoods with mitigating behavior versus 15.94% of those without, odds ratio 1.789), indicating that neighborhoods with larger proportions of non-Hispanic White population reduced exposure density after controlling for urban density. On the other hand, the proportion of Black population is found to have mixed outcomes depending on neighborhood group. The effect for Asian and Hispanic populations is found to be consistent across the different neighborhood clusters. Neighborhoods with a greater reduction in exposure density are more likely to have a higher proportion of Asian households and fewer Hispanic households. Overall, more racially diverse neighborhoods are less likely to change their mobility behavior, particularly in higher-density communities.

Educational attainment and job occupation are also found to be important determinants of social distancing and exposure density change. Neighborhoods where social distancing behavior adoption is greatest tend to have populations with higher educational attainment (based on the percentage of the population with a bachelor's degree or higher) and a larger percentage of employees working in finance, real estate, professional, scientific and management occupations. On the other hand, neighborhoods without mitigating behaviors tend to have a higher proportion of employees working in manufacturing, wholesale, retail, transportation and health care, occupations more likely to be considered essential or less conducive to work-from-home arrangements[45]. This finding is observed in the results of both the t-tests and the logistic model, where the odds ratio for essential workers is 0.519.

Income, health insurance coverage and neighborhood health characteristics are found to be among the influencing factors in the behavioral responses to social distancing guidelines, after controlling for land use and density. Neighborhoods with higher median household incomes (measured as the normalized difference from the city
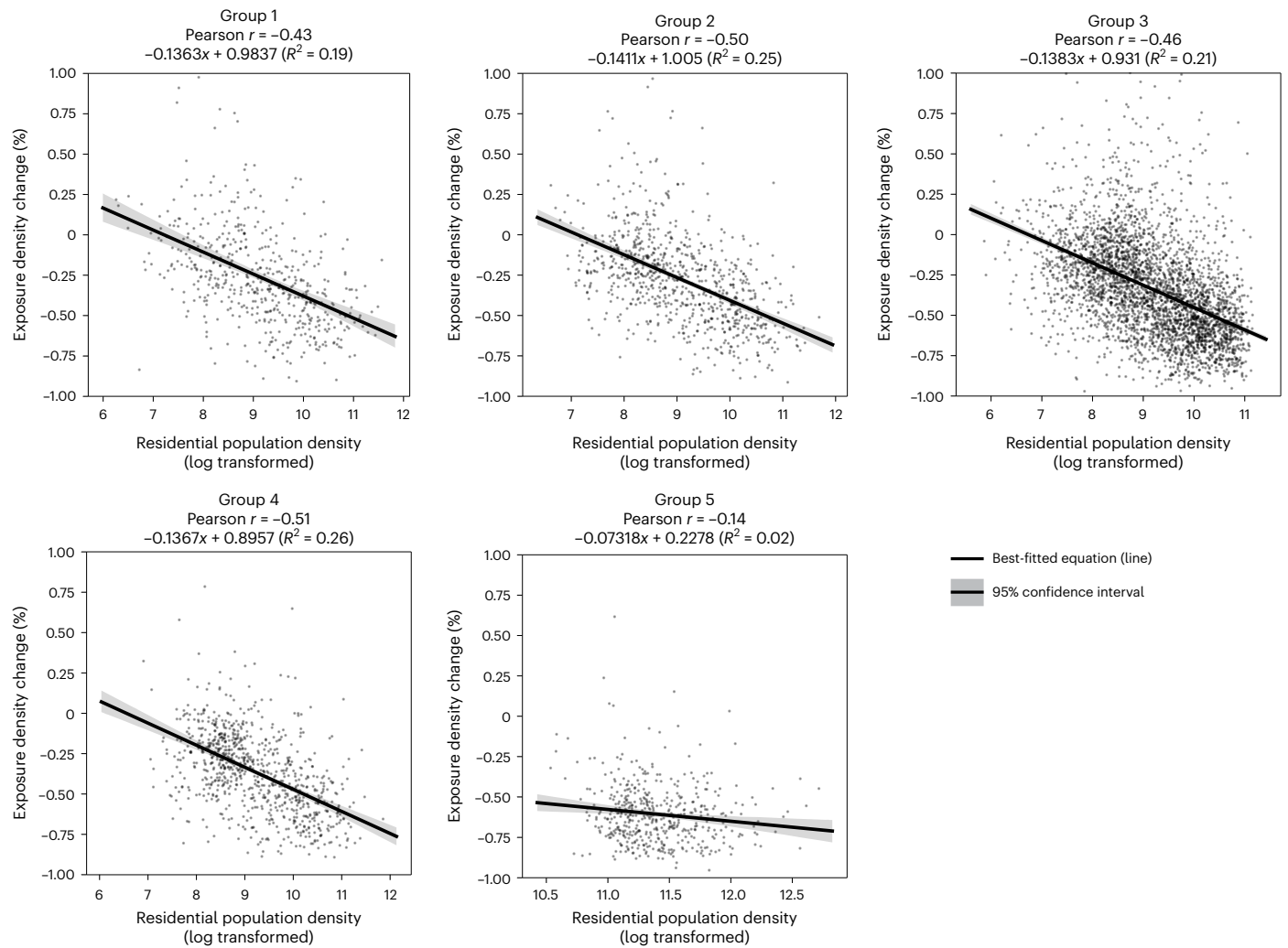
Fig. 4 | Scatter plots of residential population density and exposure density change for each cluster group. The linear best fit is shown.
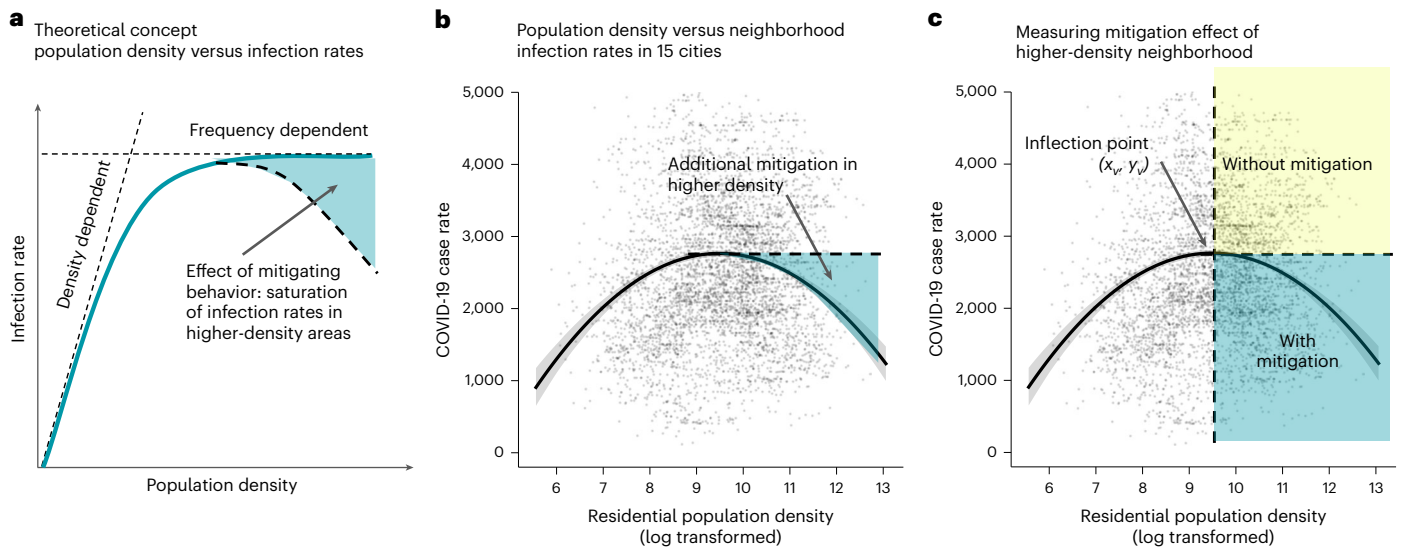
Fig. 5 | Relationship between residential population density and COVID-19 infection rate. a, Illustration of the conceptual relationship between population density and infection rate. **b**, A scatter plot of neighborhood residential population density and infection rates for the first wave of the COVID-19 pandemic. The generalized best fit curve ($y = -120.2x^2 + 2{,}282x - 8{,}062$) results in a higher R^2 (0.027) than a linear relationship ($R^2 = 0.00026$). **c**, A conceptual diagram for identifying neighborhoods with and without mitigating behaviors.

Table 1 | Demographic and socioeconomic characteristics of higher-density neighborhoods with and without mitigating behavior. Statistically significant features shown

Feature	Neighborhoods without mitigation	Neighborhoods with mitigation	t-Statistics (P value)
Non-Hispanic white population (%)	15.94%	41.65%	−26.06 (0.000)
Black population (%)	32.25%	24.96%	6.06 (0.000)
Asian population (%)	10.09%	12.73%	4.52 (0.000)
Hispanic population (%)	41.57%	19.00%	25.16 (0.000)
Racial diversity	0.55	0.47	9.00 (0.000)
Median income (z-score normalized within city)	−0.43	0.14	−15.24 (0.000)
Gini coefficient	0.45	0.48	−11.25 (0.000)
Population with Bachelor's degree (%)	15.52%	25.73%	−24.43 (0.000)
Employees working from home (%)	3.32%	5.15%	−13.81 (0.000)
Unemployment rate (%)	4.76%	4.19%	4.97 (0.000)
Essential workers (%)	25.36%	14.62%	29.00 (0.000)
Professional workers (%)	28.44%	48.27%	−29.21 (0.000)
Households without health insurance (%)	10.53%	7.06%	14.92 (0.000)
Voters who voted for Joe Biden in 2020 (%)	52.80%	61.28%	−6.72 (0.000)
Housing units with overcrowding (%)	12.26%	6.99%	17.19 (0.000)
Population with underlying health conditions (%)	14.04%	11.56%	17.16 (0.000)

mean) reduced exposure density more than relatively lower-income neighborhoods, resulting in a larger mitigation effect in high-density areas. The results for households without health insurance and residents with underlying health conditions suggest that socioeconomically disadvantaged and vulnerable neighborhoods are less likely to adopt social distancing behaviors, despite higher health risk factors, when compared with other neighborhoods with similar built environments. This is a result, in part, of the higher likelihood of households to be employed in essential occupations and/or those that cannot be done remotely. Finally, political ideology is shown to be a significant determinant of social distancing compliance across all clustered neighborhood groups. In areas with similar density and land use types, neighborhoods in Republican governed states are less likely to engage in social distancing behaviors. Additionally, neighborhoods with higher proportions of residents who voted for Democrat Joe Biden in the 2020 Presidential election are more likely to engage in social distancing behavior compared with communities with more citizens who voted for Republican Donald J. Trump.

Discussion

Our study provides evidence for the effect of urban density on COVID-19 infection rates at the scale of the neighborhood, while accounting for changes in mobility behavior and social distancing within communities. We note that our study has several limitations, including the effects of city and neighborhood testing differentials on reported COVID-19 case data, the absence of local COVID-19 case data for three cities in the study group, and the unobserved impacts of physical distancing and the use of masks. These constraints are described in detail in Methods.

We observe two critical findings. First, the risk of COVID-19 infection is, in general, positively correlated with density; however, as density increases, the likelihood of the adoption of mitigating behaviors increases, which counteracts the potential transmission risk associated with higher-density neighborhoods. The effect of neighborhood density on COVID-19 infection rates, therefore, is found to be nonlinear, influenced by socio-spatial disparities in income, racial and ethnic composition, and political ideology. Case rates increase with density to a point, measured here to be approximately 13,260 residents per square mile of residential land area, at which case rates begin to decrease. After controlling for neighborhood density and land-use characteristics, we observe that this decrease is driven, in part, by the adoption of mitigating behaviors in the form of social distancing and changes in mobility. Residents in relatively higher-density communities change their mobility behavior most, perhaps in response to the perception of higher transmission risk, and this results in a mitigating effect on infection rates below what might be expected from density factors alone. While infection rates tend to decrease after a specific population density, there is significant variation after accounting for neighborhood socioeconomic and demographic characteristics.

If social distancing behavioral responses stem solely from disparities in risk perception based on residential population density, we would expect that most—if not all—higher-density neighborhoods should exhibit a mitigation effect, either from behavior change (shifting activities from nonresidential to residential areas) or a decrease in population (out migration from higher risk areas). To the contrary, we observe that there are higher-density communities without mitigating behaviors, such as in the South Bronx and Flatbush in New York City, Dorchester in Boston and East Garfield Park in Chicago (as illustrated by communities above the horizontal line from the inflection point of the density-infection rate curve and spatially visualized in Supplementary Fig. 5). This suggests that while density influences social distancing behavior, additional factors contribute to exposure density change and infection rates more than density alone.

Second, we find that only certain groups are able to change their mobility behavior sufficiently to offset the increased risk due to density. The observed differences in mitigating behavior are directly correlated with the racial and economic composition of the neighborhood. We find that, after controlling for neighborhood population density and land-use mix, communities without mitigating mobility behaviors tend to have a greater proportion of racial and ethnic minorities, lower median incomes and more critical vulnerabilities, including overcrowded housing, lack of health insurance and higher likelihood of underlying health conditions. Furthermore, the proportion of essential workers in nonmitigating neighborhoods is almost double that of neighborhoods with mitigating behaviors. This indicates that neighborhoods facing the greatest risk from COVID-19 infection are least able to change their mobility behavior in response to social distancing mandates, putting these neighborhoods at continued risk for COVID-19 transmission despite stay-at-home orders. These findings suggest a paradox for social distancing mandates—while they are intended to protect the most vulnerable, these policies may have a disparate impact on at-risk communities that are unable, for a wide range of reasons, to meaningfully change mobility behaviors. As such, while stay-at-home guidelines and the resultant changes in behavior reduce the risk of infection for some, the ancillary negative effects on health and well-being from sheltering in place may have created additional burdens for vulnerable communities, without a considerable reduction in transmission risk.

Taken together, while infection rates increase with density, behavior change in higher-density neighborhoods appears to outweigh the intrinsic risk associated with densely populated communities. However, low-income and minority communities, facing cascading health challenges, are found to be least able to modify their mobility behavior and therefore experienced a disproportionate burden of COVID-19 infection risk during the first wave of the pandemic. The implications

of this for equitable health policy interventions are substantial. If the response to stay-at-home orders is not uniform and vulnerable communities are least able to adopt social distancing behaviors, then not only do these communities face greater risk of infection due to a higher probability of exposure, they also bear the burden of the negative impacts associated with stay-at-home mandates.

Methods

Data

This study relies on several data sources. A descriptive summary is presented in Supplementary Tables 6–8. The primary data used for calculating the exposure density index are anonymized smartphone geolocation pings obtained from VenPath, a data-marketplace company providing mobile-application data based on more than 200 smartphone applications across the United States. The dataset contains approximately 250 billion geotagged data points associated with 4.5 million unique devices hourly, and covers the period from 1 February to 13 July 2020. Due to the sensitivity of the data, raw data were collected from the dedicated Amazon Web Services S3 cloud storage service using the Amazon Web Services command line interface and processed within the secured and access-controlled environment of the New York University (NYU) High Performance Computing facility. To further ensure anonymity and data privacy, after classifying mobile activity data into one of the representative types using land-use information (as described further in Methods), geolocation data were aggregated to a 250-meter grid and further averaged at the neighborhood level (operationalized here as the census tract, but other areal units could be used with the method, as shown in Supplementary Fig. 1). The data processing and data management plan were approved by NYU's institutional review board (approval no. IRB-FY2018-1645).

For the purpose of assigning an activity type to the mobility geolocation data, we used a range of city-specific land-use data combined with building footprint data, administrative boundaries and road networks. These data were sourced from official administrative records obtained through publicly available open data platforms. Local road network data were obtained for each city from the OpenStreetMap platform using QGIS software with dedicated plugins. Each data layer was converted into the Geographic Coordinate System corresponding with the mobility data and clipped to the local municipality (city) boundary extent. A full list of ancillary data used for this rasterization process and analysis is provided in Supplementary Table 7.

To examine neighborhood-level COVID-19 infection rates, we collected available data on confirmed COVID-19 cases for 12 cities from multiple data sources at the zip code level (localized COVID-19 data were not available for three of the initial sample of 15 cities). On the basis of the number of reported COVID-19 cases for each city, we used spatial interpolation to adjust to the census tract level. Specifically, the infection rate is defined as the number of confirmed cases per 10,000 residential population. COVID-19 case data sources are presented in Supplementary Table 8. Additionally, we retrieved multiple ancillary datasets for neighborhood demographic, socioeconomic and political characteristics. We used the US Census Bureau 2019 5-year estimate American Community Survey database for census tract data on race, ethnicity, household type, housing, income and job occupation attributes[46]. Voting patterns and political affiliation were extracted from the standardized precinct 2020 presidential data, developed and shared by the *New York Times*[47]. We also obtained census tract data on neighborhood health conditions, specifically chronic disease and health indicators provided by the Centers for Disease Control and Prevention[48]. We used these data to estimate the percentage of neighborhood residents with underlying medical conditions. These datasets are publicly available and acquired from open data platforms and government databases.

Land-use rasterization and mobility behavior classification

We considered activity outside of residential land uses to be associated with a greater risk of infection. While within-household transmission rates have been found to be a important component of COVID-19 disease spread[49], our study focuses on the role of urban density in infection risk and, therefore, emphasizes the likelihood of close contacts with those outside the family or household unit.

The rasterization and mobility activity classification methodology is illustrated in Supplementary Fig. 6. To differentiate between residential and nonresidential activities, we first created a 1-meter resolution rasterized basemap with specific land-use type labels for each city. The rasterized basemap integrates city and administrative boundaries, street and sidewalk networks, land-use classifications and building footprints derived from various data sources (as presented in Supplementary Table 7). For each city, the extent of the study area was defined using the administrative boundaries of the municipality, converted into the projected coordinate system (EPSG:3857) and rasterized at a 1-square-meter resolution. Similarly, street network, land use and building footprint data were also reprojected, rasterized, encoded and aligned with the city boundaries raster. After processing, these rasters were then stacked such that each pixel represents a specific land-use/land-cover class. Land cover information is categorized into multiple integer values to represent different standardized land-use types (such as 10 for residential properties, 30 for commercial buildings, and so on; for the full list of classes used, see Supplementary Table 9). In addition to the associated land-use classification label, each pixel is identified by its positional index referenced to the bottom-left corner of the raster, as well as its geographical coordinates.

This information is used for assigning mobility geolocation data to a specific land-use grid cell. To do so, each mobility ping was spatially joined to a corresponding land-use raster pixel based on its location. Calculation of the changes in mobility behavior and exposure density are based on the activity levels aggregated to a 250-meter × 250-meter grid and normalized by the land-use distribution within each grid cell. To estimate those changes, we count the hourly number of unique devices within each land-use category based on the matched raster pixel in a grid cell. This process helps to further ensure data privacy and account for uncertainty in the device geolocation accuracy. Land-use assignment is used as a proxy for the nature of the activity occurring in the respective cell. It allows us to differentiate between activities occurring within four distinct groups, namely residential areas, nonresidential areas (for example, commercial and retail), outdoor areas (for example, park and playground) and on major roads associated predominantly with vehicular traffic (as presented in Supplementary Table 9). Activities occurring on major road networks are excluded from the analysis. To align with the spatial resolution of the ancillary datasets, the resultant activity and exposure density metrics are aggregated and reported at the census tract level.

Initial data processing of each land-use layer was performed using QGIS software followed by the rasterization process using the Python environment within NYU's Center for Urban Science and Progress Research Computing Facility (CUSP RCF). Mobile activity mapping was executed inside NYU High Performance Computing's Hadoop cluster using the PySpark environment.

Measuring social distancing behavior using exposure density

To measure neighborhood behavioral responses to social distancing across multiple cities, we updated and expanded our exposure density metric developed previously[9]. This accounts for the land-use rasterization process for each city. Exposure density is based on the probability of contact (contact rate) with others outside of residential land uses. It measures the likelihood of random contact with a potentially infected individual given a unit area, land-use category and defined time interval. This can be expressed as the neighborhood population, including both infectious and susceptible subgroups and

accounting for population changes over the study period, multiplied by the proportion of activity in nonresidential and outdoor land-use settings, normalized by the total area of nonresidential and outdoor land uses within a specified geographic boundary. Our exposure density metric accounts for not only the volume of people active in a given area, but also the extent of activity occurring outside of predominantly residential areas.

Here, we define exposure density ($E_{r(t)}$) as the population within a unit area during a given time period. ($E_{r(t)}$) is the number of unique devices in a selected geographic area multiplied by the proportion of activity occurring in nonresidential and outdoor land uses. This measure approximates a contact probability for mobility activities occurring outside of residential areas. Therefore, exposure density is specified as:

$$E_{r(t)} = \frac{N_{(t)} \times P_{\mathrm{NRO}(t)}}{A_{\mathrm{NRO}}}, \tag{1}$$

where ($E_{r(t)}$) is the exposure density for a given temporal unit, N_t is the total population in a given geographic area, P_{NRO} is proportion of activity occurring in nonresidential or outdoor land uses (as a proxy for contact probability) and A_{NRO} is the total area of nonresidential and outdoor land uses in the selected geographic unit. P_{NRO} takes into account not only how many people stay in a given area, but also how many people are active outside of their presumed place of residence. P_{NRO} is given by:

$$P_{\mathrm{NRO}(t)} = \frac{N_{\mathrm{NR}(t)} + N_{\mathrm{O}(t)}}{N_{(t)}}, \tag{2}$$

where $N_{\mathrm{NR}(t)}$ is the hourly number of unique devices in nonresidential land uses and $N_{\mathrm{O}(t)}$ is the hourly number of unique devices in outdoor land uses at the time t. If the population of a given neighborhood all remain in their presumed home locations, as measured by the mobility data, then the exposure density would be equal to zero.

We measure the change in exposure density and activity proportions in different land-use categories for each census tract before, during and after the respective stay-at-home order was implemented for each city. This is represented by:

$$E_{r\mathrm{change}(t_1)} = \frac{(E_{r(t_1)} - E_{r(t_0)})}{E_{r(t_0)}}, \tag{3}$$

$$P_{L\mathrm{change}(t_1)} = \frac{(P_{L(t_1)} - P_{L(t_0)})}{P_{L(t_0)}}, \tag{4}$$

where $E_{r(t_1)}$ is the 14-day average exposure density after the stay-at-home order, $E_{r(t_0)}$ is the 14-day average exposure density before the stay-at-home order (prepandemic period) and $P_{L(t_1)}$ is the 14-day average proportion of activities occurring in different land use types L (residential, nonresidential and outdoor) in a given temporal period. $E_{r\mathrm{change}(t_1)}$ and $P_{L\mathrm{change}(t_1)}$ provide a measure of social distancing behavior, accounting for the both the volume of activity and its nature.

Estimating the effects of neighborhood density

One hypothesis we consider is that if social distancing behavior is independent of neighborhood density, there should be no statistically significant differences in exposure density change between neighborhoods with varying built environment characteristics. Specifically, we test this hypothesis by clustering neighborhoods based on their density and land-use composition, and then examining whether exposure density changes vary across the resultant neighborhood groups.

To identify neighborhood clusters based on density and built environment characteristics, we apply a k-means clustering algorithm based on five input variables: residential population density, percentage of residential land cover, percentage of nonresidential land cover, percentage of industrial or institutional land cover and percentage of open space land cover. Residential population density is defined as the residential population divided by the area of residential land cover. The k-means algorithm is specified as:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} ||x_i^{(j)} - c_j||^2, \tag{5}$$

where J is the sum of squares of the distance of each data point to its assigned vector, x_i is each data point, c_j is a centroid for cluster j, k is the number of clusters and n is the number of data points. To find cluster centers representing certain regions of the data, k-means clustering conducts an iterative process. After optimization, it was determined that five ($k = 5$) clusters minimized the total sum of squares of distance between data points and the respective centroid of the cluster to which the data point belongs.

We applied this clustering analysis to all census tract neighborhoods (6,216 census tracts, excluding census tracts with residential land cover accounting for less than 5% of the total area), independent of city identification. The resultant clustered neighborhood groups were then integrated with exposure density measures to examine statistically significant differences in social distancing behavior adoption between groups by using a one-way analysis of variance (ANOVA) and a Tukey's test for post hoc analysis.

Mitigating behaviors in higher-density neighborhoods

We first plotted the relationship between residential population density and COVID-19 infection rates for each neighborhood, segmented by clustered group, as a descriptive analysis of neighborhood distributions (Fig. 5). To measure mitigation effects in higher-density neighborhoods, we identified a best-fit curve using a quadratic equation. The inflection point of each curve marks the point where infection rates break from the frequency-dependent constraint, as shown in Fig. 5. Neighborhoods with case rates below this threshold are considered to have lower than expected case rates, and thus demonstrate a mitigation effect. This is specified as:

$$N_i = \begin{cases} 1 & \text{if } D_i \geq x_v \text{ and } C_i \leq y_v \\ 0 & \text{if } D_i \geq x_v \text{ and } C_i > y_v \end{cases}, \tag{6}$$

where N_i is a binary variable equal to 1 if a neighborhood exhibits mitigating behavior, and 0 otherwise, D_i is the residential population density of neighborhood i, C_i is the COVID-19 case rate for neighborhood i and (x_v, y_v) is a vertex (inflection point) of the fitted curve (Fig. 5c). The mitigation effect for each neighborhood is calculated based on the vertical distance between the y value of the inflection point and the case rate for each neighborhood specified as:

$$M_i = \begin{cases} \frac{(y_v - C_i)}{y_v} & \text{if } D_i \geq x_v \text{ and } C_i \leq y_v \\ 0 & \text{if } D_i \geq x_v \text{ and } C_i > y_v \end{cases}, \tag{7}$$

where M_i is the estimated mitigation effect of neighborhood i, C_i is the COVID-19 case rate of neighborhood i and y_v is the y value of the inflection point of the fitted curve. M_i indicates the modeled COVID-19 case rate reduction resulting from mitigating behaviors.

We compared higher-density neighborhoods with mitigation ($D_i \geq x_v$ and $C_i \leq y_v$) and those without mitigation ($D_i \geq x_v$ and $C_i > y_v$) within and across clustered neighborhood groups (Fig. 5c). We analyzed neighborhood socioeconomic, demographic and political attributes, including race and ethnicity, age group, household characteristics, education and occupation, neighborhood health burdens and political affiliation. T-tests were used to identify statistically significant differences between neighborhoods.

In addition, a logistic regression model was used to identify statistically significant correlates of neighborhood characteristics and mitigating behaviors, controlling for the built environment context. Specifically, the dependent variable is a binary for mitigating behavior adoption (1: neighborhoods with mitigation, 0: neighborhoods without mitigation) and independent variables include demographic, socioeconomic and political features.

Limitations

There are several limitations to our analysis. First, we rely on case rate data at the neighborhood level for our study. It is acknowledged that testing rates varied across cities and communities during the first wave of the pandemic; however, testing data are not consistently available at the local level in the studied cities. In addition, as with all COVID-19 infection data, reported infections are based on residential location, not necessarily on the place where infection occurred. Second, our exposure density measure captures the relative density of mobile devices aggregated to the neighborhood scale. We cannot infer from these data the spatial proximity of devices to draw any conclusions regarding person-to-person physical distancing (for example, remaining 6 feet apart, as suggested by the Centers for Disease Control and Prevention guidelines). Finally, land-use and exposure density measures are collected and estimated for all 15 cities in the study group, whereas local COVID-19 infection data was only available for 12 of these cities. Thus, our study of the relationship between urban density and exposure density is based on the full sample, while the link between density and infection rates is estimated from the 12 cities with complete data are available. The study group is described in more detail in Supplementary Information.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data used to conduct the analyses described here are summarized in Supplementary Tables 6–8. Processed, aggregate data derived from publicly available, open data sources that support the findings of this study are available via the dedicated data repositories. The rasterized land use classification data for selected US cities is available in New York University's UltraViolet repository with the identifier of https://doi.org/10.58153/t7m6v-37090 (ref. 50). The dataset of zip code-level COVID-19 case rates used in this study is also available in the NYU's UltraViolet platform with the identifier of https://doi.org/10.58153/avp10-a8h86 (ref. 51). The primary mobility data that support the findings of this study are available from VenPath, Inc., but restrictions apply based on a data sharing agreement. The aggregated neighborhood-level exposure density metrics along with the corresponding code are available in the dedicated GitHub repository.

Code availability

Code and associated materials are available in a dedicated GitHub repository available at https://github.com/UrbanIntelligenceLab/socio-spatial-inequality-and-the-effects-of-density-on-covid19-transmission-in-us-cities under MIT License.

References

1. Leon, D. A. Cities, urbanization and health. *Int. J. Epidemiol.* **37**, 4–8 (2008).
2. Hu, H., Nigmatulina, K. & Eckhoff, P. The scaling of contact rates with population density for the infectious disease models. *Math. Biosci.* **244**, 125–134 (2013).
3. Sy, K. T. L., White, L. F. & Nichols, B. E. Population density and basic reproductive number of COVID-19 across United States counties. *PLoS ONE* **16**, e0249271 (2021).
4. McFarlane, C. Repopulating density: COVID-19 and the politics of urban value. *Urban Stud.* https://doi.org/10.1177/00420980211014810 (2021).
5. Benda, N. C., Veinot, T. C., Sieck, C. J. & Ancker, J. S. Broadband internet access is a social determinant of health! *Am. J. Public Health* **110**, 1123–1125 (2020).
6. Brynjolfsson, E. et al. COVID-19 and remote work: an early look at US data. *NBER* https://doi.org/10.3386/w27344 (2020).
7. Carozzi, F., Provenzano, S. & Roth, S. Urban density and COVID-19: understanding the US experience. *Ann. Region. Sci.* https://doi.org/10.1007/s00168-022-01193-z (2022).
8. Jo, Y., Hong, A. & Sung, H. Density or connectivity: what are the main causes of the spatial proliferation of COVID-19 in Korea? *Int. J. Environ. Res. Public Health* **18**, 5084 (2021).
9. Hong, B., Bonczak, B. J., Gupta, A., Thorpe, L. E. & Kontokosta, C. E. Exposure density and neighborhood disparities in COVID-19 infection risk. *Proc. Natl Acad. Sci. USA* **118**, e2021258118 (2021).
10. Muñoz-Price, L. S. et al. Racial disparities in incidence and outcomes among patients with COVID-19. *JAMA Netw. Open* **3**, e2021892 (2020).
11. Zhai, W., Fu, X., Liu, M. & Peng, Z.-R. The impact of ethnic segregation on neighbourhood-level social distancing in the United States amid the early outbreak of COVID-19. *Urban Stud.* https://doi.org/10.1177/00420980211050183 (2021).
12. Bambra, C., Riordan, R., Ford, J. & Matthews, F. The COVID-19 pandemic and health inequalities. *J. Epidemiol. Community Health* **74**, 964–968 (2020).
13. López-Gay, A. et al. Sociodemographic determinants of intraurban variations in COVID-19 incidence: the case of Barcelona. *J. Epidemiol. Community Health* **76**, 1–7 (2022).
14. Adhikari, S. 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).
15. Batty, M. Social distancing at scale. *Environ. Plan. B Urban Anal. City Sci.* **47**, 1533–1536 (2020).
16. Garnier, R., Benetka, J. R., Kraemer, J. & Bansal, S. Socioeconomic disparities in social distancing during the COVID-19 pandemic in the United States: observational study. *J. Med. Internet Res.* **23**, e24591 (2021).
17. Gostin, L. O. & Wiley, L. F. Governmental public health powers during the COVID-19 pandemic: stay-at-home orders, business closures, and travel restrictions. *JAMA* **323**, 2137–2138 (2020).
18. Gao, S., Rao, J., Kang, Y., Liang, Y. & Kruse, J. Mapping county-level mobility pattern changes in the United States in response to COVID-19. *SSRN* **12**, 16–26 (2020).
19. Berry, C. R., Fowler, A., Glazer, T., Handel-Meyer, S. & MacMillen, A. Evaluating the effects of shelter-in-place policies during the COVID-19 pandemic. *Proc. Natl Acad. Sci. USA* **118**, e2019706118 (2021).
20. Chen, C. Y.-C., Byrne, E. & Vélez, T. Impact of the 2020 pandemic of COVID-19 on families with school-aged children in the United States: roles of income level and race. *J. Fam. Issues* **43**, 719–740 (2022).
21. Tull, M. T. et al. Psychological outcomes associated with stay-at-home orders and the perceived impact of COVID-19 on daily life. *Psychiatry Res.* **289**, 113098 (2020).
22. Benita, F. Human mobility behavior in COVID-19: a systematic literature review and bibliometric analysis. *Sustain. Cities Soc.* **70**, 102916 (2021).
23. Huang, X., Li, Z., Jiang, Y., Li, X. & Porter, D. Twitter reveals human mobility dynamics during the COVID-19 pandemic. *PLoS ONE* **15**, e0241957 (2020).
24. Wellenius, G. A. et al. Impacts of social distancing policies on mobility and COVID-19 case growth in the US. *Nat. Commun.* **12**, 3118 (2021).

25. Bhadra, A., Mukherjee, A. & Sarkar, K. Impact of population density on COVID-19 infected and mortality rate in India. *Model. Earth Syst. Environ.* **7**, 623–629 (2021).

26. Lai, K. Y., Webster, C., Kumari, S. & Sarkar, C. The nature of cities and the COVID-19 pandemic. *Curr. Opin. Environ. Sustain.* **46**, 27–31 (2020).

27. Khavarian-Garmsir, A. R., Sharifi, A. & Moradpour, N. Are high-density districts more vulnerable to the COVID-19 pandemic? *Sustain. Cities Soc.* **70**, 102911 (2021).

28. Teller, J. Urban density and COVID-19: towards an adaptive approach. *Buildings Cities* **2**, 150–165 (2021).

29. Mouratidis, K. & Yiannakou, A. COVID-19 and urban planning: built environment, health, and well-being in Greek cities before and during the pandemic. *Cities* **121**, 103491 (2022).

30. Venerandi, A., Aiello, L. M. & Porta, S. Urban form and COVID-19 cases and deaths in Greater London: an urban morphometric approach. *Environ. Plan. B* **50**, 1228–1243 (2023).

31. Zhang, X. et al. Compact cities and the COVID-19 pandemic: systematic review of the associations between transmission of COVID-19 or other respiratory viruses and population density or other features of neighbourhood design. *Health Place* **76**, 102827 (2022).

32. Sharifi, A. & Khavarian-Garmsir, A. R. The COVID-19 pandemic: impacts on cities and major lessons for urban planning, design, and management. *Sci. Total Environ.* **749**, 142391 (2020).

33. Barak, N., Sommer, U. & Mualam, N. Urban attributes and the spread of COVID-19: the effects of density, compliance and socio-political factors in Israel. *Sci. Total Environ.* **793**, 148626 (2021).

34. Yin, H. et al. Association between population density and infection rate suggests the importance of social distancing and travel restriction in reducing the COVID-19 pandemic. *Environ. Sci. Pollut. Res.* **28**, 40424–40430 (2021).

35. Verma, R., Yabe, T. & Ukkusuri, S. V. Spatiotemporal contact density explains the disparity of COVID-19 spread in urban neighborhoods. *Sci. Rep.* **11**, 10952 (2021).

36. Martins-Filho, P. R. Relationship between population density and COVID-19 incidence and mortality estimates: a county-level analysis. *J. Infect. Public Health* **14**, 1087–1088 (2021).

37. Hong, A. & Chakrabarti, S. Compact living or policy inaction? Effects of urban density and lockdown on the COVID-19 outbreak in the US. *Urban Stud.* **60**, 1588–1609 (2023).

38. Zhou, L. et al. Spatial congruency or mismatch? Analyzing the COVID-19 potential infection risk and urban density as businesses reopen. *Cities* **123**, 103615 (2022).

39. Hamidi, S., Ewing, R. & Sabouri, S. Longitudinal analyses of the relationship between development density and the COVID-19 morbidity and mortality rates: early evidence from 1,165 metropolitan counties in the United States. *Health Place* **64**, 102378 (2020).

40. Hamidi, S., Sabouri, S. & Ewing, R. Does density aggravate the COVID-19 pandemic? *J. Am. Plan. Assoc.* **86**, 495–509 (2020).

41. Brodeur, A., Grigoryeva, I. & Kattan, L. Stay-at-home orders, social distancing, and trust. *J. Popul. Econ.* **34**, 1321–1354 (2021).

42. Dave, D., Friedson, A. I., Matsuzawa, K. & Sabia, J. J. When do shelter-in-place orders fight COVID-19 best? Policy heterogeneity across states and adoption time. *Econ. Inq.* **59**, 29–52 (2021).

43. Mollison, D. Spatial contact models for ecological and epidemic spread. *J. R. Stat. Soc. B* **39**, 283–313 (1977).

44. Riley, S. Large-scale spatial-transmission models of infectious disease. *Science* **316**, 1298–1301 (2007).

45. Blau, F. D., Koebe, J. & Meyerhofer, P. A. Who are the essential and frontline workers? *Bus. Econ.* **56**, 168–178 (2021).

46. 2019 5 year American community survey (ACS). *US Census Bureau* https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2019/5-year.html (2019).

47. Presidential precinct data for the 2020 general election. *GitHub* https://github.com/TheUpshot/presidential-precinct-map-2020 (2020).

48. PLACES: local data for better health, census tract data. *The Center for Disease Control and Prevention* https://chronicdata.cdc.gov/500-Cities-Places/PLACES-Local-Data-for-Better-Health-Census-Tract-D/cwsq-ngmh (2021).

49. Lewis, N. M. et al. Household transmission of severe acute respiratory syndrome coronavirus-2 in the United States. *Clin. Infect. Dis.* **73**, e1805–e1813 (2021).

50. Bonczak, B. J., Hong, B. & Kontokosta, C. E. Rasterized land use classification data for selected cities in the US. *New York University* https://doi.org/10.58153/t7m6v-37090 (2023).

51. Hong, B., Bonczak, B. J. & Kontokosta, C. E. Zipcode level COVID-19 case rate data for selected cities in the US. *New York University* https://doi.org/10.58153/avp10-a8h86 (2023).

Acknowledgements

Author contributions

C.E.K. conceived the study. C.E.K. obtained funding. B.H. and B.J.B. analyzed the data. C.E.K., B.H. and B.J.B. interpreted the results. C.E.K. and B.H. prepared the preliminary manuscript. All authors revised and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44284-023-00008-2.

Correspondence and requests for materials should be addressed to Constantine E. Kontokosta.

Peer review information *Nature Cities* thanks Mehdi Alidadi, Creighton Connolly and Laura F. White for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

nature portfolio

Corresponding author(s):   Constantine Kontokosta

Last updated by author(s):   Nov 10, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The mobility data (mobile phone geolocation data) used for this study were originally collected by VenPath, Inc. from several hundreds of smartphone applications covering more than 120 million devices every month across the U.S. The data were provided to the authors through a dedicated Amazon Web Services (AWS) S3 cloud storage service and transferred to the New York University's High Performance Computing (NYU HPC) Hadoop cluster Peel (formerly Dumbo) using AWS CLI (v1.22.82). Auxiliary datasets for the study were collected from open data sources (e.g. federal, state, or local governments) using Python programming language (v3.6 and 3.8), publicly available Application Programming Interfaces (APIs), QGIS Geographic Information System (GIS) software (v3.4), and manual data collection processes.
Data analysis	Analyses were conducted using Python version 3.8, Apache PySpark version 2.4, and QGIS version 3.4 Madeira. All of the code scripted for data collection and analysis for this study as well as the resulting models and visualizations will be available in the dedicated GitHub repository (https://github.com/UrbanIntelligenceLab/socio-spatial-inequality-and-the-effects-of-density-on-covid19-transmission-in-us-cities) before publication under the MIT License.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data used to conduct the analyses described here are summarized in Table S6, Table S7, and Table S8. Processed, aggregate data derived from publicly-available, open data sources that support the findings of this study are available via the dedicated data repositories. The rasterized land use classification data for selected US cities is available in New York University's UltraViolet repository with the identifier of https://doi.org/10.58153/t7m6v-37090. The dataset of zipcode-level COVID-19 case rates that was used in this study is also available in the NYU's UltraViolet platform with the identifier of https://doi.org/10.58153/avp10-a8h86. The primary mobility data that support the findings of this study are available from VenPath, Inc., but restrictions apply based on the data sharing agreement. The aggregated neighborhood level exposure density metrics along with the corresponding code are available in the dedicated GitHub repository.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences   ☒ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study involves quantitative research using mobile phone geolocation data and unsupervised machine learning to quantify and measure the impact of social distancing and urban density during and after the COVID-19 pandemic across 15 cities in the U.S.
Research sample	The primary data used for calculating the exposure density index are the anonymized smartphone geolocation pings obtained from VenPath, Inc. - a data-marketplace company providing mobile-application data based on more than 200 smartphone applications across the United States. The dataset of the total size of 4 TB contains approximately 100 billion data points associated with 4.5 million unique devices hourly.
Sampling strategy	The study extracts the subset of the raw data spanning the period from February 1st to July 13th of 2020 and use all devices active at least once within the geographical bounding box of the designated 15 cities. As such, there was no additional sampling strategy.
Data collection	The mobility data (mobile phone geolocation data) used for this study were originally collected by VenPath, Inc. from several hundreds of smartphone applications covering more than 120 million devices every month across the U.S. The data were provided to the authors through a dedicated Amazon Web Services (AWS) S3 cloud storage service and transferred to the New York University's High Performance Computing (NYU HPC) Hadoop cluster Peel (formerly Dumbo) using AWS CLI (v1.22.82). Auxiliary datasets for the study were collected from open data sources (e.g. federal, state, or local governments) using Python programming language (v3.6 and 3.8), publicly available Application Programming Interfaces (APIs), QGIS Geographic Information System (GIS) software (v3.4), and manual data collection processes.
Timing	The study uses data collected between February 1st, 2020 and July 13th, 2020.
Data exclusions	No data were excluded.

Non-participation	No participants
Randomization	This study is observational and no randomization was used.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging