



OPEN

Spatiotemporal contact density explains the disparity of COVID-19 spread in urban neighborhoods

Rajat Verma, Takahiro Yabe & Satish V. Ukkusuri[✉]

The rapid early spread of COVID-19 in the US was experienced very differently by different socioeconomic groups and business industries. In this study, we study aggregate mobility patterns of New York City and Chicago to identify the relationship between the amount of interpersonal contact between people in urban neighborhoods and the disparity in the growth of positive cases among these groups. We introduce an aggregate spatiotemporal *contact density index* (CDI) to measure the strength of this interpersonal contact using mobility data collected from mobile phones, and combine it with social distancing metrics to show its effect on positive case growth. With the help of structural equations modeling, we find that the effect of CDI on case growth was consistently positive and that it remained consistently higher in lower-income neighborhoods, suggesting a causal path of income on case growth via CDI. Using the CDI, schools and restaurants are identified as high contact density industries, and the estimation suggests that implementing specific mobility restrictions on these point-of-interest categories is most effective. This analysis can be useful in providing insights for government officials targeting specific population groups and businesses to reduce infection spread as reopening efforts continue to expand across the nation.

The rapid worldwide spread of the novel coronavirus disease (COVID-19) has caused significant distress to citizens and governments worldwide and has warranted unprecedented control measures on human mobility such as travel bans, mandatory quarantines, and stay-at-home/shelter-in-place orders. In the United States, like many other countries, these restrictive policies have incurred huge economic losses¹, leading to experts showing interest in targeted policy-making based on differential effects of the virus on different subject and activity types. These policies may be either supportive, such as the CARES Act relief based on household income², or restrictive, such as heightened restrictions on some business activities more than others, such as schools, bars, concerts, sports events, and indoor dining. Similarly, in some cases, specific neighborhoods within cities may be classified as “coronavirus hotspots” that are subjected to additional precautions and heightened restrictions, such as the New Rochelle community in the NYC metropolitan region in March 2020³.

It has been shown that these policies have had differential impacts on the livelihood and health of different socioeconomic groups within cities and states, including predominantly low-income⁴, lesser educated⁵, and non-White^{6,7} communities. Additionally, although age is a major such discriminant, it is inherently strongly correlated with higher susceptibility to infection by and morbidity due to the coronavirus. In considering targeted policies, such as targeting urban neighborhoods, it then becomes crucial to understand how these different social groups may respond to such policies in terms of mobility and the growth of the disease.

A crucial hindrance to this approach is a lack of high-fidelity publicly available epidemiological data related to the disease. State governments and popular COVID-19 trackers such as the Covid Tracking Project⁸ and the JHU Coronavirus Research Center⁹ generally provide the data at the county level. Epidemiological data are scarce for smaller spatial units such as zip code tabulation areas (ZCTAs) and have been only recently made publicly accessible and by only a few regional governments such as those of NYC¹⁰, Illinois¹¹, and Ohio¹².

Many studies that have exploited county-level data have shown evidence of the effectiveness of these social distancing measures and related mobility reduction in reducing the rate of daily new COVID-19 positive cases, both in the US^{13–15} as well as in other countries^{16,17}. However, there is one major concern associated with the current research focusing on the relationship between aggregate mobility and the spread of COVID-19 that we address in this study. In the absence of human movement data at the individual scale, most studies consider mobility as an aggregate entity measured by the total number of trips between counties or other large regions^{18–20} or distance-based measures²¹.

Lyles School of Civil Engineering, Purdue University, West Lafayette 47906, USA. ✉email: sukkusur@purdue.edu

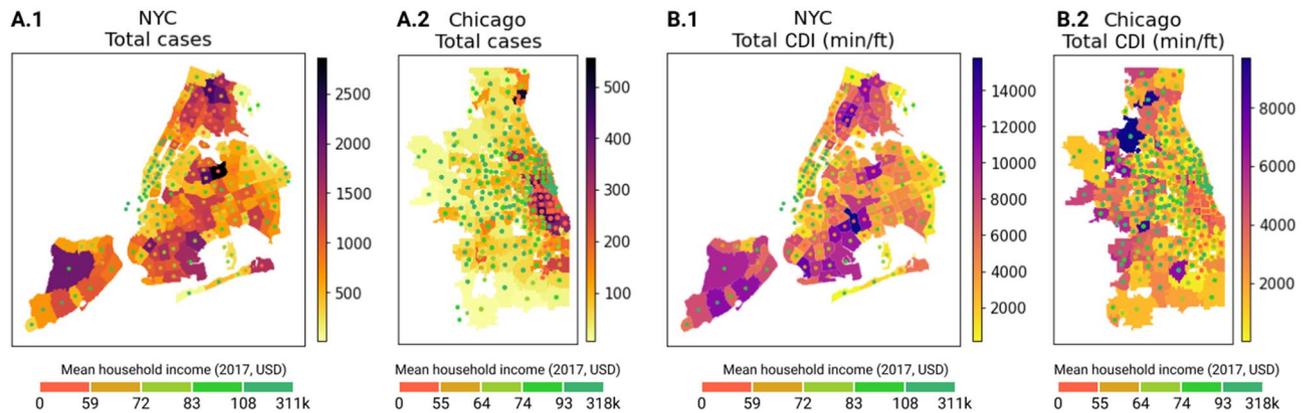


Figure 1. Map of study cities showing total cases and CDI. ZCTAs of the two study cities showing (A) total cases and (B) total contact density index (min/ft) during 20–26 April. For reference, the quintile class of the mean household income of the ZCTAs are also shown as colored centroid dots. Source of shapefiles: data.gov; Map created in Python 3.8.

Although overall movement traffic flow is a reasonable proxy of social distancing among travelers as a whole between regions, it suffers from two crucial limitations. First, social distancing inherently involves physical interaction between individuals which is not directly captured by traffic flow and distance traveled. Travelers who travel long distances alone or in small groups with little exposure to contact with others, such as in private vehicles, are misclassified as equally potential vectors of the virus as those who travel in close proximity of many others, such as public transit during the peak hours²². Second, such measures may not be a good indicator of the contact density travelers are subjected to at their destinations, where they are likely to spend more time in proximity of other visitors than the means of travel. This includes the number and physical spacing of other visitors at the destination as well as their common dwell time in contact with each other. These are important factors to be taken into account since the Centers for Disease Control and Prevention specifically highlights via its famous 6-foot-15-min rule for a close contact²³. While recent studies have considered the distribution of trips by dwell time²⁴ and crowd density based on non-residential square footage²⁵ at certain trip destinations, these have been studied in isolation as time series trends but not as a comprehensive measure of exposure to contact.

In this work, we address these two limitations—lack of a contact density-based mobility measure, and mobility analysis at the ZCTA level—in understanding the differential impact of mobility restrictions on socioeconomic groups within cities and states. We leverage anonymized origin-destination foot traffic movement patterns based on mobile phone GPS records and aggregated by SafeGraph Inc. at the level of urban neighborhoods (as ZCTAs) and the business category/industry of the trip destinations, hereby referred to as places of interest (POIs). Although there are significant concerns with using mobile phone GPS data, such as low penetration rate and privacy and data protection issues^{26,27}, they have been nonetheless used both in general mobility analysis^{28,29} as well as specifically for COVID-19^{14,15}.

Based on the insights obtained from the studies exploiting county-level trip/distance measures, we hypothesize that the number of positive coronavirus cases in residential neighborhoods is positively correlated with the amount of contact density their residents are subjected to when they travel outside their neighborhoods. Furthermore, we hypothesize that this difference in aggregate contact density emanates from, among other things, the difference in several socioeconomic factors of the residents. We test this causal pathway using structured equations models based on a panel of daily mobility and case growth data of ZCTAs in New York City (NYC) and the Chicago metropolitan area up to June 2020. Finally, we analyze the contribution of contact density to 12 major trip destination types (e.g., schools, hospitals, and restaurants) to the overall exposure to understand which industries could be targeted for tighter mobility restrictions in the event of a growth spurt in COVID-19 cases. In doing so, we also study the variation in the negative effect of income on contact exposure felt in these industries.

Results

Lower income groups had higher contact exposure and more cases. In this section, we analyze the heterogeneity of infection spread by socioeconomic status in NYC and Chicago which had been some of the worst-hit cities by COVID-19 during the first wave of the pandemic³⁰. We quantify this heterogeneity along the dimension of mean household income of the ZCTAs as well as the exposure to contact with others (not necessarily with just contagious individuals). To this end, we introduce a contact density index (CDI) which captures the spacing of visitors and their duration of stay at POIs as a proxy for the aggregate spatiotemporal density of interpersonal contacts outside home (see “Materials and methods” section).

We use an unweighted gravity model-like approach to distribute the contact density (CDI) from each POI to the ZCTA of its visitors on each day and then aggregate over all POIs to get the contact density of each ZCTA. Figure 1 shows the cumulative number of positive cases and the cumulative CDI of the two study cities in the week of 20–26 April, since the earliest date for which reliable ZCTA-level epidemiological data are available for Chicago is April 18. The dots at the centroids of the ZCTAs depict their income class, given by the quintile of the average household income distribution of the ZCTAs as of 2017.

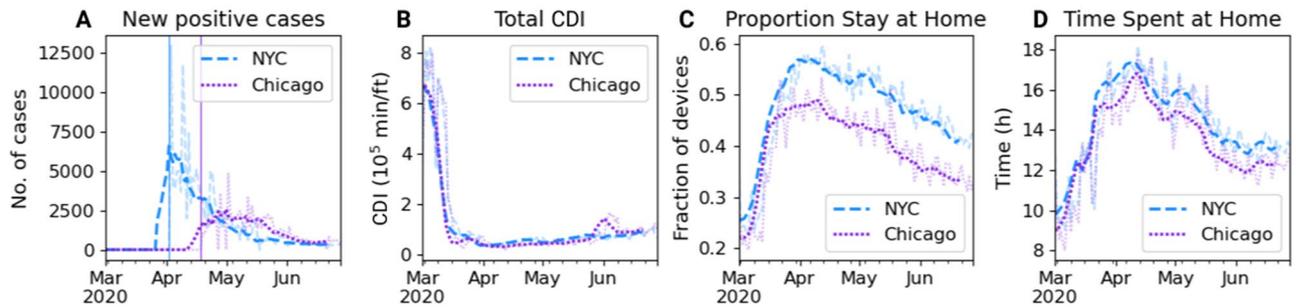


Figure 2. Trends of cases and mobility measures. Daily variation of (A) new positive cases, (B) contact density index, (C) fraction of devices registered as staying at home all day, (D) median time spent at home. The light shaded curves denote the daily trends while the dark ones depict their 7-day forward-shifted moving average. The vertical lines in panel (A) represent the first dates of available data of the number of cases.

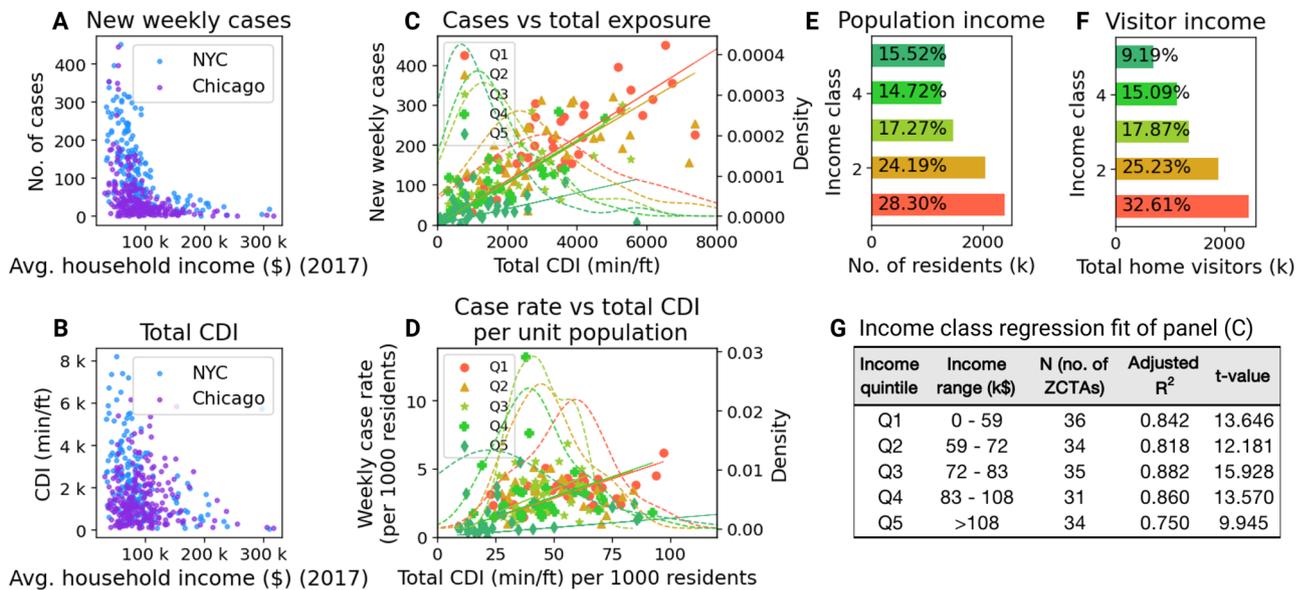


Figure 3. Relationship between cases, CDI, and income of ZCTAs during the exemplary week of 20-26 April. Income versus (A) total CDI and (B) new positive cases in NYC and Chicago; (C) New cases versus total CDI in NYC, colored by income quintile, along with the probability density distribution of ZCTAs by income quintile; (D) same as (C) but with values divided by ZCTA population. For reference, distributions of (E) population of ZCTAs of NYC and (F) the weekly number of visitors are also shown. (G) Description of income quintiles and ordinary linear regression results in panel (C).

At first glance, it can be observed that lower income regions (such as The Bronx in NYC and southern Chicago) had disproportionately more positive cases than their higher income counterparts in this week. They also experienced a higher contact density as measured by total CDI during that period. While cases were relatively more evenly distributed in NYC, with some peaks in neighborhoods in Queens and King Counties, they were much more concentrated in South and downtown Chicago, albeit much fewer than NYC in general. Moreover, the virus transmission started to decline in Chicago about two weeks after this week, while it had already started declining in NYC since at least early April (see Fig. 2A). On the contrary, though, mobility began increasing in April in both the cities after the initial phase of mobility plummet following the issuance of stay-at-home orders on March 20 in NYC³¹ and Chicago³². During the phase, the total CDI initially decreased rapidly (Fig. 2B) and more people started spending more time at home (Fig. 2C,D).

On closer inspection, the relationship between the spatiotemporal contact density (CDI), the spread of the virus, and the incomes of neighborhoods becomes more evident in Fig. 3. Panels A and B show that both total CDI and the number of cases were higher in lower income neighborhoods. The strength of linear relationship between total CDI and weekly cases (panel C) is highly distinguishable by income quintile (panel G). Even after controlling for the population of the neighborhoods, we find that CDI per capita is strongly associated with cases per capita (panel D), although the strength of this correlation reduces after controlling for population. This effect likely occurs due to a higher proportion of visitors belonging to lower-income neighborhoods after the imposition of the stay-at-home order in NYC (panel F). This observation supports the idea that lower income

Category	Name	Description	Range
Socioeconomic, S_i	Income	Natural logarithm of mean household income of ZCTA i	$[0, \infty)$
	Low edu	Fraction of population having a high school diploma or less	$[0, 1]$
	Poor	Fraction of population classified as living below the poverty line	$[0, 1]$
	Age	Fraction of population aged 65 years and above	$[0, 1]$
	Black	Fraction of population which identifies as African American (monoracial)	$[0, 1]$
	Transit	Fraction of population whose primary mode of commute to work is public transit (buses, monorail, or subway)	$[0, 1]$
Mobility, $M_{i,w}$	CDI $E_{i,w}$	Weighted total contact density index (min/ft) of POI visitors living in ZCTA i in the time window $w = [t - 7, t)$, transformed with the mapper $f(x) = \ln(1 + x)$	$[0, \infty)$
	Time home $T_{i,w}$	Fraction of day spent by the mobile devices within the Geohash7 of their owners's home in ZCTA i on day t , averaged over the window w	$[0,1]$
	Prop home $P_{i,w}$	Proportion of mobile devices registered as staying within the home Geohash7 of residents of ZCTA i all day on at least one day in the window w	$[0,1]$
	Devices $N_{i,w}$ (not used)	Number of mobile devices belonging to residents of ZCTA i on day t ; not used directly in the model but only to aggregate <i>Prop Home</i> and <i>Time Home</i> over the window w	$[0, \infty)$
Health	Cases $y_{i,t}$	Number of new positive cases in ZCTA i on day t , also transformed with $f(x)$	$[0, \infty)$

Table 1. Description of SEM variables. Description of the independent, target, and latent variables used in the daily structured equations models, along with their symbols and ranges.

people were more exposed to contact with others after the stay-at-home orders came into effect in these cities. This would be primarily because of the nature of their professions, with a higher representation of jobs requiring on-site work and/or belonging to essential services (such as nursing, grocery store operations, etc.) catering to low-income individuals⁴.

Econometric modeling. We hypothesize that the difference in the caseload of lower-income neighborhoods can be explained by the difference in the amount of exposure to contact they were subjected to during the early phase of the lockdown. We highlight the importance of measuring this exposure with CDI and other social distancing metrics instead of relying on the variation of number of trips. This is because the number of trips may misrepresent the exposure by unnecessarily counting solo trips and discounting the interaction with others at the destination.

Here, we test this hypothesis by testing the strength of the causal path of household income to virus transmission through a latent measure of exposure to close contact using a structured equations model (SEM). We specify this exposure measure to be latent so that it can take into account the effects of social interaction at places of commercial as well as non-commercial activity and compensate for the shortcomings of the used mobility variables. While exposure at commercial places is captured reasonably well with CDI at POIs (which include most major places of commercial activity except offices of private firms), the two social distancing measures—*Prop Home* and *Time Home* (Table 1) estimate the exposure due to travel outside home without considering social interaction.

We develop daily SEMs where for each day t , a causal pathway is assumed from 6 static socioeconomic measures (all considered correlated to each other), particularly the mean household income, to the daily number of new cases via latent exposure measured by the three daily mobility variables (all correlated to each other). Models of similar design have been used to show the impact of inter-county travel flow on case growth rate³³, though without considering contact density as a facilitator. The details of the model structure are provided in “[Materials and methods](#)” section and the variables are described in Table 1.

The parameter estimates of the two main relationships of interest in the daily SEMs of the two cities are shown in Fig. 4, along with the standard error of these estimates. The coefficients $\beta_S[1]$ measuring the effect of income on latent exposure (panel A) are consistently negative for both the cities. This implies that residents of lower income neighborhoods in these cities remained more likely to coming in close contact with other individuals throughout the study period. This effect is higher in the case of Chicago, although the large fluctuations in this effect are not correlated with any remarkable shift in other variables, such as lifting of the lockdown or a sudden and brief change in public response to COVID-19 that could have triggered this change. This difference in effects between NYC and Chicago could also emanate from the already large pool of infected people and higher mobilization of resources in NYC due to its uniquely intense peak of cases in mid-March³⁴.

A similar consistency is also observed in the effect of exposure on daily new cases where an increase in exposure is linked with a corresponding increase in the number of cases (Fig. 4B). In this case, however, it is observed that virus transmission in NYC was more sensitive to the exposure to close contact than in Chicago during the initial months after lockdown. After controlling for unobserved variables in these models, one could interpret this in this way—even if household income in NYC is not as considerable an indicator of exposure to contact as in Chicago, contact density contributed more substantially to the growth of cases in NYC than in Chicago.

These observations provide the core insight for understanding the causal mechanism of trip destination-based contact density on the course of COVID-19 in the early phase of lock-downs in these cities. In the next section, we discuss how contact density varied across different destination types, which could be used to identify the industries active in helping spread COVID-19 faster.

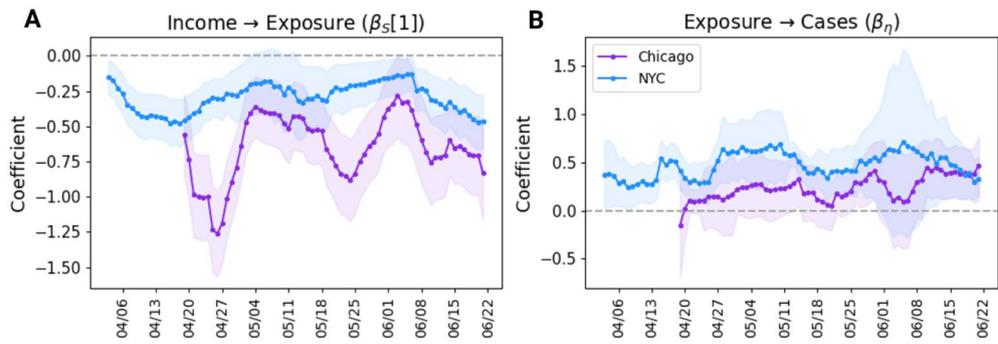


Figure 4. Estimates of important parameters of daily SEMs. Daily series of coefficient estimates of the two SEM relationships of interest for Chicago and NYC shown as 7-day moving averages: (A) effect of income on exposure, and (B) effect of exposure on new cases. The shaded region indicates the region spanned by estimate \pm standard error.

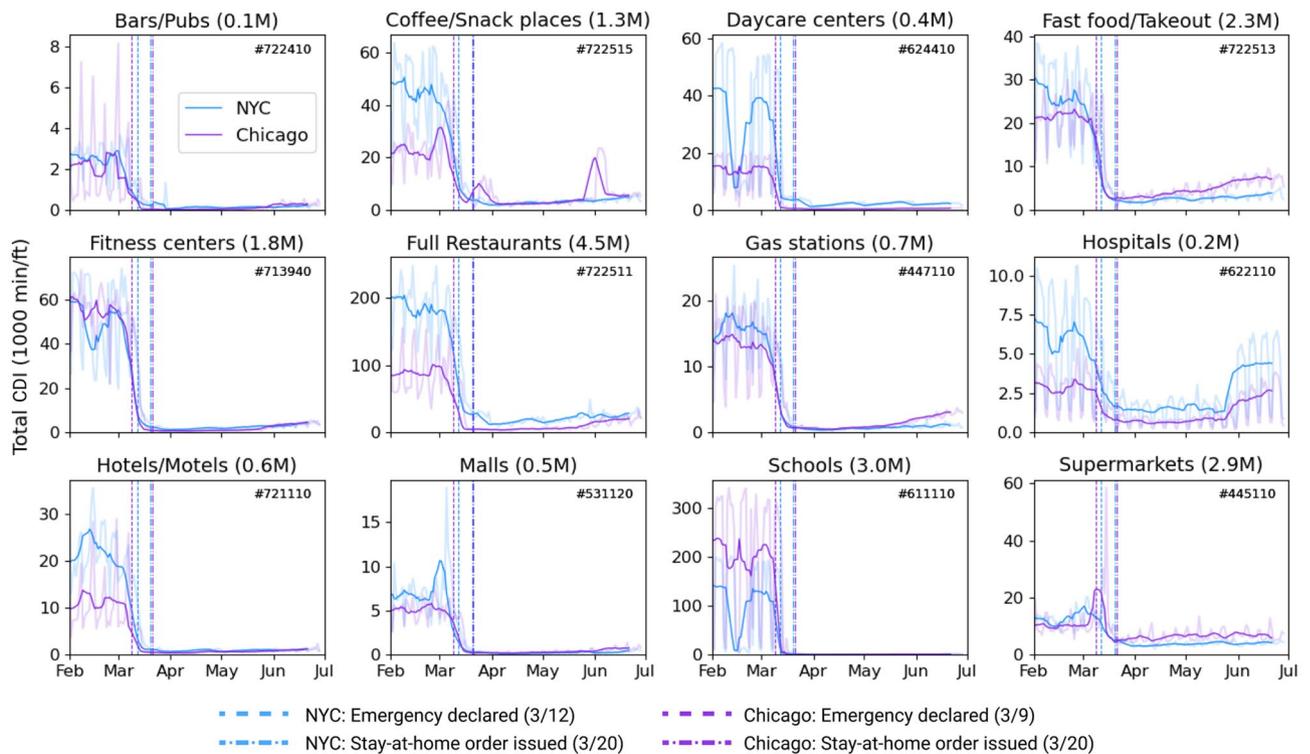


Figure 5. Trends of CDI by industry. Daily trends of CDI to POIs of 12 popular industries in NYC and Chicago, shown as 7-day moving averages. The lighter shaded curves are the daily trends. For reference, the 6-digit NAICS code and the total number of multi-person visits to POIs between February 1 and June 28. The dates of two main mobility restrictive policies in these cities are also shown.

Contact density by destination types. To better understand the characteristics of trips that contribute more to crowdedness, we next discuss the travel trends in NYC and Chicago to POIs of different industries. The publicly available Google Community Mobility Report³⁵ has been commonly used to study the differences in travel behavior across trip categories^{14,36}. However, it only provides data at the state or county level and for a select travel categories, such as home, work, groceries, etc., meaning there is limited opportunity to explore specific industries of interest, such as bars and hospitals. The SafeGraph mobility patterns are provided at the POI level, so they can be used to study these categories in detail, such as in³⁷.

We focus on 12 popular industries (by daily visits) in the study cities and label them according to their industry codes as per the North American Industry Classification System (NAICS). The trends of the total daily CDI across these industries are shown in Fig. 5, along with their NAICS codes and total number of visits to their POIs in the entire city (excluding lone hourly visits). In this figure, it can be seen that all of the industries experienced a drastic decline after the declaration of emergency in the two cities, with many categories falling

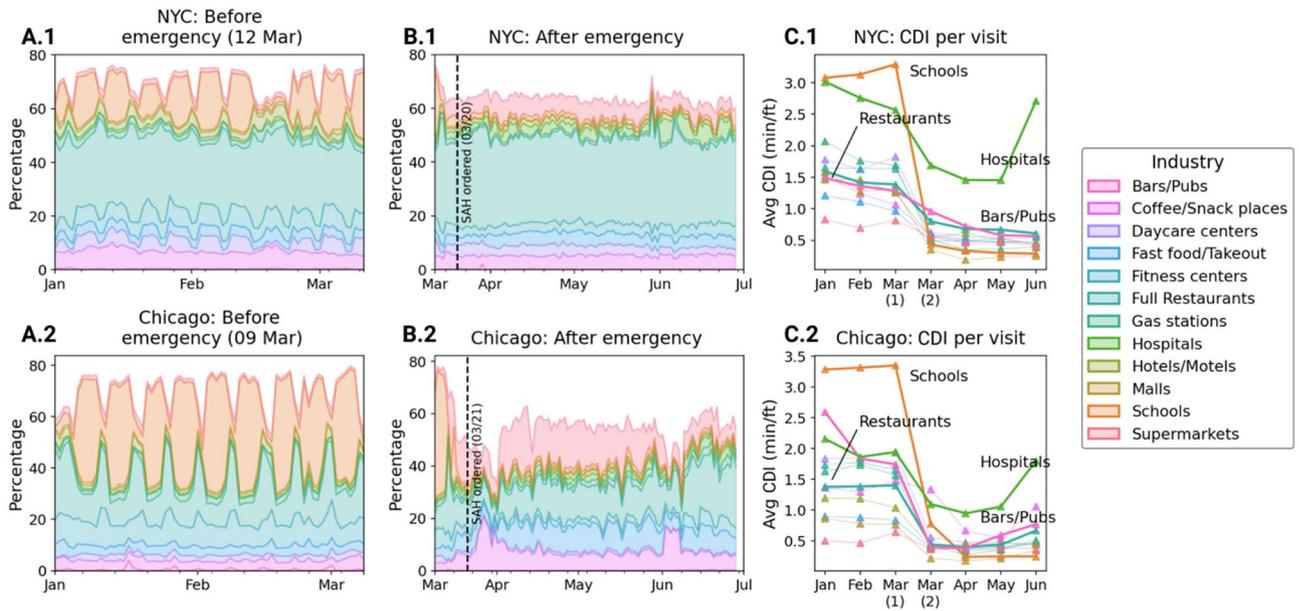


Figure 6. Contribution of CDI by industry. Proportion of total daily CDI attributable to the 12 industries of interest in the study cities (A) before and (B) after the declaration of emergency. The other industries contribute the remainder of the CDI, denoted by the empty region in the chart areas. (C) Monthly change in average CDI per visit in the two cities, with March divided into two parts (before and after March 15).

close to zero density, such as schools and malls immediately after the lockdown (stay-at-home rule). Since then, CDI has increased in hospitals and at fast food places but has largely remained negligible compared to before emergency. While schools and fitness centers have seen the biggest plummet in CDI, supermarkets have seen the lowest decline after a brief weekend surge in Chicago, likely due to panic buying.

There are a few interesting shifts in travel behavior across the two cities. CDI in the Chicago metropolitan area had been lower on average than NYC at supermarkets, fast food places, and gas stations prior to the shutdown, but this pattern started reversing afterwards. This could simply be a consequence of the severity of enforcement of lockdown practices in NYC, with several reports discussing the severe punitive actions being taken against social distancing violators during this period.

Contribution of industries on CDI. Though Fig. 5 provides an overview of mobility patterns across the major industries considered here, an important factor in the consideration of categorical restrictions is the contribution of these industries to the overall contact density at a macroscopic level. This is clarified in Fig. 6 where panels A and B show the proportion of CDI coming from POIs in the 12 important industries. The differences before and after the emergency declaration are evident in many categories. An interesting shift in the pattern of contact density is the near eradication of weekly recurring patterns of the total contribution of these 12 industries in both the cities (i.e., 100 minus the white portion in panels A and B). This can be reasonably attributed to the closure of services that typically show strong traffic variation between weekdays and weekends, such as schools, daycare centers, fitness centers, and eating places.

Schools offer a particularly interesting case in point. Public schools in NYC were closed on March 16, but many private schools and school districts had already started closing a week ago. Prior to closure, schools exhibited some of the highest crowdedness, both in terms of total CDI (Fig. 6A) and average CDI per visit (Fig. 6C). This makes sense given that visits to schools typically have much higher dwell time (typically 4–5 h) than other POIs and also have more visitors in general (see Fig. 5). The drop in CDI following school closure is even starker in Chicago where recurring periods of high CDI vanished almost overnight close to its date of issuance of the stay-at-home order. These observations support the decision of the public authorities of closing schools on the grounds of exposure to contact.

The full-service restaurant industry also stands out as being the dominant destination type for visitors over the entire study period for both the cities. Although this industry has not seen a decline in CDI as sharp as schools and daycare centers, it has had the most impact in the total reduction of CDI. This was likely facilitated by the differing degrees of closure following lockdown, with most restaurants remaining fully shut while a few others provided outdoor dining services³⁸.

The results for these two industries provide affirmation for the actions taken by public authorities in exercising special restrictions on them. However, these decisions have had implications on the disparity of contact density across neighborhoods of different income levels. This difference by industry is highlighted in Fig. 7. This figure shows the relationship between POI industry and contact density considering the income level of the people who visit these POIs and reaffirms the sharp decline in the CDI in schools, fitness centers, and bars, especially in Chicago. We chose 4 weeks representative of the different phases of mobility restrictions in the study period

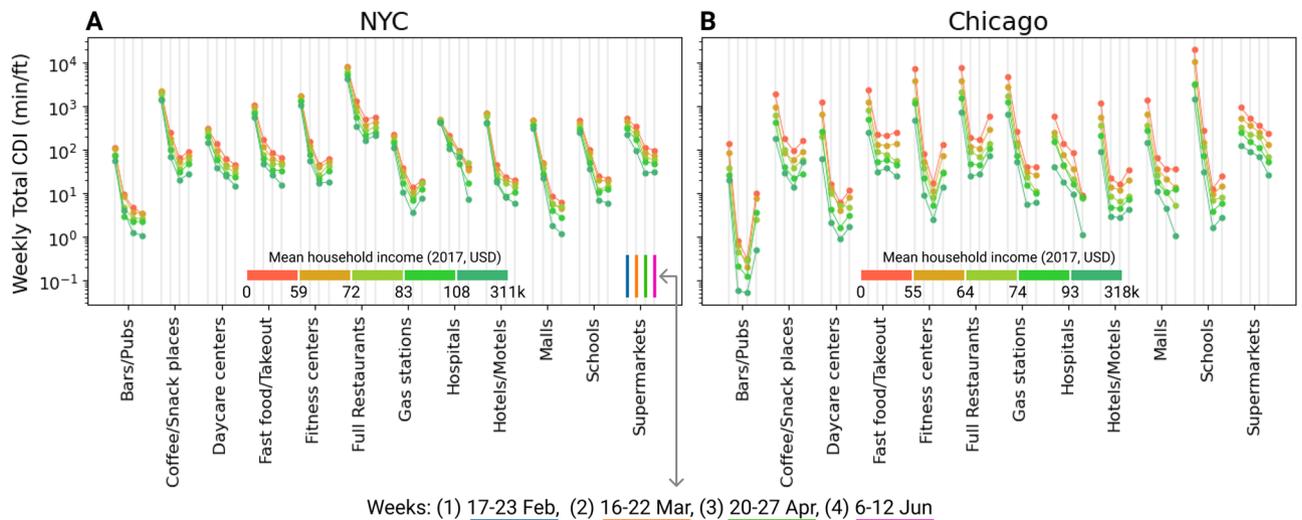


Figure 7. Phase comparison of CDI by industry and income. Variation of CDI generated by residents of neighborhoods of 5 income classes by visiting POIs of the industries of interest in 4 specific weeks representing different phases of the study period in (A) NYC and (B) Chicago.

to summarize the evolution of this disparity—mid-February (pre-lockdown), mid-March (just after lockdown), late April (a month afterwards), and early June (beginning of reopening).

A clear pattern in this figure is a consistent ordering of CDI with respect to income classes across all the categories and both the cities. Even though CDI declined very sharply after the implementation of stay-at-home, this order did not change much. Interestingly, the ratio of the CDI generated by the lower-income neighborhoods to that generated by the higher-income ones did not change substantially in Chicago but increased substantially in NYC. This can be inferred by noting the difference in the width of the bands across the industries at the left end (pre-lockdown) with the right end (lockdown and reopening), which on a linear scale represents the ratio of CDI. It could be argued that a stricter lockdown in NYC could have triggered a more polarized response from the public partly attributable to the inability of the lower-income people to stay at and work from home.

These observations provide new and confirm already accepted insights pertaining to mobility and spread of COVID-19, such as the rise of socioeconomic disparity in cities during at least the early period of the pandemic and the role of special restrictions on certain types of places.

Discussion

Social distancing policies like stay-at-home orders and closure of many services have caused widespread decline in overall mobility since mid-March due to the spread of COVID-19 in the US. While research has shown that these restrictions have been associated with an increase in socioeconomic disparity among urban neighborhoods, little work has been done on understanding the mechanism of this change. Also, while current research in this respect has often relied on macroscopic mobility measures like population flow and distance travelled, there has been limited work which seeks to understand the impact of mobility that directly relates to the crowdedness and duration of close physical contacts in public spaces. This is important because the World Health Organization and other prominent health organizations have established the importance of social distancing in reducing transmission.

In this study, we attempt to bridge the gap of relationship between human mobility involving high propensity for interpersonal contact and the disparate change of mobility among different socioeconomic groups in the US. We do this by analyzing the contact density-based mobility patterns of Chicago and New York City at the daily level in the first four months of the widespread outbreak of the pandemic. Based on aggregate mobile phone-based mobility data provided by SafeGraph Inc., we develop a spatiotemporal contact density index (CDI). This is an aggregate mobility metric based on three important factors associated with the idea of socio-physical contact - the total number of people who visit a place within the city, the area over which the visitors are spread over, and the duration of their stay there, with a special consideration of the schedule of their visits.

We observe that income is a consistently strong indicator of contact density measured by CDI in both the cities. Recognizing that CDI does not capture socio-physical interaction at places not classified as places of interest (POIs), we conceptualize an abstract combination of CDI with two social distancing metrics. These are (i) the proportion of the mobile phone-tracked population staying at home all day on a given day, and (ii) the amount of time they spend at home. We establish the negative effect of mean household income of zip code areas on this latent contact density and in turn the positive effect of the spatiotemporal contact density on the number of COVID-19 cases using a time series of structured equations models.

We then attempt to explain the composition of contact density by the destination categories (industries) of the trips generating that density, measured by CDI. We observe that heightened restrictions on mobility to POIs of certain categories, such as schools and restaurants, have contributed substantially to the decline in the overall exposure to socio-physical contact in these cities, while the effect of closure of bars has limited contribution to

this decline. This lends support to the idea of industry-specific targeted lockdown policies that the government officials have been implementing throughout the pandemic. Finally, we also observe that the disparity in contact density by income class considerably increased over all of the important industries after lockdown in Chicago but not much in New York.

Given the practical importance of these insights, we also recognize the numerous limitations with this macroscopic approach of quantifying exposure to close contact. First, the CDI we propose does not consider contact with infected or contagious people, rather contact with other visitors in general. We will consider introducing an epidemiological model to account for this limitation in a future study. Second, it is based on assumptions about the spatiotemporal positioning of visitors within POIs that may be highly ideal and likely uncommon. Third, the scale of this measure is highly dependent on the true number of visitors at POIs which under the currently available data is a valid concern due to issues related to low representative coverage of the mobile devices tracked by SafeGraph. For more details on this, the reader may refer Coston et al.³⁹.

We recognize that the lack of epidemiological data before April can be problematic for establishing a causal pathway. Having said that, we assert that this measure nevertheless provides more pertinent information about COVID-19 transmission and is more comprehensive than flow and distance-based measures and should be pursued as a tool of monitoring the progress of policies pertaining to mobility restriction. We hope to extend this study to provide a sound basis to the validity and practical applications of this measure and the insights in this study in the future.

Materials and methods

Data description. *Mobility.* Two mobility datasets were obtained from SafeGraph Inc. whose foot traffic records have been used in many studies related to mobility during the COVID-19 pandemic^{5,15}. The first dataset provides information about trips originating from different CBGs as defined in the American Community Survey (ACS) of 2013–2017. It includes relevant information such as the number and duration of devices staying in their home CBGs and their destination CBGs.

The second dataset contains information of about 4 million POIs across the US, including their unique 6-digit NAICS code (defining the industry of the POIs), floor area, enclosing CBG, hourly count of trips to these POIs, weekly distributions of trip distance, weekly trip dwell time, and origin CBGs of the visitors. We excluded the POIs located inside hospitals (mostly fast food restaurants) because of classification error due to the surge in visits to hospitals in the study cities during the peak of the pandemic.

Epidemiological data. Information about daily new tests, positive cases, deaths, and hospitalizations due to COVID-19 was obtained from the respective government health department websites—NYC¹⁰ and Chicago¹¹. We considered the Chicago metro region as the ZCTAs included in 5 counties—Cook, DuPage, Lake, Kane, and Will, totaling 253 ZCTAs. The data for these counties were derived from the Illinois dataset which has data available starting from 18 April 2020. The NYC health dataset spans 178 ZCTAs across the five boroughs in the NYC area, and has daily updated data starting from 3 April 2020. The resultant dataset has missing information about tests between 18 May and 6 June, so testing rate data was not considered in this study. We recognize the lack of ZCTA-level epidemiological data before and during the peak of infection spread in mid-March as a major data limitation.

Socioeconomic factors. The 2017 ACS was used to obtain socioeconomic variables of interest at the CBG level which were then aggregated at the ZCTA level. A principal component analysis of these variables resulted in the selection of 6 main measures of socioeconomic standing which are listed in Table 1.

Mobility metrics. Exposure-based mobility was measured with three metrics—two social distancing metrics pertaining to residents' movement outside of their neighborhoods, and a POI-based contact density index (CDI).

Duration and proportion of stay at home. We used two measures of the degree of compliance to the stay-at-home orders issued in NYC and Chicago in March 2020 using the dwell time composition of mobile devices—*PropHome* and *TimeHome*, which are described in Table 1, considering them reasonable indirect measures of social distancing practices³⁷. These metrics are measured by SafeGraph based on their estimated assignment of device owners to their home neighborhoods and provided at the CBG level, which we aggregated to the ZCTA level. For more details about these metrics, see <https://docs.safegraph.com/docs/social-distancing-metrics>.

Contact density index (CDI). We define a contact density index (CDI) that estimates the amount of exposure to contact with others subjected to an individual during their visit to a place of interest (e.g. school, hospital). It takes into account three key attributes of such trip-making—the number of people coming in contact with each other, the spacing between them at the time of contact, and the duration of this contact.

The first component is directly measured by the number of trips to a given place in a given day and has been used extensively in estimating the effect of mobility on community transmission^{6,18,19}. The other two components require microscopic details for exact computation which are generally not available on a large scale. Hence, we approximate these by making some general assumptions on the trips to POIs. These assumptions are that (i) visitors are uniformly spread out on the floor area of the POI at any given time, so that they can be assumed to be arranged in a square grid, (ii) all visitors arrive at the POI at the beginning of the hour of their trip, (iii) as in

the worst case, every visitor comes into contact with every other visitor for the entire duration of their stay at the POI, which may vary individually.

We define contact density index of a given POI in a given hour as the worst-case total contact duration of its visitors divided by the square root of the POI floor square footage. It is measured in minutes/foot. It can be easily aggregated over higher scopes (such as POI-daily level or ZCTA-daily level) by simply summing over hours. Its expressions for the 3 scopes considered here are given below.

$$\text{POI-hourly: } E_{p,h} = \frac{\tau_{p,h}}{\sqrt{A_p}}; \quad \text{POI-daily: } E_{p,t} = \sum_{h=1}^{24} E_{p,h}; \quad \text{ZCTA-daily: } E_{i,t} = \sum_{p \in z} E_{p,t} \quad (1)$$

Here, $\tau_{p,h}$ is the total contact duration (in minutes) of POI p (lying in ZCTA z) during hour h of day t and A_p is the floor area of the POI p in squared feet, which excludes parking lots but may include unusable space such as for fixtures. $\tau_{p,h}$ is the sum of contact duration of each pair of visitors in hour h , given by the minimum of their dwell times, since both visitors have to be physically present at the POI to come into contact with each other.

This can be illustrated with an example. Suppose we have 6 persons (say, A–F) visiting a POI between 1:00 and 2:00 p.m. with the following dwell times (minutes): A:10, B:10, C:20, D:20, E:20, F:40. According to our assumptions, visitors A and B come into contact with everyone else for 10 min, so the total contact duration of both A and B is $10 * 5 = 50$ min. Visitors C, D, E come into contact for 10 min with A and B and 20 min with everyone else, so their contact duration is $10 * 2 + 3 * 20 = 80$ min. Finally, F contacts A and B for 10 min and with C, D, and E for 20 min, so its contact duration is $10 * 2 + 3 * 20 = 80$. Since each pair has to be counted once, the total contact duration is half of the sum of these individuals' contact duration, i.e., $0.5 * (50 + 50 + 80 + 80 + 80 + 80) = 210$ min.

When the dwell time distribution is discrete, such as in the SafeGraph data, it has to be assumed that each trip in a given bucket has a dwell time equal to the representative point of that bucket. For a k -bucketed distribution, the expression of total contact duration is

$$\tau_{p,h} = \frac{1}{2} \sum_{i=1}^k \mu_i n_i \left(n_i - 1 + \sum_{j=i+1}^k n_j \right) \quad (2)$$

Here, n_i is the number of trips to POI p in hour h whose dwell time lies in the i^{th} bucket, with μ_i denoting the representative point of that bucket. It can be seen that visits in higher duration buckets dominate this measure, making it more realistic in terms of the compounding effect of trip duration on contact density. It also follows from this expression that a higher value of k (corresponding to finer intervals) provides a more accurate estimate of total contact duration. The dwell time distribution in the SafeGraph data is given by the $k = 4$ buckets: $[0, 5)$, $[5, 20)$, $[20, 60)$, $[60, \infty)$, so we chose $\mu = [2.5, 12.5, 40, 60]$ min.

Structured equations model. The model form used in the daily SEMs is represented by the following system of equations. The variables are described in Table 1.

$$y_{i,t} = \alpha_y + \beta_0 \cdot y_{i,t-1} + \beta_\eta \cdot \eta_{i,w} + \varepsilon_{y_{i,t}} + \mu_{y_i} + \nu_{y_t} \quad (3a)$$

$$\eta_{i,w} = \alpha_S + \beta_S^T \mathbf{S}_i + \varepsilon_{S_{i,t}} + \mu_{S_i} + \nu_{S_t} \quad (3b)$$

$$E_{i,w} = \sum_{z=t-7}^{t-1} E_{i,z} = \alpha_E + 1 \cdot \eta_{i,w} + \varepsilon_{E_{i,t}} + \mu_{E_i} + \nu_{E_t} \quad (3c)$$

$$P_{i,w} = \sum_{z=t-7}^{t-1} N_{i,z} P_{i,z} = \alpha_P + \beta_P \cdot \eta_{i,w} + \varepsilon_{P_{i,t}} + \mu_{P_i} + \nu_{P_t} \quad (3d)$$

$$T_{i,w} = \sum_{z=t-7}^{t-1} N_{i,z} T_{i,z} = \alpha_T + \beta_T \cdot \eta_{i,w} + \varepsilon_{T_{i,t}} + \mu_{T_i} + \nu_{T_t} \quad (3e)$$

For each day t , this form assumes a causal impact of static socioeconomic variables (\mathbf{S}_i , all mutually correlated) on the total latent exposure, $\eta_{i,w}$, measured by the daily mobility variables ($\mathbf{M}_{i,w}$, all mutually correlated) which itself influences the number of cases on that day, $y_{i,t}$. The effects of unmeasured contributory factors, such as testing rate and human behavior (better hygiene, use of protective face masks, personal motivation to travel, etc.) are captured by the number of cases in the neighborhood on the previous day, $y_{i,t-1}$. Also, we consider the total mobility of the past 7 days ($w = [t - 7, t)$) as contributing to the growth of cases on day t instead of the CDI on that day, based on a manifestation period of 7 days for COVID-19 (5 days for incubation⁴⁰ + 2 days for reporting). In these equations, ε , μ , and ν respectively denote the random spatiotemporal, fixed spatial, and fixed temporal error terms.

Data availability

The COVID-19 cases data are available on the state department of health websites of NYC¹⁰ and Chicago¹¹. The ACS (census) data are available from <http://www.census.gov>. The mobility data from SafeGraph are available for free on request from <http://www.safegraph.com/covid-19-data-consortium>. The datasets generated during this study and the code used to generate them are available in the repository http://www.github.com/umnlib/covid19_safegraph.

Received: 5 January 2021; Accepted: 5 May 2021

Published online: 26 May 2021

References

1. Thunström, L., Newbold, S. C., Finnoff, D., Ashworth, M. & Shogren, J. F. The benefits and costs of using social distancing to flatten the curve for COVID-19. *J. Benefit-Cost Anal.* **11**, 179–195. <https://doi.org/10.1017/bca.2020.12> (2020).
2. US-Congress. The Coronavirus Aid, Relief, and Economic Security (CARES) Act (2020).
3. Nir, S. M. Coronavirus in N.Y. Inside New Rochelle's 'Containment Area' (2020). <http://www.nytimes.com/2020/03/12/nyregion/coronavirus-new-rochelle-containment.html>.
4. Weill, J. A., Stigler, M., Deschenes, O. & Springborn, M. R. Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. *Proc. Natl. Acad. Sci. USA* **117**, 19658–19660. <https://doi.org/10.1073/PNAS.2009412117> (2020).
5. Brough, R., Freedman, M. & Phillips, D. Understanding socioeconomic disparities in travel behavior during the COVID-19 pandemic. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.3624920> (2020).
6. Jia, J. S. *et al.* Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* **582**, 389–394. <https://doi.org/10.1038/s41586-020-2284-y> (2020).
7. Webb Hooper, M., Nápoles, A. M. & Pérez-Stable, E. J. COVID-19 and Racial/Ethnic Disparities. *JAMA* **323**, 2466–2467. <https://doi.org/10.1001/jama.2020.8598> (2020).
8. The COVID Tracking Project (2020). <https://covidtracking.com>.
9. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1) (2020).
10. New York City Government. NYC Coronavirus Disease 2019 (COVID-19) Data (2020). <http://github.com/nychealth/coronavirus-data>.
11. Illinois Department of Public Health, Illinois COVID-19 Statistics. (2020). <http://dph.illinois.gov/covid19/covid19-statistics>.
12. Ohio Department of Health. Ohio COVID-19 Dashboard (2020). <http://coronavirus.ohio.gov/wps/portal/gov/covid-19/dashboards/key-metrics/cases-by-zipcode>.
13. Courtemanche, C., Garuccio, J., Le, A., Pinkston, J. & Yelowitz, A. Strong social distancing measures in the United States reduced the COVID-19 growth rate. *Health Affairs* <https://doi.org/10.1377/hlthaff.2020.00608> (2020).
14. Abouk, R. & Heydari, B. The immediate effect of COVID-19 policies on social distancing behavior in the United States. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.3571421> (2020).
15. Andersen, M. Early evidence on social distancing in response to COVID-19 in the United States. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.3569368> (2020).
16. Gatto, M. *et al.* Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proc. Natl. Acad. Sci. USA* **117**, 10484–10491. <https://doi.org/10.1073/pnas.2004978117> (2020).
17. Lai, S. *et al.* Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* <https://doi.org/10.1038/s41586-020-2293-x> (2020).
18. Badr, H. *et al.* Social distancing is effective at mitigating COVID-19 transmission in the United States. *medRxiv* <https://doi.org/10.1101/2020.05.07.20092353> (2020).
19. Kraemer, M. U. *et al.* The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497. <https://doi.org/10.1126/science.abb4218> (2020).
20. Linka, K., Peirlinck, M., Sahli Costabal, F. & Kuhl, E. Outbreak dynamics of COVID-19 in Europe and the effect of travel restrictions. *Comput. Methods Biomech. Biomed. Eng.* **23**, 710–717. <https://doi.org/10.1080/10255842.2020.1759560> (2020).
21. Yabe, T. *et al.* Non-compulsory measures sufficiently reduced human mobility in Tokyo during the COVID-19 epidemic. *Sci. Rep.* **10**, 1–9. <https://doi.org/10.1038/s41598-020-75033-5> (2020).
22. Labonté-Lemoyne, É., Chen, S. L., Coursaris, C. K., Sénécal, S. & Léger, P. M. The unintended consequences of COVID-19 mitigation measures on mass transit and car use. *Sustainability* **12**, 1–13. <https://doi.org/10.3390/su12239892> (2020).
23. Centers for Disease Control and Prevention. COVID-19 (Coronavirus Disease). Public Health Guidance for Community-Related Exposure (2020).
24. Huang, X. *et al.* Time-series clustering for home dwell time during COVID-19: What can we learn from it?. *ISPRS Int. J. Geo-Inf.* **9**, 675. <https://doi.org/10.3390/ijgi9110675> (2020).
25. Wang, Y., Chen, H., Ngo, V. & Luo, X. Crowdedness as the Missing Link between Shelter-In-Place and the Spread of COVID-19. <https://doi.org/10.2139/ssrn.3634613> (2020).
26. Oliver, N. *et al.* Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Sci. Adv.* **6**, 1–7. <https://doi.org/10.1126/sciadv.abc0764> (2020).
27. de Montjoye, Y. A. *et al.* Comment: On the privacy-conscious use of mobile phone data. *Sci. Data* **5**, 1–6. <https://doi.org/10.1038/sdata.2018.286> (2018).
28. Siła-Nowicka, K. *et al.* Analysis of human mobility patterns from GPS trajectories and contextual information. *Int. J. Geogr. Inf. Sci.* **30**, 881–906. <https://doi.org/10.1080/13658816.2015.1100731> (2016).
29. Vazquez-Prokopec, G. M. *et al.* Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment. *PLoS ONE* **8**, 1–10. <https://doi.org/10.1371/journal.pone.0058802> (2013).
30. Glaeser, E. L., Gorbach, C. S. & Redding, S. J. How much does Covid-19 increase with mobility—Evidence from New York and four other US cities. Tech. Rep., National Bureau of Economic Research (2020). <https://doi.org/10.3386/w27519>.
31. Governor Cuomo Signs the 'New York State on PAUSE' Executive Order (2020). <http://www.governor.ny.gov/news/governor-cuomo-signs-new-york-state-pause-executive-order>.
32. Illinois-Government. Executive Order 2020-10: Stay at Home in Illinois (2020). <http://www2.illinois.gov/pages/executive-orders/executiveorder2020-10.aspx>.
33. Xiong, C., Hu, S., Yang, M., Luo, W. & Zhang, L. Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections. *Proc. Natl. Acad. Sci.* <https://doi.org/10.1073/pnas.2010836117> (2020).
34. Goyal, P. *et al.* Clinical characteristics of Covid-19 in New York city. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMc2010419> (2020).
35. COVID-19 Community Mobility Reports (2020). <http://www.google.com/covid19/mobility/>.
36. Wang, H. & Yamamoto, N. Using a partial differential equation with Google Mobility data to predict COVID-19 in Arizona. *Math. Biosci. Eng.* **17**, 4891–4904. <https://doi.org/10.3934/mbe.2020266> (2020).

37. Jay, J. *et al.* Neighborhood income and physical distancing during the COVID-19 pandemic in the U.S. *medRxiv* **4**, 4. <https://doi.org/10.1101/2020.06.25.20139915> (2020).
38. Imbruce, V. Fostering food equity in an immigrant neighborhood of New York City during COVID-19. *J. Agric. Food Syst. Community Dev.* **10**, 1–5. <https://doi.org/10.5304/jafscd.2020.101.028> (2020).
39. Coston, A. *et al.* Leveraging administrative data for bias audits: Assessing disparate coverage with mobility data for COVID-19 policy. *arXiv* 173–184, <https://doi.org/10.1145/3442188.3445881> (2020).
40. Lauer, S. A. *et al.* The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Intern. Med.* **172**, 577–582. <https://doi.org/10.7326/M20-0504> (2020).

Acknowledgements

This study was conducted as a part of CRISP project funded by the National Science Foundation grant CMMI 1638311. However, the facts and opinions presented in this article are the personal statements of the authors.

Author contributions

S.U. and R.V. conceived the study design. R.V. and T.Y. analyzed the results. R.V. wrote the manuscript text. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.V.U.

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

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021