CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Segregated mobility patterns amplify neighborhood disparities in the spread of COVID-19

Andras Gyorgy[1], Thomas Marlow[2], Bruno Abrahao[3,4] and Kinga Makovi[5]*

[1]Engineering Division, New York University Abu Dhabi, Abu Dhabi, UAE, [2]Center for Interacting Urban Networks (CITIES), New York University Abu Dhabi, Abu Dhabi, UAE, [3]Leonard N. Stern School of Business, New York University, New York, NY, USA, [4]Information Systems and Business Analytics, NYU Shanghai, Shanghai, China, and [5]Social Science Division, New York University Abu Dhabi, Abu Dhabi, UAE
*Corresponding author. Email: km2537@nyu.edu

Action Editor: Nicola Perra

Abstract
The global and uneven spread of COVID-19, mirrored at the local scale, reveals stark differences along racial and ethnic lines. We respond to the pressing need to understand these divergent outcomes via neighborhood level analysis of mobility and case count information. Using data from Chicago over 2020, we leverage a metapopulation Susceptible-Exposed-Infectious-Removed model to reconstruct and simulate the spread of SARS-CoV-2 at the ZIP Code level. We demonstrate that exposures are mostly contained within one's own ZIP Code and demographic group. Building on this observation, we illustrate that we can understand epidemic progression using a composite metric combining the volume of mobility and the risk that each trip represents, while separately these factors fail to explain the observed heterogeneity in neighborhood level outcomes. Having established this result, we next uncover how group level differences in these factors give rise to disparities in case rates along racial and ethnic lines. Following this, we ask what-if questions to quantify how segregation impacts COVID-19 case rates via altering mobility patterns. We find that segregation in the mobility network has contributed to inequality in case rates across demographic groups.

Keywords: COVID-19; SEIR; mobility networks; inequality; segregation

1. Introduction

The spread of the SARS-CoV-2 virus during 2020 was uneven not only across countries and cities but also across neighborhoods within cities. The consequences are stark racial, ethnic, and class inequalities both in incidence and mortality rates (Adhikari et al., 2020; Almagro & Orane-Hutchinson, 2022; Chen & Krieger, 2021; Hatef et al., 2020; Hawkins et al., 2020; Krieger et al., 2020; van Dorn et al., 2020; Wang & Tang, 2020; Torrats-Espinosa, 2021; Levy et al., 2022). Latinx communities, in particular, have experienced sweeping outbreaks in multiple large US cities (Benitez et al., 2020; Kim et al., 2020; Reitsma et al., 2021), such as New York City (Adhikari et al., 2020; Ogedegbe et al., 2020; Pathela et al., 2021) and Chicago (Kim & Bostwick, 2020; Bryan et al., 2021). Similar unequal patterns have also been observed in cities around the globe (Zhang et al., 2020; Gozzi et al., 2021; Mena et al., 2021). This heterogeneity results in mounting challenges in policy design aimed at curbing the progression of the pandemic without exacerbating already existing inequalities (Gozzi et al., 2021; van Dorn et al., 2020; Hunter et al., 2021; Sheng et al., 2022; Cevik & Baral, 2021). To date, a comprehensive explanation for racial disparities

in the contagion and incidence of COVID-19 remains elusive (Levy et al., 2022; Tizzoni et al., 2022).

In this work, we contribute to the epidemiological modeling literature by building more systematic explanations of sociodemographic case rate inequality among segregated urban neighborhoods (Tizzoni et al., 2022; Zelner et al., 2022). In the early phase of the pandemic, researchers turned their attention to international travel and between place mobility to understand how COVID-19 spread from city to city (Brinkman & Mangum, 2022; Hâncean et al., 2021; Wells et al., 2020), while follow-up work documented that reduced mobility limits the size of outbreaks (Badr et al., 2020; Glaeser et al., 2020; Schlosser et al., 2020; Wellenius et al., 2021). Furthermore, residential segregation along race, ethnic, and class lines emerges as a link between pandemic dynamics and mobility (Acevedo-Garcia, 2000), as it spatially structures where people can and cannot reduce mobility. For instance, communities with more essential workers who were required to commute to work tended to have more severe outbreaks (Almagro & Orane-Hutchinson, 2022; Glaeser et al., 2020). These different mobility patterns provided one possible explanation for racial and ethnic inequalities (Selden & Berdahl, 2020; Almagro & Orane-Hutchinson, 2022) and highlighted that working from home is a privilege enjoyed predominantly by white Americans (Blow, 2020; Carrión et al., 2021; Gould & Shierholz, 2020).

Given the tight correlation between mobility and case rates, policymakers implemented various measures aimed at curtailing the spread of the virus (Gostin & Wiley, 2020; Zheng et al., 2020). Lockdowns, curfews, school and business closures, and limiting travel to essential trips have been some of the most important public health measures that cities implemented to "flatten the curve" prior to the availability of pharmaceutical interventions (Chinazzi et al., 2020; Courtemanche et al., 2020; Kraemer et al., 2020; Maier & Brockmann, 2020; Tian et al., 2020; Wellenius et al., 2021). Despite these steps, cities exhibited ample variation in case rates, which calls for a more nuanced explanation of how within- and between-neighborhood mobility affect epidemic progression.

Using Chicago as an exemplar major US city, we perform a set of measurements using mobility (SafeGraph, 2021) and case count data (The City of Chicago, 2021) at the ZIP Code level. Our analysis indicates that daily travel, and thus exposure to COVID-19, is contained mostly within neighborhoods and demographic groups. Although this localization of daily mobility was amplified in 2020, it also exists independent of pandemic conditions. Underpinned by this observation, we reveal how the volume of mobility and the risk that each trip represents together as a composite metric shape contagion progression in a meta-population SEIR model (Chang et al., 2020). While neither of these two factors alone can reliably explain the observed heterogeneity across neighborhoods, our composite metric does. This approach allows us to perform data-driven simulations to ask what-if questions related to neighborhood inequality in counterfactual scenarios probing the effects of shifts both in mobility patterns and in the risk that each trip represents. For instance, redirecting 25 percentage points of trips from Latinx majority ZIP Codes to Latinx majority ZIP Codes to other ZIP Codes yields approximately 20 percentage point case rate reduction in the same group without significantly affecting the outcomes in other demographic groups. Through simulations we provide policymakers with critical information about how COVID-19 spreads in cities, and what short- and long-term levers may be available for creating more equitable outcomes.

2. Methods

We first describe the datasets and the mobility network we derive from these data underpinning our analysis. Then, we discuss the SEIR model governing epidemic progression over the mobility network, and the algorithm to calculate the time-varying risk profile of each ZIP Code. Finally, we detail how we create various counterfactual scenarios to reveal the role of network structure and demographic disparities.

2.1 Datasets

2.1.1 Demographic data

Demographic data comes from the US Census Bureau, American Community Survey, 2014–2018 5-year estimates (United States Census Bureau, 2020). We label ZIP Codes by their majority racial or ethnic group using a 50% threshold. If a ZIP Code does not meet this threshold for any race or ethnicity, then we label it "Mixed." Additional variables from the Census used for analysis include the median household size, median income, percentage of employed individuals, percentage of insured individuals, percentage of households with four or more people (i.e., overcrowded households), and percentage of buildings over 50 years as a proxy for the quality of ventilation. These additional variables are used to represent the trade-offs between socio-demographic characteristics and mobility metrics to achieve city-average case rates by demographic groups.

2.1.2 Mobility data

We rely on the SafeGraph *Social Distancing Metrics* dataset for mobility data (SafeGraph, 2021). SafeGraph aggregates anonymous cellphone-based movement data and provides estimates of (i) devices residing within a census block group and (ii) trips between each block group pair. We aggregate these data to the ZIP Code level to match the geographic resolution of the data available on COVID-19 cases in Chicago. Where the boundaries of census block groups do not perfectly fit within ZIP Codes, we split trips between ZIP Codes using weights provided by the US Housing and Urban Development (Wilson & Din, 2018). Finally, we scale trips between ZIP Codes by applying a ZIP Code specific scaling factor. We calculate scaling factors by dividing a ZIP Code's census population by its average number of devices between January 15 and February 15, 2020. We select this period as it precludes early January holiday travel and major COVID-19 disruptions in the United States. The average scaling factor in our data is 25 (Tables S1–S4). For each day in 2020, our final movement data is a set of 58-by-58 matrices with cells containing the estimated daily trips between and within each ZIP Code.

2.1.3 Population data

Daily population estimates are based on the number of smartphone devices residing within a ZIP Code in a 24-h period. We multiply the number of devices using the same ZIP-specific factors used to scale mobility data to estimate the daily population residing in a ZIP Code (Figure S1). Population size is allowed to vary over time to account for temporal fluctuations (e.g., people leaving/returning to the metro area), which recent census data shows were significant during the first year of the pandemic (Frey, 2022).

2.1.4 COVID-19 case and test data

In Chicago, longitudinal ZIP Code level COVID-19 case data is provided weekly by the Department of Public Health for all 58 ZIP Codes starting March 1, 2020 (The City of Chicago, 2021). We first distribute these case counts uniformly across the days of the given week to obtain reported daily case counts, then the reported case count is corrected by considering testing disparities as well as the data from the *covidestim* project that provides daily estimates of the case count in Cook county (*https://covidestim.org/*) (Chitwood et al., 2022). To our knowledge, we rely on the most recent and highest quality data available to correct for the fact that not all COVID-19 cases are diagnosed (Pitzer et al., 2021; Bilal et al., 2021). Our approach (detailed in Supplementary Materials (SM) section S1) also reflects effects of disparities in testing across demographic groups, thus offering a significant improvement over prior research that used a time-invariant scaling factor agnostic to demographic differences (Chang et al., 2020).

In Milwaukee, longitudinal ZIP Code level COVID-19 case and test data are provided daily by Wisconsin Department of Health Services (2022), thus no daily attribution from weekly reports

was necessary. For more details, see SM section S7. These data are used for robustness analyses to reproduce the main results obtained in Chicago.

2.1.5 Data ethics

COVID-19 confirmed case counts for Chicago are aggregated to the ZIP Code level and published online by the Chicago Department of Public Health (The City of Chicago, 2021). SafeGraph mobility data covers individuals who consent to share their location data in third party mobile-phone applications. SafeGraph aggregates the location information of users who opt-in to the collection of anonymous geo-spatial data. The description of our project was reviewed by the New York University Abu Dhabi Institutional Review Board and a determination of "non-human subjects research" was issued as (i) the researchers were not engaged in collecting any data used in this paper, and thus were not interacting with human subjects; and (ii) the data used do not contain identifiable information, nor could be re-identified.

2.2 Mobility network

We consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with time-varying edges. The vertices $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ represent the n nodes of interest (i.e., ZIP Codes). Each node has both time-invariant and time-varying attributes. Time-invariant attributes include the area and demographics composition among other characteristics from US Census data. Time-varying attributes at time t include the population size $N_i^{(t)}$, the number of people in different disease states, and the parameter $\psi_i^{(t)}$ characterizing the average risk of exposure during a trip (the latter two are discussed in "Model dynamics"). Directed edges are weighted, where the weight $w_{i,j}^{(t)}$ represents our estimate of the number of individuals from node v_i visiting node v_j on the tth day of our simulation. Furthermore, the nodes belong to a demographic group G_g for $g = 1, 2, \ldots, K_G$ (e.g., $K_G = 4$ in Chicago yielding the following demographic groups: Majority Black, Majority Latinx, Majority White, and Mixed).

2.3 Model dynamics

To model the spread of SARS-CoV-2, we overlay a metapopulation disease transmission model on the mobility network defined earlier (Chang et al., 2020; Parino et al., 2021). This model builds on prior work focusing on the transmission of SARS-CoV-2 incorporating a fine-grained and time-varying mobility network into the calculations of the transmission rate to address recent calls for integrating data on the contact structure with epidemiological models (Cevik & Baral, 2021).

We use an SEIR model with susceptible (S), exposed (E), infectious (I), and removed (R) compartments for each node v_i (SM section S2). Susceptible individuals have never been infected, hence they can acquire the virus and enter the exposed state upon contact with an infectious individual. During the exposed state, individuals carry the disease but are unable to infect others. In the removed state, individuals can no longer be infected or infect others (e.g., recovered, self-isolated, or deceased individuals). While reinfection with COVID-19 has been possible, but rare, during the course of 2020 (Qureshi et al., 2022), in our model no individual may be reinfected. Note also that our analysis covers 2020, that is, prior to the widespread roll-out of COVID-19 vaccines that began on the 15th of December (The City of Chicago, 2020). As a consequence, the SEIR models considered here shall not be further complicated with additional compartments, such as vaccinated individuals. The $E \to I$ and $I \to R$ transition rates are inversely proportional to the mean latency and infectious periods, respectively.

Each node v_i maintains its own SEIR instantiation with $S_i^{(t)}$, $E_i^{(t)}$, $I_i^{(t)}$, and $R_i^{(t)}$ denoting the number of individuals in each disease state on day t. The total size of the population at node v_i

at time t is given by $N_i^{(t)} = S_i^{(t)} + E_i^{(t)} + I_i^{(t)} + R_i^{(t)}$. To reflect the standard transitions between consecutive disease states, we update them at each time step t as follows: $S_i^{(t+1)} = \eta^{(t)}(S_i^{(t)} - N_{S_i \rightarrow E_i}^{(t)})$, $E_i^{(t+1)} = \eta^{(t)}(E_i^{(t)} + N_{S_i \rightarrow E_i}^{(t)} - N_{E_i \rightarrow I_i}^{(t)})$, $I_i^{(t+1)} = \eta^{(t)}(I_i^{(t)} + N_{E_i \rightarrow I_i}^{(t)} - N_{I_i \rightarrow R_i}^{(t)})$, and $R_i^{(t+1)} = \eta^{(t)}(R_i^{(t)} + N_{I_i \rightarrow R_i}^{(t)})$, where $\eta^{(t)} = N_i^{(t+1)}/N_i^{(t)}$ ensures that population size at time $t+1$ will scale from its size at time t.

2.3.1 New exposures

To compute the number of new exposures $N_{S_i \rightarrow E_i}^{(t)}$ at node v_i at time t, we assume that any susceptible visitor to node v_j at time t has the same independent probability $\lambda_j^{(t)}$ of being infected and transitioning from the susceptible to the exposed state. That is, we do not consider heterogeneity in transition rates by demographic groups. As there are $w_{i,j}^{(t)}$ visitors from node v_i to node v_j at time t, and we assume that $S_i^{(t)}/N_i^{(t)}$ fraction of them are susceptible, the number of new exposures among them is distributed according to $\text{Binom}(w_{i,j}^{(t)} S_i^{(t)}/N_i^{(t)}, \lambda_j^{(t)}) \approx \text{Pois}(\lambda_j^{(t)} w_{i,j}^{(t)} S_i^{(t)}/N_i^{(t)})$. Rather than including heterogeneous mixing within each node as seen in other studies (Gozzi et al., 2021; Melegaro et al., 2017; Mossong et al., 2008; Zhang et al., 2020), our approach assumes homogeneous mixing of visitors. Thus, the number of new exposures among those from node v_i is distributed as the sum of the above expression over all nodes, so that new exposures are distributed as

$$N_{S_i \rightarrow E_i}^{(t)} \sim \text{Pois}\left(\frac{S_i^{(t)}}{N_i^{(t)}} \sum_{j=1}^{n} w_{i,j}^{(t)} \lambda_j^{(t)}\right), \qquad \lambda_j^{(t)} = \frac{\psi_j^{(t)}}{a_j} \sum_{k=1}^{n} w_{k,j}^{(t)} \frac{I_k^{(t)}}{N_k^{(t)}}. \qquad (1)$$

Here, the rate of infection $\lambda_j^{(t)}$ at node v_j at time t decreases with the area a_j of the ZIP Code and increases with the inflow to node v_j, the probability of infected people visiting, and the time-dependent and ZIP-specific parameter $\psi_j^{(t)}$, capturing the risk that trips carry on average. This risk depends on many factors, for example, typical length, nature of exposure, use of protective equipment, such as mask wearing, and the possibilities for and adherence to social distancing guidelines and norms, etc. See Figure S8 for the temporal distribution of this parameter. The calculation of $\psi_j^{(t)}$ is detailed when model calibration is discussed below.

2.3.2 New infectious and removed cases

Exposed individuals become infectious at a rate that is inversely proportional to the mean latency period δ_E, which is assumed to be identical for all nodes. Similarly, infectious individuals transition to the removed state at a rate that is inversely proportional to the mean infectious period δ_I which is also considered to be identical for all nodes. Therefore, we assume that at each time step t each exposed individual has a time-independent probability of first becoming infectious then of transitioning to the removed state, given by $N_{E_i \rightarrow I_i}^{(t)} \sim \text{Binom}\left(E_i^{(t)}, 1/\delta_E\right)$ and $N_{I_i \rightarrow R_i}^{(t)} \sim \text{Binom}(I_i^{(t)}, 1/\delta_I)$, respectively. According to previous studies (Chang et al., 2020), estimates for the mean latency and infectious periods are $\delta_E = 4$ days and $\delta_I = 3.5$ days, respectively. These are the parameters we adopt.

2.3.3 Initialization

We first identify the first day with non-zero estimated case count: March 7, 2020, the 67th day of the year. We take this as day 0 ($t = 0$) in the computational model. Additionally, we approximate

the infectious and removed compartments at $t = 0$ as initially empty, that is, all infected individuals are in the exposed compartment (Chang et al., 2020). Finally, we assume that the same initial prevalence occurs in every ZIP Code, that is, every individual in each demographic group has the same independent probability p_0 of being exposed (Chang et al., 2020). Accordingly, the model is initialized as $S_i^{(0)} = (1 - p_0)N_i^{(0)}$, $E_i^{(0)} = p_0 N_i^{(0)}$, $E_i^{(0)} = R_i^{(0)} = 0$ with $p_0 = 0.001$. Studying the impact of different seeding probabilities are outside of the scope of our study.

2.3.4 Calibration of the model

Calibration of the SEIR model is performed by estimating the value of the unknown parameter $\psi_i^{(t)}$ for each day t based on the new infections $C_i^{(t)}$ occurring on that specific day among the people residing in the ZIP Code. The time-dependent and ZIP-specific parameter $\psi_j^{(t)}$ captures the risk that trips carry on average (see the "New exposures" section above for further details).

For each ZIP Code, the estimate $C_i^{(t)}$ is the realization of the random variable $N_{S_i \to E_i}^{(t)}$. Since $N_{S_i \to E_i}^{(t)}$ follows a Poisson distribution, its expected value can be decomposed as

$$\mathbb{E}\left[N_{S_i \to E_i}^{(t)}\right] = \frac{S_i^{(t)}}{N_i^{(t)}} \sum_{j=1}^{n} w_{i,j}^{(t)} \lambda_j^{(t)} = \frac{S_i^{(t)}}{N_i^{(t)}} \sum_{j=1}^{n} w_{i,j}^{(t)} \frac{\psi_j^{(t)}}{a_j} \sum_{k=1}^{n} w_{k,j}^{(t)} \frac{I_k^{(t)}}{N_k^{(t)}},$$

$$= \frac{S_i^{(t)}}{N_i^{(t)}} \left(\frac{\psi_i^{(t)}}{a_i} w_{i,i}^{(t)} \sum_{k=1}^{n} w_{k,i}^{(t)} \frac{I_k^{(t)}}{N_k^{(t)}} + \sum_{j \neq i} w_{i,j}^{(t)} \frac{\psi_j^{(t)}}{a_j} \sum_{k=1}^{n} w_{k,j}^{(t)} \frac{I_k^{(t)}}{N_k^{(t)}} \right),$$

thus yielding

$$\mathbb{E}\left[N_{S_i \to E_i}^{(t)}\right] = \frac{S_i^{(t)}}{N_i^{(t)}} \left(\psi_i^{(t)} \rho_i^{(t)} \frac{w_{i,i}^{(t)}}{N_i^{(t)}} \sum_{k=1}^{n} \frac{w_{k,i}^{(t)}}{N_k^{(t)}} I_k^{(t)} + \sum_{j \neq i} \psi_j^{(t)} \rho_j^{(t)} \frac{w_{i,j}^{(t)}}{N_j^{(t)}} \sum_{k=1}^{n} \frac{w_{k,j}^{(t)}}{N_k^{(t)}} I_k^{(t)} \right), \quad (2)$$

where $\rho_i^{(t)} = N_i^{(t)}/a_i$ is the population density of node v_i on day t. When estimating $\psi_i^{(t)}$, we assume that $\psi_j^{(t)} \approx \psi_j^{(t-1)}$, corresponding to the assumption that contributions from nodes are not expected to change drastically on a daily basis. With this, we estimate $\psi_i^{(t)}$ as

$$\hat{\psi}_i^{(t)} = \frac{C_i^{(t)} \frac{N_i^{(t)}}{S_i^{(t)}} - \sum_{j \neq i}^{n} \hat{\psi}_j^{(t-1)} \rho_j^{(t)} \frac{w_{i,j}^{(t)}}{N_j^{(t)}} \sum_{k=1}^{n} \frac{w_{k,j}^{(t)}}{N_k^{(t)}} I_k^{(t)}}{\rho_i^{(t)} \frac{w_{i,i}^{(t)}}{N_i^{(t)}} \sum_{k=1}^{n} \frac{w_{k,i}^{(t)}}{N_k^{(t)}} I_k^{(t)}} \quad (3)$$

to ensure that $C_i^{(t)} = \mathbb{E}\left[N_{S_i \to E_i}^{(t)}\right]$, where we relied on the estimate $\hat{\psi}_j^{(t-1)}$ from the previous time step of $\psi_j^{(t-1)}$. Therefore, the estimation algorithm is as follows. First, we initialize $\hat{\psi}_j^{(0)} = 0$ for $j = 1, 2, \ldots, n$. Then, for $t = 1, 2, \ldots, t_{stop}$ we compute $\hat{\psi}_i^{(t-1)}$ from (3) for $i = 1, 2, \ldots, n$. Although this approach yields a slightly suboptimal solution (as we are solving decoupled scalar optimization problems instead of solving them jointly), it produces negligible prediction error (Figure S9) as the slight imprecisions due to substituting the unknown quantity $\psi_j^{(t)}$ with $\psi_j^{(t-1)}$ calculated in the previous round are further attenuated as contributions from between-ZIP trips are typically dominated by those from within-ZIP trips (Tables S1–S4).

While the fit produces almost no error at the group level, there are 8 small ZIP Codes belonging to the same group (Majority White) in close proximity to each other where error is considerably higher than in all other ZIP Codes. Further analysis reveals that in all these ZIP Codes the movement pattern shows significantly lower homophily (here the percentage of trips occurring within the same ZIP Code) than the population average (Figure S9), thus a possible explanation is misattribution of trips among neighboring ZIP Codes. Therefore, we next considered combining these 8 ZIP Codes, leading to no significant changes in the distribution of trips, population, area, movement pattern, and epidemic progression (Figure S10), while ZIP Code level performance of the model accuracy increases dramatically, virtually eliminating all error (Figure S9). Therefore, in what follows, we combine these ZIP Codes (60601, 60602, 60603, 60604, 60605, 60606, 60654, and 60661), and treat this aggregate as a single node.

2.4 Analysis details

2.4.1 Source of exposures

Considering new exposures among people residing in ZIP Code i, it follows from (1) that those that are due to trips to ZIP Code j and to ZIP Codes in group G_k are distributed according to Poisson processes with parameters $S_i^{(t)}/N_i^{(t)} w_{i,j}^{(t)} \lambda_j^{(t)}$ and $S_i^{(t)}/N_i^{(t)} \sum_{j \in G_k} w_{i,j}^{(t)} \lambda_j^{(t)}$, respectively. Therefore, the group level average probabilities that exposure in group G_i happens due to trips within one's own ZIP Code and one's own group, respectively, are given at time t by

$$ p_i^{(t)} = \frac{1}{n_i} \sum_{j \in G_i} \frac{w_{j,j}^{(t)} \lambda_j^{(t)}}{\sum_{k=1}^{n} w_{j,k}^{(t)} \lambda_k^{(t)}}, \qquad q_i^{(t)} = \frac{1}{n_i} \sum_{j \in G_i} \frac{\sum_{k \in G_i} w_{j,k}^{(t)} \lambda_k^{(t)}}{\sum_{k=1}^{n} w_{j,k}^{(t)} \lambda_k^{(t)}}, $$

where n_i is the number of nodes in group G_i. For more details, see SM Tables S1–S4.

2.4.2 Eliminating case rate disparities

Considering (1), the number of new exposures in ZIP Code i at time t follows a Poisson distribution with parameter

$$ \frac{S_i^{(t)}}{N_i^{(t)}} \sum_{j=1}^{n} w_{i,j}^{(t)} \lambda_j^{(t)} = \frac{S_i^{(t)}}{N_i^{(t)}} \sum_{j=1}^{n} w_{i,j}^{(t)} \frac{\psi_j^{(t)}}{a_j} \sum_{k=1}^{n} w_{k,j}^{(t)} \frac{I_k^{(t)}}{N_k^{(t)}}. \tag{4} $$

Introduce

$$ \beta_i^{(t)} = \psi_i^{(t)} \frac{N_i^{(t)}}{a_i} \left(\frac{w_{i,i}^{(t)}}{N_i^{(t)}} \right)^2 \qquad \text{and} \qquad \gamma_i = \frac{I_i^{(t)}}{N_i^{(t)}}, \tag{5} $$

and note that from (4) it follows that the number of new exposures $N_{S_i \to E_i}^{(t)}$ in ZIP Code i at time t can be approximated as a Poisson distribution with parameter $\beta_i^{(t)} S_i^{(t)} I_i^{(t)}/N_i^{(t)}$ since

$$ \frac{S_i^{(t)}}{N_i^{(t)}} \sum_{j=1}^{n} w_{i,j}^{(t)} \lambda_j^{(t)} \approx \beta_i^{(t)} \gamma_i^{(t)} S_i^{(t)}. $$

if $w_{i,j}^{(t)}/w_{i,i}^{(t)} \ll 1/n$, $\gamma_i^{(t)}/\gamma_j^{(t)} \approx 1$, and $\beta_i^{(t)}/\beta_j^{(t)} \approx 1$, which are confirmed in Figure S10b, Figure S3, and Figure S12a, respectively. Therefore, epidemic progression in ZIP Code i is largely determined by $\beta_i^{(t)}$, encompassing the evolution of three major ZIP-specific factors: the parameter $\psi_i^{(t)}$, population density $N_i^{(t)}/a_i$, and within-ZIP trip rate $w_{i,i}^{(t)}$ according to (5). While separately none

of these factors displays a strong correlation with case rate, their combination in $\beta_i^{(t)}$ defined in Equation (5) has significant predictive power regarding case rate (Figure S12).

While between-ZIP trips are crucial at the beginning of the pandemic (Brinkman & Mangum, 2022; Hâncean et al., 2021; Wells et al., 2020), once started, within-ZIP trips largely drive the epidemic progression, as they are significantly more frequent (Tables S1–S4). As a result, we can approximate the transmission rate at ZIP Code i as $\beta_i^{(t)}$ neglecting the effects of between-ZIP trips in two ways. First, by omitting trips leading to other ZIP Codes, we do not take into account the corresponding exposures. Second, by discarding incoming trips, thus the risk they carry, we also underestimate the infection rate at ZIP Code i. Therefore, $\beta_i^{(t)}$ quantifies the average number of individuals that a disease carrier would infect daily (Kermack & McKendrick, 1927) in a network of disconnected nodes.

When matching a group level characteristic, within the selected group we modify both the mean and temporal evolution of this characteristic such that they match the group level mean and temporal evolution in the target group.

2.4.3 Structural effects

To analyze the effects of network segregation among demographic groups, we manipulate the levels of homophily in the movement matrices. We measure homophily as the percentage of trips occurring within the same ZIP Code and the same group.

We reduce homophily (i.e., reduce segregation in the mobility network) via Laplace smoothing (Manning et al., 2008) on all outgoing trips originating in nodes from a given group, focusing on one group at a time, while all other trips (originating in different groups) remain unchanged. This is achieved by transforming the movement matrices via the following rescaling of the weights (i.e., redirecting trips among ZIP Codes): $\bar{w}_{i,j}^{(t)} \leftarrow (w_{i,j}^{(t)} + \alpha nd^2)/(1 + \alpha nd)$ with $d = \frac{1}{n}\sum_{j=1}^{N} w_{i,j}^{(t)}$, where n is the number of nodes in the graph (i.e., number of ZIP Codes). This scaling has four important properties. First, for $\alpha = 0$ we have $\bar{w}_{i,j}^{(t)} = w_{i,j}^{(t)}$, thus we recover the original mobility patterns. Second, as $\alpha \to \infty$ we have $\bar{w}_{i,j}^{(t)} \to d$, that is, outgoing edges will have uniform weights, corresponding to homogeneous movement. Third, for all values of α we have $\sum_{j=1}^{n} \bar{w}_{i,j}^{(t)} = nd = \sum_{j=1}^{n} w_{i,j}^{(t)}$, so that the overall movement volume (outgoing trip rate) for any of the nodes remains unchanged. Finally, for all values of α the ordering of trips remain unchanged, i.e., if $w_{i,j}^{(t)} < w_{i,k}^{(t)}$ then $\bar{w}_{i,j}^{(t)} < \bar{w}_{i,k}^{(t)}$.

We increase homophily for a selected group in two ways. In the first scenario, we isolate each demographic group, one at a time. When isolating group G_i, we remove all trips between the selected group and other groups, and rescale the remaining edge weights according to

$$w_{j,k}^{(t)} \leftarrow \begin{cases} w_{j,k}^{(t)} \dfrac{\sum_{h=1}^{n} w_{j,h}^{(t)}}{\sum_{h \notin G_i} w_{j,h}^{(t)}} & j, k \notin G_i, \\[3ex] w_{j,k}^{(t)} \dfrac{\sum_{h=1}^{n} w_{j,h}^{(t)}}{\sum_{h \in G_i} w_{j,h}^{(t)}} & j, k \in G_i, \\[2ex] 0 & \text{otherwise.} \end{cases}$$

so that all (outgoing) trip rates remain unchanged and preserved outgoing trips maintain the same relative importance for each node. In the second scenario, we isolate nodes within each demographic group, one at a time according to

$$
w_{j,k}^{(t)} \leftarrow
\begin{cases}
\sum_{h=1}^{n} w_{j,h}^{(t)} & j = k \in G_i, \\
w_{j,k}^{(t)} \dfrac{\sum_{h=1}^{n} w_{j,h}^{(t)}}{\sum_{h \notin G_i} w_{j,h}^{(t)}} & j, k \notin G_i, \\
0 & \text{otherwise.}
\end{cases}
$$

This way, we remove edges between nodes in the selected group, as well as trips that lead to ZIP Codes in other demographic groups, and we increase the frequency of within-ZIP trips to preserve the trip rate. The above transformation also rescales the remaining edges in the network so that all (outgoing) trip rates remain unchanged, and the preserved outgoing trips maintain the same relative importance for each node in all other groups.

To reiterate, all manipulations of the movement matrices preserve the overall movement volume (outgoing trip rate) for all ZIP Codes. This is crucial as the counterfactual computational experiments we conduct provide us with results that relate to the structure of mobility, rather than to its volume. For more details, see SM section S5.

2.4.4 Achieving equal group outcomes in case rate

We first generate counterfactuals by modifying both vulnerability and trip rate (while maintaining their temporal evolution) in each group according to $\psi_j^{(t)} \leftarrow f_\psi \psi_j^{(t)}$ and $w_{j,k}^{(t)} \leftarrow f_w w_{j,k}^{(t)}$ where the scaling factors f_ψ and f_w are selected (for each group separately) as follows: f_ψ is sampled from a uniform distribution over $[0.5, 2]$, whereas f_w is computed as $\sqrt{f_\beta/f_\psi}$ where $f_\beta = f_\psi f_w^2$ is chosen randomly from the range $[0.05, 1]$. This way the group level β is determined randomly over the range $[0.05, 1]$, and this change is achieved randomly by modifying both the vulnerability via $\psi_j^{(t)}$ and the movement quantity via $w_{j,k}^{(t)}$. For $\beta < 0.6$, the relationship between group level mean β and case rate is approximately linear (Figure S12). With this, we can compute not only how β should have changed for a given group to achieve a certain case rate outcome, but we can also translate these to changes into vulnerability and trip rate according to (5). Similarly, for six of the most relevant socio-economic differences (median household size, median income, percentage of employed, percentage of insured, overcrowdedness, and percentage of buildings over 50 years old), we first calculate the correlation between them and both the case rate and β, then relying on linear regression analysis we compute the required change in these socio-demographic factors to achieve a particular group level case rate outcome. For more details, see SM section S6.

2.4.5 Other counties

As our approach is underpinned by the relative isolation of ZIP Codes (high homophily), we consider 15 of the largest metropolitan areas: Chicago, IL; Columbus, OH; Dallas, TX; Detroit, MI; Fort Worth, TX; Houston, TX; Indianapolis, IND; Los Angeles, CA; Las Vegas, NV; Miami, FL; New York, NY; Philadelphia, PA; Phoenix, AZ; San Diego, CA; Seattle, WA. In these cities, there are 5 demographic groups (not necessarily in every city): Majority Asian, Majority Black, Majority Latinx, Majority White, and Mixed. The movement patterns in all these counties display a high degree of homophily (Figures S4–S7, Table S5). Therefore, we expect that our analytical approach can be extended to other major US cities provided that the required data become available. At present, a major bottleneck is the lack of geographically fine-grained data on COVID-19 cases and testing frequency.

(a)

(b)
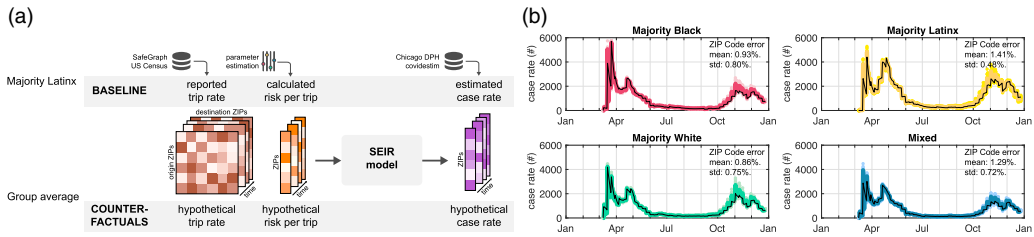
Figure 1. A compartmental SEIR model overlaid on a mobility network faithfully captures the progression of the pandemic in Chicago over 2020 including multiple peaks. (a) We combine reported case count with testing data to obtain estimates of case rates given the importance of undetected infections in sustaining the epidemic spread (Giordano et al., 2020), which is then leveraged to calculate the risk of exposure over time, considering mobility patterns. This approach enables us not only to reveal the composition of where infections occur over time, but also to evaluate various counterfactual scenarios. (b) Group level estimated case rate is depicted in black, simulation data from 10 independent runs are depicted with dots, their average is displayed with thick curves of the same color (SM section S2).

3. Results

We begin by parameterizing the compartmental SEIR model overlaid on a mobility network to faithfully reconstruct the progression of the pandemic in Chicago over 2020 including multiple peaks. Figure 1 presents the general analytical approach and the results of model fitting. Using this model as a baseline, our analysis then proceeds through four additional steps: (i) we uncover the source and distribution of exposures (e.g., within ZIP Codes and demographic groups versus between ZIP Codes and across demographic groups); (ii) we quantify the average risk that each trip represents in each demographic group and its link to case rate disparities during the progression of the epidemic; (iii) we investigate how the structure of the mobility network shapes epidemic progression; and (iv) we reveal the trade-offs between behavioral (e.g., trip rate) and socio-economic factors (e.g., income, household size) when establishing group-level outcomes in COVID-19 case rates. In addition to these main steps, we evaluate the robustness of our approach by recreating some results in 15 counties containing large cities and all the main analysis for the much smaller city of Milwaukee. The following sections describe in detail each step of our analysis.

3.1 A compartmental SEIR model overlaid on a mobility network faithfully captures the progression of the pandemic

We represent the movement of individuals among ZIP Codes as a directed network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where weight $w_{i,j}^{(t)}$ between node v_i and v_j represents the number of people traveling from ZIP Code i to ZIP Code j on day t. In 2020, the network comprises 1,211,040 daily edges among the 58 ZIP Codes in Chicago, and the mobility network displays weak connections and a spatial arrangement roughly analogous to the actual city layout Figure 2. We overlay a compartmental SEIR model on the mobility network (Figure 1(a)) to reconstruct the progression of the COVID-19 pandemic in Chicago over 2020 at the ZIP Code level (Figure 1(b)), where each ZIP Code's population is distributed over susceptible (S), exposed (E), infectious (I), and removed (R) states. New infections occur daily according to the movement within and between ZIP Codes, without the possibility of reinfections (Chang et al., 2020).

3.2 The overwhelming majority of exposures occur due to within-ZIP trips

Despite the central role of interconnectivity in mobility networks, our observations show that much of human mobility remained local in Chicago (Figure 2), echoing prior findings (Alessandretti et al., 2018; Schläpfer et al., 2021). While marked decreases in urban mobility
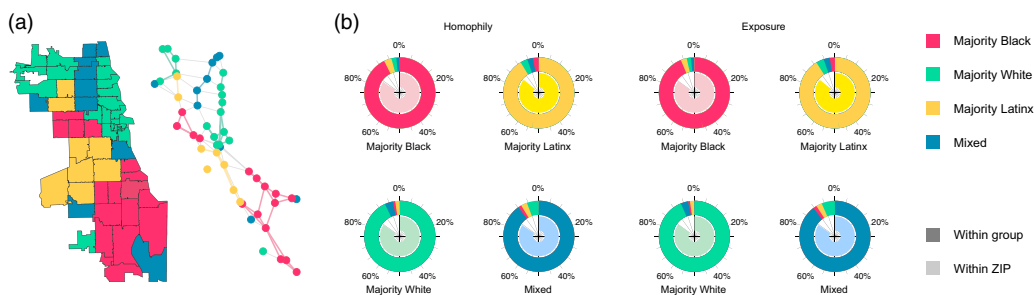
Figure 2. Daily mobility is concentrated within one's own ZIP Code and demographic group, fundamentally governing the spread of COVID-19. (a) Chicago ZIP Codes together with the observed mobility network (colored by majority demographic group), displayed using an inverse edge weighting (Fruchterman & Reingold, 1991) of trip probability between ZIP Codes (low probability and within-ZIP trips are not displayed). For more details, see Figure S2. (b) Group level averages of outgoing trip distribution and estimated sources of exposure across demographic groups are displayed in the outer rings, whereas inner rings represent trips to own ZIP and exposures within own ZIP.

occurred, especially after the declaration of national emergency in the US, the structure of mobility remained relatively stable and became even more geographically concentrated (Marlow et al., 2021). Examining the distribution of trips during the 2020 COVID-19 pandemic (Figure 2(b)) reflects this pattern: over 80% of an individual's trips are contained within their own ZIP Code (Majority Black: 85%; Majority Latinx: 86%; Majority White: 85%; Mixed: 84%, see Tables S1–S4 for ZIP Code level data). The rest of the trips highlight isolation along demographic lines: over 90% occur within group (Majority Black: 94%; Majority Latinx: 91%; Majority White: 94%; Mixed: 90%), reflecting patterns of neighborhood segregation (Wang et al., 2018; Sampson, 2019), leaving less than 10% for trips connecting different groups (Majority Black: 6%; Majority Latinx: 9%; Majority White: 6%; Mixed: 10%). From an epidemic progression perspective, this means that instead of having a well-connected network composed of intertwined ZIP Codes, the mobility network, in reality, comprises almost entirely isolated ZIP Codes with a low level of between-node mobility, mostly confined to trips within the same demographic group. This pattern extends to other major US cities as well (Figures S4–S7).

The mathematical model underpinning our analysis allows us not only to accurately reconstruct epidemic progression at the ZIP Code level (Figure S9), but also to deconstruct the source of exposures (Tables S1–S4). In particular, we estimate (Figure 2(b)) that over 80% of infections occurred within a person's own ZIP Code in 2020 (Majority Black: 83%; Majority Latinx: 86%; Majority White: 84%; Mixed: 84%), and over 90% within the same demographic group (Majority Black: 93%; Majority Latinx: 91%; Majority White: 93%; Mixed: 90%). These surprisingly high numbers are a direct consequence of the significant isolation of ZIP Codes highlighted earlier. As the overwhelming majority of exposures happen during trips confined to one's own ZIP Code (within-ZIP trips), trips connecting ZIP Codes (between-ZIP trips) play a relatively minor role in the emergence of case rate disparities. However, in line with our understanding of the diffusion of simple contagions via bridging ties in networks (Macy, 1991; Park et al., 2018), between-ZIP travel is crucial in the early phase when seeding a local outbreak (Brinkman & Mangum, 2022; Hâncean et al., 2021; Wells et al., 2020; Kuchler et al., 2022).

3.3 Vulnerability and trip rate together explain differences in epidemic progression between demographic groups

The daily infection rate depends on the volume of movement, the probability that traveling individuals are infected, the area a_i of each ZIP Code, and a time-varying parameter $\psi_i^{(t)}$ that captures
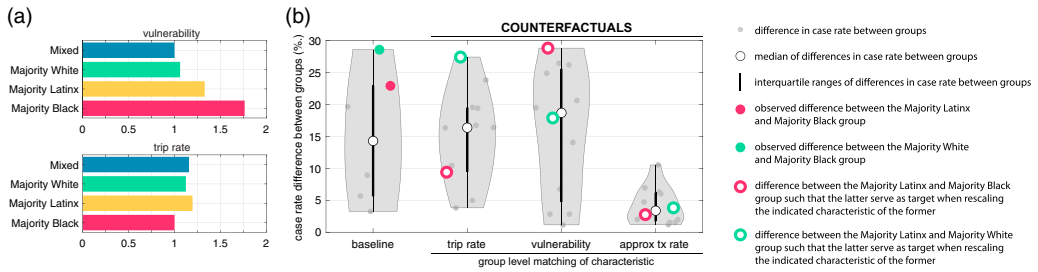
Figure 3. Vulnerability and trip rate together govern epidemic progression. (a) Group level averages of vulnerability (defined as the product of $\psi_i^{(t)}$ and population density $N_i^{(t)}/a_i$) and trip rate (both normalized to the smallest). (b) Difference in case rate between groups (grey dots). The baseline scenario refers to the observed outcome in Chicago over 2020 (total of six observed differences considering the four groups). For each counterfactual scenario, the indicated characteristic (trip rate, vulnerability, approximate transmission rate) of each ZIP Code that belongs to a select group is scaled by the same factor, until the group average reaches that of a target group, at which point the case rate difference is calculated between these two groups (three rescalings for each of the four groups, yielding a total of 12 differences). White circles represent the median of these differences, black segments denote the interquartile ranges, and grey shaded regions indicate the probability density of the data (smoothed by a kernel density estimator, see (Hintze & Nelson, 1998) for more details). Approximate transmission rate is scaled via trip rate, results are similar when scaled via vulnerability (SM section S4). Full red/green circles denote the observed difference between the Majority Latinx and Majority Black/White groups in the baseline scenario. Empty red/green circles denote the difference between the Majority Latinx and Majority Black/White groups in the counterfactuals such that the Majority Black/White groups serve as targets when rescaling the indicated characteristic of the Majority Latinx group. For more details, see SM section S4 for a pairwise comparison between demographic groups, including matching not only group level means, but also the temporal evolution of trip rate, vulnerability, and approximate transmission rate.

the risk of exposure to COVID-19 due to trips leading to ZIP Code i. We estimate $\psi_i^{(t)}$ for each ZIP Code using Maximum Likelihood estimation to calibrate our SEIR model using the estimated case count for each ZIP Code. This approach not only yields an accurate reconstruction of COVID-19 progression at the ZIP Code level capturing multiple peaks, but it also enables the construction and analysis of a wide range of counterfactual scenarios.

We define *vulnerability* as the product of $\psi_i^{(t)}$ with population density $N_i^{(t)}/a_i$. The average number of individuals that a disease carrier would infect daily (Kermack & McKendrick, 1927) is given by the approximate transmission rate $\beta_i^{(t)}$ in (5), connecting the quality and quantity of mobility via the interplay of vulnerability and trip rate, and also exposing key differences between demographic groups. For instance, while the trip rate of the Majority Black group is 11% lower than that of the Majority White group (6.77 vs 7.60), the former group has a 18% greater case rate (39% vs 33%), owing to their 67% higher vulnerability (0.0175 vs 0.0105). Similarly, while the Majority Latinx and Mixed groups have almost identical trip rates (8.10 vs 7.84), their case rates differ by 48% (62% vs 42%) as the vulnerability of the former exceeds by 33% that of the latter (0.0132 vs 0.0099).

Leveraging the observation that within-ZIP trips tend to concentrate most of the population's mobility, we may compare different demographic groups through the lens of vulnerability and trip rate (Figure 3(a)). The derivation of $\beta_i^{(t)}$ underscores that focusing on either vulnerability or trip rate to explain differential outcomes across different demographic groups is not sufficient. To isolate the unique contribution of vulnerability and mobility to discrepancies in COVID-19 outbreaks across demographic groups, we estimate the impact of these factors by constructing counterfactual scenarios where we change either the vulnerability or the trip rate of all ZIP Codes of one demographic group to closely resemble the average observed in another demographic group (Figure 3(b)). For instance, throughout 2020 the Majority Latinx group has 23 percentage points higher case rate than Majority Black group and 29 percentage points higher than Majority White group. This is due to the fact that the Majority Latinx group has both high vulnerability

and trip rate (Figure 3(a)). When decreasing their trip rate to match that of the Majority Black and Majority White groups, the discrepancy decreases (to 9 and 27 percentage points, respectively), yet significant gaps remain due to differences in vulnerability. Next, while the difference in case rates decreases when decreasing the vulnerability of the Majority Latinx group to match the group level average of the Majority White group (to 17 percentage points), it further increases when increasing it to match the group level average of the Majority Black group (to 29 percentage points). These results highlight that differences in vulnerability and trip rate alone do not provide sufficient explanation of observed group level disparities.

Once the combined effect of vulnerability and trip rate is matched via the group level average of the approximate transmission rate $\beta_i^{(t)}$, the differences in case rates virtually disappear (Figure 3(b)). For instance, the difference between the Majority Latinx and Majority Black group decreases to three percentage points, and the difference between the Majority Latinx and Majority White group decreases to four percentage points. Our approach thus allows us to attribute the differences in case rate between ZIP Codes of different demographic composition to differences in vulnerability and trip rate. This further highlights that both the volume of movement (trip rate) and the risk of exposure per trip (vulnerability) are essential in understanding and mitigating epidemic progression, and inequality in either can and will lead to inequality in case rate among demographic groups.

3.4 Decreased network segregation reduces case rate inequality

After revealing how vulnerability and trip rate together lead to inequalities in outcomes, we next investigate the contribution of movement patterns to reveal the role that the structure of the mobility network plays. To this end, we focus on both further increasing segregation of demographic groups (i.e., pushing the mobility network towards higher levels of homophily), as well as decreasing segregation (i.e., increasing uniformity of outgoing between-ZIP trips, thereby pushing the network towards lower levels of homophily). Thus, we generate a set of counterfactual movement networks (SM section S5), keeping both trip rate and vulnerability unchanged across all demographic groups at all times to isolate the impact of the structure of the mobility network.

From both sets of analysis, a similar pattern emerges: decreasing homophily leads to reduced inequality in case rates, while increasing segregation could significantly exacerbate outcomes in already vulnerable groups (see also (Laumann & Youm, 1999) documenting a substantively similar pattern in case of sexually transmitted diseases). Specifically, the case rate in the Majority Latinx group would rise from 63% to approximately 72% when we isolate the group, and to 71% when we further isolate all nodes in this group (Figure 4). In turn, when we decrease segregation for the Majority Latinx ZIP Codes, a 25 percentage point reduction in homophily yields approximately 20 percentage point case rate reduction in the same group without significantly affecting the outcomes in other groups. Conversely, when we apply Laplace smoothing on the outgoing trips originating in Majority White ZIP Codes, case rates increase mildly in this group, balanced by a comparable drop in the Majority Latinx group (Figure S26).

These results are the direct consequence of the most moderate and pronounced outbreaks occurring in the Majority White and Majority Latinx groups, respectively. Redirecting trips from the former to other groups (including the Majority Latinx group) has two main consequences: (i) trips from the Majority White group now lead to neighborhoods with higher case rates; and (ii) the probability of encountering someone infected in the Majority Latinx group decreases due to the influx from the Majority White group (with lower average case rates). We can interpret redirecting trips from the Majority Latinx group similarly, and as expected, rewiring the mobility pattern in the other groups yield similar but smaller changes (Figure S26).

These observations underscore a secondary effect of policies that aim to reduce the volume of mobility. As prior work has already demonstrated, trip rate reduction primarily impacts
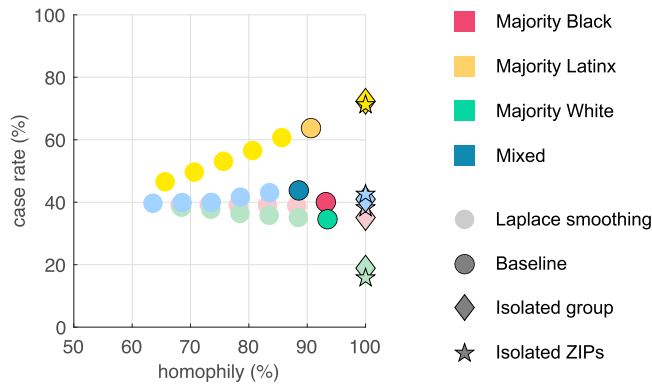
Figure 4. Decreased homophily reduces inequality in COVID-19 case rate. Baseline refers to the observed mobility data (dark circles). Homophily is reduced by rescaling outgoing trips using Laplace smoothing for all nodes within a select group in 5%. increments (light circles), keeping both the trip rate and the ordering of weights (i.e., importance of connections) unchanged, without affecting trips originating in other groups. Homophily is increased either by isolating a select group from the rest of the network (diamonds), or by further isolating nodes within the group (stars). Case rate is presented only for the group with altered mobility patterns (SM section S5).

longer-distance trips and trips leading to demographically dissimilar ZIP Codes (Marlow et al., 2021). These changes thus yield an increase in segregation and may "lock in" privilege, as well as exacerbate inequalities. In fact, keeping the level of mobility and vulnerability unchanged but reducing group-level homophily in the mobility network to its pre-pandemic levels, the city level case rate remains practically identical, but the gap between the worst faring (Majority Latinx) and best faring (Majority White) groups reduces by 4.16 percentage points, or 15% (Table S6). Specifically, while the former sees a reduction in case rate of 2.95 percentage points, the latter experiences an increase of 1.21 percentage points. This highlights that even subtle increases in the group-level homophily in the mobility network could meaningfully contribute to inequality in outcomes by protecting and further increasing the privilege of some communities at the expense of already vulnerable ones. Thus, policy-makers should consider such unintended and detrimental effects, and balance these with the beneficial impact of mobility reduction (Badr et al., 2020; Glaeser et al., 2020; Wellenius et al., 2021; Chinazzi et al., 2020; Courtemanche et al., 2020; Kraemer et al., 2020).

3.5 The trade-off between mobility, vulnerability, and socio-demographic factors

A large body of prior work documented an association between COVID-19 case rates and socio-demographic characteristics, such as race, ethnicity, socio-economic status, and occupational structures (Torrats-Espinosa, 2021; Glaeser et al., 2020; Chen et al., 2021; Levy et al., 2022). While these associations are crucial in understanding the resource needs of various communities, they are notoriously hard to change via short-term interventions. Reducing mobility or decreasing vulnerability (e.g., introducing and enforcing mask mandates, or providing quarantining facilities to avoid exposing family members in densely populated households) are thus more suitable candidates for short-term intervention. Via the approximate transmission rate ($\beta_i^{(t)}$) we can associate variations in each of its components with changes in case rates, as well as translate these variations in vulnerability and trip rate to shifts in the socio-demographic composition of ZIPs using a simple linear regression.

To this end, we first generate counterfactual scenarios to explore the relationship between group level approximate transmission rates and case rates as follows. We start with randomly rescaling both the trip rate and vulnerability of each node within all groups, then we simulate the
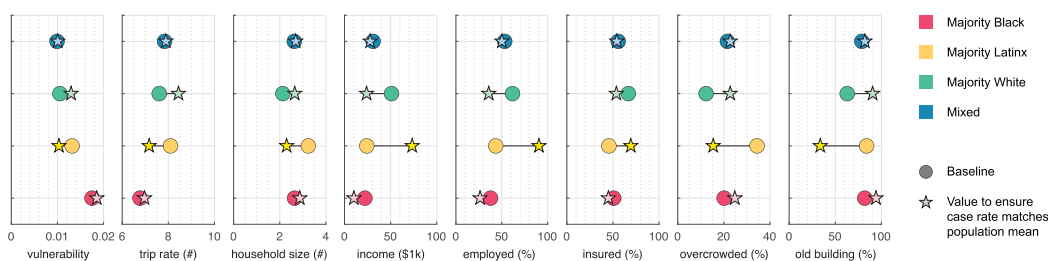
Figure 5. Trade-offs between vulnerability, mobility, and sociodemographic factors. Circles correspond to the observed baseline, stars represent the value that would ensure that the case rate within the selected group would match the city level average (43%).

corresponding scenarios using the network model to reveal the connection between case rate and approximate transmission rate. Considering the effect that changing the group level vulnerability and trip rate would have on the approximate transmission rate according to (5), we can select a target case rate in any of the groups, translate it to the corresponding approximate transmission rate, and through that identify the required group level vulnerability or trip rate, while leaving the other unchanged. This exposes inequalities in case rates across demographic groups in a new light (Figure 5). For instance, those living in Majority Latinx ZIPs would need to cut 0.93 daily trips on average to achieve the city level average case rate, while those living in Majority White ZIPs could afford 0.84 more daily trips—a difference of almost 2 daily trips (with a population average of 7.5).

We can further translate these changes in mobility and vulnerability to changes in socio-economic composition associated with case rates. We consider six factors that have been associated with COVID-19 case rate: mean household size, median household income, percentage of individuals employed, percentage of individuals insured, percentage of overcrowded households (i.e., households with 4 or more people), and the percentage of old buildings (i.e., buildings older than 50 years, which may impact respiratory health and exposure to environmental contaminants such as lead (Shaw, 2004)). These factors show an expected strong correlation with case rate and approximate transmission rate (SM section S6). Our results reveal that due to their low case rate, the Majority White group could afford a 25% increase in median household size or a 50% drop in median household income to reach the city level average case rate, instead of increasing their daily trip rate by 0.84. Conversely, as the Majority Latinx group suffers from a case rate significantly greater than the city level average, the required socioeconomic changes are of the opposite sign (as an alternative to 0.93 fewer daily trips) to ensure the same case rate outcome (Figure 4). Taken together, these results provide a cross-walk between changes in vulnerability, trip rate, and socio-demographic factors via the case rate.

3.6 Limitations

The most important limitations of our work are centered around (i) data availability constraints in other cities; (ii) granularity of our analysis; and (iii) modeling assumptions.

First, our approach requires temporally and spatially fine-grained mobility, testing, and case data. While mobility data are available in each of the 15 major US cities, case and testing data are only available for Chicago, New York, and Seattle. In New York, ZIP Code level testing/case data are made public for only after the first major peak of the pandemic. In Seattle, ZIP Codes are classified as either Majority White or Mixed, failing to provide a context for the rich analysis that we carried out in Chicago to uncover disparities across demographic groups. Notably, in Milwaukee, where the data required for the analysis are available, our key findings hold about

how segregated mobility patterns amplify neighborhood disparities in the spread of COVID-19 (SM section S7).

Second, we perform our analysis at the ZIP Code level, treating them as if they were homogeneous units. Similarly to other recent work (Bluhm et al., 2022; Crawford et al., 2022; Levy et al., 2022; Pei et al., 2021; Wikle et al., 2022), we assume that SafeGraph mobility data are representative of the population with no major discrepancies across demographic groups. Additionally, we consider test and case counts at the ZIP Code level, however, majority group membership of a ZIP Code may not accurately reflect the demographic composition of the population actually tested or infected. Finally, while our analysis uncovers ZIP-to-ZIP variability in the average risk that trips represent, it may mask heterogeneity across trips within each ZIP (e.g., due to duration). To address these limitations, subsequent analysis would require (currently unavailable) higher resolution data, such as (i) mobility patterns disaggregated at the ZIP Code level by demographic background; (ii) the racial and ethnic composition of those who are tested or diagnosed within each ZIP Code; and (iii) detailed information about individual trips (e.g., duration).

Third, our model makes several assumptions about the dynamics of COVID-19 spread which decrease the realism of the model. First, while we capture heterogeneity in contact between socio-demographic groups based on their trips between ZIP Codes, we assume homogeneous mixing within ZIP Codes, due to lack of finer grained data. Second, our approach does not incorporate either compartments for the vaccinated population or the possibility of reinfections. Similarly, while we assume that the period between exposed and infected states does not appreciably change over 2020, this might need to be revisited with the emergence of new variants. While these were appropriate choices for 2020, they ought to be revised for modeling the COVID-19 pandemic beyond 2020 by introducing additional compartments or more complex transition probabilities between compartments.

4. Discussion

The analysis we present in this paper relies on longitudinal mobility, case, and test data at the ZIP Code level in Chicago. Our approach presents a significant improvement over prior work, given the fine-grained geographic resolution of case data, and the integration of novel estimates for the number of infected but asymptomatic individuals even accounting for testing disparities across demographic groups. We build on the observation that ZIP Codes are essentially isolated from one another in the mobility network, enabling us to interpret epidemic progression in Chicago over 2020 through the lens of the approximate transmission rate $\beta_i^{(t)}$ encompassing ZIP Code level characteristics alone: volume of within-ZIP trips and the average risk that each of these trips represents. In addition to pinpointing where exposures occur, we also demonstrate that differences in either of these factors alone do not explain the discrepancy in case rates across demographic groups. Obtaining qualitatively similar results in Chicago and Milwaukee despite the dramatic difference in population size suggests the generalizability of our findings (SM section S7). As the mobility network of 15 major US cities display a pattern similar to those observed in Chicago and Milwaukee (i.e., ZIP Codes and groups are essentially isolated), we expect similar results in those contexts, provided that longitudinal ZIP Code level case and test data become available.

Over the course of 2020 urban mobility networks changed significantly (Schlosser et al., 2020; Marlow et al., 2021), as mobility has become more localized, and neighborhoods have become more isolated. We demonstrate that this isolation, which likely extends beyond 2020, reinforces inequalities in case rates across communities, and our observations suggest that these effects are further amplified by localized and concentrated mobility upon the onset of an outbreak. By leveraging actual mobility data to better approximate contact between demographic groups and neighborhoods, we reinforce the findings of exploratory SEIR models highlighting

that demographic stratification of contact can increase inequality in case rates (Ma et al., 2021). Homophily thus likely protects already privileged communities where the material conditions of a neighborhood (e.g., the availability of protective equipment, better building conditions, and minimal overcrowding) already contribute to reducing the risk that each trip represents. Additionally, segregation in mobility could further privilege places where residents can more easily adopt behaviors that limit their exposure, such as forming close-knit social bubbles (Block et al., 2020) or strategically accessing essential services to avoid crowding (Chang et al., 2020; Nishi et al., 2020).

It is important to recognize that our results are consistent with COVID-19 spreading as a simple, rather than a complex contagion (Centola, 2020), where weak ties connecting neighborhoods remain important in seeding localized epidemics. However, as we show, within-neighborhood mobility fundamentally shapes the long-term trajectory of local epidemics. In the future, inequalities in vaccination rates (Agarwal et al., 2021) could further exacerbate these divergent pandemic experiences by socio-demographic background. While reducing mobility has proven to be effective to "flatten the curve," the next frontier of interventions must focus on reducing the likelihood of infection upon contact, which includes vaccination campaigns and mask mandates that have become hot-button political issues (DeMora et al., 2021; Romer & Jamieson, 2020). Upon designing and implementing these strategies, policy makers must also strive to ensure that their beneficial effects are enjoyed equitably across demographic groups, instead of further widening the gap among them.

References

Acevedo-Garcia, D. (2000). Residential segregation and the epidemiology of infectious diseases. *Social Science & Medicine*, *51*(8), 1143–1161.

Adhikari, S., Pantaleo, N. P., Feldman, J. M., Ogedegbe, O., Thorpe, L., & Troxel, A. B. (2020). Assessment of community-level disparities in coronavirus disease 2019 (COVID-19) infections and deaths in large US metropolitan areas. *JAMA Network Open*, *3*(7), e2016938.

Agarwal, R., Dugas, M., Ramaprasad, J., Luo, J., Li, G., & Gao, G. (2021). Socioeconomic privilege and political ideology are associated with racial disparity in COVID-19 vaccination. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(33), e2107873118.

Alessandretti, L., Sapiezynski, P., Sekara, V., Lehmann, S., & Baronchelli, A. (2018). Evidence for a conserved quantity in human mobility infections and deaths in large US metropolitan areas. *Nature Human Behaviour*, *2*(7), 485–491.

Almagro, M., & Orane-Hutchinson, A. (2022). JUE Insight: The determinants of the differential exposure to COVID-19 in New York City and their evolution over time. *Journal of Urban Economics*, *127*, p. 103293. Available at: https://doi.org/10.1016/j.jue.2020.103293.

Badr, H. S., Du, H., Marshall, M., Dong, E., Squire, M. M., & Gardner, L. M. (2020). Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *The Lancet Infectious Diseases*, *20*(11), 1247–1254.

Benitez, J., Courtemanche, C., & Yelowitz, A. (2020). Racial and ethnic disparities in COVID-19: Evidence from six large cities. *Journal of Economics, Race, and Policy*, *3*(4), 243–261.

Bilal, U., Tabb, L. P., Barber, S., & Roux, A. V. D. (2021). Spatial inequities in COVID-19 testing, positivity, confirmed cases, and mortality in 3 U.S. cities: An ecological study. *Annals of Internal Medicine*, *174*(7), 936–944.

Block, P., Hoffman, M., Raabe, I. J., Dowd, J. B., Rahal, C., Kashyap, R., & Mills, M. C. (2020). Social network–based distancing strategies to flatten the COVID-19 curve in a post–lockdown world. *Nature Human Behaviour*, *4*(6), 588–596.

Blow, C. M. (2020). Opinion — social distancing is a privilege. *The New York Times*. 5 April. Available at: https://www.nytimes.com/2020/04/05/opinion/coronavirus-social-distancing.html (Accessed: 11 July 2021).

Bluhm, R., Polonik, P., Hemes, K. S., Sanford, L. C., Benz, S. A., Levy, M. C. & Burney, J. A. (2022). Disparate air pollution reductions during California's COVID-19 economic shutdown. *Nature Sustainability*, *5*(6), 509–517.

Brinkman, J., & Mangum, K. (2022). Jue insight: The geography of travel behavior in the early phase of the COVID-19 pandemic. *Journal of Urban Economics*, *127*, 103384.

Bryan, M. S., Sun, J., Jagai, J., Horton, D. E., Montgomery, A., Sargis, R., & Argos, M. (2021). Coronavirus disease 2019 (COVID-19) mortality and neighborhood characteristics in Chicago. *Annals of Epidemiology*, *56*, 47–54.e5.

Carrión, D., Colicino, E., Pedretti, N. F., Arfer, K. B., Rush, J., DeFelice, N., & Just, A. C. (2021). Neighborhood-level disparities and subway utilization during the COVID-19 pandemic in New York City. *Nature Communications*, *12*(1), 3692.

Centola, D. (2020). Considering network interventions. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(52), 32833–32835.

Cevik, M., & Baral, S. D. (2021). Networks of SARS-CoV-2 transmission. *Science*, *373*(6551), 162–163.

Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2020). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, *589*(7840), 82–87.

Chen, J. T., & Krieger, N. (2021). Revealing the unequal burden of COVID-19 by income, race/ethnicity, and household crowding: US county versus ZIP Code analyses. *Journal of Public Health Management and Practice*, *27*(Supplement 1), S43–S56.

Chen, M. K., Chevalier, J. A., & Long, E. F. (2021). Nursing home staff networks and COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(1), e2015455118.

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S. & Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, *368*(6489), 395–400.

Chitwood, M. H., Russi, M., Gunasekera, K., Havumaki, J., Klaassen, F., Pitzer, V. E. & Menzies, N. A. (2022). Reconstructing the course of the COVID-19 epidemic over 2020 for US states and counties: Results of a Bayesian evidence synthesis model. *PLOS Computational Biology*, *18*(8), e1010465.

Courtemanche, C., Garuccio, J., Le, A., Pinkston, J., & Yelowitz, A. (2020). Strong social distancing measures in the United States reduced the COVID-19 growth rate. *Health Affairs*, *39*(7), 1237–1246.

Crawford, F. W., Jones, S. A., Cartter, M., Dean, S. G., Warren, J. L., Li, Z. R. & Morozova, O. (2022). Impact of close interpersonal contact on COVID-19 incidence: Evidence from 1 year of mobile device data. *Science Advances*, *8*(1), eabi5499.

DeMora, S. L., Merolla, J. L., Newman, B., & Zechmeister, E. J. (2021). Reducing mask resistance among white evangelical Christians with value-consistent messages. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(21), e2101723118.

Frey, W. H. (2022). *New census data shows a huge spike in movement out of big metro areas during the pandemic*.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, *21*(11), 1129–1164.

Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A. & Colaneri, M. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, *26*(6), 855–860.

Glaeser, E. L., Gorback, C., & Redding, S. J. (2020). Jue insight: How much does COVID-19 increase with mobility? Evidence from New York and four other U.S. cities. *Journal of Urban Economics*, *127*, 103292.

Gostin, L. O., & Wiley, L. F. (2020). Governmental public health powers during the COVID-19 pandemic: Stay-at-home orders, business closures, and travel restrictions. *JAMA*, *323*(21), 2137–2138.

Gould, E., & Shierholz, H. (2020). *Not everybody can work from home: Black and Hispanic workers are much less likely to be able to telework*. Washington, DC: Blog post. Economic Policy Institute.

Gozzi, N., Tizzoni, M., Chinazzi, M., Ferres, L., Vespignani, A., & Perra, N. (2021). Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile. *Nature Communications*, *12*(11), 2429.

Hatef, E., Chang, H.-Y., Kitchen, C., Weiner, J. P., & Kharrazi, H. (2020). Assessing the impact of neighborhood socioeconomic characteristics on COVID-19 prevalence across seven states in the United States. *Frontiers in Public Health*, *8*, 571808.

Hawkins, R. B., Charles, E. J., & Mehaffey, J. H. (2020). Socio-economic status and COVID-19–related cases and fatalities. *Public Health*, *189*, 129–134.

Hintze, J. L., & Nelson, R. D. (1998). Violin plots: A box plot-density trace synergism. *The American Statistician*, *52*(2), 181–184.

Hâncean, M.-G., Slavinec, M., & Perc, M. (2021). The impact of human mobility networks on the global spread of COVID-19. *Journal of Complex Networks*, *8*(6), cnaa041.

Hunter, R. F., Garcia, L., de Sa, T. H., Zapata-Diomedi, B., Millett, C., Woodcock, J. & Moro, E. (2021). Effect of COVID-19 response policies on walking behavior in US cities. *Nature Communications*, *12*(1), 3652.

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, *115*(772), 700–721.

Kim, N. H., Lan, K. F., Nkyekyer, E., Neme, S., Pierre-Louis, M., Chew, L. & Duber, H. C. (2020). Assessment of disparities in COVID-19 testing and infection across language groups in Seattle. *Washington JAMA Network Open*, *3*(9), e2021213.

Kim, S. J., & Bostwick, W. (2020). Social vulnerability and racial inequality in COVID-19 deaths in Chicago. *Health Education & Behavior*, *47*(4), 509–513.

Kraemer, M. U. G., Yang, C-H., Gutierrez, B., Wu, C-H., Klein, B., Pigott, D. M. & Scarpino, S. V. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, *368*(6490), 493–497.

Krieger, N., Waterman, P. D., & Chen, J. T. (2020). COVID-19 and overall mortality inequities in the surge in death rates by ZIP Code characteristics. *American Journal of Public Health*, *110*(12), e1–e3.

Kuchler, T., Russel, D., & Stroebel, J. (2022). JUE Insight: The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook. *Journal of Urban Economics*, *127*, p. 103314. Available at: https://doi.org/10.1016/j.jue.2020.103314.

Laumann, E. O., & Youm, Y. (1999). Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: A network explanation. *Sexually Transmitted Diseases*, *26*(5), 250–261.

Levy, B. L., Vachuska, K., Subramanian, S. V., & Sampson, R. J. (2022). Neighborhood socioeconomic inequality based on everyday mobility predicts COVID-19 infection in San Francisco, Seattle, and Wisconsin. *Science Advances*, *8*(7), eabl3825.

Ma, K. C., Menkir, T. F., Kissler, S., Grad, Y. H., & Lipsitch, M. (2021). Modeling the impact of racial and ethnic disparities on COVID-19 epidemic dynamics. *Elife*, *10*, e66601.

Macy, M. W. (1991). Chains of cooperation: Threshold effects in collective action. *American Sociological Review*, *56*(6), 730–747.

Maier, B. F., & Brockmann, D. (2020). Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science*, *368*(6492), 742–746.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. USA: Cambridge University Press.

Marlow, T., Makovi, K., & Abrahao, B. (2021). Neighborhood isolation during the COVID-19 pandemic. *Sociological Science*, *8*(9), 170–190.

Melegaro, A., Fava, E. D., Poletti, P., Merler, S., Nyamukapa, C., Williams, J. & Manfredi, P. (2017). Social contact structures and time use patterns in the Manical and province of Zimbabwe. *PLOS One*, *12*(1), e0170459.

Mena, G. E., Martinez, P. P., Mahmud, A. S., Marquet, P. A., Buckee, C. O., & Santillana, M. (2021). Socioeconomic status determines COVID-19 incidence and related mortality in Santiago. *Chile Science*, *372*(6545), eabg5298.

Mossong, J. B., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R. & Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine*, *5*(3), e74.

Nishi, A., Dewey, G., Endo, A., Neman, S., Iwamoto, S. K., Ni, M. Y. & Young, S. D. (2020). Network interventions for managing the COVID-19 pandemic and sustaining economy. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(48), 30285–30294.

Ogedegbe, G., Ravenell, J., Adhikari, S., Butler, M., Cook, T., Francois, F. & Horwitz, L. I. (2020). Assessment of racial/ethnic disparities in hospitalization and mortality in patients with COVID-19 in New York City. *JAMA Network Open*, *3*(12), e2026881.

Parino, F., Zino, L., Porfiri, M., & Rizzo, A. (2021). Modelling and predicting the effect of social distancing and travel restrictions on COVID-19 spreading. *Journal of the Royal Society Interface*, *18*(175), 20200875.

Park, P. S., Blumenstock, J. E., & Macy, M. W. (2018). The strength of long-range ties in population-scale social networks. *Science*, *362*(6421), 1410–1413.

Pathela, P., Crawley, A., Weiss, D., Maldin, B., Cornell, J., Purdin, J. & NYC Serosurvey Team (2021). Seroprevalence of severe acute respiratory syndrome coronavirus 2 following the largest initial epidemic wave in the United States: Findings from New York City, 13 May to 21 July 2020. *The Journal of Infectious Diseases*, *224*, 196–206.

Pei, S., Yamana, T. K., Kandula, S., Galanti, M., & Shaman, J. (2021). Burden and characteristics of COVID-19 in the United States during 2020. *Nature*, *598*(7880), 338–341.

Pitzer, V. E., Chitwood, M., Havumaki, J., Menzies, N. A., Perniciaro, S., Warren, J. L. & Cohen, T. (2021). The impact of changes in diagnostic testing practices on estimates of COVID-19 transmission in the united states. *American Journal of Epidemiology*, *190*(9), 1908–1917.

Qureshi, A. I., Baskett, W. I., Huang, W., Lobanova, I., Naqvi, S. H., & Shyu, C.-R. (2022). Reinfection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in patients undergoing serial laboratory testing. *Clinical Infectious Diseases*, *74*(2), 294–300.

Reitsma, M. B., Claypool, A. L., Vargo, J., Shete, P. B., McCorvie, R., Wheeler, W. H. & Goldhaber-Fiebert, J. D. (2021). Racial/ethnic disparities in COVID-19 exposure risk, testing, and cases at the subcounty level in California. *Health Affairs*, *40*(6), 870–878.

Romer, D., & Jamieson, K. H. (2020). Conspiracy theories as barriers to controlling the spread of COVID-19 in the U.S. *Social Science & Medicine*, *263*, 113356.

SafeGraph. (2021). *Social distancing metrics*. https://docs.safegraph.com/docs/social-distancing-metrics.

Sampson, R. J. (2019). Neighbourhood effects and beyond: Explaining the paradoxes of inequality in the changing American metropolis. *Urban Studies*, *56*(1), 3–32.

Schlosser, F., Maier, B. F., Jack, O., Hinrichs, D., Zachariae, A., & Brockmann, D. (2020). COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(52), 32883–32890.

Schläpfer, M., Dong, L., O'Keeffe, K., Santi, P., Szell, M., Salat, H. & West, G. B. (2021). The universal visitation law of human mobility. *Nature*, *593*(7860), 522–527.

Selden, T. M., & Berdahl, T. A. (2020). COVID-19 and racial/ethnic disparities in health risk, employment, and household composition. *Health Affairs*, *39*(9), 1624–1632.

Shaw, M. (2004). Housing and public health. *Annual Review of Public Health*, *25*(1), 397–418.

Sheng, J., Malani, A., Goel, A., & Botla, P. (2022). JUE Insights: Does mobility explain why slums were hit harder by COVID-19 in Mumbai, India? *Journal of Urban Economics*, *127*, p. 103357. Available at: https://doi.org/10.1016/j.jue.2021.103357.

The City of Chicago (2020). *Vaccine distribution phases*. https://www.chicago.gov/content/city/en/sites/covid19-vaccine/home/vaccine-distribution-phases.html. Accessed 19 August 2021.

The City of Chicago (2021). *COVID-19 cases, tests, and deaths by ZIP Code*, https://data.cityofchicago.org/Health-Human-Services/COVID-19-Cases-Tests-and-Deaths-by-ZIP-Code/yhhz-zm2v. Accessed 12 January 2021.

Tian, H., Liu, Y., Li, Y., Wu, C-H., Chen, B., Kraemer, M. U. G. & Dye, C. (2020). An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science*, *368*(6491), 638–642.

Tizzoni, M., Nsoesie, E. O., Gauvin, L., Karsai, M., Perra, N., & Bansal, S. (2022). Addressing the socioeconomic divide in computational modeling for infectious diseases. *Nature Communications*, *13*(11), 2897.

Torrats-Espinosa, G. (2021). Using machine learning to estimate the effect of racial segregation on COVID-19 mortality in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(7), e2015577118.

United States Census Bureau (2020). *American community survey, 2014-2018 5-year estimates*. https://www.census.gov/programs-surveys/acs.html.

van Dorn, A., Cooney, R. E., & Sabin, M. L. (2020). COVID-19 exacerbating inequalities in the US. *Lancet*, *395*(10232), 1243–1244.

Wang, Q., Phillips, N. E., Small, M. L., & Sampson, R. J. (2018). Urban mobility and neighborhood isolation in America's 50 largest cities. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(30), 7735–7740.

Wang, Z., & Tang, K. (2020). Combating COVID-19: Health equity matters. *Nature Medicine*, *26*(44), 458–458.

Wellenius, G. A., Vispute, S., Espinosa, V., Fabrikant, A., Tsai, T. C., Hennessy, J. & Gabrilovich, E. (2021). Impacts of social distancing policies on mobility and COVID-19 case growth in the US. *Nature Communications*, *12*(11), 3118.

Wells, C. R., Sah, P., Moghadas, S. M., Pandey, A., Shoukat, A., Wang, Y. & Galvani, A. P. (2020). Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(13), 7504–7509.

Wikle, N. B., Tran, T. N–A., Gentilesco, B., Leighow, S. M., Albert, E., Strong, E. R. & Boni, M. F. (2022). SARS–CoV–2 epidemic after social and economic reopening in three U.S. states reveals shifts in age structure and clinical characteristics. *Science Advances*, *8*(4), eabf9868.

Wilson, R., & Din, A. (2018). Understanding and enhancing the U.S. department of housing and urban development's ZIP Code crosswalk files. *Cityscape*, *20*(2), 277–294.

Wisconsin Department of Health Services (2022). *COVID-19 historical data by ZIP Code tabulation area v2*. https://data.dhsgis.wi.gov/datasets/wi-dhs::covid-19-historical-data-by-zip-code-tabulation-area-v2/about.

Zelner, J., Masters, N. B., Naraharisetti, R., Mojola, S. A., Chowkwanyun, M., & Malosh, R. (2022). There are no equal opportunity infectors: Epidemiological modelers must rethink our approach to inequality in infection risk. *PLOS Computational Biology*, *18*(2), e1009795.

Zhang, J., Litvinova, M., Liang, Y., Wang, Y., Wang, W., Zhao, S. & Yu, H. (2020). Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science*, *368*(6498), 1481–1486.

Zheng, Q., Jones, F. K., Leavitt, S. V., Ung, L., Labrique, A. B., Peters, D. H. & Singh, N. (2020). Hit-COVID, a global database tracking public health interventions to COVID-19. *Scientific Data*, *7*(1), 286.