UC Riverside

UC Riverside Previously Published Works

Title

Health disparity in the spread of COVID-19: Evidence from social distancing, risk of interactions, and access to testing

Permalink

Authors

Wei, Ran
Zhang, Yujia
Gao, Song
et al.

Publication Date

2023-07-01

DOI

10.1016/j.healthplace.2023.103031

Peer reviewed

Health disparity in the spread of COVID-19: Evidence from social distancing, risk of interactions, and access to testing

Ran Wei [a,*], Yujia Zhang [a], Song Gao [b], Brandon J. Brown [c], Songhua Hu [d], Bruce G. Link [a]

[a] School of Public Policy, University of California, Riverside, CA, 92521, USA
[b] GeoDS Lab, Department of Geography, University of Wisconsin, Madison, WI, 53706, USA
[c] Department of Social Medicine, Population and Public Health, University of California, Riverside, CA, USA
[d] Maryland Transportation Institute, Department of Civil and Environmental Engineering, University of Maryland, College Park, MD, 20742, USA

A B S T R A C T

Objective: – To identify and assess whether three major risk factors that due to differential access to flexible resources might help explain disparities in the spread of COVID-19 across communities with different socioeconomic status, including socioeconomic inequalities in social distancing, the potential risk of interpersonal interactions, and access to testing.
Methods: Analysis uses ZIP code level weekly COVID-19 new cases, weekly population movement flows, weekly close-contact index, and weekly COVID-19 testing sites in Southern California from March 2020 to April 2021, merged with the U.S. census data to measure ZIP code level socioeconomic status and cofounders. This study first develops the measures for social distancing, the potential risk of interactions, and access to testing. Then we employ a spatial lag regression model to quantify the contributions of those factors to weekly COVID-19 case growth.
Results: Results identify that, during the first COVID-19 wave, new case growth of the low-income group is two times higher than that of the high-income group. The COVID-19 case disparity widens to four times in the second COVID-19 wave. We also observed significant disparities in social distancing, the potential risk of interactions, and access to testing among communities with different socioeconomic status. In addition, all of them contribute to the disparities of COVID-19 incidences. Among them, the potential risk of interactions is the most important contributor, whereas testing accessibility contributes least. We also found that close-contact is a more effective measure of social distancing than population movements in examining the spread of COVID-19.
Conclusion: – This study answers critically unaddressed questions about health disparities in the spread of COVID-19 by assessing factors that might explain why the spread is different in different groups.

1. Introduction

Declared by the World Health Organization (WHO) as a pandemic on March 11, 2020, the coronavirus 2019, or COVID-19, has swept the world and raised global awareness of the threat that current and future pandemics hold for human populations (Altindiş and Ghafour, 2021; Cascella et al., 2022; Watson et al., 2022). While among unvaccinated populations everyone might be equally likely to be infected with COVID-19 from a biological perspective, substantial literature documents significant sociodemographic inequalities in the spread of the disease (Hooper et al., 2020; Clouston et al., 2021; Hu et al., 2022; Lee et al., 2022). For example, Hooper et al. (2020) reported that the rates of

COVID-19 are greatest among Latino and African American populations. Clouston et al. (2021) found that the growth of COVID-19 incidence is slower in counties with higher socioeconomic status after public health policy led to lockdowns. Lee et al. (2022) observed a 48% higher risk of infection among Hispanic people compared with white people.

There are also many studies aiming to explain what lead to such inequality. Clouston et al. (2021) apply the theory of fundamental social causes (FCT) to explain the negative relationship between the COVID-19 incidence and socioeconomic status at the county level after social distancing was implemented. FCT was developed to explain the persistence of socioeconomic and racial health disparities in disease and death across space and time, regardless of the intervening mechanisms and the

---

particular health outcome(s) in question (Link and Phelan, 1995; Phelan et al., 2010). This theory proposes that the reproduction of health inequality emerges when people use "flexible" resources such as knowledge, money, power, prestige, and beneficial social connections to strive, individually or collectively, for advantageous health circumstances. These resources can be used to avoid risks and adopt protective strategies no matter what the most prevalent diseases in a particular place or time happen to be. Based on the FCT theory, Clouston et al. (2021) suggested that the socioeconomic inequalities lead to differentiated access to flexible resources to implement social distancing, which might cause the disproportionate distribution of COVID-19 incidence across counties with different socioeconomic status.

While Clouston et al. (2021) did not explicitly measure the disparities of social distancing across different socioeconomic status groups, much work measured social distancing disparities using a variety of human mobility metrics, such as point of interest (POI) visits (Jay et al., 2020; Weill et al., 2020; Hu et al., 2022), time spent at or away from home (Borgonovi and Andrieu 2020; Jay et al., 2020; Weill et al., 2020; Hu et al., 2022), number of gathering (Lee et al., 2022), and mobility flows (Chang et al., 2021; Hou et al., 2021). These studies all consistently show significant disparities in social distancing across different socioeconomic status groups. However, it remains unclear how disparities in social distancing contribute to COVID-19 incidence during different stages of the pandemic. Hu et al. (2022) even found that counties with better social distancing implementations, in terms of reducing their movement and increasing their staying home percentage, do not always have lower infection rates, indicating that disparities in social distancing cannot fully explain the disparities in COVID-19. Then the question is what other disparities resulting from different access to flexible resources are shaping the health disparity in the spread of COVID-19 during different stages of the pandemic?

In this paper, we fill this research gap by identifying and assessing three factors that due to differential access to flexible resources might help explain disparities in the spread of COVID-19 across communities with different socioeconomic status. The first are factors related to the implementation of social distancing guidelines which were only incompletely adhered to perhaps in part because of social circumstances limiting the capacity of some people to do so. As mentioned earlier, many different measures of social distancing have been used in the literature. Noi et al. (2022) analyzed and compared twenty-six mobility and contact-related indices across nine various sources and suggested that any single measure might not describe all aspects of mobility. As a result, we propose to use two distinct social distancing measures to more accurately capture social distancing and COVID-19 transmission dynamics among individuals. The first measure is the mobility flow matrix that summarizes the amount of population movement between any two communities within a certain time interval, e.g., daily or weekly. It has a strong potential for tracking temporal changes in COVID-19 transmissions risk, providing in-depth insights into how movement patterns affect health disparities in COVID-19 incidence (Kang et al., 2020b; Schlosser et al., 2020; Tan et al., 2021). The second measure is the close-contact index, a measure that assesses the number of close-contact interactions using anonymized mobile phone device location data. We approximate the close-contact index by calculating how many devices come within 5 m of the target device within a 5-min window based on their GPS locations (Ye and Gao 2022). Crawford et al. (2022) created the 2-m close-contact index to model Connecticut's town-level COVID-19 transmission dynamics. They suggested that the close-contact index resulted in a better model fit than other mobility metrics. However, no research to date has analyzed whether there are socioeconomic inequalities in close-rate contact.

The second risk factor we will measure is the potential risk of interpersonal interactions. The various measures of social distancing quantify the likelihood of interpersonal interactions but do not fully capture the potential risk of such interactions. When COVID-19 first emerged on January 2020, communities with high socioeconomic status

(SES) had the highest number of COVID-19 cases because their international travels had high risk of getting infected (Clouston et al., 2021). Then when COVID-19 spread across communities the risk shifted. High SES groups secluded themselves whereas lower SES groups work and residential circumstances enforced a greater degree of interpersonal contact (Shin et al., 2021; Fiske et al., 2022). We propose two measures to approximate the potential risk of interpersonal interactions. One is the number of COVID-19 cases in the community two weeks prior to the current assessment an indicator that assessed the potential risk of interactions with infected people in the community. The other is the number of COVID-19 cases in neighboring communities in the two weeks prior to the current assessment, which can signal the risk of interactions occurring with neighboring communities.

Third we evaluate inequalities in communities' access to COVID-19 testing across different socioeconomic status groups to explore whether they are contributing to disparities in the spread of COVID-19. Early testing is a crucial protective factor to prevent COVID-19 spread. However, states across the United States experienced insufficient testing access in the early stages of the pandemic and uneven geographic distribution of testing sites even after test volume increased (Rader et al., 2020; Schmitt-Grohé et al., 2020; Tao et al., 2020). Existing studies mainly focused on the first two or three months of COVID-19 (Tao et al., 2020; Kang et al., 2020a; Schmitt-Grohé et al., 2020), whereas the time-varying effect of testing accessibility in different stages of the pandemic, among different socioeconomic status, in particular, is largely unknown.

Specifically, we will investigate the dynamics of health disparities in the spread of COVID-19 at the ZIP code level in Southern California from April 2020 to April 2021. The following research questions guide this effort: 1) Do disparities in the spread of COVID-19 appear and evolve across communities with different socioeconomic statuses during different stages of pandemic, 2) Are there inequalities in social distancing, the potential risk of interpersonal interactions, and access to testing across communities with different socioeconomic status, and if so how do such disparities evolve during different stages of pandemic? 3) How do any inequalities in these risk and protective factors contribute to the production of health disparity in the spread of COVID-19 during different stages of the COVID-19 pandemic?

2. Study area and data

Our study area focuses on six counties of Southern California, including Los Angeles County, Orange County, San Diego County, Riverside County, Imperial County, and San Bernardino County, including 556 ZIP code areas and a total population of 21.23 million in the year 2020 (Fig. 1). This study area is selected not only because of data availability but also due to its socioeconomic and geographic diversity of the population. The average population of the 556 ZIP code areas is 34,256, with a standard deviation of 22,855.

To analyze the disparity dynamics in COVID-19 incidence and related disparities from 2020 to 2021, we obtained the daily COVID-19 confirmed cases at the ZIP code level and the locations of testing sites between March 23, 2020, and April 5, 2021, from the California Department of Public Health (CDPH). The first specialized COVID-19 testing sites were opened on March 26, 2020. As of April 11, 2021, 1933 active testing sites were identified in Southern California, which is 15 times the number that were open in April 2020 (126 sites). Among them, 71% of the sites required appointments before testing, further restricting accessibility. To estimate testing capacity, we collected the weekly county-level COVID-19 testing data from the CDPH. In addition, we acquired the latest 2020 American Community Survey (ACS) 5-year estimates from U.S. Census to analyze variations in socioeconomic status at the ZIP code level.
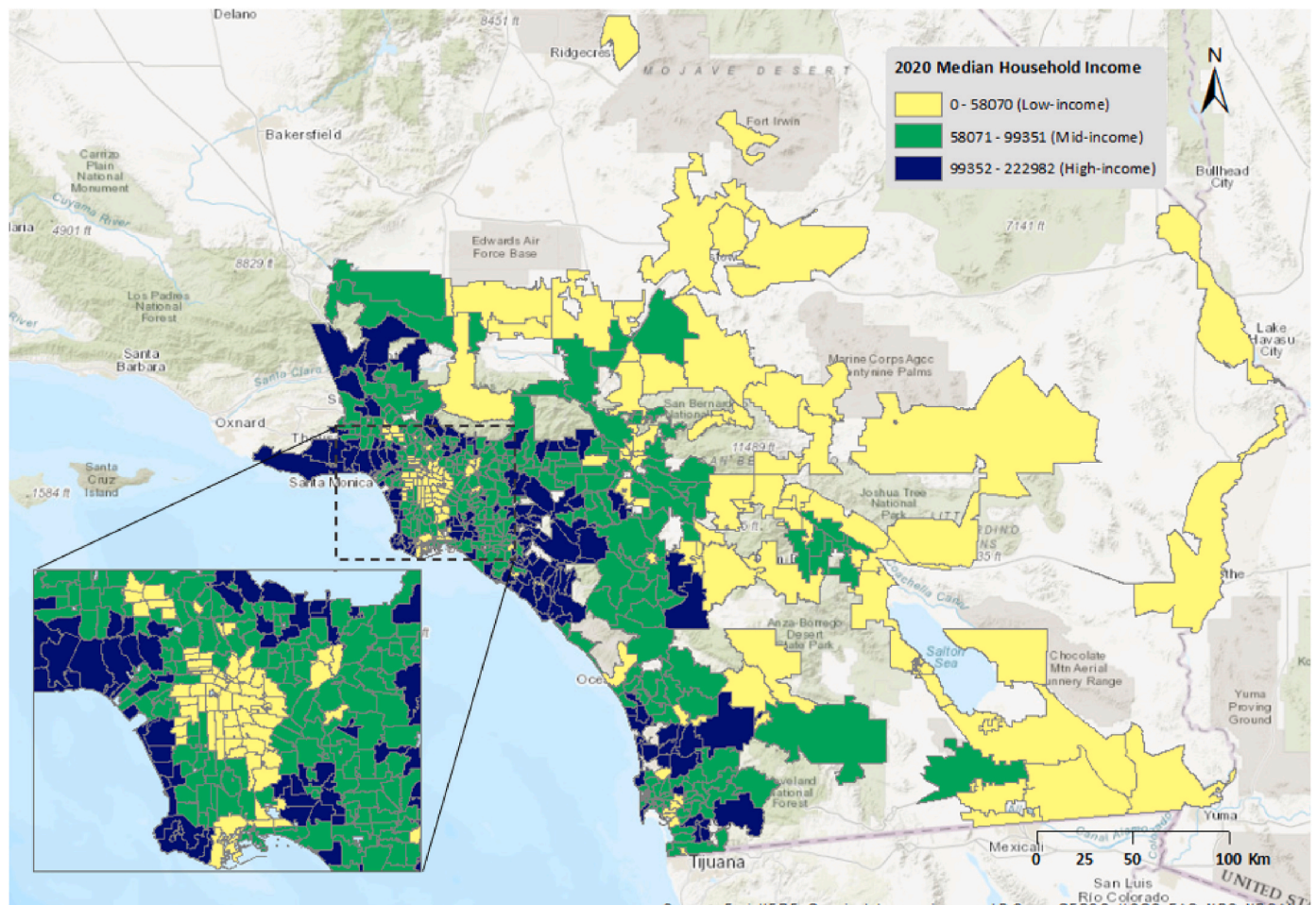
Fig. 1. Study area ZIP codes and median household income in 2020.

3. Method

In this section, we first present the methods used to measure the disparities in COVID-19 incidences, social distancing, potential risk of interactions, and testing accessibility across communities with different socioeconomic status. Then we develop a spatial regression model to evaluate the contributions of the socioeconomic inequalities in social distancing, the potential risk of interactions, and testing accessibility to the production of health disparities in the spread of COVID-19 during different stages of the COVID-19 pandemic.

3.1. Measure disparities in COVID-19 incidence

To evaluate how health disparities in the spread of COVID-19 emerged and evolved across communities with different socioeconomic status during different stages of pandemic, we classified the 556 ZIP codes into three groups (high, medium, and low-income groups) based on their median household income, and assess how the weekly COVID-19 incidences vary across high and low-income groups at each week from March 23, 2020, to April 5, 2021. As shown in Fig. 1, the 139 ZIP codes within the bottom 25% quantile median household income are considered as the low-income group. Their corresponding median household income varies from 0 to $58,070. Similarly, the 139 ZIP codes within the top 25% quantile of median household income was considered as high-income group, with median household income varying from $99,352 to $222,982. We summarize the weekly COVID-19 incidences by adding up the reported daily COVID-19 cases for each day of the week. In total, 54 weekly COVID-19 incidence are analyzed.

3.2. Socioeconomic inequalities in social distancing

As mentioned earlier, we used two measures to track social distancing behavior using the mobility flow matrix and the close-contact rate. Based on SafeGraph data that include millions of anonymous mobile phone users 'visits to various places, Kang et al. (2020b) presented a method to generate a daily and weekly multiscale human mobility flow dataset in the U.S. Following this method, we created a mobility flow matrix for each week between March 23, 2020, and April 5, 2021 to estimate the amount of population movement between any two ZIP codes within a given week. With the mobility flow matrix we can calculate the total number of population inflows to each ZIP code as a measure of interactions with other communities. Mathematically, if we use F_{ji}^t as the number of population movements from ZIP code j to i at week t, the total number of population inflows to ZIP code i at week t, denoted by TF_i^t, can be represented as:

$$TF_i^t = \sum_j F_{ji}^t \tag{1}$$

In addition to the mobility flow matrix, we also employed a large-scale anonymized mobile phone device location panel data that we acquired from UberMedia that covers 70% of the U.S. population. The proxy of close contacts of each mobile device is measured by calculating how many devices come within 5 m of the target device within a 5-min window based on their GPS locations (Ye and Gao 2022). In this way, a close-contact index (S_i^t) is created for each ZIP code at each week between March 23, 2020, and April 5, 2021, to approximate the average number of close-contacts per person at the ZIP code. It is important to

note that both mobility measures are created based on a sample of mobile device data, and then population weighting is used to approximate the measure for the entire population at ZIP code.

3.3. Socioeconomic inequalities in potential risk of interactions

As discussed previously, we use two measures to approximate the potential risk of interpersonal interactions. One is the number of COVID-19 cases at the community two weeks ago (C_i^{t-2}) to indicate the potential risk of interactions that are occurring inside of the community. A 14-day or 2-week time lag is applied to capture the incubation period of COVID-19. Ahouz and Golabpour (2021) examined intervals between 1 and 14 days and found that 14 days is the optimal time period to predict the incidence of COVID-19.

We also develop a measure to estimate the potential risk of interactions occurring among neighboring communities. Specifically, let C_i^t represent the number of COVID-19 cases at ZIP code i at week t, then the potential risk of interactions with its neighboring communities at week t, denoted by R_i^t, can be approximated by the following equation:

$$R_i^t = \sum_j C_j^{t-2} * \frac{F_{ji}^{t-2}}{TF_i^{t-2}} \tag{2}$$

The R_i^t is essentially a weighted sum of the number of COVID-19 cases at neighboring communities two weeks ago. The weight used here is the proportion of the population movements from j to i in the total movements into i (TF_i^{t-2}). By using such a weighted sum we can take into account both the number of cases at neighboring communities and the number of interactions between i and its neighboring communities. The potential risk of interaction with neighboring communities will be high when there are a large number of COVID-19 cases in its neighboring communities and a large number of population movements between them two weeks ago. Such risk of interaction will be minimal if there are very few COVID-19 incidences in its neighboring communities and very few population movements between them two weeks ago.

3.4. Socioeconomic inequalities in testing accessibility

We adopted the enhanced two-step floating catchment method (2SFCA) to assess testing accessibility. First proposed by Radke and Mu (2000) and later enhanced by Luo and Wang (2003) and Luo and Qi (2009), it is one of the most widely used methods to measure access to health care providers. The index takes into account geographic proximity and is essentially a form of physician-to-population ratio, which is easy to interpret. Following the 2SFCA, the community's access to COVID-19 testing sites will be computed using the following two steps:

Step 1: For each testing site k, we search all ZIP codes that are within its 30-min drive time (this is the service area of testing site k). 30-min drive time is a commonly used rational threshold in health care-related accessibility studies (Wang and Luo, 2005; Tao et al., 2020). Then we compute the testing-to-population ratio, V_k, within the catchment area:

$$V_k^t = \frac{T_k^t}{\sum\limits_{i \in \{d_{ik} < 30\}} P_i} \tag{3}$$

Where T_k^t is the testing capacity of testing site k at week t, estimated using the county-level tests count (the weekly number of tests in a county divided by its number of active testing sites). P_i is the population at ZIP code i whose centroid falls within service area of testing site k ($d_{ik} < 30$), d_{ik} is the travel time between ZIP code i and testing site k. The testing-to-population ratio V_k^t is essentially a ratio of facility capacity to demand, integrating both the demand (P_i) and the capacity at testing sites (T_k^t).

Step 2: For each ZIP code i, we search all testing sites k that are within

its 30-min drive time (that is the catchment area of ZIP code i), and sum up the testing-to-population ratios V_k^t (derived in step 1), at these locations as the testing accessibility (A_i^t) of ZIP code i at week t as follows:

$$A_i^t = \sum_{k \in \{d_{ik} < 30\}} V_k^t \tag{4}$$

In this way the testing accessibility A_i^t considers the access to all testing sites within 30-min driving time. It is important to note that A_i^t is a summation of testing-to-population ratios so there is no bounding value for it. It is a relative measure. A higher value of A_i^t indicates better access to testing.

3.5. Analyzing contributions of socioeconomic inequalities in risk and protective factors to disparities in COVID-19 incidence

To assess the contributions of the risk and protective factors described above to disparities in the spread of COVID-19 during different stages of the pandemic, we constructed a cross-sectional spatial lag regression model for each week between March 23, 2020, and April 5, 2021. The spatial lag regression model is adopted because it can capture the neighboring effect of COVID-19 transmission and address the issue of spatial autocorrelation (Anselin, 1988). Following previous notation, for each week t a cross-sectional spatial lag regression model is constructed as follows:

$$C^t = \rho^t W^t C^t + \beta^t X^t + \gamma^t Z + u^t \tag{5}$$

The dependent variable is the number of COVID-19 cases at each ZIP code at week t (C^t). The number of COVID-19 cases is used as the dependent variable rather than incidence rates (number of cases divided by the population) because the mobility measures quantify the possible number of interpersonal interactions and are essentially associated with case growth instead of rate growth, as demonstrated in previous literature (Xiong et al., 2020; Weill et al., 2020; Hu et al., 2021). The independent variables include spatial autoregressive term ($W^t C^t$) time-varying variables (X^t), and non-time-varying independent variables (Z). The X^t consist of the measures of social distancing that include the number of total population inflows into each ZIP code two weeks ago (TF^{t-2}) and the close-contact index at each ZIP code two weeks ago (S^{t-2}), the measures of potential risk of interactions that include the number of cases two weeks ago (C^{t-2}) and the weighted sum of the number of COVID-19 cases at neighboring communities two weeks ago (R^t), and the testing accessibility measure two weeks ago (A^{t-2}). The non-time-varying independent variables (Z) include population at each ZIP code to control the impacts of population on the COVID-19 cases and dummy variables indicating the income groups. Logarithmic transformation is used for these variables to mitigate the nonnormality issues (Xiong et al., 2020). For each week t we construct a spatial regression model and estimate the coefficients (β^t) of the time-varying variables (X^t) to assess the potential impacts of these factors to COVID-19 incidence. Then we rely upon how β^t vary at each week t to reflect how the impacts of these variables to COVID-19 incidence might change across different stages of pandemic. The spatial regression modeling was performed using the latest open-source Python Spatial Analysis Library (PySAL).

4. Results

4.1. Socioeconomic inequalities in the incidence of COVID-19

Fig. 2 illustrated the trend of the weekly new COVID-19 cases, in which the center line indicates the mean of new cases counts among all the ZIP codes. The shaded areas represent two standard deviations from the mean. For the COVID-19 case, two major peaks can be observed. The first peak appeared from June to July 2020, followed by a more substantial second surge from November 2020 to January 2021 (Fig. 2).
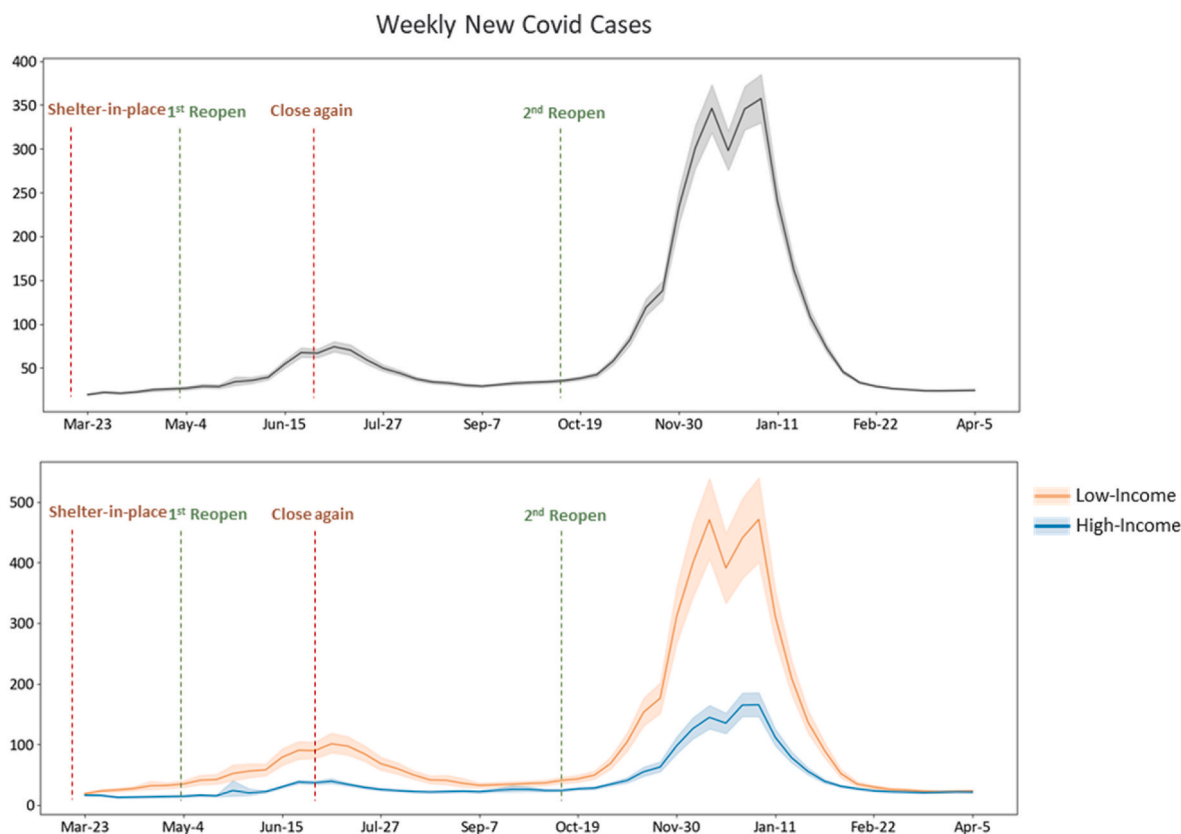
Fig. 2. COVID-19 cases (C_i^t) trend, overall (above) and high-vs. low-income groups (below).

Governor Gavin Newsom was among the first state governors to declare a state of emergency on March 4, 2020, and issued the statewide shelter-in-place order on March 19 (Friedson et al., 2021). A direct response to this was a slow case growth between March and April 2020, with less than 50 new cases weekly per ZIP code. In May, California followed a gradual reopening plan to resume economic activities. With this reopening, however, the number of confirmed cases increased notably through mid-June, resulting in a delay of the state's reopen plan and enforcing another governor's order on June 27 to close all businesses again, in the midst of the first wave. The new case counts continued to rise and reached the first peak in July, approaching 100 cases per ZIP code. Then the number of new cases started to decline in mid-July and remained stable until the advent of the second wave. California's second attempt at reopening started in mid-October 2020, which was then followed by a second outbreak in early November, when COVID-19 case counts reached unprecedented levels that dwarfed the summer wave in July, soaring from 100 to nearly 400 new cases per ZIP code. After January 2021, the growing pattern of new cases started to level off again and returned to the 2020 pre-outbreak levels from February 2021.

According to the 2020 Census, high- and low-income groups share a similar population size, representing 21% and 24% of the total population, respectively. However, striking differences in COVID-19 infections between the high- and low-income groups were observed in both magnitude and variations (Fig. 2). The COVID-19 case growth of the low-income group protruded almost two times higher than that of high-income groups before November 2020, indicating a faster virus spread in low-income communities. The inter-(income) group gap widened to four times in the second wave, which peaked in December 2020. Meanwhile, considerable variations in COVID-19 cases were noted among low-income groups as opposed to relatively minor intra-group discrepancies among high-income communities. We also computed the COVID-19 incidences per 10,000 population at each zip code throughout the pandemic periods and the results are presented in

Appendix Fig. A1. As it shows, similar disparities are observed in the incidence rate as well.

4.2. Socioeconomic inequalities in social distancing

Fig. 3 revealed the trend of population inflow among the 556 ZIP codes in Southern California. Likewise, the center line was the average population inflow of all the ZIP codes. The shaded areas indicated two standard deviations from the mean. Overall, the population movements experienced a slight drop in late March 2020 and reached the lowest point in early April in response to California's "shelter-in-place" order enforcement. Despite minor fluctuations, population movements grew gradually with the statewide reopening and remained stable from mid-June to late November 2020. The Thanksgiving holiday in late November 2020 was associated with a sharp rise in inter-community travel. In contrast to the declining trend of COVID-19 infection cases, the population movements kept growing moderately during the post-holiday periods. The comparison of high- and low-income groups revealed different dynamics. In the initial "stay-at-home" phase (from late March to early May 2020), the population movements of the high-income group reduced more than the low-income group. However, beginning with the reopening in May, the high-income group reversed the trend and had more movements than the low-income group, despite a roughly similar inter-group movement pattern. Both groups displayed similar intra-group variations, which largely overlapped in a consistent way, as illustrated in Fig. 3.

The close-contact index illustrated a different aspect of social distancing, compared to population inflow, especially after July 2020 (Fig. 4). The close-contact rate maintained a low level from March to April 2020. It started to increase in May 2020 during the reopening and rapidly formed a spike in mid-July 2020 (from 60 to 100 close contact per person). A sharp drop occurred in late July 2020 after enforcing the "shut-down-all-business" mandate and the number of close contacts per
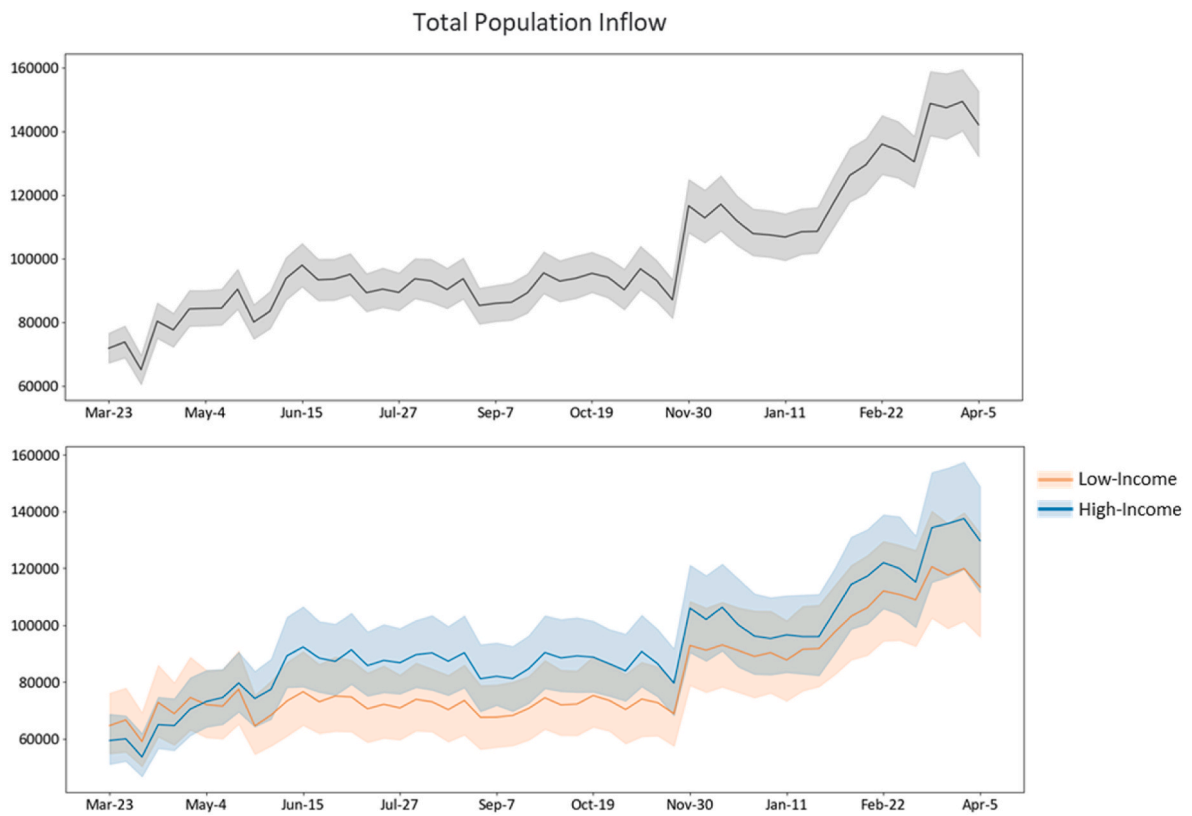
Total Population Inflow

Fig. 3. Population inflow (TF_i^t), overall (above) and high-vs. low-income groups (below).
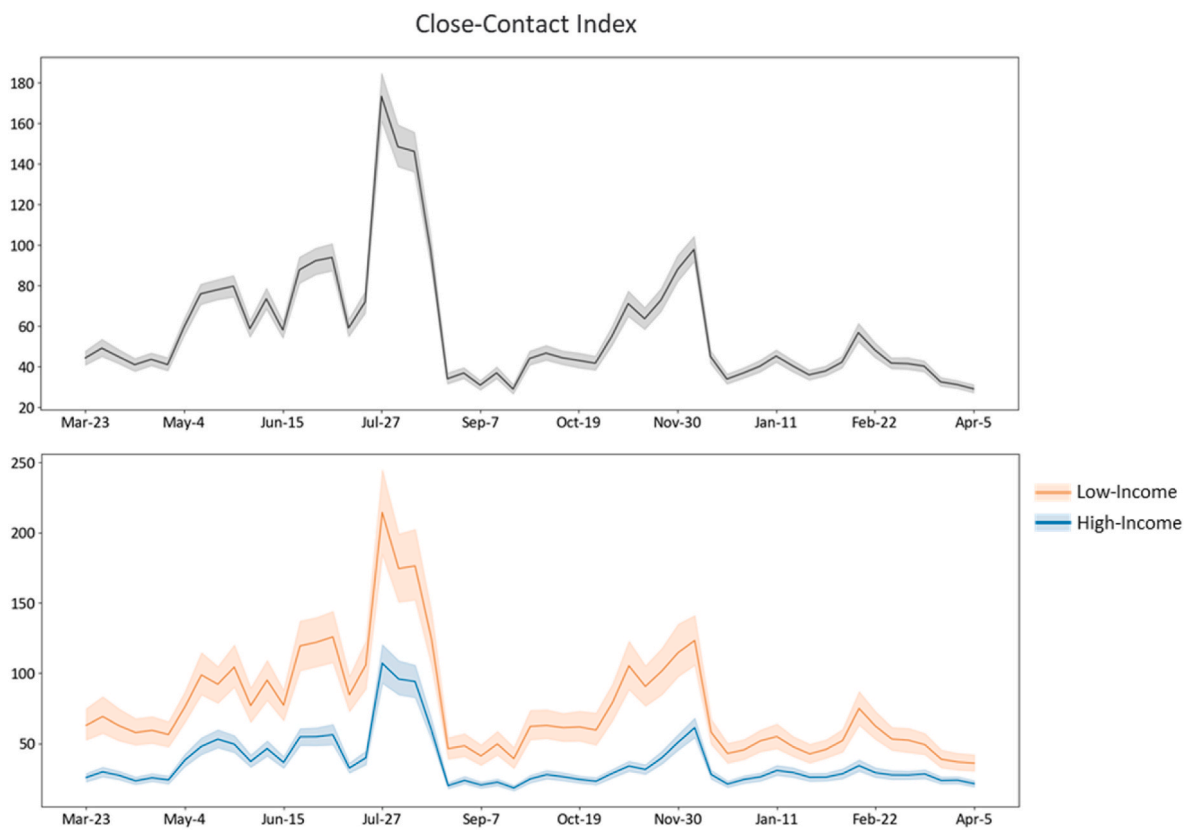
Close-Contact Index

Fig. 4. Close-contact index (S^t) trend, overall (above) and high-vs. low-income groups (below).

person touched the bottom of the below-40 level in August. After the 2020 summer "extreme event," the close-contact index, despite several fluctuations, remained at relatively low levels, below 100 per person. It is important to note that the low-income group maintained a consistently higher close-contact index-two times higher than the high-income group.

4.3. Socioeconomic inequalities in the potential risk of interactions

Fig. 5 revealed the trend of the neighboring cases weighted by population inflow, which allows us to better capture the inter-community COVID-19 transmission risk. Derived from the neighboring cases, its overall temporal trend is very similar to that of the COVID-19 cases but with higher magnitude and less variation. Overall, the weighted neighboring cases of low-income ZIP codes were consistently higher than those of the high-income, reflecting higher inter-community COVID-19 transmission risk of the low-income areas, even though the gap between the high and low-income groups was not as large as its counterpart in Fig. 2, especially during second COVID-19 wave.

4.4. Socioeconomic inequalities in testing accessibility

As described in section 3.4, the testing accessibility measure is a summation of testing-to-population ratios of all testing sites that are within 30-min drive time of a ZIP code, so it is a relative measure without bounding values. A higher value of the measure indicates better access to testing. Fig. 6 shows that the overall testing accessibility varies from 0.9 to 1.3 and exhibited the highest variations between May and July 2020. Testing accessibility of the low-income group was slightly higher than the high-income group. Specifically, the testing accessibility of the low-income group, which started to increase one week ahead of the high-income group in April 2020, gained a continuous rise initially and remained above 1.25 after July 2020. In contrast, the high-income

group's testing accessibility stayed slightly below 1. However, there are much more significant variations in testing accessibility among the low-income groups than the high-income groups, as suggested by the shaded areas that represent two standard deviations from the mean.

4.5. Contributions to disparities in COVID-19 incidence

The initial correlation analysis shows that the close-contact index and population inflow are strongly correlated with the Pearson coefficient of 0.75. In addition, they were both strongly correlated to the total population, with a correlation coefficient of 0.61 and 0.4, respectively. As mentioned earlier, this is due to the fact that both mobility measures are created based on a sample of mobile device data and population weighting is used to approximate the measure for the entire population at ZIP code. Given the strong correlation between these two mobility measures, two sets of spatial lag regression models are constructed for each week between April 6, 2020 and April 5, 2021, one including close-contact index and other independent variables listed above, and the other including population inflow and other independent variables listed above. The results show that close-contact models have overall better model fits. As a result we report the results of close-contact models here but include the coefficients of population inflow in the population inflow regression models for comparative purpose. The weekly Pseudo R^2 for close-contact models range from 0.34 to 0.97, with an average value of 0.7, as shown in Fig. 7, indicating that the independent variables provide a good explanation of the variations in COVID-19 incidences. Pseudo R^2 is the squared correlation between the observed and predicted values for the dependent variable. Please note that it's not a true R^2 since it does not correspond to the share of the variance explained by the model. The trend of the Pseudo R^2 corresponded with the new COVID-19 incidence, with better model fits in the two COVID-19 spikes than in the low tides.

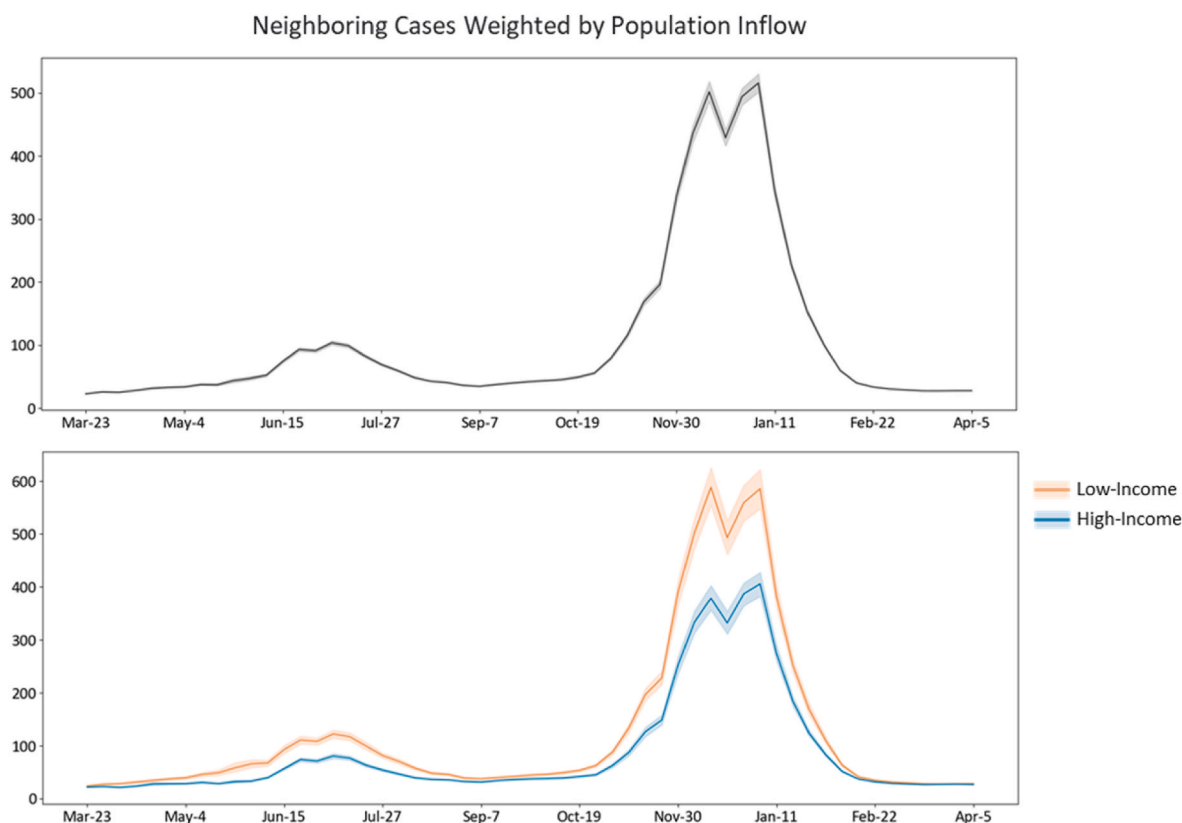Most coefficients of the spatial lag model were significant at the 0.05

Fig. 5. Neighboring cases weighted by population inflow (R_i^t), overall (above) and high-vs. low-income groups (below).
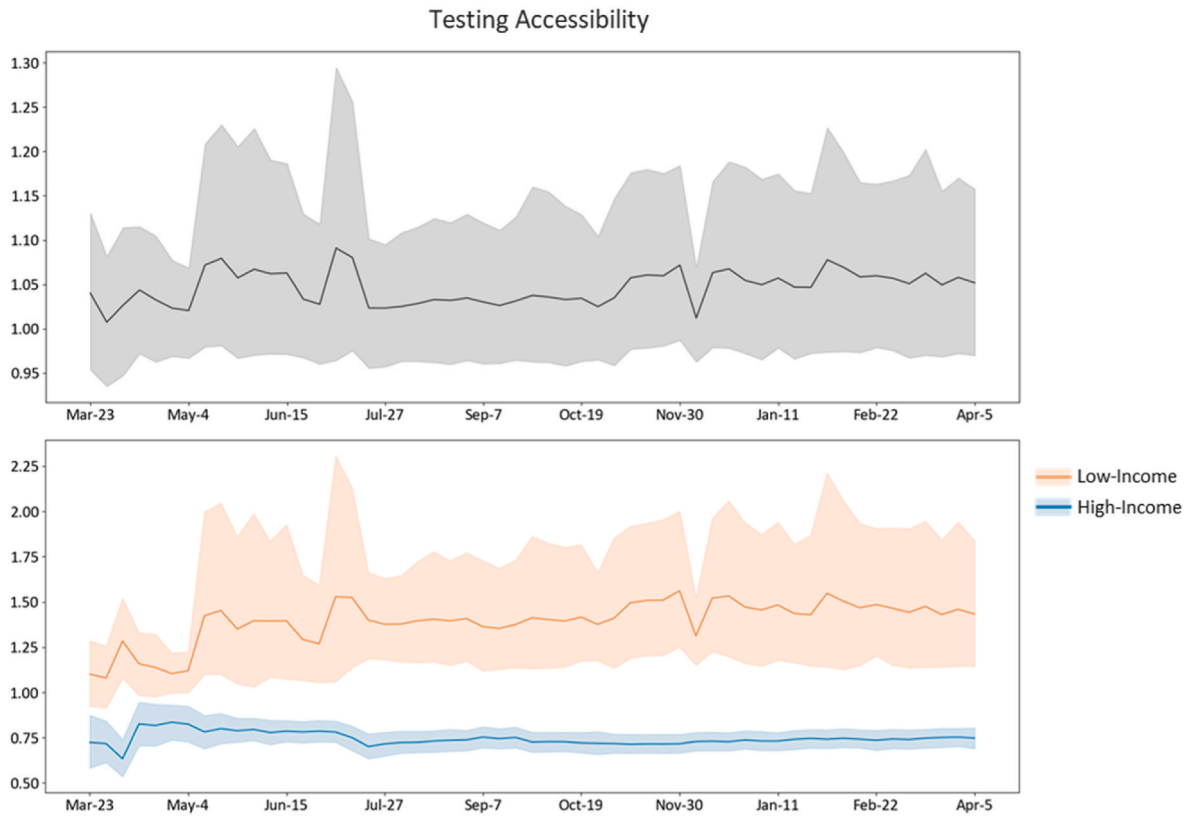
Testing Accessibility

Fig. 6. Testing accessibility (A_i^t) trend, overall (above) and high-vs. low-income groups (below).

Fig. 7. Weekly Pseudo R^2 of spatial-lag models.

level. Fig. 8 shows significant coefficients of the 2-weeks-ago close-contact index and population inflow, where the dot markers indicate the coefficient estimates and error-bars represent their corresponding standard errors. As the results shows, both showed consistent positive associations with COVID-19 case changes. Both positive notwithstanding, the close-contact index had a more substantial effect on the COVID-19 case changes than population flow. The close-contact index is significant over most of the time. Its coefficients range from 0.03 to 0.18 with an average value of 0.11, indicating that a 10% increase in the close-contact index could lead to a 1.05% $((1.1^{0.11} - 1) * 100\%)$ increase in the new case count two weeks later. Instead, the population inflow has a weaker effect with its coefficients ranging from 0.02 to 0.09 averaged at a value of 0.05, suggesting that a 10% increase in the close-contact index could lead to a 0.48% $((1.1^{0.05} - 1) * 100\%)$ increase in the new case count two weeks later.

Fig. 9 shows significant coefficients of 2-weeks-ago local cases and 2-weeks-ago weighted neighboring cases. From April 2020 to April 2021, these two variables had the most potent positive effects on the COVID-19 spread. Both positive notwithstanding, the 2-weeks-ago local case has a more substantial effect on the COVID-19 case changes than 2-weeks-ago weighted neighboring cases, with corresponding average coefficients of 0.55 and 0.27, respectively, while controlling for each other and other covariates. On average, a 10% increase in 2-weeks-ago local cases and 2-weeks-ago weighted neighboring cases could lead to a 5% and 3% rise in new case count two weeks later, respectively.

In the first two weeks of April 2020, the weighted neighboring cases initially showed a stronger effect on case growth than 2-weeks-ago local cases. During this time, COVID-19 incidences at a ZIP code were dominated by the neighboring cases, where a 10% rise in weighted neighboring cases could trigger a 4%–7% increase in new cases two
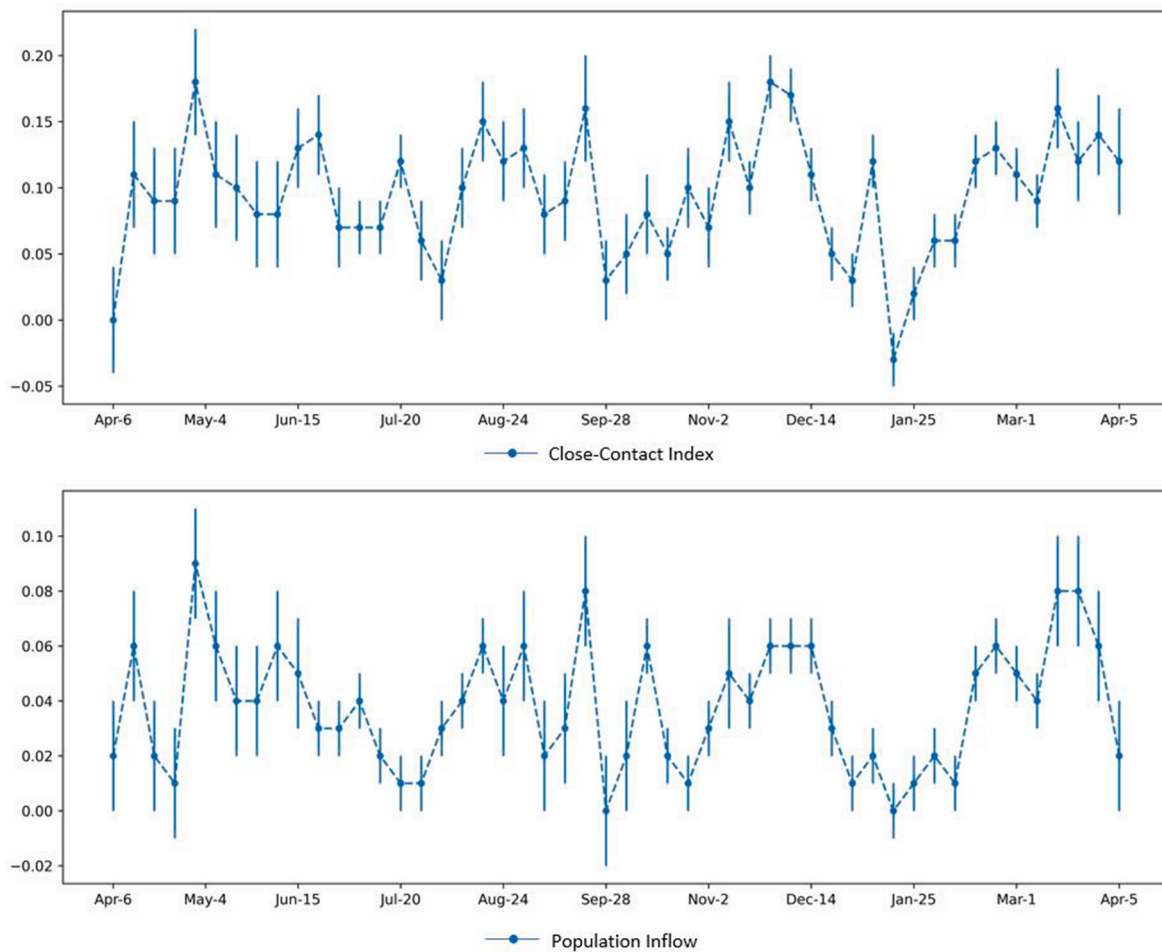
Fig. 8. Spatial-lag model coefficients, 2-week-ago close-contact rate (S^{t-2}) and population inflow (TF^{t-2}).
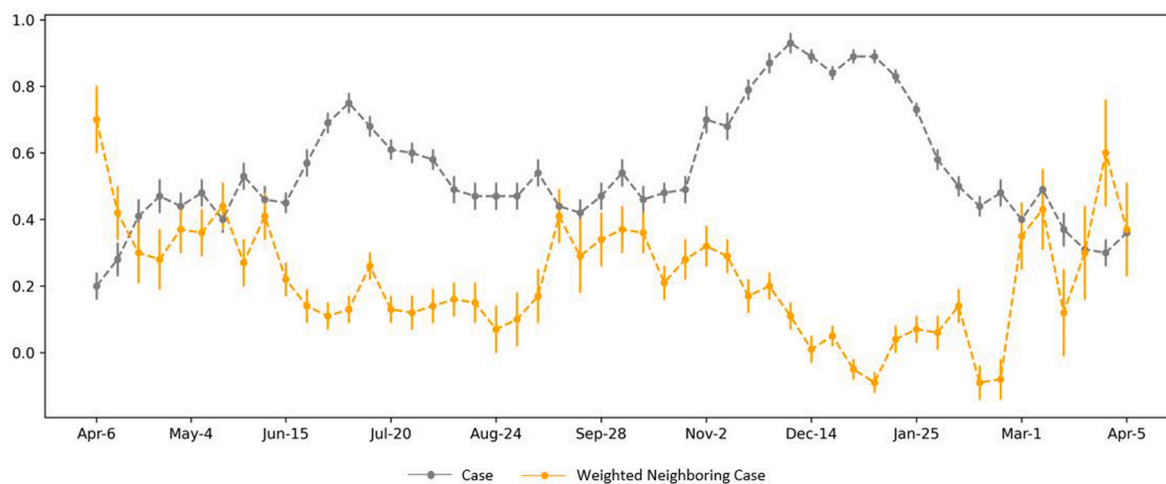

Fig. 9. Spatial-lag model coefficients, 2-week-ago local cases (C^{t-2}) and 2-week-ago weighted neighboring cases (R^t).

weeks later. After two weeks, the local case became stronger than neighboring cases, and kept growing during the first COVID-19 spike, leading to a 7% increase in new cases two weeks later as the result of a 10% growth in local cases.

Meanwhile, weighted neighboring cases diminished and became insignificant after mid-August 2020. At this stage, however, the cumulative growth of local cases dominated the virus spread, whereas the effects of the neighboring cases became negligible. After the first COVID-

19 spike, the effects of local cases gradually declined but raised again during the second spike, when a 10% increase in local cases could lead to a 9% increase in new cases two weeks later. These two variables exhibited similar alternating trends to the second COVID-19 wave. Unlike in the first wave, however, the coefficients of weighted neighboring cases were consistently lower than local cases in the second wave, suggesting that local cases are more dominant than neighboring cases as a factor affecting the formation of the second COVID-19 wave.

Unlike the social distancing and potential risk of interaction measures, the coefficients of testing accessibility bounced between positive and negative throughout the study period and it was only significant over 28% of the time.

5. Discussion

In this paper, we analyze three potential social inequalities in risk and protective factors and use them to help explain the disparities in the spread of COVID-19 across communities with different socioeconomic status during various stages of the pandemic. There are several major findings that worth further discussion.

First, we identified striking differences in COVID-19 infections between the high- and low-income groups in both magnitude and variations. During the first COVID-19 wave, weekly new cases from the low-income group were two times the size of the high-income. In the second wave, this gap widened to four times. This suggests the powerful impacts of socioeconomic status on health outcomes.

Second, when measuring the inequalities in the implementation of social distancing, we found that people from high-income areas sharply reduced and remained substantially below their low-income area counterparts with respect to close contact during every stage of the pandemic, even though there are higher population movements in the high-income areas than the low-income areas after May 2020.

Third, inequalities in social distancing, the potential risk of interactions, and testing accessibility go a long way in explaining variations in COVID-19 incidence. Specifically, the potential risk of interactions contributes most to the spread of CVID-19, whereas testing accessibility has the least impacts. While both 2-weeks-prior local case and 2-weeks-prior weighted neighboring cases have significant effects on COVID-19 incidences, their strength exhibits alternating rises and falls during the two COVID-19 waves. The neighboring cases dominate in the early stage of community spread, but local cases become dominant during the peak season. The two-week time lag could allow early warning of an incoming peak. In terms of the two social distancing measures, the close-contact rate is more impactful than the population inflows. This helps explain why people in high-income areas have lower infections even though their population movements are higher than those in low-income areas. People in high-income areas avoid the more impactful factor – close contact, and did not limit themselves by avoiding the one we found less impactful – mobility. The testing accessibility was negatively associated with case changes for most of the time, illustrating that 2-weeks-prior testing could effectively identify infected populations and trigger behavior changes like (self-) quarantines, isolation, or mask-wearing (Li et al., 2020; Skoll et al., 2020). However, testing accessibility had not become significant until May 2020, indicating the inadequate supply at the beginning of the COVID-19 spread (Schmitt-Grohé et al., 2020; Tao et al., 2020).

There are also several limitations in our research that call for further research. First, our analysis only includes the data from March 23, 2020 to April 5, 2021. During this period, the vaccines are not widely available so that the impacts of vaccines might be minimal. However vaccination data and further analysis are needed to account for the impacts of vaccines. Second, the number of COVID-19 cases is used as the dependent variable in our spatial regression models rather than incidence rates (number of cases divided by the population) because the mobility measures quantify the possible number of interpersonal interactions and are essentially associated with case growth instead of rate growth, as demonstrated in previous literature (Xiong et al., 2020; Weill et al., 2020; Hu et al., 2021). While we include population as a controlling variable in our regression model to mitigate the impacts of population on the number of cases, the strong correlations between mobility measures and population also undermine such efforts. Third, information on the weekly testing site's capacity was derived from the county-level COVID-19 Cases Tests data from the California Department of Public Health (CDPH), which did not capture the heterogeneity of testing accessibility at the ZIP code level. Additionally, we could not track the accessibility to testing sites through public transit as our accessibility index was limited to car access only. This might not cause significant inaccuracies because public transit was not available in many counties due to pandemic during 2020 and driving is the major community method in southern California. We also use the ZIP code centroid as the proxy when measuring the travel time between ZIP codes and testing sites. This could also cause inaccuracies in the accessibility measure. Third, confirmed cases and positive test data alone might not accurately reflect the severity of income-related health disparities. Studies have demonstrated that the addition of COVID-19-induced death data can overcome the well-documented inconsistencies in reported cases and tests. Thus, better illuminating the inequality embedded among the income-divided communities coping with the COVID-19 infection (e.g., Chen and Krieger, 2021). Despite these limitations, this research contributes significantly to understanding health disparities in the spread of COVID-19 because of its testing of different possible mechanisms through which such disparities may have arisen.

6. Conclusion

Since the COVID-19 outbreak in the U.S., California has one of the highest number of COVID-19 cases. We set out to answer critically unaddressed questions about health disparities in the spread of COVID-19 in Southern California and the reasons for these spread patterns. Our study identified the emergence and persistence of health disparities in the spread of COVID-19 across communities with different socioeconomic status from April 2020 to April 2021. To better understand what contributes to such health disparities, we further examined three major socioeconomic inequalities that could plausibly result from differential access to flexible resources, including social distancing, disparities in potential risk of interactions, and disparities in testing accessibility. We observed significant disparities in all of them among communities with different socioeconomic status. In addition, all of them contribute to the disparities of COVID-19 incidence. Among them the potential risk of interactions is the most important contributor, whereas testing accessibility contributes least. We also found that close-contact is a much more important measure of social distancing than population movements in examining the spread of COVID-19.

Data availability

The authors do not have permission to share data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.healthplace.2023.103031.

References

Altindiş, M., Ghafour, H., 2021. Is blood type linked to COVID19 risk? Sakarya Tıp Dergisi 11 (1), 207–217.

Anselin, L., 1988. Spatial Econometrics: Methods and Models, vol. 4. Springer Science & Business Media.

Borgonovi, F., Andrieu, E., 2020. Bowling together by bowling alone: social capital and Covid-19. Soc. Sci. Med. 265, 113501.

Cascella, M., Rajnik, M., Aleem, A., Dulebohn, S.C., Di Napoli, R., 2022. Features, Evaluation, and Treatment of Coronavirus (COVID-19) *(Statpearls [internet])*.

Chang, M.C., Kahn, R., Li, Y.A., Lee, C.S., Buckee, C.O., Chang, H.H., 2021. Variation in human mobility and its impact on the risk of future COVID-19 outbreaks in Taiwan. BMC Publ. Health 21 (1), 1–10.

Chen, J.T., Krieger, N., 2021. Revealing the unequal burden of COVID-19 by income, race/ethnicity, and household crowding: U.S. county versus ZIP code analyses. J. Publ. Health Manag. Pract. 27 (1), S43–S56.

Clouston, S.A., Natale, G., Link, B.G., 2021. Socioeconomic inequalities in the spread of coronavirus-19 in the United States: a examination of the emergence of social inequalities. Soc. Sci. Med. 268, 113554.

Crawford, F.W., Jones, S.A., Cartter, M., Dean, S.G., Warren, J.L., Li, Z.R., et al., 2022. Impact of close interpersonal contact on COVID-19 incidence: evidence from 1 year of mobile device data. Sci. Adv. 8 (1), eabi5499.

Friedson, A.I., McNichols, D., Sabia, J.J., Dave, D., 2021. Shelter-in-place orders and public health: evidence from California during the Covid-19 pandemic. J. Pol. Anal. Manag. 40 (1), 258–283.

Fiske, A., Galasso, I., Eichinger, J., McLennan, S., Radhuber, I., Zimmermann, B., Prainsack, B., 2022. The second pandemic: examining structural inequality through reverberations of COVID-19 in Europe. Soc. Sci. Med. 292, 114634.

Hooper, M.W., Nápoles, A.M., Pérez-Stable, E.J., 2020. COVID-19 and racial/ethnic disparities. JAMA 323 (24), 2466–2467.

Hou, X., Gao, S., Li, Q., Kang, Y., Chen, N., Chen, K., et al., 2021. Intracounty modeling of COVID-19 infection with human mobility: assessing spatial heterogeneity with business traffic, age, and race. Proc. Natl. Acad. Sci. USA 118 (24), e2020524118.

Hu, S., Xiong, C., Yang, M., Younes, H., Luo, W., Zhang, L., 2021. A big-data driven approach to analyzing and modeling human mobility trend under non-pharmaceutical interventions during COVID-19 pandemic. Transport. Res. C Emerg. Technol. 124, 102955.

Hu, S., Xiong, C., Younes, H., Yang, M., Darzi, A., Jin, Z.C., 2022. Examining spatiotemporal evolution of racial/ethnic disparities in human mobility and COVID-19 health outcomes: evidence from the contiguous United States. Sustain. Cities Soc. 76, 103506.

Jay, J., Bor, J., Nsoesie, E.O., Lipson, S.K., Jones, D.K., Galea, S., Raifman, J., 2020. Neighbourhood income and physical distancing during the COVID-19 pandemic in the United States. Nat. Human Behav. 4 (12), 1294–1302.

Kang, J.Y., Michels, A., Lyu, F., Wang, S., Agbodo, N., Freeman, V.L., Wang, S., 2020a. Rapidly measuring spatial accessibility of COVID-19 healthcare resources: a case study of Illinois, USA. Int. J. Health Geogr. 19 (1), 1–17.

Kang, Y., Gao, S., Liang, Y., Li, M., Rao, J., Kruse, J., 2020b. Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic. Sci. Data 7 (1), 1–13.

Lee, H., Andrasfay, T., Riley, A., Wu, Q., Crimmins, E., 2022. Do social determinants of health explain racial/ethnic disparities in COVID-19 infection? Soc. Sci. Med., 115098

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., Shaman, J., 2020. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 368 (6490), 489–493.

Link, B.G., Phelan, J., 1995. Social conditions as fundamental causes of disease. J. Health Soc. Behav. 80–94.

Luo, W., Qi, Y., 2009. An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. Health Place 15 (4), 1100–1107.

Luo, W., Wang, F., 2003. Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region. Environ. Plann. Plann. Des. 30 (6), 865–884.

Noi, E., Rudolph, A., Dodge, S., 2022. Assessing COVID-induced changes in spatiotemporal structure of mobility in the United States in 2020: a multi-source analytical framework. Int. J. Geogr. Inf. Sci. 36 (3), 585–616.

Phelan, J.C., Link, B.G., Tehranifar, P., 2010. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. J. Health Soc. Behav. 51, S28–S40.

Rader, B., Astley, C.M., Sy, K.T.L., Sewalk, K., Hswen, Y., Brownstein, J.S., Kraemer, M. U., 2020. Geographic access to United States SARS-CoV-2 testing sites highlights healthcare disparities and may bias transmission estimates. J. Trav. Med.27(7) 076.

Radke, J., Mu, L., 2000. Spatial decompositions, modeling and mapping service regions to predict access to social programs. Geogr. Inf. Sci. 6 (2), 105–112.

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. Proc. Natl. Acad. Sci. USA 117 (52), 32883–32890.

Schmitt-Grohé, S., Teoh, K., Uribe, M., 2020. *COVID-19: Testing Inequality in New York City* (No. W27019). National Bureau of Economic Research.

Shin, S.H., Ji, H., Lim, H., 2021. Heterogeneity in preventive behaviors during COVID-19: health risk, economic insecurity, and slanted information. Soc. Sci. Med. 278, 113944.

Skoll, D., Miller, J.C., Saxon, L.A., 2020. COVID-19 testing and infection surveillance: is a combined digital contact-tracing and mass-testing solution feasible in the United States? Cardiovasc. Digit. Health J. 1 (3), 149–159.

Tan, S., Lai, S., Fang, F., Cao, Z., Sai, B., Song, B., et al., 2021. Mobility in China, 2020: a tale of four phases. Natl. Sci. Rev. 8 (11), 148.

Tao, R., Downs, J., Beckie, T.M., Chen, Y., McNelley, W., 2020. Examining spatial accessibility to COVID-19 testing sites in Florida. Spatial Sci. 26 (4), 319–327.

Wang, F., Luo, W., 2005. Assessing spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas. Health Place 11 (2), 131–146.

Watson, O.J., Barnsley, G., Toor, J., Hogan, A.B., Winskill, P., Ghani, A.C., 2022. Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. Lancet Infect. Dis. 22 (9), 1293–1302.

Weill, J.A., Stigler, M., Deschenes, O., Springborn, M.R., 2020. Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. Proc. Natl. Acad. Sci. USA 117 (33), 19658–19660.

Xiong, C., Hu, S., Yang, M., Luo, W., Zhang, L., 2020. Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections. Proceedings of the National Academy of Sciences 117 (44), 27087–27089.

Ye, W., Gao, S., 2022. Understanding the spatiotemporal heterogeneities in the associations between COVID-19 infections and both human mobility and close contacts in the United States. In: Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology, pp. 1–9.