Cities 131 (2022) 103892

Contents lists available at ScienceDirect

Cities

journal homepage: www.elsevier.com/locate/cities

Do the determinants of COVID-19 transmission differ by epidemic wave? Evidence from U.S. counties

Jaehyun Ha[a], Sugie Lee[b,*]

[a] Sol Price School of Public Policy, University of Southern California, Los Angeles, CA, USA
[b] Department of Urban Planning & Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 04763, Republic of Korea

ARTICLE INFO

Keywords:
COVID-19
Epidemic wave
Density
Crowding
Spatial regression analysis

ABSTRACT

This paper uses data from the United States to examine determinants of the spread of COVID-19 during three different epidemic waves. We address how sociodemographic and economic attributes, industry composition, density, crowding in housing, and COVID-19-related variables are associated with the transmission of COVID-19. After controlling for spatial autocorrelation, our findings indicate that the percentage of people in poverty, number of restaurants, and percentage of workers teleworking were associated with the COVID-19 incidence rate during all three waves. Our results also show that dense areas were more vulnerable to the transmission of COVID-19 after the first epidemic wave. Regarding the density of supermarkets, our study elaborates the negative aspects of wholesale retail stores, which likely provide a vulnerable place for virus transmission. Our results suggest that sociodemographic and economic attributes were the determinants of the early phase of the pandemic, while density showed positive association with the transmission during subsequent waves. We provide implications for regions serving as gateway cities with high density and number of population. To add, we further provide evidence that non-pharmaceutical interventions in the early stage may mitigate the virus transmission.

1. Introduction

Since the first outbreak of COVID-19 in 2020, >450 million cases had been reported worldwide by early March 2022. In the United States, there were >80 million confirmed cases and nearly one million deaths as of March 2022 according to the Centers for Disease Control and Prevention (2021a). During the pandemic, the number of daily new cases continued its rise and fall forming small and large waves at different times across space. Since the latest and largest wave that occurred in December 2021, the outlook in the United States is quite optimistic as the majority of population completed their vaccination including the booster shot. To add, the latest variant of the coronavirus is known to be less fatal than the ones we experienced earlier which had led policymakers to lift social distancing policies. However, it is yet unclear about when the global pandemic will end since other countries such as South Korea, Germany, and Vietnam are experiencing a dramatic increase in the number of daily confirmed COVID-19 cases recently. Because the pandemic has had tragic effects on the economy as well as people's daily lives and the housing market (Kang et al., 2020; Martin, Markhvida, Hallegatte, & Walsh, 2020), it is imperative that we identify the main determinants of the new coronavirus.

The determinants of COVID-19 transmission have been widely reviewed in diverse perspectives, including socio-demographic and economic factors (Hamidi, Ewing, & Sabouri, 2020; Andersen, Harden, Sugg, Runkle, & Lundquist, 2020; Stankowska & Stankowska-Mazur, 2022), built environment factors measured in small and large scales (Sun & Zhai, 2020; Hamidi, Sabouri, & Ewing, 2020; Hu, Roberts, Azevedo, & Milner, 2021; Kashem, Baker, González, & Lee, 2021; Kwon, Oh, Choi, & Kim, 2022), health resources (Wang et al., 2021), as well as the mobility of people with regards to social distancing policies (Huang & Li, 2022; Wei et al., 2021). Compared to when we first confronted the pandemic, we now have more empirical evidence on the characteristics of population and areas vulnerable to the COVID-19 transmission as well as whether the COVID-19 related policies have been effective. The takeaways from the studies provide implications on the population group who are vulnerable to the risk of infection, promising measures to mitigate the magnitude of subsequent pandemic waves, and ways to address the negative effect of pandemics we may face in the future.

Most studies, however, focused on the early stage of the pandemic to examine the underlying factors of the COVID-19 transmission. The studies were successful in terms of suggesting implications for reducing the risk of transmission at subsequent waves. However, the main

* Corresponding author.
E-mail addresses: jaehyunh@usc.edu (J. Ha), sugielee@hanyang.ac.kr (S. Lee).

https://doi.org/10.1016/j.cities.2022.103892
Received 25 October 2021; Received in revised form 11 July 2022; Accepted 31 July 2022
Available online 4 August 2022

attributes associated with the spread of the virus during the initial stage could be related to the possibility of the virus entering a region. Moreover, we have witnessed that the subsequent waves are generally larger, and that the spatial distribution of confirmed COVID-19 cases are different across waves. A couple of recent works were able to address the differences in the determinants of COVID-19 transmission across waves; for instance, Kim, Lee, and Gim (2021) showed that the association between COVID-19 outcomes and racial/ethnic minorities differ across the pandemic waves. With these findings in mind, the determinants of the spread of COVID-19 could differ by time due to the changes in cultural and social environments, mobility of population, as well as non-pharmaceutical intervention measures. Furthermore, this could partially explain the conflicting results in the literature.

Therefore, we explore the determinants of the spread of COVID-19 in three different waves that occurred in the United States in 2020. In particular, we focus on the time period before the vaccination was approved and provided to people. We focus on the effects of socio-demographic and economic attributes, industry composition, density, crowding in housing, and COVID-19-related variables on the virus incidence rate at the county level. Meanwhile, because the transmission of the coronavirus is likely to have stronger effects on adjacent areas, we address the spatial autocorrelation issue in our analysis. Based on our analytical results, we elaborate the differences among the three waves and suggest implications for managing the pandemic. To the best of our knowledge, this paper is the first to examine the determinants of the spread of COVID-19 in three different phases from the built environment perspective. We expect that this study will contribute to understanding how the COVID-19 was spread out during the first year of the pandemic.

2. Literature review

Recent scientific studies have explored factors associated with the infection and death rates of the new coronavirus. In particular, socio-economic attributes such as age, income, race, and sex have been widely tested in previous studies (Hamidi, Ewing, & Sabouri, 2020; Andersen et al., 2020). Among those factors, income levels, which is related to deprivation, turned out to be a critical factor creating disparities in COVID-19 outcomes (Baena-Díez, Barroso, Cordeiro-Coelho, Díaz, & Grau, 2020). Low-income populations are more expected to be exposed to the virus because they experience overcrowding in low-quality housing and transit and typically work in sectors incapable of teleworking (Gibson et al., 2011; Baena-Díez et al., 2020; Hamidi, Sabouri, & Ewing, 2020). Similarly, Black population showed higher infection and death rates because they are more likely to work in essential service sectors and have low access to health-related facilities than the White population (Andersen et al., 2020). As addressed in the literature, the sociodemographic and economic attributes related to the profile of essential workers are associated with high-risk of COVID-19 transmission (Hawkins, 2020). In this perspective, the composition of industries at the county level is presumably a significant determinant of COVID-19 transmission as certain jobs cannot be done at home even during the lockdown.

From both planning and environmental perspectives, one ongoing interest is whether population density significantly affects the incidence rate of COVID-19. Although recent literature has reported conflicting results, most studies have suggested a positive relationship between density and the number of confirmed cases per capita (Arif & Sengupta, 2020; Duhon, Bragazzi, & Kong, 2021; Niu, Yue, Zhou, & Zhang, 2020; Sy, White, & Nichols, 2021; Verma, Yabe, & Ukkusuri, 2021). This is well in line with theory, which posits that higher density produces either intended or unintended interactions among people (Hamidi, Sabouri, & Ewing, 2020). However, several studies also have shown that the association between density and the spread of coronavirus could be insignificant or even negative, possibly as a result of enlightened social distancing policies (Sun & Zhai, 2020; Hamidi, Ewing, & Sabouri, 2020; Hamidi, Sabouri, & Ewing, 2020; Khavarian-Garmsir, Sharifi, & Moradpour, 2021). Because the effect of density on the COVID-19 incidence rate has significant implications for urban design, management, and planning, more scientific findings are needed to deepen our knowledge and plan for the post-coronavirus era.

Indoor crowding is also an important factor that aggravates the spread of viruses (Hu, Roberts, Azevedo, & Milner, 2021; Seidlein, Alabaster, Deen, & Knudsen, 2020). Because the new coronavirus is an airborne disease, higher levels of crowding cause intense exposure to the virus, which produces a higher possibility of infection (Seidlein et al., 2020). Although it is difficult to quantify crowding, recent studies have tried to estimate it as a situation in which the number of occupants per household exceeds the number of rooms (Hu, Roberts, Azevedo, & Milner, 2021). Household size has also been suggested as a variable to address crowding levels (Saadat, Rawtani, & Hussain, 2020). To elaborate, neighborhoods with larger households are likely to be vulnerable to transmission by creating opportunities for interactions at home (Borjas, 2020; Hamidi & Hamidi, 2021). Those studies showed a positive relationship between crowding and the COVID-19 incidence rate, calling further attention to deprived areas such as slums and immigrant communities. Moreover, the effect of crowding on the infection rate has significant implications because it relates to the effectiveness of social distancing policies and teleworking.

Related to lockdown policies implemented in cities, the change in population mobility is also known as an underlying factor of COVID-19 transmission. One early study by Glaeser, Gorback, & Redding, 2022 showed that a 10 % reduction in mobility has led to a 20 % decrease in the incidence rate of COVID-19 during February and May 2020. This result suggests that the transmission of COVID-19 is associated with social activities and gatherings, further implying the importance of policies that regulate population mobility (Badr et al., 2020; Wang et al., 2021). While a number of studies indicated that social distancing policies were successful in reducing the risk of transmission (Kraemer et al., 2020; Noland, 2021; Wei et al., 2021), Huang and Li (2022) reported that the effect of lockdown policy could differ across neighborhoods depending on the capability and willingness to reduce mobility. To add, Liu, Gross, and Ha (2021) showed that the frequency and distance of trips during the pandemic are related to financial capacity and access to grocery stores. Other works were also able to examine the relationship between transit ridership and COVID-19 transmission as individuals are exposed to overcrowded and confined spaces during travel (Kim et al., 2021; Hamidi & Hamidi, 2021).

Non-pharmaceutical interventions (NPIs) such as face mask mandates, six-feet distancing, hand washing, and self-isolation (i.e. quarantine) are also significant factors of the spread of COVID-19. A study by Lyu and Wehby (2020) showed that face mask orders at the state level could have significantly reduced the number of COVID-19 cases during the early phase of the pandemic. After experiencing a large increase in COVID-19 cases, the White House and CDC recommended people wear a face mask as of April 3, 2020 (Fisher, Barile, Guerin, et al., 2020). Since then, most of the states initiated the mask requirement, while several states did not issue the requirement during the pandemic. However, estimating the impact of NPIs on COVID-19 transmission is challenging due to the lack of data on people's actual behavior.

Scientists from diverse fields have attempted to explain the transmission mechanism of the new coronavirus (Huang et al., 2020; Jamshidi, Baniasad, & Niyogi, 2020; Perone, 2020; Sahoo, Powell, Mittal, & Garg, 2020). Their studies have successfully provided significant findings that imply practical methods for managing and alleviating the pandemic. In sum, low-income people, people of color, the elderly, and employees in frontline industries are the most vulnerable population groups during the current pandemic. Among attributes of the built environment, household structure, housing quality, population density, and the service level of facilities such as restaurants and grocery stores (i.e. points of interests) are known to be associated with the virus incidence rate (Hamidi & Hamidi, 2021; Lai, Webster, Kumari, & Sarkar, 2020). All those factors relate to the degree of exposure and the possibility of infection because those socioeconomic and environmental attributes force people to interact with or encounter people and thus face the risk of infection. However, most studies are based on the early phase of the pandemic (Hamidi, Ewing, & Sabouri, 2020; Hamidi, Sabouri, & Ewing, 2020; Hu, Roberts, Azevedo, & Milner, 2021; Li, Ma, & Zhang, 2021; Ma, Li, & Zhang, 2021), and thus their results could be limited in explaining the subsequent waves of the pandemic. Because it is likely that the response and behavior of people, as well as policymakers, have changed since the first outbreak of the pandemic, more efforts should be made to understand the factors underlying the recent spread, with a focus on discrepancies from the earlier periods.

Due to the complexity of COVID-19 transmission, a number of studies were able to address the difference in the determinants of confirmed cases across pandemic waves. Kim, Lee, and Gim (2021) showed that the positive association between the percentage of Black/African American and COVID-19 incidence rate decreased over time in the United States. To add, the authors showed that the positive relationship between non-Hispanic White and COVID-19 emerged during the later phases of the pandemic. On the other hand, Gaisie, Oppong-Yeboah, and Cobbinah (2022) reported that the relationship between density and infection cases are not consistent across pandemic waves based on their analysis of metropolitan Melbourne. In detail, the authors showed that density had a significant impact on COVID-19 transmission only during the first wave, while it changed to either insignificant (second wave) or a negative (third wave) relationship. Similarly, Arauzo-Carod, Domènech, and Gutiérrez (2021) also showed that the effects of percentage of younger population and distance to the nearest intermodal station on COVID-19 cases were inversed between two different phases of the pandemic. As addressed in recent works, the transmission of COVID-19 is affected by different factors across time, requiring further investigation.

Existing studies have been able to provide fruitful evidence for understanding how attributes of population and built environment are associated with the incidence rate of COVID-19. However, the studies are limited in two ways. First, most studies focus on the early stage of the pandemic while assuming that the relationship between either population or built environment related factors and COVID-19 transmission would be consistent. We have recently witnessed a couple of case studies which reported that the relationship could differ across waves, implying the necessity of further investigation (Kim, Zanobetti, & Bell, 2021; Gaisie et al., 2022). To the best of our knowledge, however, there is yet no literature on how the relationship between built environment and COVID-19 transmission differ by pandemic waves in the United States. Second, the majority of the literature were not able to address the holistic view regarding the determinants of COVID-19 transmission. While these studies provided significant insights into understanding the factors that are associated with the number of confirmed cases, the results could change when other factors are considered. In this study, we intend to explore the determinants of COVID-19 transmission across three pandemic waves while addressing the following attributes: sociodemographic and economic factors, industry composition, density and the potential for crowding and daily activities, population mobility, and NPIs.

3. Data and methods

3.1. Study area

The study area of this paper is the mainland of the United States (Fig. 1). As of March 21, 2021, the total number of COVID-19 cases in the area was 29.1 million, and the mortality rate was approximately 1.8 %. Whereas recent literature about COVID-19 in the United States limited the study area to counties that contain metropolitan areas, we analyzed all 3108 counties on the mainland. Because the new coronavirus has spread to all parts of the country, we found no rationale to exclude counties outside metropolitan areas. The unit of analysis is the county because data on COVID-19 cases and other attributes are available at that level. Fig. 1 illustrates the spatial pattern of the number of COVID-19 cases per 10,000 people from March 1, 2020 to March 21, 2021. We excluded January and February 2020 because the number of COVID-19 cases during that period was likely underreported, which could create bias (Hamidi, Ewing, & Sabouri, 2020). We divided the study area into four regions: Midwest, South, West, and Northeast.

3.2. Research approach

The United States saw three waves of the pandemic in 2020 and early 2021 (Drake, 2020): first wave (beginning of the epidemic–late May

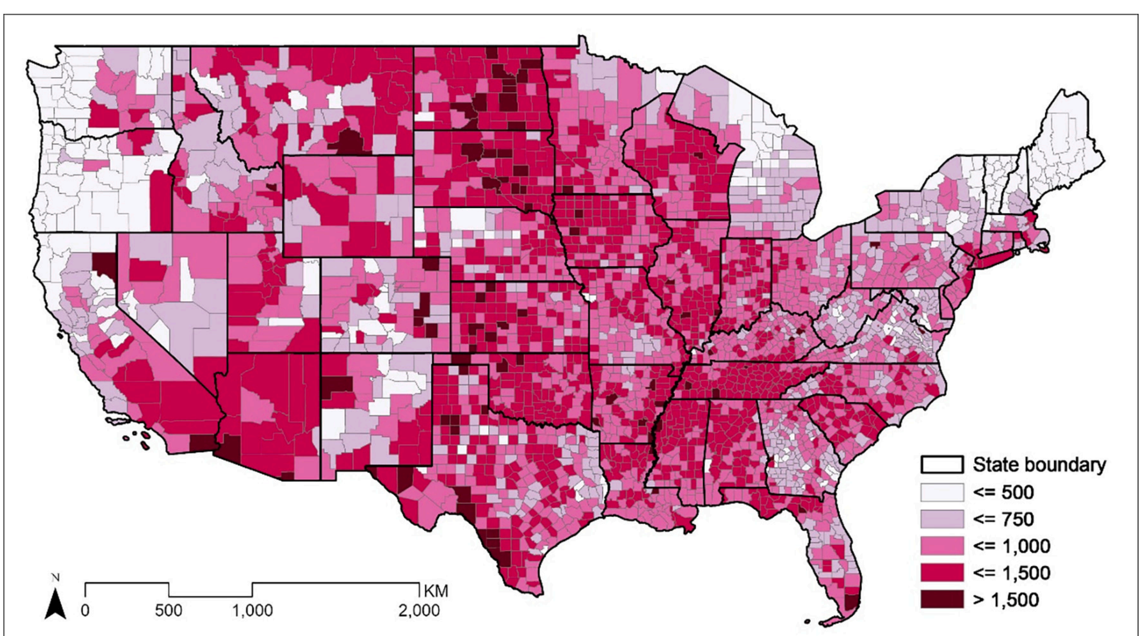

Fig. 1. Number of COVID-19 cases per 10,000 people in U.S. counties (March 1, 2020–March 21, 2021).
(Source: USA Facts).

2020), second wave (June–August 2020), and third wave (September 2020–March 2021). The third wave was much larger than the first two waves, which created 0.2–0.3 million new cases and 2000–3000 deaths per day. Also, whereas the first and second waves were predominantly limited to specific regions, the current wave has affected almost every region. It was thus ever more critical for policymakers to find ways to flatten the current wave and prevent future surges. In particular, we here set four hypotheses on how the determinants of the spread of COVID-19 would differ by pandemic waves.

First, we hypothesize that people from different sociodemographic groups would show a different association with the spread of COVID-19 by pandemic waves. Regarding that anxiety leads people to adopt preventive behaviors, older adults are likely to stay safe from the COVID-19 as a new pandemic wave occurs (Riad, Huang, Zheng, & Elavsky, 2020). On the other hand, low-income people are likely to suffer during the pandemic waves as they have less capacity to avoid public transportation or to purchase daily goods online (Jiao & Azimian, 2021). Second, we hypothesize that job composition characteristics would also show a different association with the spread of the virus during different pandemic waves. In particular, jobs from different industries have different feasibility of remote work, in which it is likely to lead frontline workers suffer more from the COVID-19 (Lan, Wei, Hsu, Christiani, & Kales, 2020). In addition, employees in the health care and social assistance would have been exposed to higher risk of COVID-19 transmission. Third, we hypothesize that the relationship between density and the spread of COVID-19 could differ according to different pandemic waves. While several works implied that higher density does not lead to higher virus transmission rates (Hamidi, Ewing, & Sabouri, 2020; Hamidi, Sabouri, & Ewing, 2020; Khavarian-Garmsir et al., 2021), most of those studies were based on the initial stage of the pandemic. In the initial stage of the pandemic, attributes related to either the entrance of the virus or implementing social distancing policies could be more important rather than the density. In the subsequent pandemic waves, however, higher density is likely to relate to higher virus transmission rates as dense areas provide environments where viruses can easily spread. Finally, we hypothesize that the impact of density of points of interests (POIs) on the spread of COVID-19 would differ by pandemic waves. For instance, restaurants, bars, and grocery stores have taken different measures to mitigate the spread of the virus during each pandemic wave, in which it is likely to lead to different outcomes. Therefore, we need advanced knowledge about the main factors that have caused each wave, which will allow us to devise practical measures to end or at least minimize the transmission of the new coronavirus. Moreover, a better understanding of the characteristics of each wave will provide insights about managing potential future pandemics in the light of lessons learned from COVID-19.

3.3. Data and variables

The outcome variables for this paper are the total number of confirmed cases of the new coronavirus per 10,000 people during the surge of each wave. The population of each county was determined using 2015–2019 American Community Survey (ACS) data. The variables were transformed using the natural log function to maintain consistency with other previous studies and ensure normality (Andersen et al., 2020; Hamidi, Ewing, & Sabouri, 2020; Hamidi, Sabouri, & Ewing, 2020). Unlike other studies, which considered the number of COVID-19 cases during arbitrary time periods, we focus on the increasing interval of the three epidemic waves that occurred between March 1 and December 15, 2020 (Fig. 2). Because every decreasing interval is a consequence of each increasing interval, we decided to focus on the time periods in which the number of COVID-19 cases surged. Each wave was defined based on the date showing local minimum values, and each increasing interval was identified as the time period between the start of a wave and the date on which the local maximum number of cases was reported within each wave (Fig. 2). The number of daily confirmed COVID-19 cases was obtained from USA Facts (2022). As a result of those analyses, we defined three increasing intervals: March 1–April 3, June 9–July 17, and September 8–November 21, 2020. It should be noted that we defined the third wave before the Thanksgiving week in 2020 since it could have impacted the spread of COVID-19 as well as the number of COVID-19 testing. Moreover, we did not consider the COVID-19 cases that occurred after December 15, 2020, since it was the day when vaccination doses were approved and distributed in the United States. Fig. 2 illustrates the time periods used for our dependent variables.

For the explanatory variables, we considered sociodemographic and economic attributes, employment composition, density and crowding factors, level of services of daily facilities, COVID-19-related aspects, and control factors at the county level. First, using the 2015–2019 ACS data, we computed the percentage of males, population older than 65 years, and Black, Hispanic, and Asian people. Those are the basic factors that address sociodemographic attributes because recent works have reported that males and people of color are likely to have higher incidence rates than others (Garg, Kim, Whitaker, et al., 2020; Hu, Roberts, Azevedo, & Milner, 2021). Although older adults clearly have a higher mortality rate than younger people (Hamidi, Ewing, & Sabouri, 2020), the association between age and the incidence rate is unclear. For instance, two studies conducted by Hamidi, Ewing, and Sabouri (2020) and Hamidi, Sabouri, and Ewing (2020) reported conflicting results

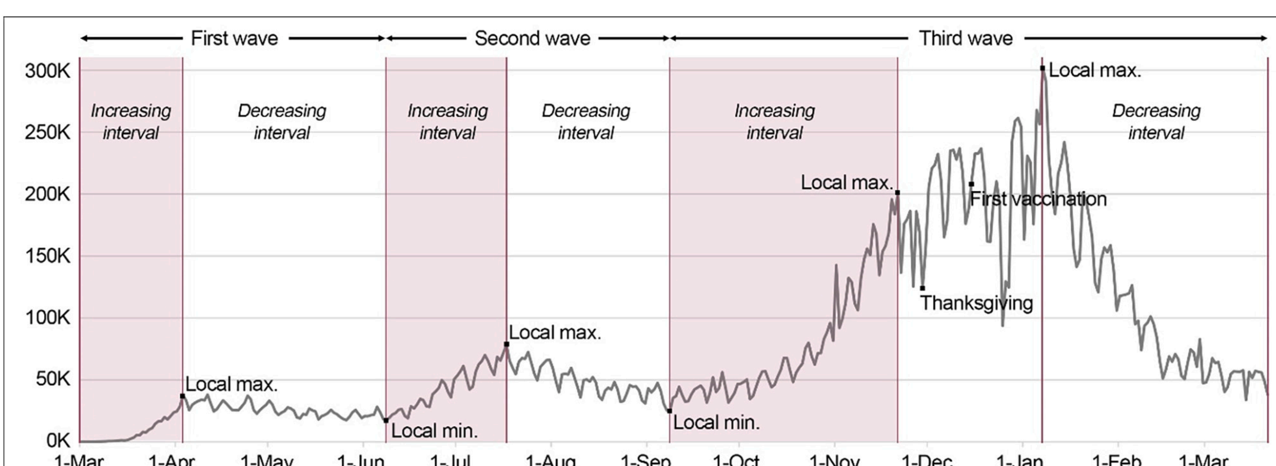

Fig. 2. Number of daily COVID-19 cases (1000) in the U.S. during three epidemic waves.
(Source: USA Facts).

about how the percentage of the population older than 60 affected the infection rate. We also considered the percentage of people with a bachelor's degree or higher to control for the educational attainment level. Finally, we included the percentage of people who earn <150 % of the poverty threshold. The 2015–2019 ACS data we used are the latest data available and were published in December 2020.

Next, we considered the industry composition of employment in each county, again using the 2015–2019 ACS data. In detail, we computed the percentage of employers per industry type based on the idea that the incidence rate of COVID-19 would vary according to the availability of telework. To elaborate, the probability of working at home would be lower for people who work in frontline industries such as construction, manufacturing, and retail, than for people who typically work in offices (Andersen et al., 2020). In addition, the frontline sectors would aggravate the spread of COVID-19 because its employees travel and interact more than people working from home and have difficulty staying home when they are sick (Harrington, 2020). For this analysis, we created a variable that considers three sectors: secondary, tertiary, and quaternary. For the secondary sector, we computed the percentage of employees working in industries such as construction, manufacturing, transportation, and utilities. This group is likely to experience the highest difficulty in working at home during the implementation of social distancing policies. The tertiary sector is mostly related to the service industry, including wholesale and retail trade, finance and insurance, and accommodation and food services. The quaternary sector contains information, professional, and scientific services; educational services; health care and social assistance; and public administration. However, we created a variable for the percentage of employees in health care and social assistance industry since it is intimately associated with the risk of COVID-19 transmission. Here, we intended to simplify the types of industry based on recent works that have shown the difference in impact of COVID-19 and employees' travel behavior among secondary, tertiary, and quaternary sectors (Jiechang, 2021; Xu & Wei, 2021; Li, Wang, et al., 2021).

To address population density, we use population-weighted density, which is known to have a positive relationship with the possibility of contacting other people and facing a risk of infection (Baser, 2020). Unlike conventional population density, which simply divides the number of people by the total area, population-weighted density weights the density per subarea according to its population, using the formula below. Baser (2020) explained that population-weighted density better explains the spread of coronavirus than the conventional formulation because it considers variations in density across an area. Although the effects of density on the COVID-19 incidence rate are unclear, we adopt Baser's (2020) idea in this analysis. We also include the total population per county, by which we attempt to control for the overall size of the population, in response to the results of Hamidi, Sabouri, and Ewing (2020), which suggested that populated metropolitan areas are particularly vulnerable to virus transmission. Therefore, we controlled for the population of each county because more populous counties would be especially susceptible to the inflow of the virus. The calculation method for the population-weighted density of a county is as follows.

$$D_p(Population\ Weigthed\ Density) = \frac{1}{P}\sum\left(p_i \times \frac{p_i}{a_i}\right)$$

where p_i and a_i indicate the number of people and area of census tract i within a county, and P is the total number of people in the county.

To account for indoor crowding, we use two variables, which mainly address crowding at home and the potential of interaction between family members. Using the 2015–2019 ACS data, we measured the percentage of large households (equal to or larger than four members). As stated in previous works, large households have a higher possibility of bringing the new coronavirus home (Hamidi & Hamidi, 2021; Saadat et al., 2020). For instance, Hamidi and Hamidi (2021) showed that twice larger households are associated with >30 % increase in COVID-19 infection rate. We also used the 2015–2019 ACS data to calculate the percentage of households in which the average number of occupants per room exceeds 1.0 because a high number of occupants per room is likely to increase transmissibility among family members at home. To elaborate, households with large size and high density would experience difficulties in practicing social distancing during stay-at-home orders.

Another important aspect of the urban environment affecting the spread of COVID-19 is the level of services of daily facilities such as bars, cafes, restaurants, and supermarkets. Those daily facilities are places in which the virus can be transmitted (Lai et al., 2020; Lu et al., 2020). Even during lockdown, such daily facilities can lead to undesirable interactions among people. Therefore, we estimated the number of facilities per 1000 people to address this aspect. Because the numbers of bars, cafes, and restaurants are highly correlated, producing multicollinearity issues, we considered only the number of restaurants. To add, bars were subjected to shut down during the pandemic in numerous states including Arizona and Washington DC, while restaurants were subjected to lighter regulation by restricting the capacity at 50 %. We considered the number of supermarkets per 1000 people because supermarkets have different characteristics from the other three facilities. To collect the locations of the daily facilities, we used the Business Search application programming interface developed by Yelp.

We also considered the attributes of COVID-19-related policies and people's behavior. We collected data on social distancing levels for two weeks prior to the start of each wave. This is based on the finding by Abouk and Heydari (2020) that the effect of a social distancing policy on the number of COVID-19 cases takes 10–15 days to appear. In detail, we used an index developed by the University of Maryland that computes the social distancing level per county based on people's mobility, such as the percentage of people staying at home and those reducing their trips, by considering changes in work trips and unemployment claims (Maryland Transportation Institute, 2020; Zhang et al., 2020). We included the percentage of employees working at home as an explanatory variable. Because it is likely that people's behavior and response to COVID-19 changed throughout the pandemic, we expected to find discrepancies in the effects of these factors on the spread of virus across waves. In addition, we controlled for people's behavior related to NPIs by considering the number of cumulative days when face mask was mandated at the state level before each pandemic wave. We here assume that the face mask policy would have had a significant impact on COVID-19 transmission as ten states including Florida, Georgia, and Idaho never issued the mandate.

Finally, we added the number of people tested for the new coronavirus per 1000 people at the state level. The number of people tested for COVID-19 is only available at the state level and is provided through the COVID Tracking Project website (The COVID Tracking Project, 2020). Because we have three different time periods, we calculated this variable for the increasing interval of each wave. We also added a regional dummy, four regions provided by the Census Bureau of the United States (Midwest, Northeast, South, and West), to account for differences in temperature, humidity, and other geographical variances.

3.4. Analytical method

As addressed above, the dependent variable is the number of COVID-19 cases per 10,000 people in three different time periods. Most of the independent variables are invariant, but the social distancing level, test rate for COVID-19, and percentage of people working from home differ by time period. We first constructed three models based on the ordinary least squares (OLS) method to confirm their appropriateness of the models. However, we found that the OLS method produced spatial autocorrelation issues in the regression residuals (Anselin, 1988): the Moran's I value of the regression residuals for the three models were 0.329, 0.305, and 0.486, respectively, which implies the presence of spatial autocorrelation in either the dependent variable or the error term. Those results were computed in the GeoDa program with the queen's weight matrix.

To address the spatial autocorrelation issue, we checked the results of the Lagrange multiplier test and the robust Lagrange multiplier test (Anselin, Syabri, & Kho, 2010). Overall, the results indicated that the spatial lag model provided a better model fit than the spatial error model. In other words, the number of COVID-19 cases per 10,000 people is spatially autocorrelated. Because our study includes all counties on the mainland of the United States, every sample was a neighbor to one or more other samples. To confirm that we have addressed the spatial autocorrelation issue by using the spatial lag model, we have estimated the changes in Moran's I of residuals, Akaike Information Criterion (AIC), and Schwarz Criterion (SC). All indicators suggested that the spatial autocorrelation issue is solved with the spatial lag model and that the fit of the models have improved compared to when using OLS. In addition, we scaled our variables to obtain standardized coefficients for comparing the magnitude of impact of the explanatory variables.

4. Results

4.1. Descriptive analysis

As shown in Table 1 and Fig. 3, the spatial pattern of the number of COVID-19 cases per 10,000 people varied by epidemic wave. During the increasing interval of the first wave, the spread of COVID-19 was concentrated in areas such as New York, New Jersey, Louisiana, Idaho, and Colorado. During that first period, there were 268,900 confirmed cases overall, and Blaine County, Idaho, recorded the highest COVID-19 incidence rate (181.0 cases per 10,000 people). In the second increasing interval, areas in the sun belt region such as California, Arizona, Texas, Louisiana, and Florida showed high virus rates. Arizona (151.7 cases per 10,000 people) and Florida (119.9 cases per 10,000 people) showed the highest virus incidence rates, whereas the virus rates in New York and New Jersey fell to approximately 13.3 cases per 10,000 people. During the increasing interval of the third epidemic wave, the total number of COVID-19 cases was approximately 5.419 million, and northern states such as North Dakota, South Dakota, Wisconsin, and Montana showed the largest rate of COVID-19 cases (Fig. 3). As shown in Fig. 3, the number of COVID-19 cases per 10,000 people was spatially autocorrelated, implying hotspots. Furthermore, the average number of COVID-19 cases per 10,000 people increased with each wave, from 2.90 for the first wave to 35.67 and 244.98 in the second and third waves, respectively.

The descriptive statistics for the sociodemographic and economic attributes show substantial differences between counties. For instance, the percentage of the population older than 65 years varied from 3.2 to 56.7, with an average of 18.85. The percentage of people of color also varied widely, from 0 to >70 % (Table 1). The composition of the industry sector also showed large differences. At the state level, Indiana, Alabama, and Tennessee showed the largest proportion of secondary sector industries, and the District of Columbia, Maryland, and Massachusetts showed the highest percentage of quaternary sector industries.

In the population-weighted density calculation, counties in New York and California, such as New York, Bronx, Kings, Queens, and San Francisco, showed the highest values. The number of restaurants and supermarkets per 1000 people also varied widely by county, as shown in the descriptive statistics (Table 1). For the COVID-19-related attributes, the social distancing level and percentage of people working at home were lower during the third epidemic stage than in the second one. This implies that human behavior might have contributed to the large magnitude of the third wave. Overall, the descriptive statistics clearly show that the variables applied in our study have great variances.

4.2. Determinants of the COVID-19 incidence rate

The results of our regression analysis are shown in Table 2. We have three models, each representing one epidemic wave. The regression

Table 1
Descriptive statistics (N = 3108).

Variables	Mean (n)	Std. dev	Min.	Max.
Outcome variables				
COVID-19 incidence rate (no. of cases per 10,000 people)*				
Increasing interval of the 1st wave (March 1–April 3)	2.90	7.88	0.00	181.02
Increasing interval of the 2nd wave (June 9–July 16)	35.67	44.21	0.00	875.57
Increasing interval of the 3rd wave (Sep. 8–Nov. 20)	244.98	182.15	7.07	1865.59
Independent variables				
Sociodemographic and economic attributes				
% of males	50.05	2.32	42.81	72.72
% of population aged 65 or older	18.85	4.61	3.20	56.71
% with a bachelor's degree or higher	21.95	9.58	0.00	77.56
% of Blacks or African Americans	9.01	14.48	0.00	87.23
% of Hispanics or Latinos	9.46	13.93	0.00	99.17
% of Asians	1.28	2.37	0.00	36.28
% of population under 150 % of poverty threshold	25.33	8.56	4.84	68.30
Industry composition attributes				
% of workers in the secondary sector	25.27	7.46	0.00	55.82
% of workers in the tertiary sector	26.37	5.03	3.54	60.61
% of workers in the quaternary sector	23.13	5.66	0.00	72.36
% of workers in health care and social assistance	13.59	3.80	0.00	39.81
Density and crowding attributes				
Population-weighted density (Baser, 2020)*	346.98	1209.46	0.06	40,817.35
Total population (10,000)*	10.38	33.24	0.01	1008.16
% of large households (no. member ≥4)	20.16	4.86	4.22	51.12
% of crowded households (occupants per room ≥1.0)*	2.32	1.90	0.00	34.66
Level of services of daily facilities				
Number of restaurants per 1000 people	1.50	1.15	0.00	27.16
Number of supermarkets per 1000 people	0.21	0.24	0.00	4.85
COVID-19-related attributes				
Social distancing index before the start of each wave				
Avg. of two weeks prior to the 1st wave	21.73	3.54	13.00	51.00
Avg. of two weeks prior to the 2nd wave	25.81	6.78	6.00	70.00
Avg. of two weeks prior to the 3rd wave	24.94	6.31	7.00	63.00
% of workers working at home				
During the increasing interval of the 1st wave	17.36	2.54	12.50	30.70
During the increasing interval of the 2nd wave	37.55	3.55	31.30	58.00
During the increasing interval of the 3rd wave	33.28	3.66	26.60	51.50
Face-mask mandate policy				
# of days with face-mask mandate before the 1st wave at the state-level**	n/a	n/a	n/a	n/a
# of days with face-mask mandate before the 2nd wave at the state-level	6.87	16.11	0.00	62.00
# of days with face-mask mandate before the 3rd wave at the state-level	51.38	46.28	0.00	153.00
Control attributes				

(continued on next page)

Table 1 (*continued*)

Variables	Mean (n)	Std. dev	Min.	Max.
Number of COVID-19 tests per 10,000 people (state-level)				
During the increasing interval of the 1st wave	42.98	26.39	5.85	139.86
During the increasing interval of the 2nd wave	706.55	226.09	356.24	1641.31
During the increasing interval of the 3rd wave	2537.13	1153.26	929.97	7883.01
Regional dummy				
Midwest	(1135)			
Northeast	(237)			
South	(1327)			
West	(409)			

* Variables were transformed using the natural log function in the final regression model.

** During the first wave, there were no face mask mandate at the state-level.

results were obtained using the spatial lag model (SLM), and we report both the unstandardized beta (*B*) and the standardized beta (β). As shown in Table 2, the Moran's I value for the regression residuals based on the OLS models were statistically significant, indicating the presence of spatial autocorrelation. On the other hand, the Moran's I values for the SLM models were not significant, which shows that we successfully addressed the spatial autocorrelation. The spatial lag variables (Rho) of all three models were statistically significant, further indicating that we controlled the spatial autocorrelation problem. In addition, the Akaike info criterion (AIC) and Schwarz criterion (SC) values were lower in the SLM models than in the OLS models, implying that we obtained a better model fit by applying spatial regression models.

Regarding the sociodemographic and economic attributes, the percentage of males was positively associated with the COVID-19 incidence rate only during the third wave. Although recent studies have shown that the percentage of males correlates with the spread of COVID-19, our results do not support those findings for the first and second waves (Hu, Roberts, Azevedo, & Milner, 2021). The percentage of older adults had an insignificant relationship with the incidence rate during the first wave and had a negative effect on the incidence rate during the second and third waves. Because it is known that the COVID-19 mortality rate is excessively high for the elderly (Hamidi, Ewing, & Sabouri, 2020), our results imply that older adults might have followed social distancing policies and stay-at-home guidelines well to protect their vulnerable immune systems. The percentage of people with a high educational attainment level was positively associated with the COVID-19 incidence rate during the first and second waves, which does not conform with previous results (Baser, 2020). The racial composition of each county's population was also significantly associated with the COVID-19 incidence rate. Although the percentage of Asians correlated negatively with the incidence rate during the second and third waves, the percentage of Hispanics or Latinos correlated positively with the incidence rate. On the other hand, the percentage of Blacks or African Americans was positively associated with the incidence rate during the first and second waves but showed a negative relationship in the third wave. The percentage of the population under 150 % of the poverty threshold was positively associated with the incidence rate in all three waves. People who are experiencing poverty are likely to work in contagious places, have difficulties teleworking, reside in low-quality housing, and struggle to acquire facial masks. This result is in line with other studies, which have shown that poverty and deprivation well explain high incidence rates (Gibson et al., 2011; Baena-Díez et al., 2020; Hamidi, Sabouri, & Ewing, 2020).

Variables related to industry composition were mostly significant only during the first and second waves. As shown in Table 2, the percentage of workers in either the secondary or tertiary sector was positively associated with the incidence rate. In particular, the secondary sector includes construction, manufacturing, and transportation-related jobs, in which employees cannot telework. Furthermore, the tertiary sector contains mostly service-related jobs, and those workers must also continue their work in-person. These results imply that certain industry compositions are likely to be associated with higher virus transmission. The coefficient for the percentage of workers in the quaternary sector in the first wave showed a negative sign. Therefore, it is likely that workers in the quaternary sector have low COVID-19 incidence rates. It is noteworthy that the percentage of workers in the health care and social assistance showed a positive sign in the third wave. The employees could have been exposed to higher risk of infection.

The effects of population-weighted density and the overall number of people differed by epidemic wave. We controlled for population size because Hamidi, Ewing, and Sabouri (2020) and Hamidi, Sabouri, and Ewing (2020) implied that the number of people itself correlates directly with the risk of infection. During the first wave, the coefficient for density was not significant, but a high overall number of people was associated with higher incidence rates. This result conforms with Hamidi, Sabouri and Ewing (2020), in which they explained that more-populated areas are more likely to be strongly hit by the virus because they have higher connectivity with other areas. On the other hand, population-weighted density showed a positive sign during the second and third waves, but the overall number of people was significant only during the second wave. Those results imply that the COVID-19 incidence rates are higher in dense areas, especially after the first epidemic wave. To elaborate, areas with larger populations were likely to have more inflow of the virus during the start of the pandemic, but substantial spread of the virus was most likely in densely populated areas.

The indoor-crowding-related variables also showed interesting results. During the first wave, only the percentage of larger households was significant and had a positive sign: larger households have a higher possibility of bringing the virus home, which leads to further transmission. On the other hand, the percentage of both large households and crowded households had positive and significant signs in the second wave, and neither variable was significant in the third wave. We assume that this result is associated with the social distancing level, which was higher during the second wave than the third. In other words, the higher social distancing level during the second wave would have resulted in higher transmission of the virus at home. Moreover, the results imply the importance of housing quality, with crowded housing a likely place for higher transmissibility. The results for the third wave suggest that transmission of the new coronavirus among family members was not a significant issue.

For the level of services of daily facilities, the number of restaurants per 1000 people showed a positive, significant sign in all three epidemic waves. Thus, restaurants, where people gather and interact with one another, increase the transmission of the virus. Even though some states, such as New York, limited restaurants to pick-up or delivery, the number of restaurants remained a significant predictor of COVID-19 cases. On the other hand, the number of supermarkets per 1000 people showed a significant negative sign during the third wave. Because the number of COVID-19 cases peaked during the third wave, this result has significant implications. Specifically, a larger number of supermarkets per 1000 people apparently reduces unintended interactions among people by reducing the service population of each store. Insofar as supermarkets are daily facilities for people's living, a greater number of supermarkets could distribute the population and further decrease the risk of transmission. The results also imply that restaurants and supermarkets affect the COVID-19 incidence rate differently.

Regarding the COVID-19-related attributes, the social distancing level during the two weeks before the start of each wave showed a significant negative sign only during the second wave. As shown in Table 1, the average social distancing level during the second wave was the highest, further implying that social distancing practice is effective only above a certain threshold. The percentage of workers working at home showed a negative sign in all three epidemic waves. According to this

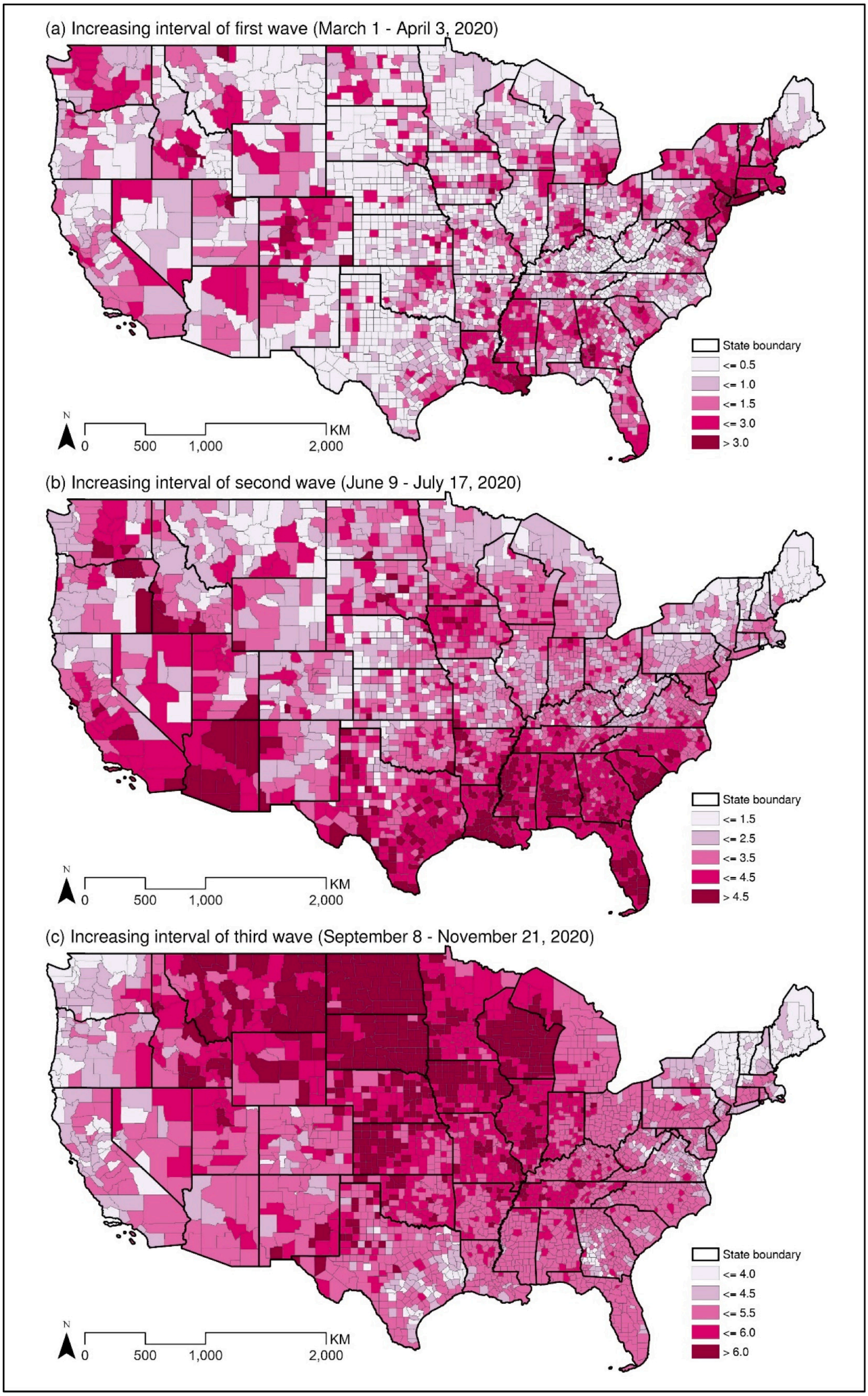

Fig. 3. Number of COVID-19 cases per 10,000 people in U.S. during three epidemic waves.

Table 2
Determinants of COVID-19 cases per 10,000 people.

Variables	Model 1: First wave				Model 2: Second wave				Model 3: Third wave			
	coef.		*z*	*β*	*coef.*		*z*	*β*	*coef.*		*z*	*β*
Sociodemographic and economic attributes												
% of males	0.000		0.02	0.000	0.007		1.16	0.014	0.018	***	5.43	0.055
% of population aged 65 or older	0.002		0.45	0.009	−0.011	**	−2.39	−0.044	−0.011	***	−4.29	−0.065
% with a bachelor's degree or higher	0.019	***	9.03	0.219	0.009	***	3.56	0.079	0.000		0.15	0.003
% of Blacks or African Americans	0.010	***	10.16	0.170	0.013	***	11.23	0.172	−0.003	***	−4.18	−0.049
% of Hispanics or Latinos	0.001		1.11	0.019	0.012	***	9.76	0.151	0.002	***	2.72	0.034
% of Asians	−0.002		−0.30	−0.005	−0.018	**	−2.40	−0.037	−0.008	**	−2.12	−0.026
% of population under 150 % of poverty threshold	0.003	*	1.73	0.034	0.009	***	3.96	0.070	0.004	***	2.83	0.040
Industry composition attributes												
% of workers in the secondary sector	0.006	**	2.46	0.051	0.009	***	3.13	0.058	−0.001		−0.56	−0.008
% of workers in the tertiary sector	0.006	**	2.17	0.038	0.010	***	2.74	0.043	0.001		0.63	0.008
% of workers in the quaternary sector	−0.006	**	−2.29	−0.044	−0.004		−1.11	−0.019	−0.001		−0.58	−0.008
% of workers in health care and social assistance	0.003		0.93	0.015	−0.003		−0.63	−0.009	0.008	***	3.27	0.038
Density and crowding attributes												
Population-weighted density†	−0.002		−0.17	−0.005	0.041	***	3.23	0.076	0.039	***	5.85	0.111
Total number of people (10,000)†	0.076	***	4.82	0.139	0.113	***	5.64	0.148	−0.018	*	−1.68	−0.035
% of large households (no. member ≥4)	0.008	**	2.45	0.050	0.015	***	3.55	0.065	0.003		1.20	0.018
% of crowded households (occupants per room ≥1.0)††	0.035		1.40	0.024	0.117	***	3.73	0.057	−0.021		−1.28	−0.016
Level of services of daily facilities												
Number of restaurants per 1000 people	0.022	*	1.75	0.031	0.033	**	2.12	0.033	0.040	***	4.79	0.060
Number of supermarkets per 1000 people	−0.038		−0.71	−0.012	−0.050		−0.73	−0.011	−0.219	***	−5.91	−0.071
COVID-19-related attributes												
Social distancing index before the start of each wave	−0.001		−0.32	−0.005	−0.012	***	−4.51	−0.074	−0.001		−1.01	−0.011
% of workers working at home	−0.008	*	−1.74	−0.026	−0.014	***	−3.57	−0.045	−0.020	***	−8.57	−0.096
# of days with face-mask mandate at the state-level	n/a†††				−0.006	***	−4.87	−0.078	−0.000	**	−2.50	−0.028
Control attributes												
# of COVID-19 tests per 10,000 people (state-level)	0.003	***	7.09	0.101	0.000	***	2.91	0.033	0.000	**	1.97	0.019
Regional dummy (Northeast)	0.074		1.60	0.024	−0.204	***	−3.15	−0.048	−0.281	***	−8.24	−0.099
Regional dummy (South)	−0.022		−0.77	−0.013	0.102	***	2.78	0.044	−0.179	***	−8.40	−0.117
Regional dummy (West)	−0.031		−0.80	−0.013	−0.169	***	−3.49	−0.050	−0.172	***	−6.19	−0.077
Lag (Rho)	0.566	***	30.40	0.566	0.519	***	28.31	0.519	0.747	***	54.24	0.747
Constant	−1.392	***	−3.39	−0.006	−0.709		−1.45	−0.005	1.241	***	4.42	−0.003
Model summary: No. of observations	3108				3108				3108			
Moran's I (OLS method)	0.329		***		0.305		***		0.486		***	
Moran's I (SLM method)	−0.010				0.003				−0.009			
AIC (OLS method)	6016.12				7329.83				4773.93			
AIC (SLM method)	5248.36				6662.74				3032.00			
SC (OLS method)	6161.12				7480.88				4924.97			
SC (SLM method)	5399.41				6819.82				3189.08			

† Variables were transformed using the natural log function in the final regression model.
†† The reference of the regional dummy variables is Midwest.
††† During the first wave, there were no face mask mandate at the state-level.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

result, encouraging telework is a powerful, effective tool for reducing transmission. The COVID-19 test rate was also significant for all three waves, indicating that some COVID-19 cases might be unknown due to insufficient testing. In addition, the number of days with face-mask mandate showed a significant and negative sign, further implying the importance of NPI measures. Finally, the regional dummy variables were mostly significant, implying that the four regions have significant differences in the number of COVID-19 cases.

In this paragraph, we make some comparison between counties with large number of population to provide statistics that could help readers understand our result. New York County, for instance, showed a large number of COVID-19 cases during the early phase (3.97 cases per 1000). The number did not grow much in the third pandemic wave (4.05 cases per 1000). On the other hand, the number of COVID-19 cases in the Los Angeles County almost tripled from 1.71 to 4.65 per 1000. While the two regions share similar attributes in some aspect, Los Angeles County has higher percentage of workers in the secondary sector, large households, as well as crowded households, while having lower density of daily facilities (e.g., restaurants, supermarkets). To add, counties such as Cook, Santa Clara, and Essex also showed a steep increase in COVID-19 cases along the pandemic wave, in which they show similar aspects with Los Angeles in terms of industry composition and crowding attributes. A single factor may be insufficient to explain the growth of COVID-19 cases but it provides an example that supports our regression results.

5. Discussion

5.1. Changes in the effects of related factors across pandemic waves

Our paper identified how the determinants of COVID-19 transmission varied during the three pandemic waves, further implying the necessity of unraveling the complexity of the pandemic across time. While a couple of recent papers were able to address the discrepancies among the waves (Gaisie et al., 2022; Kim, Zanobetti, and Bell, 2021), our paper is the first to provide a holistic view based on the COVID-19 cases in the United States. In this section, we mainly address the changes in the effects of related factors across the three pandemic waves focusing on our four hypotheses.

Regarding our first hypothesis on the difference among sociodemographic and economic groups, it is remarkable that areas with higher poverty levels experienced severe COVID-19 transmission regardless of the pandemic waves. Because low-income people are likely to stay in-person at work during the pandemic, additional measures should have taken place in the relevant areas. This result also supports the findings in other works (Baena-Díez et al., 2020). While it is evident that areas with a higher percentage of the Black and Hispanic populations experienced higher transmission rates, our results imply that it was not the case in the third pandemic wave. This result is aligned with Kim, Lee, and Gim (2021), which showed that the COVID-19 outcomes by racial groups are inconsistent over time. On the other hand, our results for older adults showed that they were negatively associated with the COVID-19 transmission, perhaps due to their willingness and capacity to stay safe.

Related to our second hypothesis, the association between industry composition and COVID-19 incidence rate also differed across waves at the county level. In particular, employees from the secondary and tertiary sectors are unlikely to work at home and abide to social distancing policies. It could be also the case that the employees have more interaction with others at workplace, leading to the transmission of the virus. It should be also highlighted that a larger portion of workers in the health care and social assistance industry were positively associated with the COVID-19 transmission during the third wave. The results suggest that their high risk of exposure to COVID-19 had actually increased the transmission rate. Moreover, the employees could have brought the virus home and increased the incidence rate. Our results imply that areas with most employments in the secondary and tertiary sector as well as the health care sector should have proactively taken measures during the pandemic.

The result on the relationship between density and the incidence rate of COVID-19 also supports our third hypothesis. During the first wave, the effect of density on the COVID-19 incidence rate was statistically insignificant, which is in line with some other recent work (Hamidi, Sabouri, & Ewing, 2020). Considering the positive relationship between the number of people in an area and the incidence rate in the first wave, it is likely that the possibility of virus inflow was more important than density itself. However, density showed a positive relationship with the COVID-19 incidence rate during the second and third waves. In other words, the risk of transmission after the first wave was higher in dense areas, which has further implications for planners. Although the relationship between density and the spread of the virus is a contentious topic, our study shows that density has a significant effect on virus transmission, especially after the first epidemic wave. Because other works that have conducted similar analyses have focused on the early periods of the pandemic (Hamidi, Ewing, & Sabouri, 2020; Hamidi, Sabouri, & Ewing, 2020; Sun and Zhai, 2020), our findings from later waves notably contribute to the literature.

Finally, the results for the number of restaurants and supermarkets also imply that neighborhood design matters during the pandemic. A larger number of restaurants aggravated the spread of COVID-19, presumably by providing places for gathering and social interactions among people. People who visit restaurants are likely to take off their facial masks, making those areas contagious. Although local and state authorities in the United States have provided guidelines for restaurants and bars, many restaurants allowed on-site dining, even indoors (Centers for Disease Control and Prevention, 2021b). Our results indicate that the operation of restaurants can aggravate the spread of the virus. On the other hand, a larger number of supermarkets was negatively associated with the COVID-19 incidence rate during the third wave. According to this result, areas with one wholesale store (i.e., Costco, Walmart) instead of multiple small supermarkets were vulnerable to the virus. This result indicates that making supermarkets spatially well-distributed could reduce the risk of virus transmission. From a planning perspective, our study suggests the importance of designing neighborhood-serving retail because it is likely to provide sufficient resilience regarding virus transmission. To add, this result is also aligned with the previous literature that showed that the availability of supermarkets is associated with the mobility of population (Liu et al., 2021).

5.2. Characterizing the three early pandemic waves in US

Based on our findings on the determinants of the COVID-19 transmission during the three pandemic waves, we characterized the transmission at each phase. Fig. 4 shows how the effects of our explanatory variables have changed across the three waves. To add, Fig. 4 displays the standardized coefficients which allow us to compare the magnitude of the effect. In the first wave, the transmission of COVID-19 was largely dominated by the percentage of highly educated people as well as the total number of population. These two factors could be associated with the possibility of bringing the new virus to the area, further suggesting that the gateway cities in the United States could have experienced severe incidence rates. For instance, counties such as New York, Queens, Bronx, Bergen, and Middlesex showed a large number of COVID-19 cases during the first phase. At the local level, counties with a higher percentage of Blacks/African Americans as well as a larger portion of employees in the secondary and tertiary sector experienced a larger number of COVID-19 cases in the first wave. In sum, the sociodemographic and economic attributes of the population were the significant determinants of COVID-19 cases in the early stage.

In the second wave, the highlighted factors (e.g. highly educated people, Black/African American, industry composition) during the first wave also showed a significant impact on COVID-19 transmission. However, it is remarkable that factors related to the built environment were significant determinants during the second wave. Unlike the first wave, density had a significant positive effect on COVID-19 transmission. To add, factors related to crowding at home (i.e. % of large and crowded households) showed a significant effect during the second wave, which is in line with the literature (Hu, Roberts, Azevedo, & Milner, 2021; Seidlein et al., 2020). Overall, the results suggest that the built environment played an important role in aggravating the pandemic in the second wave. In particular, counties such as Hidalgo, San Joaquin, Fresno, and San Bernardino showed a large increase in COVID-19 cases during the second wave, in which they have a high percentage of large households. While the transmission in the early stage of the pandemic is associated with the factors that bring in the virus, built environment factors facilitate the transmission of the virus within local areas.

As shown in Fig. 4, the effect of sociodemographic and economic attributes on COVID-19 transmission mostly diminished in the third wave. Instead, factors such as density, number of restaurants and supermarkets were highly associated with the spread of the virus. While numerous works suggested an insignificant relationship between density and COVID-19 transmission, our results provide evidence that counties with higher density have led to higher incidence rates. On the other hand, the number of supermarkets played a significant role in reducing the transmission; more supermarkets imply less unintended interaction between people during their daily activities (Liu et al., 2021). For instance, counties such as El Paso, Oklahoma, and Salt Lake showed a large number of COVID-19 cases in the third wave, in which they had less than one supermarket for 5000 population. From the results, it is clear that the built environment factors related to the possibility of interaction between people had largely contributed to the continuing pandemic.

Among the COVID-19-related policies, encouraging telework turned out to be an effective tool during the pandemic. Although we acknowledge that telework is limited to certain industries and occupations, implementing telework is expected to reduce transmission. The standardized coefficient value for telework increased steadily between the first and third waves, further indicating that areas with a high percentage of workers working at home were less vulnerable than other areas during the third wave. Social distancing levels were associated with incidence rates only during the second wave. Because social distancing levels decreased in the third wave, this result implies that social distancing practices were ineffective in the third wave. Moreover, the results imply that policymakers should ensure that social distancing is practiced well and widely. Lastly, the results show that the number of

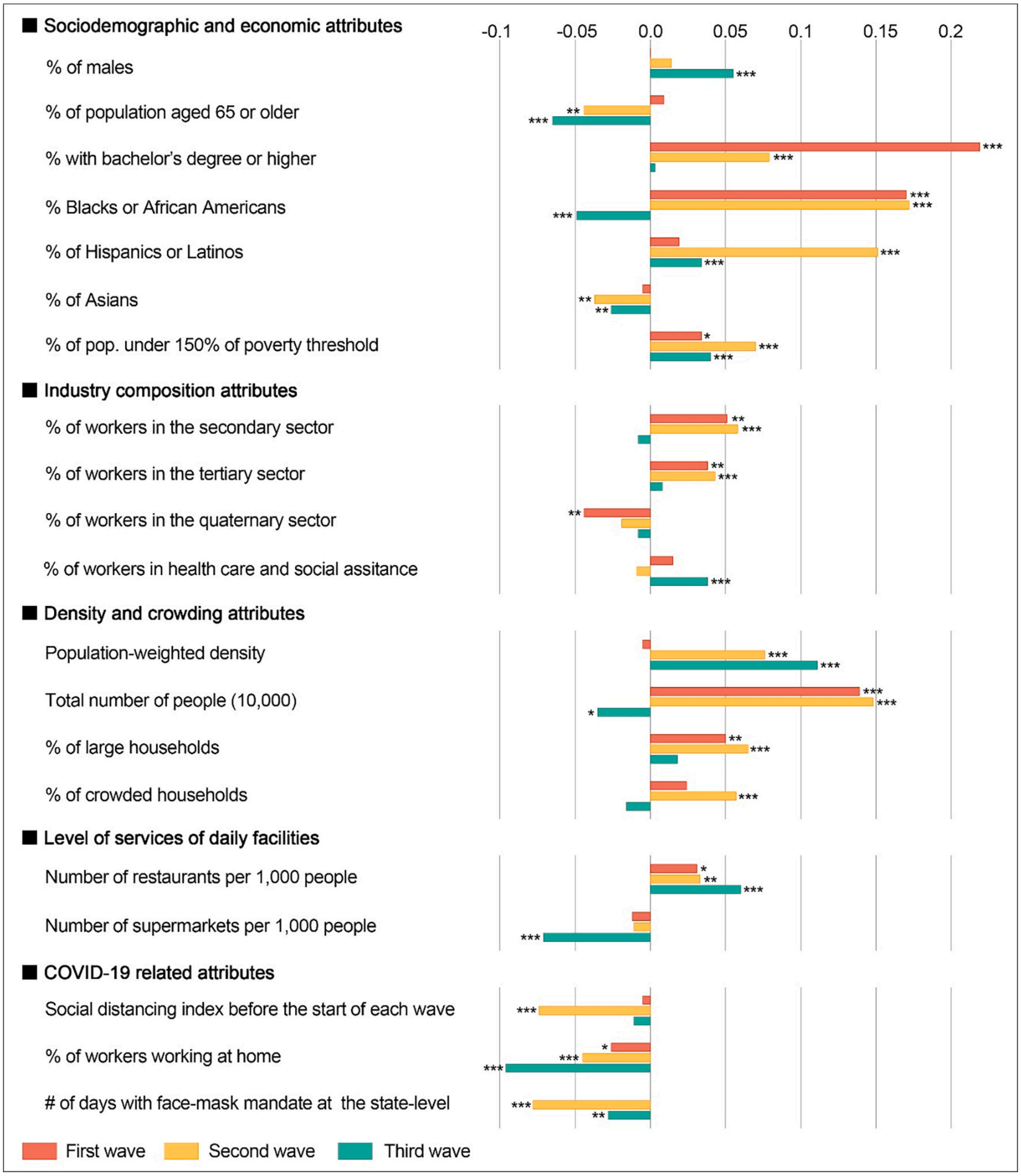

Fig. 4. Effects of explanatory variables on COVID-19 incidence rate during the three pandemic waves.
(Note: Standardized coefficients are presented (*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$)).

days with face-mask mandate at the state-level have significantly contributed to mitigating the COVID-19 transmission. This result calls attention to the states which did not issue the mandate.

6. Conclusion

Using evidence from all 3108 counties on the mainland United States, this paper confirmed that the characteristics of the spread of COVID-19 varied by epidemic wave. After controlling for spatial autocorrelation, we provided a fruitful analysis that enhances our understanding of the pandemic. Most importantly, we have found that the percentage of people in poverty, the number of restaurants, and the percentage of workers teleworking were associated with the COVID-19 incidence rate in all three epidemic waves. Based on the result, policymakers should focus on the low-income population to ensure that they can prevent potential infection. We also found that dense areas were vulnerable to the spread of COVID-19 virus during the second and third epidemic waves. Another important contribution of this study is related to the effects of daily facilities on the incidence rate: we found that restaurants are strongly contagious and that a paucity of supermarkets can lead to higher transmission rates.

This research provides implications related to addressing the transmission of viruses including the COVID-19. First, the result suggests that gateway cities could be the areas experiencing a large number of infection cases during the first wave of the pandemic. With this in mind, policymakers should be active in implementing mitigation measures, especially in gateway cities with a large number of population and highly educated people. From the global perspective, the implication could be limited to countries that are large enough where gateway cities are distant from other regions. Second, our paper shows that built environment factors including density, crowding in home, and the distribution of daily facilities are significant determinants during the subsequent waves of the pandemic. This result provides implications for planners in general, while cities with high density and larger households should pay additional attention. For instance, cities with a large number of shared housing should also be aware of transmission at home. Third, policymakers in cities with a large portion of employees in the secondary or tertiary sector should also note that they should be able to manage the pandemic during its early stage. Finally, COVID-19 related policies promoting work-at-home and mandating face masks contributed to less transmission even after controlling for factors associated with population and the built environment. This result further suggests policymakers adopt NPIs in the early phase of the pandemic to mitigate any possible negative impacts.

Despite the clear contributions of this paper, our study also has a few limitations that should be addressed in future works. First, we were unable to address the determinants of mortality. Although the incidence rate itself is important because it leads to higher demand on healthcare facilities, future works could address the determinants of mortality in different epidemic waves. Second, future works could focus on discrepancies in epidemic waves between different areas. The epidemic waves defined in this research are based on the overall trend of the United States; future studies could focus on epidemic waves at the local level. Third, future works might also want to elaborate more on our explanatory variables. For instance, the composition of industry or the type of daily facilities could be analyzed in more detail. Factors related to transportation networks and people's behavior, such as facial mask usage, could also be added to the explanatory variables. Fourth, the determinants of COVID-19 transmission could have again changed in recent months due to the distribution of vaccinations. Further investigations focusing on the pandemic phase after the vaccination process are required. Finally, the results of our study could be limited to the context of the United States as the pattern of the transmission widely differed across countries.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Jaehyun Ha: Conceptualization, Methodology, Formal analysis, Writing – original draft. **Sugie Lee:** Supervision, Writing – review & editing.

Declaration of competing interest

None.

References

Abouk, R., & Heydari, B. (2020). *The immediate effect of COVID-19 policies on social distancing behavior in the United States*. https://doi.org/10.2139/ssrn.3571421. Available at SSRN.

Andersen, L. M., Harden, S. R., Sugg, M. M., Runkle, J. D., & Lundquist, T. E. (2020). Analyzing the spatial determinants of local COVID-19 transmission in the United States. *Science of the Total Environment, 754*, Article 142396.

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht: Kluwer.

Anselin, L., Syabri, I., & Kho, Y. (2010). GeoDa: An introduction to spatial data analysis. In *Handbook of applied spatial analysis* (pp. 73–89). Berlin, Heidelberg: Springer.

Arauzo-Carod, J. M., Domènech, A., & Gutiérrez, A. (2021). Do local characteristics act in a similar way for the first two waves of COVID-19? Analysis at intraurban level in Barcelona. *Journal of Public Health, 43*(3), 455–461.

Arif, M., & Sengupta, S. (2020). Nexus between population density and novel coronavirus (COVID-19) pandemic in the south Indian states: A geo-statistical approach. *Environment, Development and Sustainability*, 1–29.

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.

Baena-Díez, J. M., Barroso, M., Cordeiro-Coelho, S. I., Díaz, J. L., & Grau, M. (2020). Impact of COVID-19 outbreak by income: Hitting hardest the most deprived. *Journal of Public Health, 42*(4), 698–703.

Baser, O. (2020). Population density index and its use for distribution of COVID-19: A case study using Turkish data. *Health Policy*. https://doi.org/10.1016/j.healthpol.2020.10.003. Available online.

Borjas, G. J. (2020). *Demographic determinants of testing incidence and COVID-19 infections in New York City neighborhoods (No. w26952)*. National Bureau of Economic Research.

Centers for Disease Control and Prevention. (2021a). *COVID-19 cases, deaths, and trends in the US*. CDC COVID Data Tracker Accessed July. 10. 2021 https://covid.cdc.gov/covid-data-tracker/.

Centers for Disease Control and Prevention. (2021b). Considerations for restaurant and bar operators. https://www.cdc.gov/coronavirus/2019-ncov/community/organizations/business-employers/bars-restaurants.html Accessed June. 15. 2020.

Drake, J. (2020). *The real cause of America's third wave of COVID-19*. Forbes. https://www.forbes.com/sites/johndrake/2020/12/07/the-real-cause-of-americas-third-wave-of-covid-19/?sh=7b35f76a12fd. (Accessed 30 December 2022).

Duhon, J., Bragazzi, N., & Kong, J. D. (2021). The impact of non-pharmaceutical interventions, demographic, social, and climatic factors on the initial growth rate of COVID-19: A cross-country study. *Science of the Total Environment, 760*, 1–9.

Fisher, K. A., Barile, J. P., Guerin, R. J., et al. (2020). Factors associated with cloth face covering use among adults during the COVID-19 pandemic — United States, April and May 2020. *MMWR. Morbidity and Mortality Weekly Report, 69*, 933–937. https://doi.org/10.15585/mmwr.mm6928e3external icon

Gaisie, E., Oppong-Yeboah, N. Y., & Cobbinah, P. B. (2022). Geographies of infections: Built environment and COVID-19 pandemic in metropolitan Melbourne. *Sustainable Cities and Society, 81*, 1–15.

Garg, S., Kim, L., Whitaker, M., et al. (2020). Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019: COVID-NET, 14 states, March 1–30. *Morbidity and Mortality Weekly Report (MMWR), 69*, 458–464.

Gibson, M., Petticrew, M., Bambra, C., Sowden, A. J., Wright, K. E., & Whitehead, M. (2011). Housing and health inequalities: A synthesis of systematic reviews of interventions aimed at different pathways linking housing and health. *Health & Place, 17*(1), 175–184.

Glaeser, E. L., Gorback, C., & Redding, S. J. (2022). JUE insight: How much does COVID-19 increase with mobility? Evidence from New York and four other US cities. *Journal of Urban Economics, 127*, 1–11.

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

Hamidi, S., & Hamidi, I. (2021). Subway ridership, crowding, or population density: Determinants of COVID-19 infection rates in New York City. *American Journal of Preventive Medicine, 60*(5), 614–620.

Hamidi, S., Sabouri, S., & Ewing, R. (2020). Does density aggravate the COVID-19 pandemic? Early findings and lessons for planners. *Journal of the American Planning Association, 86*(4), 495–509.

Harrington, J. (2020). *Poor paid sick leave, proximity to people make these tough jobs in coronavirus crisis*. WLST. https://www.usatoday.com/story/money/2020/04/18/american-jobs-with-the-biggest-sick-leave-problems-right-now/111535588/. (Accessed 25 December 2022).

Hawkins, D. (2020). Differential occupational risk for COVID-19 and other infection exposure according to race and ethnicity. *American Journal of Industrial Medicine, 63*(9), 817–820.

Hu, M., Roberts, J. D., Azevedo, G. P., & Milner, D. (2021). The role of built and social environmental factors in COVID-19 transmission: A look at America's capital city. *Sustainable Cities and Society, 65*, 1–14.

Huang, J., Kwan, M. P., Kan, Z., Wong, M. S., Kwok, C. Y. T., & Yu, X. (2020). Investigating the relationship between the built environment and relative risk of COVID-19 in Hong Kong. *ISPRS International Journal of Geo-Information, 9*(11), 624.

Huang, Y., & Li, R. (2022). The lockdown, mobility, and spatial health disparities in COVID-19 pandemic: A case study of New York City. *Cities, 122*, Article 103549.

Jamshidi, S., Baniasad, M., & Niyogi, D. (2020). Global to USA county scale analysis of weather, urban density, mobility, homestay, and mask use on COVID-19. *International Journal of Environmental Research and Public Health, 17*(21), 7847.

Jiao, J., & Azimian, A. (2021). Exploring the factors affecting travel behaviors during the second phase of the COVID-19 pandemic in the United States. *Transportation Letters, 13*(5–6), 331–343.

Jiechang, X. (2021). Tertiary industry under the COVID-19 pandemic: Impact and response. In *Economics of the pandemic* (pp. 118–140). Routledge.

Kang, M., Choi, Y., Kim, J., Lee, K. O., Lee, S., Park, I. K., Park, J., & Seo, I. (2020). COVID-19 impact on city and region: What's next after lockdown? *International Journal of Urban Sciences, 24*(3), 297–315.

Kashem, S. B., Baker, D. M., González, S. R., & Lee, C. A. (2021). Exploring the nexus between social vulnerability, built environment, and the prevalence of COVID-19: A case study of Chicago. *Sustainable Cities and Society, 75*, Article 103261.

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

Kim, M. H., Lee, J., & Gim, T. H. T. (2021). How did travel mode choices change according to coronavirus disease 2019? Lessons from Seoul, South Korea. *International Journal of Urban Sciences, 25*(3), 437–454.

Kim, H., Zanobetti, A., & Bell, M. L. (2021). Temporal transition of racial/ethnic disparities in COVID-19 outcomes in 3108 counties of the United States: Three phases from January to December 2020. *Science of the Total Environment, 791*, Article 148167.

Kraemer, M. U., 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.

Kwon, D., Oh, S. E. S., Choi, S., & Kim, B. H. (2022). Viability of compact cities in the post-COVID-19 era: Subway ridership variations in Seoul Korea. *The Annals of Regional Science*, 1–29.

Lai, K. Y., Webster, C., Kumari, S., & Sarkar, C. (2020). The nature of cities and the COVID-19 pandemic. *Current Opinion in Environmental Sustainability, 46*, 27–31.

Lan, F. Y., Wei, C. F., Hsu, Y. T., Christiani, D. C., & Kales, S. N. (2020). Work-related COVID-19 transmission in six Asian countries/areas: A follow-up study. *PloS one, 15* (5), Article e0233588.

Li, S., Ma, S., & Zhang, J. (2021). Association of built environment attributes with the spread of COVID-19 at its initial stage in China. *Sustainable Cities and Society, 67*, Article 102752.

Li, T., Wang, J., Huang, J., Yang, W., & Chen, Z. (2021). Exploring the dynamic impacts of COVID-19 on intercity travel in China. *Journal of Transport Geography, 95*, Article 103153.

Liu, J., Gross, J., & Ha, J. (2021). Is travel behaviour an equity issue? Using GPS location data to assess the effects of income and supermarket availability on travel reduction during the COVID-19 pandemic. *International Journal of Urban Sciences, 25*(3), 366–385.

Lu, J., Gu, J., Li, K., Xu, C., Su, W., Lai, Z.Yang, Z., … (2020). COVID-19 outbreak associated with air conditioning in restaurant, Guangzhou, China, 2020. *Emerging Infectious Diseases, 26*(7), 1628–1631.

Lyu, W., & Wehby, G. L. (2020). Community use of face masks and COVID-19: Evidence from a natural experiment of state mandates in the US: Study examines impact on COVID-19 growth rates associated with state government mandates requiring face mask use in public. *Health Affairs, 39*(8), 1419–1425.

Ma, S., Li, S., & Zhang, J. (2021). Diverse and nonlinear influences of built environment factors on COVID-19 spread across townships in China at its initial stage. *Scientific Reports, 11*(1), 1–13.

Martin, A., Markhvida, M., Hallegatte, S., & Walsh, B. (2020). Socio-economic impacts of COVID-19 on household consumption and poverty. *Economics of Disasters and Climate Change, 4*(3), 453–479.

Maryland Transportation Institute. (2020). *University of Maryland COVID-19 impact analysis platform*. College Park, USA: University of Maryland. https://data.covid.umd.edu, Accessed on Dec. 24. 2020.

Niu, X., Yue, Y., Zhou, X., & Zhang, X. (2020). How urban factors affect the spatiotemporal distribution of infectious diseases in addition to intercity population movement in China. *ISPRS International Journal of Geo-Information, 9*(11), 615.

Noland, R. B. (2021). Mobility and the effective reproduction rate of COVID-19. *Journal of Transport & Health, 20*, Article 101016.

Perone, G. (2020). The determinants of COVID-19 case fatality rate (CFR) in the Italian regions and provinces: An analysis of environmental, demographic, and healthcare factors. *Science of the Total Environment, 755*, Article 142523.

Riad, A., Huang, Y., Zheng, L., & Elavsky, S. (2020). *COVID-19 induced anxiety and protective behaviors during COVID-19 outbreak: Scale development and validation*. Available at SSRN 3594370.

Saadat, S., Rawtani, D., & Hussain, C. M. (2020). Environmental perspective of COVID-19. *Science of the Total Environment, 728*, Article 138870.

Sahoo, P. K., Powell, M. A., Mittal, S., & Garg, V. K. (2020). Is the transmission of novel coronavirus disease (COVID-19) weather dependent? *Journal of the Air & Waste Management Association, 70*(11), 1061–1064.

Seidlein, L., Alabaster, G., Deen, J., & Knudsen, J. (2020). Crowding has consequences: Prevention and management of COVID-19 in informal urban settlements. *Building and Environment, 188*, Article 107472.

Stankowska, A., & Stankowska-Mazur, I. (2022). The third wave of COVID-19 versus the residential preferences in Poland: An assessment of economic factors and psychological determinants. *Sustainability, 14*(3), 1339.

Sun, C., & Zhai, Z. (2020). The efficacy of social distance and ventilation effectiveness in preventing COVID-19 transmission. *Sustainable Cities and Society, 62*, Article 102390.

Sy, K. T. L., White, L. F., & Nichols, B. E. (2021). Population density and basic reproductive number of COVID-19 across United States counties. *PloS One, 16*(4), Article e0249271.

The COVID Tracking Project. (2020). Data download. https://covidtracking.com/data/download Accessed Dec. 18. 2020.

USA Facts. (2020). Download data for daily known COVID-19 cases per county. https://static.usafacts.org/public/data/covid-19/covid_confirmed_usafacts.csv. (Accessed 18 December 2022).

Verma, R., Yabe, T., & Ukkusuri, S. V. (2021). Spatiotemporal contact density explains the disparity of COVID-19 spread in urban neighborhoods. *Scientific Reports, 11*(1), 1–11.

Wang, L., Zhang, S., Yang, Z., Zhao, Z., Moudon, A. V., Feng, H.Cao, B., … (2021). What county-level factors influence COVID-19 incidence in the United States? Findings from the first wave of the pandemic. *Cities, 118*, Article 103396.

Wei, Y., Wang, J., Song, W., Xiu, C., Ma, L., & Pei, T. (2021). Spread of COVID-19 in China: Analysis from a city-based epidemic and mobility model. *Cities, 110*, Article 103010.

Xu, J., & Wei, W. (2021). The effects of tax and fee reduction policy on mitigating shock of the COVID-19 epidemic in China. *Applied Economics, 53*(46), 5303–5318.

Zhang, L., Ghader, S., Pack, M., Darzi, A., Xiong, C., Yang, M., Sun, Q., Kabiri, A., & Hu, S. (2020). *An interactive COVID-19 mobility impact and social distancing analysis platform. medRxiv 2020*. https://doi.org/10.1101/2020.04.29.20085472 (preprint).