Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Urban environments and COVID-19 in three Eastern states of the United States

Whanhee Lee, Honghyok Kim, Hayon Michelle Choi, Seulkee Heo, Kelvin C. Fong, Jooyeon Yang, Chaerin Park, Ho Kim, Michelle L. Bell

School of the Environment, Yale University, New Haven, CT, USA
Department of Public Health Science, Graduate School of Public Health, Seoul National University, Seoul, Republic of Korea

HIGHLIGHTS
• The association of urban environments with COVID-19 spread and mortality are unclear.
• We found a positive association between population density and COVID-19 outcomes (transmission and mortality).
• The % of Black or Hispanic residents showed the highest positive association with COVID-19 incidence and mortality %.
• Higher % of households that overcrowded and % of uninsured people were positively associated with COVID-19 outcomes.
• In addition, green space was negatively associated with COVID-19 outcomes.

GRAPHICAL ABSTRACT

ABSTRACT

The United States has the highest numbers of confirmed cases and deaths during the novel coronavirus disease 2019 (COVID-19) pandemic. Previous studies reported that urban residents are more vulnerable to the spread and mortality of COVID-19 than rural residents. However, the pathways through which urban environments affect COVID-19 spread and mortality are unclear. We collected daily data on the number of confirmed cases and deaths of COVID-19 from Mar. 01 to Nov. 16, 2020 for all 91 counties in New York, New Jersey, and Connecticut in the United States. We calculated the COVID-19 incidence %, daily reproduction number, and mortality %, then estimated the associations with urban environment indicators using regression models. COVID-19 outcomes were generally highest in areas with high population density, and this pattern was evident in the early period of epidemic. Among the area-level demographic variables, the percentage of Black or Hispanic residents showed the strongest positive association with COVID-19 outcomes. Higher risk of COVID-19 outcomes was also associated with higher percentage of overcrowded households, uninsured people, and income inequality. The percent elderly, sex ratio (the ratio of males to females), and greenness were negatively associated with risk of COVID-19 outcomes. The results of this study could indicate where resources are most needed.

© 2021 Published by Elsevier B.V.
1. Introduction

After a rapid increase in atypical pneumonia cases caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was observed in Wuhan in Dec 2019 (Zhang et al., 2020a), the novel coronavirus disease, better known as COVID-19, rapidly spread to over 200 countries (World Health Organization, 2020a). As of Nov. 08, 2020, more than 50 million confirmed cases were reported worldwide and more than one million people have died due to COVID-19 (World Health Organization, 2020b). In the United States (U.S.), more than 10 million individuals were diagnosed with COVID-19 with more than 230,000 related deaths, representing the largest number of confirmed and death cases in the world (World Health Organization, 2020b).

In the U.S., New York State emerged as an epicenter in the early phase of the epidemic with over 10,000 daily new confirmed cases in early April (Goyal et al., 2020). Around the same time, New Jersey and Connecticut, neighboring states to New York, also experienced the severe COVID-19 epidemic, and in response governments of these three states announced “stay-at-home” (i.e. lockdown) orders in late March (Gostin and Wiley, 2020). After the stay-at-home orders, spread of the virus in these states plateaued since around late April–May and then stabilized to Sep. 2020 (The New York Times, 2020). However, along with the record-breaking wave of pandemic since Oct. 2020 across all the U.S., the re-spread pattern of COVID-19 has been observed in the three states of New York, New Jersey, and Connecticut (as of Nov. 08).

Urban cities are generally regarded as a more vulnerable area to infectious diseases (Matthew and McDonald, 2006; Wade, 2020), and recent studies also reported that urban areas have been more susceptible to COVID-19 (Connolly et al., 2020; Sharifi and Khavarian-Garmaroudi, 2020). Moreover, several previous studies observed a positive association between urban population density and transmission of COVID-19 (Rashed et al., 2020; Rocklöv and Sjödin, 2020; Rubin et al., 2020). Living in crowded conditions and greater usage of mass transit and/or public facilities can be a major risk factor for infectious disease, thus population density has been interpreted as a proxy for the increased likelihood of crowded living environments and corresponding high infectious risk at the regional level (Rubin et al., 2020).

Nevertheless, the association between urban density and vulnerability to viral diseases should be examined more carefully, especially in this severe COVID-19 pandemic. Although more-urbanized areas can directly relate to more crowded living conditions, they also can represent higher inequalities on various factors. Previous studies reported that economically and socially marginalized persons in urban cities were more vulnerable to COVID-19, internationally (Laborde Debuquett et al., 2020; Patel et al., 2020). In particular, based on data from the New York City Department of Health and Mental Hygiene on May 18, 2020, the highest number of COVID-19 cases are concentrated in lower-income areas in the city, and the age-adjusted death % due to COVID-19 was higher in Black or Hispanic people compared to other race/ethnic groups (New York City Department of Health and Mental Hygiene, 2020). This result may indicate that regional disparities of income and race should be considered in examining the relationship between urbanization and COVID-19. Moreover, the level of urbanization and relevant inequalities could be associated with disparities in accessibility to medical services and natural environments (Vlahov and Galea, 2002), which can considerably affect the transmission and fatality of COVID-19; however, studies that investigated the potential effects of regional disparities in medical service and environment are sparse.

Therefore, this study investigated the complex effects of urban environments on the transmission and fatality of COVID-19, in all 91 counties in three Eastern U.S. states (New York, New Jersey, and Connecticut). First, using a two-stage analytic approach, we examined the association between population density and the spread of COVID-19 and fatality % at the county level. Second, we aimed to discern the urban environment indicators that can explain the effects of regional disparities (covering overcrowding, income inequality, medical service, race, and greenness) on the spread and fatality of COVID-19. In addition, we explored whether the association between urban environments on COVID-19 changed over the stages of the epidemic (the first wave, stabilized period, and the second wave stages).

2. Materials and methods

2.1. COVID-19 case and population data

We collected daily counts of the cumulative confirmed COVID-19 cases and deaths in 91 counties in New York, New Jersey, and Connecticut from Mar. 01 to Nov. 16, 2020, from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (Dong et al., 2020). In addition, we collected data on county-specific 2019 1-year estimated population from the Annual Resident Population Estimates provided by the U.S. Census Bureau.

2.2. Population density

We used the county-level population density (persons per km²) data from the 2018 American Community Survey. In addition, to perform sub-group analysis, we divided counties into three levels of urbanicity based on population density: low-density (population density under 1000 people per km²), mid-density (population density from 1000 to 5000 people per km²), and high-density areas (population density over 5000 people per km²).

2.3. Urban environment indicators

In order to examine the spatial difference in the transmission and fatality of COVID-19 by population density, we collected data on eight county-level indicators, which have a high correlation with population density (≥0.3) or medical services known to be important for the COVID-19 spread and mortality (King, 2020; Moghadas et al., 2020). These indicators could cover crowded living conditions, demographic characteristics, income inequality, insurance and hospitalization services, and vegetation level for each county.

First, we collected the % of households that are overcrowded from the Comprehensive Housing Affordability survey data for 2012–2016. Overcrowding is defined as more than one person per room, and the average value during the collection period was used in the statistical analysis. We also obtained three demographic indicators: the % of the Black and Hispanic population, the % of the population that is elderly (people age 65 or more years), and sex ratio. These demographic indicators were the 5-year estimates provided from the 2014–2018 American Community Survey (ACS). Second, we obtained an income inequality indicator, which is the ratio of household income at the 80th percentile to income at the 20th percentile (hereafter, P80/P20). The income inequality indicator was obtained from the American Community Survey for 2014–2018, and the average value during the period was used. Third, we collected the % of the population under age 65 without health insurance from the 2017 Small Area Health Insurance Estimates (SAHIE) program. In addition, we obtained the number of hospital beds from the 2017 American Hospital Association (AHA) Survey Database and calculated the number of hospital beds per 1000 persons.

Finally, we collected the county-level annual population-weighted vegetation index using the Enhanced Vegetation Index (EVI) for 2018–2019 from the Moderate Resolution Imaging Spectroradiometer (MODIS) product MOD13Q1 to quantify the level of vegetation in a country (Fensholt, 2004; Heo and Bell, 2019). We calculated average EVI based on the monthly measurements (Jan. to Dec.). Data were at the county level. More detailed information on data collection of these urban environment indicators is described in the Supplementary Materials (1. Data collection details).
2.4. Sub-period and sub-area classification

First, we used the COVID-19 data collected from the whole study period (Mar. 01 to Nov. 16, 2020). Also, in order to consider the changes in transmission pattern of COVID-19, we divided the study period into three sub-periods: Period 1 (the first wave; Mar. 01 to Apr. 30), Period 2 (stabilized period; May. 01 to Sep. 30), and Period 3 (the second wave; Oct. 01 to Nov. 16). To classify the periods, we referred to the time trend of daily transmission patterns. First, the COVID-19 incidence declined across all of our study areas at the end of April, thus we set the period through Apr. 30 as the first wave period (Period 1). Then, we set the period between May 1 and Sep.30 as a stabilized period (Period 2), because the incidence remained low (approximately <10 cases per 100,000 people) until the end of September. Finally, since early October, the rapid increase in the incidence across all study areas was observed, therefore we defined this period as the second wave (Period 3).

For some analysis, we divided the study area into four sub-areas: New York City, New York state excluding New York City, New Jersey, and Connecticut.

2.5. Statistical modelling

2.5.1. Cases and deaths of COVID-19

We calculated the total counts of confirmed cases (people who molecular diagnostic test positive for the coronavirus) and deaths (deceased patients with a positive diagnostic test) during the whole study period and for each sub-period. We computed the total counts of confirmed cases per 100,000 persons (i.e. incidence) and deaths per 100,000 persons.

2.5.2. Daily reproduction numbers

We estimated the daily reproduction number (Rt; interpreted as the instantaneous reproduction number) following Cori et al. (2013) across the whole study period. Rt indicates the average number of new infections caused by a single infected individual at time t in the population susceptible to infection (i.e., not infected or immunized population). Unlike the total counts of confirmed cases per people that indicate the magnitude of viral spread, the Rt indicates the power of transmission (i.e. transmissibility) and can guide control plans and provide insight into the effectiveness of intervention policies (Cori et al., 2013; Thompson et al., 2019). Based on previous studies, we calculated Rt with the serial interval assumed to follow a gamma distribution with mean of 4.98 days and standard deviation of 3.22 days (Lee et al., 2020; Zhang et al., 2020a) with a 14-day time window (Cori et al., 2013). All Rt calculations were performed using the EpiEstim R software package.

First, the Rt was calculated in the total study areas and for sub-areas (each area divided by the level of population density and each state) during the whole study period. Second, we calculated county-specific Rt. However, to mitigate estimation problems originated from the insufficient sample size, we applied two conditions when the county-specific Rt was computed. First, we excluded counties with an average of daily confirmed cases less than three. Thus, in the main analysis, a total of 60 counties were included to calculate the county-specific Rt. Second, we performed the calculation from Mar. 07 or the initial date when the first confirmed case was reported, whichever is earlier.

As representative indicators to show the difference in COVID-19 transmissibility across counties, we calculated the county-specific initial Rt (which can be interpreted as the basic reproduction number) (Lee et al., 2020), as well as county-specific average Rt across the whole study period and for each sub-period.

2.5.3. Association with the population density

In order to find the association between population density and the patterns of COVID-19 spread and mortality, possibly nonlinear, we applied regression models with a spline function; for all models, we included the log-transformed population density as the explanatory variable (due to the large skewness in population density for the largest counties) and modeled basis for the density variable using a natural cubic spline with three equally-spaced internal knots. Thus, we derived the relationship as the change in COVID-19 spread or mortality outcomes according to population density, allowing this relationship to be non-linear. The median value of log population density (6.35) was considered as a centering point.

First, we fitted a Poisson regression model with an overdispersion parameter using the total confirmed cases as the outcome variable. The population size for each county was included as an offset, thus the association with population density could be interpreted as the relative change in the COVID-19 incidence proportion (%) associated with a one-unit increase in the population density, compared to the centering point. Second, we performed a regression model with the county-specific Rt indicators (the initial and average Rs) as the outcome. In this model, the association with population density could be interpreted as the change in Rt indicator by a one-unit increase in the population density, compared to the centering point.

Finally, we also fitted the same modeling framework as the analysis for the total confirmed cases to estimate the association between the total death cases (mortality proportion, %) and population density. In addition, all analytic procedures were performed for the total study period and repeated for each sub-period.

2.5.4. Association with the urban environments

In order to explore whether spatial differences in the spread of COVID-19 related to the population density, we applied regression models with the urban environment indicators. For all models, a total of six urban environment indicators were included simultaneously as the explanatory variables of interest. Further, we standardized all indicators to a range of 0–1 by the Min-Max standardization method to avoid biased results due to a difference in scale, and the change in outcomes per interquartile increase in each explanatory variable was calculated to compare the magnitude of the associations.

In addition, to consider the high correlation among these urban environment indicators, we applied ridge regression. All indicators were considered as linear explanatory variables simultaneously in the ridge regression. For analyses considering the incidence and mortality as the outcomes, we fitted a ridge regression model with Poisson distribution with the population size for each county as an offset. We also applied ridge regression model with the Normal distribution to assess the association between urban environments and Rt indicators. The 95% confidence interval was estimated empirically through the bootstrap method. Finally, we considered county-specific latitude and longitude as covariates to explore the potential spatial autocorrelation. More analytic details on the ridge regression are reported in the Supplementary Materials (2. Ridge Regression).

2.6. Sensitivity analysis

Several sensitivity analyses were performed to examine whether our results are consistent to the modeling specifications (serial interval, time window, and exclusion criteria) that can affect the estimated Rt and to alternate greenness index (summer season greenness index and the Normalized Difference Vegetation Index) across the entire study period. Details of the sensitivity analysis are included in the Supplementary Materials (4. Sensitivity analysis).

3. Results

Table 1 shows descriptive information on COVID-19 cases, deaths, population density, and urban environment indicators. Across the study period, averages of the total number of confirmed cases and
Table 1: Descriptive information on the total number of COVID-19 confirmed and death cases (from Mar. 01 to Nov. 16, 2020), population, population density, and urban environment indicators during the period 2014–2019. Values: mean (range).

| Indicator                                      | New York City | New York (except New York City) | New Jersey | Connecticut | Total       |
|------------------------------------------------|---------------|--------------------------------|------------|-------------|-------------|
| Confirmed cases                                | 56,668.2 (19,749–82,775) | 4918.4 (26–54,309) | 13,372.6 (1367–30,423) | 11,607.9 (1951–31,212) | 10,300.8 (26–82,775) |
| Deaths                                         | 4822.8 (1107–7453) | 172 (0–2233) | 789.6 (91–2174) | 595 (24–1538) | 607.2 (0–7453) |
| Population (unit: 1000)                        | 1667.4 (476.1–2559.9) | 195 (4.4–1476.6) | 423 (62.4–932.2) | 445.7 (116.8–943.3) | 350.6 (4.4–2539.9) |
| Population density (persons per km2)           | 34,338.5 (8122.3–71,510.0) | 362.3 (2.7–4764.6) | 2266.8 (190.8–14,475.4) | 723.9 (198.8–1511.2) | 2700.4 (2.7–71,510) |
| % of households that are overcrowded           | 8.2 (4.0–12.0) | 1.8 (1.0–6.0) | 2.6 (1.0–8.0) | 1.6 (1.0–3.0) | 2.3 (1.0–12.0) |
| % of the elderly (age ≥ 65)                    | 15.0 (12.8–16.5) | 19.0 (13.7–31.3) | 16.9 (11.8–26.6) | 17.8 (15.8–21.3) | 18.2 (11.8–31.3) |
| Sex ratio (# of males/100 females)             | 91.3 (890.0–942.0) | 100 (92.2–121.6) | 96.1 (92.6–105.3) | 96.8 (93.1–100.5) | 98.3 (89–121.6) |
| % of population that is Black or Hispanic      | 49.4 (28.1–85.6) | 11.0 (2.5–38.9) | 29.3 (9.3–62.1) | 19.2 (8.3–31.7) | 18.1 (2.5–85.6) |
| Household income P80/P20                       | 6.7 (4.9–9.2) | 4.5 (3.8–6.2) | 4.8 (3.8–6.8) | 4.8 (4.2–5.9) | 4.7 (3.8–9.2) |
| % of population under age 65 without health insurance | 7.6 (5.0–11.0) | 5.4 (4.0–8.0) | 8.5 (5.0–11.0) | 5.5 (4.0–9.0) | 6.3 (4.0–13.0) |
| Number of hospital beds per 1000               | 3.3 (1.3–6.7) | 3.1 (0.0–8.1) | 2.7 (0.9–4.9) | 2.3 (0.7–5.4) | 3.0 (0.0–8.1) |
| EVI                                            | 0.1 (0–0.2) | 0.3 (0.2–0.3) | 0.3 (0.0–0.4) | 0.3 (0.2–0.4) | 0.3 (0.0–0.4) |

Note. P80/P20: the ratio of household income at the 80th percentile to income at the 20th percentile; EVI: population-weighed Enhanced Vegetation Index.

Deaths were highest in New York City and lowest in the New York State with New York City excluded. Correlations with population density were considerably high across all urban environment indicators (all absolute values of correlation >0.5; Table S2), except for the medical service indicators (<0.25). Fig. S1 shows correlations among the indicators. Fig. S2 shows the distribution of population density across the study area, showing the highest densities of counties in New York City and neighboring counties. Table S7 displays descriptive information of the 31 counties included in the R₀ analyses.

Fig. 1 shows the time trend of daily new confirmed cases of COVID-19 during the whole study area with two rapidly increasing trends observed in Periods 1 and 2 (the first and second waves). Fig. 2A displays daily new cases per 100,000 persons by study areas (New York City, New York state excluding New York City, New Jersey, and Connecticut). In Period 1, New York City showed the largest and fastest increase in the incidence, and the lowest and slowest increase was observed in Connecticut. Whereas, in Period 3, the reversed incidence pattern was observed; Connecticut showed the highest incidence and New York City showed the lowest incidence. Fig. 2B shows the daily reproduction number (R₀) by study areas. The initial R₀ was highest in New York City (5.7) and lowest in Connecticut (3.3). After the beginning of the epidemic, Connecticut and New Jersey generally showed the highest R₀.

Fig. 2C presents daily cases per 100,000 by sub-areas divided by high, mid, low population density areas. In Period 1, high density areas showed the highest incidence compared to mid and low density areas; however, in Period 3, mid density areas and low density areas showed higher incidence than high density area. Fig. 2D presents R₀ by high, mid, low population density areas. Except for the early part of the pandemic, which showed the highest R₀ in high density areas, all areas showed similar R₀ during the study period.

Fig. 3A shows the spatial distribution of the number of COVID-19 cases per 100,000 persons during the total study period, and Fig. 3B displays the relative increase in the incidence % associated with the population density. In the total period, the incidence % was prominently highest in the areas with high density, and a rapid increase in the incidence was observed in areas with a population density over 403.4 (log population density = 6). In addition, the peak of incidence was observed in counties where the population density was near 9–10 persons per km² (Richmond County [Staten Island], and Bronx and Queens boroughs in New York City). Fig. 3C–D presents the spatial distribution (3C) of the initial R₀ (i.e. R₀) and the association between the initial R₀ and population density (3D). There was a nearly linear relationship of the initial R₀ with the population density.

Figs. S3–4 shows the sub-period-specific (first wave, stabilized period, second wave) results corresponding to Fig. 3, and the spread pattern among sub-periods was heterogeneous. The positive association between population density and COVID-19 transmission was more prominent in Period 1 (Figs. S3D and S4D). Whereas, during Period 2 (Figs. S3E and S4E) and Period 3 (Figs. S3F and S4F), the relationship with population density was different from that of Period 1, and incidence % and average R₀ was generally higher in mid or low-density areas, compared to high-density areas.
Table 2 shows the associations between the COVID-19 incidence % and urban environment indicators. In the total period, the % of the population that was Black or Hispanic and the % of the population under age 65 without health insurance showed the highest positive association with COVID-19 incidence %. The percent of households that are overcrowded household was also positively associated with incidence %. While, % elderly and EVI were negatively associated with COVID-19 incidence %.

Table 2 also displays the relationship of the initial $R_t$ with the urban environment indicators. The initial $R_t$ showed the largest positive association with percent of the population that is Black and Hispanic and P80/P20, and it also positively associated with the percentage of households that are overcrowded, the population under age 65 without health insurance, and the number of hospital beds per 1000 persons. Whereas, the initial $R_t$ showed the strongest negative association with sex ratio, and EVI and elderly % were also significantly associated with COVID-19 incidence %.

Table 2 also displays the sub-period results corresponding to Table 2. In the early epidemic period (Period 1), the associations with the urban environment indicator were generally similar to the results of Table 2. Whereas, in Periods 2 or 3, and the association pattern was changed and varied, and some urban environment indicators (e.g., % of households that are overcrowded and EVI) showed an opposite directionality (positive) compared to the results of Period 1 (negative).
of the U.S., which was the major epicenter of the early epidemic in the United State. We found the incidence % of COVID-19 in the total study period was generally higher in counties with high population density, and this pattern was most evident in the early phase of the pandemic for this region. The initial reproduction number (\( R_0 \)) was also positively associated with population density at the county/borough level. Furthermore, the higher COVID-19 transmission was associated with higher % Black and Hispanic, % non-elderly without health insurance and overcrowding, and it was also associated with lower greenness, % elderly, and sex ratio indicating lower transmission with higher sex/ ratio (men per 100 women). The mortality risk of COVID-19 was also generally higher in high density areas. In addition, urban environment indicators related to mortality risk were generally similar to the indicators associated with the COVID-19 transmission, and the number of hospital beds showed a negative association with the mortality risk.

Positive associations between urban density and the spread of COVID-19 have been reported in multiple studies (Coşkun et al., 2020; Rashed et al., 2020; Rubin et al., 2020), and our findings are generally consistent with the existing literature. Previous studies suggested that the underlying mechanism for the association with population density is related to increased transmission of saliva, droplet, and/or aerosol between individuals when people are in close physical proximity (Bahl et al., 2020; Rubin et al., 2020; Van Doremalen et al., 2020). Furthermore, our results found that overcrowded living conditions were associated with the higher COVID-19 transmission. Particularly, being in an indoor environment with overcrowded conditions can increase the likelihood of exposure to coughs, sneeze, and food sharing, which have been suggested as dominant risk factors for the spread of coronavirus (Centers for Disease Control and Prevention, 2020a; Cook, 2020). Co-housing and apartments are common residential situations in urban areas (especially,
metropolitan areas) (Tummers, 2015), therefore improvement of overcrowded living conditions for urban residents within urban development would reduce the high transmission of COVID-19 as well as other communicable diseases.

In addition to overcrowding in households, community level sociodemographic factors such as racial composition were linked to COVID-19 outcomes, demonstrating the complex relationships between the urban environment and COVID-19 transmission and mortality. In our study area, % Black or Hispanic was associated with COVID-19 transmission and mortality; this result is consistent with other studies conducted in the U.S. (Moore et al., 2020; Stokes et al., 2020). This could be related to intersectionality of many factors such as patterns of access to health care and baseline health status (Krishnan et al., 2020; Pan et al., 2020). Further, minorities are more likely to be employed in industries providing essential services such as healthcare, food service, retail, and public transportation (U.S. Bureau of Labor Statistics, 2018). Coupled with the shortage of personal protective equipment in the pandemic’s early stages, underrepresented racial/ethnic populations likely had a higher chance of contracting COVID-19 due to increased contact with more people as a result of employment in essential services that did not allow telework or adequate sick leave (Centers for Disease Control and Prevention, 2020b).

Our study also found that COVID-19 transmission was associated with inequities in medical insurance coverage. In the U.S., economic status and health insurance are very closely related; only about half of Americans receive health insurance through their employment, and the other half have limited access to affordable healthcare (Hamel and Brodie, 2019; King, 2020). Among those who do not have access to affordable healthcare, Blacks and Hispanics are overrepresented. Due to the costs and shortages of COVID-19 diagnostic tests, those of inadequate health care insurance were likely less able to be tested for COVID-19. Our finding that county/borough % of population under 65 years without health insurance and COVID-19 incidence and mortality is consistent with the hypothesis that COVID-19 outcomes are more severe in areas where a higher proportion of uninsured persons.

The COVID-19 pandemic has affected human behaviors, including interactions with surrounding green space. In our study, we found that COVID incidence, transmission, and mortality was lower in areas of higher greenness, as indicated by higher EVI, a result consistent with a prior study (Klompmaker et al., 2020). First of all, higher greater greenness is strongly associated with lower population density (Table S2). Moreover, green spaces provide more opportunities for

---

**Table 2**

| Incidence (%) | Change per IQR (95% eCI) |
|---------------|--------------------------|
| % of households that are overcrowded | 1.52 (1.34, 1.72) |
| % of population that is elderly (age 65 or higher) | 1.59 (1.51, 0.61) |
| Sex ratio | 1.08 (1.10, 0.43) |
| % of population that is Black or Hispanic | 4.91 (7.92, 14.79) |
| Household income P80/P20 | 0.13 (0.05, 2.46) |
| % of population under age 65 without health insurance | 4.83 (4.04, 6.31) |
| Number of hospital beds per 1000 | 0.29 (−0.88, 0.07) |
| EVI | −1.33 (−1.42, −1.22) |

Initial Rt

| % of households that are overcrowded | 0.038 (0.034, 0.040) |
| % of the population that is elderly (age 65 or higher) | 0.065 (−0.066, −0.065) |
| Sex ratio | −0.133 (−0.142, −0.132) |
| % of population that is Black or Hispanic | 0.121 (0.110, 0.126) |
| Household income P80/P20 | 0.090 (0.087, 0.097) |
| % of population under age 65 without health insurance | 0.020 (0.014, 0.024) |
| Number of hospital beds per 1000 | 0.023 (0.020, 0.024) |
| EVI | −0.091 (−0.092, −0.083) |

Note. eCI: empirical confidence interval; IQR: inter quartile range; P80/P20: the ratio of household income at the 80th percentile to income at the 20th percentile; EVI: population-weighted Enhanced Vegetation Index.

**Table 3**

| Mortality (%) | Change per IQR (95% eCI) |
|---------------|--------------------------|
| % of households that are overcrowded | 4.51 (4.38, 4.61) |
| % of population that is elderly (age 65 or higher) | −1.31 (−2.20, −0.71) |
| Sex ratio | −4.90 (−5.31, −1.21) |
| % of population that is Black or Hispanic | 9.83 (9.03, 10.49) |
| Household income P80/P20 | −0.06 (−1.15, 0.99) |
| % of population under age 65 without health insurance | 8.41 (7.06, 9.41) |
| Number of hospital beds per 1000 | −2.98 (−3.29, −2.40) |
| EVI | −8.21 (−9.66, −6.71) |

Note. eCI: empirical confidence interval; IQR: inter quartile range; P80/P20: the ratio of household income at the 80th percentile to income at the 20th percentile; EVI: population-weighted Enhanced Vegetation Index.
outdoor activities, which carry lower transmission risks compared to indoor activities with limited ventilation. Further, green spaces have other health consequences such as improved mental health, physical activity, and social cohesion. Counties/boroughs with higher greenness also have more open space, allowing aerosols to disperse (Allen and Marr, 2020; Coşkun et al., 2020), and there is increasing evidence that COVID-19 is transmitted through the air (Zhang et al., 2020b). In addition, previous studies reported that availability of green space is generally higher in areas with higher socioeconomic level (Astell-Burt et al., 2014; Lakes et al., 2014; Wen et al., 2013), thus our green space results could partly reflect the effect of the higher socioeconomic status on COVID-19 outcomes. However, the recent Google Community Mobility Reports showed increases in park visitations in New Jersey and Connecticut during the pandemic (Aktay et al., 2020), and it can imply that green space may bring people into contact with others. We conjecture that this increase in park usage may be related to the non-negative association between greenness and incidence % in Period 2–3, although it should be investigated further.

Our finding also showed that areas with lower % elderly were more vulnerable to COVID-19 spread. These findings were consistent with previous studies, and more active social activities and lower severity awareness of the pandemic compared to older people were suggested as an important factor (Boehmer et al., 2020; Lee et al., 2020; Zhang et al., 2020a). On the other hand, our study showed that areas with higher male % showed lower COVID-19 transmission and mortality, however, these patterns differed from those of previous studies (Griffith et al., 2020). Previous studies reported that males have less avoided large social gatherings, close physical contact with others, or alcohol consumption, compared to females (Griffith et al., 2020; Wenham et al., 2020). We conjecture that relatively higher women economic activity in our study area (the 2018 national rank of women’s labor force participation: 29 [New York], 19 [New Jersey], and 13 [Connecticut]) and the poverty of women in New York, which was substantially higher than in the other two states (the 2018 national rank of women’s poverty rate: 42 [New York], 3 [New Jersey], and 2 [Connecticut]) (Institute for Women’s Policy Research, 2018) might be related to our results; however, we cannot fully explore these results with this data, and further research should investigate how sex/gender differences in COVID-19 outcomes vary regionally.

Over the span of the 9-month study period, the association of the COVID-19 transmission with population density and urban environments changed over time. In the first epidemic wave (Period 1), more densely populated areas, especially New York City, had higher COVID-19 transmission. Whereas in the second wave (Period 3), this pattern was diluted, and the highest incidence was observed in Connecticut, which showed the lowest incidence % in Period 1. Moreover, the afore-mentioned associations with urban environment indicators were most evident in Period 1, however, these associations diminished in Period 3. We postulate that this pattern may be related to behavior or lifestyle changes and public health interventions after the first epidemic. In particular, New York experienced the first epidemic with severe impacts and legislated to wear a face-covering in public space since Apr. 17, and earlier work found that the policy did play a dominant role to flatten the transmission in New York City (Zhang et al., 2020b). Therefore, we postulate that the role of urban environments on the spread of the virus is most significant in the early phase of the epidemic, and it can be diluted or changed over time; although more studies are needed.

We also found that urban environments affect the mortality % of COVID-19, as well as the transmission. Most of the factors that were associated with the increase in transmission of the virus also be associated with the increase in mortality % in the same direction and this pattern can be interpreted in two plausible ways. The first hypothesis is the in-oculum effect. Previous studies reported that patients exposed to a higher viral load at the infection will have a more severe illness (Chu et al., 2004; Poulsen et al., 2002), thus it is possible that people living at high risk of infection are more likely to be exposed to a higher viral load of the coronavirus. Another hypothesis is that factors generally related to better general health conditions (e.g., better income and medical services) could affect the lower mortality risk of COVID-19. Moreover, earlier studies found that the number of hospital beds per person had a negative association with mortality risk. Hospital beds are vital to the treatment of patients with severe illness (Lee et al., 2020; Moghadam et al., 2020), and the higher number of hospital beds can indicate the higher capacity for treatment of severe patients, such as beds in the intensive care unit, negative pressure beds, and ventilators (Lee et al., 2020). Our results show that the limited accessibility to hospital beds may be a significant factor that can increase the mortality % and suggest the importance of securing sufficient hospital beds in urban areas where a greater spread of viral infection is expected.

There are several limitations to this study. First, because of data limitations, we could not examine the spread pattern of more specific sub-populations, such as race/ethnicity, age, and sex, which might be associated differently with urban environments. Therefore, our study results should be interpreted as results for the total population. Second, although we considered multiple indicators and additional confounders, we could not fully adjust for potential confounding variables. In particular, when assessing the association between viral transmission and urban environment indicators, we could not consider possible variables that reflect indoor activities (e.g., types of work, working hours indoors, indoor athletic or leisure activities), public transportation, and social contact (e.g., number of people and duration of social gatherings), which can affect the viral contagion. In addition, a recent study conducted in U.S. 25 urban counties reported that the decreased mobility has a significant, positive association with reduced confirmed case growth (Badr et al., 2020). Therefore, the relationships between different types of urban population mobility and COVID-19 outcomes can be examined more carefully in future studies. Third, our study used ecological urban environment variables, thus the results of this study can only suggest ecological associations between urban environments and COVID-19 outcomes, but cannot be interpreted as the associations or causalities at the individual level. Fourth, due to the insufficient sample size, this study could not consider a total of 31 counties in the R–urban environment association analyses. Thus, the results could be biased towards counties with a large number of confirmed cases. Finally, we could not consider the possibility of underreporting deaths related to COVID-19. In particular, the U.S. Centers for Disease Control and Prevention has reported that some deceased cases due to COVID-19 can be assigned to other death causes (not diagnosed or not mentioned on the death certificate) (Centers for Disease Control and Prevention, 2021), which implies that the data on COVID-19 deaths used in this study might be undercounted and our results related to mortality could be biased towards counties with fewer underreported deaths due to COVID-19. Further investigation examining the role of urban environments on COVID-19 or other infectious diseases should consider these limitations.

Despite these limitations, our study has several strengths. First, we examined the complex effects of urban environments on COVID-19 transmission and compared the magnitude of their effects at the county/borough level, using advanced statistical modeling. Second, using the relatively long time-series data (Mar. to mid-Nov. 2020), we investigated the spatiotemporal changes in the transmission and mortality patterns of COVID-19, as well as the corresponding changes in the associations with urban environment indicators. In particular, our findings indicate the importance of initial response strategies in high density areas. Third, we our analysis of the association between urban environments and mortality due to COVID-19 suggests that specific regional indicators may play an important role to reduce fatality of COVID-19. Our results provide scientific evidence for the prioritization of resource allocation and effective intervention policies considering the location-specific urban environments.
5. Conclusion

This study suggests that populations living in high density areas may be more vulnerable to COVID-19 spread as well as mortality. It addresses the comprehensive role of urban environment indicators, which can be related to mitigating the spread and mortality of the viral disease. We also found that the relationship between urbanicity and COVID-19 changed over time. Our findings have implications for public health and policies in urban areas on the priority of resource distribution for managing this unprecedented pandemic.

CRediT authorship contribution statement

Whanee Lee: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft. Hongsyok Kim: Methodology, Data curation, Writing – review & editing. Hayon Michelle Choi: Methodology, Data curation, Writing – review & editing. Seulkeee Heo: Data curation, Writing – review & editing. Kelvin C. Fong: Writing – review & editing. Jooyeon Yang: Writing – review & editing. Chaerin Park: Writing – review & editing. Ho Kim: Writing – review & editing. Michelle L. Bell: Supervision, Funding acquisition.

Declaration of competing interest

None.

Acknowledgement

This paper was developed under Assistance Agreement No. RD835871 awarded by the U.S. Environmental Protection Agency to Yale University. It has not been formally reviewed by EPA. The views expressed in this document are solely those of Michelle L. Bell and other co-authors and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this publication.

References

Aktay, A., Baradekar, S., Cossoul, G., Davis, J., Desfontaines, D., Fabrikant, A., et al., 2020. Google COVID-19 Community Mobility Reports: Anonymization Process Description (Version 1.0).
Allen, J.G., Marr, L.C., 2020. Recognizing and controlling airborne transmission of SARS-CoV-2 in indoor environments. Indoor Air 30, 557.
Astell-Burt, T., Feng, X., Mavoa, S., Badland, H.M., Giles-Corti, B., 2014. Do low-income neighbourhoods have the least green space? A cross-sectional study of Australia's most populous cities. BMC Publ. Health 14, 1–11.
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. Lancet Infect. Dis. 20, 1247–1254.
Bahl, P., Doolan, C., de Silva, C., Chughtai, A.A., Bourouiba, L., Macnrythe, C.R., 2020. Airborne or droplet precautions for health workers treating COVID-19? J. Infect. Dis., jia189 https://doi.org/10.1093/infdis/jia189.
Boehmer, T.K., DeVies, J., Caruso, E., van Santen, K.L., Tang, S., Black, C.L., et al., 2020. Boehmer, T.K., DeVies, J., Caruso, E., van Santen, K.L., Tang, S., Black, C.L., et al., 2020. Factors that influence COVID-19 transmission: a mathematical model. J. Urban Health 97, 1655–1669.
Brown, M., Rabin, D., Huang, J., Fisher, B.T., Gasparrini, A., Tam, V., Song, L., et al., 2020. Association of absolute humidity, temperature, and population density on COVID-19 spread and delay durations: Multi-prefecture study in Japan. J. Hosp. Infect. 107, 47–53.
Broome, L., 2015. Understanding co-housing from a planning perspective: why and how? J. Urban Practice 8, 64–82.
Brown, M., Rabin, D., Huang, J., Fisher, B.T., Gasparrini, A., Tam, V., Song, L., et al., 2020. Association of absolute humidity, temperature, and population density on COVID-19 spread and delay durations: Multi-prefecture study in Japan. J. Hosp. Infect. 107, 47–53.
Broome, L., 2015. Understanding co-housing from a planning perspective: why and how? J. Urban Practice 8, 64–82.
World Health Organization, 2020a. Coronavirus disease 2019 (COVID-19) situation report — 90. In: WHO (Ed.), Coronavirus Disease 2019 Situation Report. Situation Report — 90. World Health Organization (WHO).

World Health Organization, 2020b. WHO Coronavirus Disease (COVID-19) Dashboard.

Zhang, J., Litvinova, M., Wang, W., Wang, Y., Deng, X., Chen, X., et al., 2020a. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. Lancet Infect. Dis. 20 (7), 793–802. https://doi.org/10.1016/S1473-3099(20)30230-9.

Zhang, R., Li, Y., Zhang, A.L., Wang, Y., Molina, M.J., 2020b. Identifying Airborne Transmission as the Dominant Route for the Spread of COVID-19. 117 pp. 14857–14863.