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.
County-level demographic, social, economic, and lifestyle correlates of COVID-19 infection and death trajectories during the first wave of the pandemic in the United States

Natalie C. DuPre⁎, Seyed Karimib,c, Charlie H. Zhangd, Lyndsey Blaira, Arushi Guptae, Lamyaa Mousa A. Alharbif,M a r i y a M al l u h i b if, Riten Mitraf, W. Paul McKinneyg, Bert Littleb

a University of Louisville School of Public Health and Information Sciences, Department of Epidemiology and Population Health, United States of America
b University of Louisville School of Public Health and Information Sciences, Department of Health Management and System Sciences, United States of America
c Louisville Metro Department of Public Health and Wellness, Center for Health Equity, United States of America
d University of Louisville, College of Arts and Sciences, Department of Geography and Geosciences, United States of America
e University of Louisville, School of Public Health and Information Sciences, United States of America
f University of Louisville School of Public Health and Information Sciences, Department of Bioinformatics and Biostatistics, United States of America
g University of Louisville School of Public Health and Information Sciences, Department of Health Promotion and Behavioral Sciences, United States of America

HIGHLIGHTS
• Group-based trajectories identified county clusters with similar epidemic curves.
• Communities with higher risk profiles had higher COVID-19 epidemic curves.
• Younger, smoker, female, & populations of color are higher risk counties.
• Prioritize communities of color, youth, & smokers for vaccination & resources.
• Vaccinate and support healthcare & private industry employees with close contacts.

ABSTRACT

Background: The US COVID-19 epidemic impacted counties differently across space and time, though large-scale transmission dynamics are unclear. The study’s objective was to group counties with similar trajectories of COVID-19 cases and deaths and identify county-level correlates of the distinct trajectory groups.

Methods: Daily COVID-19 cases and deaths were obtained from 3141 US counties from January through June 2020. Clusters of epidemic curve trajectories of COVID-19 cases and deaths per 100,000 people were identified with Proc Traj. We utilized polytomous logistic regression to estimate Odds Ratios for trajectory group membership in relation to county-level demographics, socioeconomic factors, school enrollment, employment and lifestyle data.

Results: Six COVID-19 case trajectory groups and five death trajectory groups were identified. Younger counties, counties with a greater proportion of females, Black and Hispanic populations, and greater employment in private sectors had higher odds of being in worse case and death trajectories. Percentage of counties enrolled in grades 1–8 was associated with earlier-start case trajectories. Counties with more educated adult populations had lower odds of being in worse case trajectories but were generally not associated with worse death trajectories. Counties with higher poverty rates, higher uninsured, and more living in non-family households had lower...
1. Introduction

The pandemic of coronavirus disease 2019 (COVID-19), caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is responsible for over 2 million deaths worldwide as of February 2021 and has drastically increased hospitalizations (WHO, 2020). In the United States between January 21, 2020 to February 21, 2021 there were over 445,000 deaths and over 26.3 million cases of COVID-19 (CDC, 2020a). COVID-19 deaths within the US continue to disproportionately affect older ages and racial and ethnic groups (Bassett et al., 2020); for example, 81.2% of COVID-19 related deaths were ≥65 years (2) and age-adjusted COVID-19 mortality rates are approximately two-times higher in Latino populations and Black populations than White populations (APM Research Lab, 2021).

The first wave of the epidemic in the United States affected communities with different intensities over space and time. While the initial outbreaks in the United States occurred in large coastal cities with high connectivity to the original epicenter of Wuhan, China (Hamidi et al., 2020), the SARS-CoV-2 virus spread throughout the nation regardless of urban/rural status and with differing magnitudes and timing. Place-based spatial analyses of COVID-19 have observed correlations between county-level demographic, social and economic factors and high incidence geographic clusters early on in the epidemic (Andersen et al., 2021; Karaye and Horney, 2020; Sun et al., 2020a; Zhang and Schwartz, 2020; Mollalo et al., 2020; Marvel et al., 2021). While it is not ideal to conduct ecological studies because they are limited in their causal understanding of transmission dynamics, there may be value in ecological place-based studies at this moment in time when there is community-wide spread and there is a need to generate hypotheses to understand large-scale population transmission dynamics. The previous ecological studies of place-based characteristics examined intensity of the epidemics in US counties (Andersen et al., 2021; Karaye and Horney, 2020; Sun et al., 2020a; Zhang and Schwartz, 2020; Mollalo et al., 2020; Marvel et al., 2021), but largely did not consider the start time of the epidemic that is an important feature of all epidemics and varied across the US. By examining the start time of the epidemic, insight may be gained on what ecological features were ripe for viral transmission and others that helped to delay the onset of the epidemic. Few studies examined COVID-19 deaths (Andersen et al., 2021; Zhang and Schwartz, 2020) and most had investigated COVID-19 burden through May 2020 (Karaye and Horney, 2020; Zhang and Schwartz, 2020; Mollalo et al., 2020). The focus of the current place-based ecological study is to expand our understanding of the timing and magnitude of the pandemic across US counties and their county-level correlates, particularly related to COVID-19 mortality that is a more accurate measure of the disease burden than case ascertainment due to the lack of testing during the first wave of the pandemic. To our knowledge, only one published study has used clustering of COVID-19 patterns over time considering a quadratic form of the epidemic curve (Vahabi et al., 2021).

In this analysis, the objectives were to 1) identify counties that had similar epidemic curves in terms of magnitude and timing during the first wave, and 2) examine county-level correlates including social, economic, demographic and lifestyle factors associated with COVID-19 case and death trajectories in the US from January through June 2020.

2. Methods

2.1. Data sources

Laboratory-confirmed and probable cases of COVID-19 and COVID-19 deaths for every county in the United States from January 21, 2020 through June 30, 2020 were obtained from the New York Times GitHub center (The New York Times, 2021). This time period was used to reflect the first wave of COVID-19 in the United States as the number of daily deaths reached its first peak in mid-April and declined through June 2020 approaching a nadir in late June 2020, indicating the first wave; this was quickly followed by an increase throughout July 2020 (CDC, 2020b). The number of daily COVID-19 cases and COVID-19 deaths were derived from this data set, and events per 100,000 people were calculated utilizing the 2018 county population size from the US Census Bureau estimates (U.S. Census Bureau, 2021).

County-level demographic factors (e.g., age, sex, race, Hispanic ethnicity), school enrollment, education level of the adult population, poverty, health insurance coverage, and non-family households were obtained from the US Census Bureau's American Community Survey 2014–2018 five-year estimates (2014–2018) (U.S. Census Bureau, 2018). Data on the proportion of the adult population who smoke and who are obese were obtained from the County Health Rankings from the Robert Wood Johnson Foundation for 2020 (County Health Rankings and Roadmaps, 2021). Employment information from the first quarter of 2020 were obtained from the Bureau of Labor Statistics using the North American Industry Classification System (NAICS) categories of employment sectors (Databases, 2021; U.S. Census Bureau, 2021), including the percentage of the county population employed in the private service-producing industry (e.g., trade, transportation, and utilities; information; financial activities; professional and business services; education and health; leisure and hospitality; and other services), private goods-producing industry (e.g., natural resources and mining; construction; and manufacturing), and the public sector (i.e. federal, state, and local governments). The aforementioned county-level covariates were linked to county-level COVID-19 cases and deaths per 100,000 to compile the analytic dataset of 3141 US counties in the 50 states and the District of Columbia.

2.2. Statistical analyses

To examine whether demographic, socioeconomic and other characteristics were associated with COVID-19 epidemic trajectories over time, we first used a group-based trajectory modeling approach that identifies groups with similar trajectories of COVID-19 and then we applied a label to each of the identified groups that reflected their epidemic curve. Secondly, we used polytomous logistic regression with the outcome as trajectory group membership and the independent variables were the county-level demographic, socioeconomic and other characteristics (see details below).

2.2.1. Identification and description of trajectories of COVID-19 cases and COVID-19 deaths between January and June 2020

SAS (SAS Institute Inc., Cary, NC, USA; Version 9.4) and Proc Traj (Jones et al., 2001; Jones and Nagin, 2007) were used to identify counties with
similar COVID-19 case trajectories and with similar COVID-19 death trajectories. Proc Traj is a group-based trajectory modeling approach that estimates multiple statistical measures describing longitudinal data (e.g., measures of change over time and variability in change over time such as the range, average over time, standard deviation, ratio of change across time periods, slopes across time periods, maximum change, and more) and uses latent class modeling to classify each observation into mutually exclusive groups (Lefkondré et al., 2004; Sylvestre et al., 2006). This approach is advantageous in that the researchers are not forced to make arbitrary cut-points to determine trajectories, but rather the Proc Traj algorithm groups observations with similar measures of change over time. This group-based trajectory modeling approach has been used in multiple health contexts to identify groups with similar changes in health outcomes over time (e.g., changes in BMI or glucose levels) (Song et al., 2016a; Song et al., 2016b; Yuan et al., 2018). For example, this method may identify a group of individuals whose BMI has stayed stable over their adult lifetime, another mutually-exclusive group whose BMI was always very high, another group whose BMI started off lean and then increased, and another group whose BMI started off as overweight and increased to obese; while the method groups the individuals based on their BMI trajectories, the labels applied to each group is determined by the researchers. To our knowledge, this group-based trajectory modeling method has not been applied in the context of infectious disease epidemics to identify groups of places that have similar changes in disease rates over time.

Instead of identifying trajectories based on daily rates across 182 days that may be particularly unstable, we more broadly summarized the COVID-19 data into rates per 100,000 in each county within seven time periods: January through March 31, 2020, April 1–15, April 16–30, May 1–15, May 16–31, June 1–15, and June 16–30. We observed that 32% of the counties had 0 cases per 100,000 during January–March 2020; thus, using SAS PROC RANK, we calculated tertiles of rates per 100,000 across these seven time periods. In the model of death trajectory modeling, the ranges of the tertile categories within each time period are presented in Supplemental Table 1. In Proc Traj, one must decide the number of groups to identify and the appropriate model specification based on the outcome variable. We used a censored normal model to identify six trajectory groups based on the tertiles of rates per 100,000 in each US county within each time period to include in the trajectory modeling. The ranges of the tertile categories within each time period are presented in Supplemental Table 1. In Proc Traj, one must decide the number of groups to identify and the appropriate model specification based on the outcome variable. We used a censored normal model to identify six trajectory groups based on the tertiles of rates per 100,000 across these seven time periods. In the model of death trajectories, we identified five trajectory groups because the sixth group included a small percentage of counties (≤1%). Identification of more than six trajectories was not considered due to the difficulty in interpreting seven or more trajectory groups. We assumed a polynomial form for time in all trajectory models.

After Proc Traj classified each county into a specific trajectory group, researchers assign a label to each group to describe the trajectory pattern. We described each trajectory group by analyzing the daily data of each identified trajectory group. A label was given to each group by examining the average number of cases and deaths per 100,000 across the seven time periods (Table 1), and by examining their visual patterns for daily cases per 100,000 and daily deaths per 100,000 in RStudio (Figs. 1 & 4); the label given to each group was based on the magnitude of the epidemic as well as the initial start time of the epidemic described in Table 1, Fig. 1, and Fig. 4. Based on the observations from Table 1 and Fig. 1 that are discussed in more detail in the Results section, we applied the following labels to the six identified COVID-19 case trajectories: “Least impacted by COVID-19 cases,” “April start to the epidemic,” “Smaller death peak in May,” “Early surge but brought down the curve,” “June surge to the epidemic,” and “Worst rates from the beginning.” The counties that were labeled “Least impacted by COVID-19 cases” had low rates of COVID-19 cases in each of the seven time periods ranging from an average of 4.3 cases per 100,000 to 22.7 cases per 100,000, whereas the peak case rates of the other five groups identified were much larger: 40.7, 89.4, 103.6, 130.6, and 179.8 cases per 100,000. The worst case group was labeled as the worst because the case rates were highest across all seven time periods ranging from an average of 30.8 to 179.8 cases per 100,000. Based on the observations from Table 1 and Fig. 4, we applied the following labels to the five identified COVID-19 death trajectories: “Least impacted by COVID-19 deaths,” “Smaller death peak in May,” “Smaller death peak in April and in late June,” “Larger death peak in May,” and “Most impacted by COVID-19 deaths with an April death peak.” The counties that were labeled “Least impacted by COVID-19 deaths” had low rates of COVID-19 deaths in each of the seven time periods ranging from an average of 0 deaths per 100,000 to 0.94 deaths per 100,000, whereas the peak death rates of the other four groups identified were much larger: 3.3, 5.1, 11.0, and 11.4 deaths per 100,000. The worst death group was labeled as the worst because the death rates were highest across all seven time periods ranging from an average of 1.2 to 11.4 deaths per 100,000. To visualize the spatial distribution of the trajectory membership, each county’s group membership was mapped using ArcGIS, version 10.6 (ESRI, Environmental Systems Research Institute Inc., Redlands, CA, USA).

2.2.2. Analysis of county-level correlates of COVID-19 trajectories
Following identification and characterization of COVID-19 trajectories in US counties, we utilized polytomous logistic regression to estimate Odds Ratios (OR) and 95% Confidence Intervals (CI) because the categorical outcome variable was the trajectory groups. In the analysis of COVID-19 cases, the odds of being in counties with an “April start to the epidemic,” “Early surge but brought down the curve,” “June surge to the epidemic,” and “Worst rates from the beginning” were in the numerators of the ORs and the odds of being in counties in the group that was “Least impact by COVID-19 cases” was the denominator. In the analysis of COVID-19 deaths, the odds of being in counties with a “Smaller death peak in May,” “Smaller death peak in April and in late June,” “Larger death peak in May,”

| Table 1 | Mean COVID-19 daily cases per 100,000 (SD) and mean COVID-19 deaths per 100,000 (SD) for each of the identified epidemic trajectories across the seven time periods. |
|--------|-------------------------------------------------|
| **COVID-19 case trajectory groups** | **n** | January 1–March 31 | April 1–April 15 | April 16–April 30 | May 1–May 15 | May 16–May 31 | June 1–June 15 | June 16–June 30 |
| Least impacted by COVID-19 cases | 333 | 10.7 (17.0) | 22.7 (34.6) | 30.0 (21.4) | 6.7 (10.2) | 7.1 (13.0) | 4.3 (5.2) | 11.1 (9.7) |
| April start to the epidemic | 310 | 0.1 (0.3) | 9.5 (11.9) | 39.0 (77.3) | 14.7 (18.9) | 19.7 (24.6) | 12.1 (24.6) | 30.6 (258.1) |
| May start to the epidemic | 374 | 0.01 (0.1) | 1.70 (3.3) | 0.99 (2.3) | 5.12 (17.9) | 8.30 (19.6) | 16.4 (31.6) | 40.7 (79.1) |
| Early surge, but brought down the curve | 275 | 19.7 (38.5) | 56.9 (87.2) | 89.4 (309.8) | 54.1 (59.0) | 43.0 (61.2) | 22.2 (21.1) | 15.7 (9.7) |
| June surge to the epidemic | 561 | 13.5 (22.2) | 23.9 (39.7) | 13.5 (19.5) | 11.9 (21.2) | 21.7 (27.9) | 43.3 (86.2) | 103.6 (158.1) |
| Worst rates from the beginning | 3288 | 30.8 (66.7) | 103.8 (168.6) | 143.0 (259.8) | 165.9 (432.8) | 155.0 (207.3) | 148.4 (260.6) | 179.8 (188.1) |
| **COVID-19 death trajectory groups** | | | | | | | | |
| Least impacted by COVID-19 deaths | 1394 | 0.17 (1.7) | 0 | 0 | 0 | 0.73 (5.1) | 0.94 (5.1) |
| Smaller death peak in May | 245 | 0.18 (1.2) | 0.003 (0.03) | 0.005 (0.06) | 3.1 (7.4) | 5.1 (8.7) | 3.2 (10.6) | 2.7 (7.2) |
| Smaller death peak in April and late June | 503 | 0.47 (2.6) | 3.3 (6.5) | 2.9 (6.9) | 0.96 (3.0) | 0.16 (0.9) | 0.84 (3.7) | 1.26 (4.2) |
| Larger death peak in May | 311 | 0.16 (0.8) | 0.71 (2.2) | 5.0 (10.5) | 8.1 (14.4) | 11.0 (20.2) | 5.6 (8.7) | 4.4 (10.7) |
| Most impacted by COVID-19 deaths with an April peak | 688 | 1.2 (3.0) | 8.1 (12.4) | 11.4 (15.7) | 10.0 (12.0) | 6.7 (8.3) | 4.3 (5.8) | 4.1 (6.6) |
Most impacted by COVID-19 deaths with an April death peak were relative to the odds of being in counties in the group that were “Least impacted by COVID-19 deaths.”

The independent variables were the county-level demographic factors (e.g., median age in the county, % female, % White, % Black, % Hispanic of any race), socioeconomic factors (e.g., % with a high school education, % below the poverty level), percentage of the county’s population enrolled in grades 1–8, high school, and in college, percentage without health insurance, percentage employed in the private service-producing industry and in the private goods-producing industry. In the COVID-19 death analyses, we examined the aforementioned variables as well as the percentage of the county’s adult population who smoke and who were obese. The Pearson’s correlation coefficients across these variables are presented in Supplemental Table 2. For variables that had correlation coefficients >0.5, we selected one of the two to include in the models to avoid multicollinearity issues. We estimated ORs of group membership for a 1 percentage-point increase in the independent variables, except for median age that was modeled for a 5-year increase in county median age. Sensitivity analyses were conducted to consider variables that were not selected due to high correlations. These sensitivity analyses are presented in Supplemental Figs. 1–5, and sensitivity analysis results adjusted for population size are presented in Supplemental Figs. 6 & 7.

3. Results

3.1. COVID-19 case trajectories and their correlates

Table 1 presents the average number of COVID-19 cases per 100,000 in each time period during January–June 2020 across counties in each of the six COVID-19 case groups. 333 counties were grouped together and labeled as “Least impacted by COVID-19 cases” with average cases per 100,000 starting at 10.7 in January–March and remaining rather stable through late June with 11.1 cases per 100,000. 310 counties were labeled as “April start to the epidemic” because the average number of cases per 100,000 went from 0.05 in January–March 2020 to 9.5 in early April then 39.0 in late April and continued to increase through June. 374 counties were grouped into a category with a “May start to the epidemic” because there were 0.01 cases per 100,000 in January–March, 1.7 in early April, 1.0 in late April, but 5.1 in early May, 8.3 in late May, 16.4 in early June and 40.7 cases per 100,000 in late June. 561 counties were labeled as having a “June surge to the epidemic” because the cases per 100,000 ranged from 11.9 to 23.9 cases per 100,000 between January through May but steeply increased in June to an average of 103.6 cases per 100,000. 275 counties were classified as having an “Early surge, but brought down the curve” because cases peaked in late April with an average of 89.4 cases per 100,000 that declined to 15.7 cases per 100,000 in late June. The majority of the counties (n = 1288) were labeled as having the “Worst rates from the beginning” because rates started with an average of 30.8 cases per 100,000 with a steep incline in April at 103.8 cases per 100,000 then continued to increase to 179.8 cases per 100,000 in late June. Histograms are presented in Fig. 1 showing the daily COVID-19 cases of US counties in each of the six epidemic trajectory groups, which supports the labels assigned to the groups identified by Proc Traj based on tertiles of rates. For example, in Fig. 1 the group of counties “Least impacted by COVID-19 cases” have the lowest daily case rates throughout January–June 2020, while the counties with the “Worst rates from the beginning” have the highest daily case rates starting in March and remains high throughout June.

Fig. 2 shows the spatial distribution of the case trajectory groups. The trajectory groups appear heterogeneously distributed across the United States, though the trajectory group with worst case rates were mainly
concentrated in densely populated South, the Northeast, and Midwest regions. Sparsely populated western parts of the country were mainly categorized as either April or May start to the epidemic or June surge, except for areas in the southwest border that suffered the worst infection rates.

Descriptive statistics of the counties’ demographic, social, and economic characteristics by COVID-19 trajectory group are presented in Supplemental Table 3. Adjusted OR estimates comparing odds of group membership (“April start to the epidemic,” “May start to the epidemic,” “Early surge but brought down the curve,” “June surge to the epidemic,” and “Worst rates from the beginning”) to the odds of being in counties “Least impacted by COVID-19 cases” for every 1% increase in each covariate are presented in Fig. 3 and Supplemental Table 4. In general, several covariates including older median age, higher poverty, more non-family households and higher high school graduates were associated with lower odds of being in the case trajectory groups “April start to the epidemic,” “May start to the epidemic,” “Early surge but brought down the curve,” “June surge to the epidemic,” and “Worst rates from the beginning.” Female sex, Black race, Hispanic ethnicity, lack of health insurance, and employment in the private sector industries had generally positive associations with case trajectory groups. Meanwhile, covariates that were generally not associated with case trajectory groups included high school enrollment, college enrollment, White race, and employment in the public sector.

More specifically, counties with an older median age of the population had lower odds of being in counties with the worst rates trajectory, the June surge, an early surge but brought down the curve, and having an April start to the epidemic (Fig. 3 and Supplemental Table 4). This may be highlighting the role of younger populations in the spread and decline of COVID-19 during the first wave. Interestingly, age was not associated with counties that had a May start to the epidemic (OR = 1.06 95% CI 0.88, 1.27) that may be reflecting the May benefits of the April stay-at-home orders. For counties with 1% greater proportion of the population who were female, there were higher odds of being in counties with an April start to the epidemic (OR = 1.10 95% CI 1.02, 1.18), the June surge to the epidemic (OR = 1.12 95% CI 1.04, 1.21), and the worst county rates (OR = 1.25 95% CI 1.16, 1.35) but sex was not associated with being in counties with a May start to the epidemic when the benefits of the April stay-at-home orders were in effect (OR = 1.00 95% CI 0.93, 1.07). Similarly, counties with a higher proportion of Black and Hispanic populations had higher odds of being in counties with an April start to the epidemic, June surge to the epidemic, and counties with the worst rates (Fig. 3 and Supplemental Table 4), which may reflect the persistent racial/ethnic health disparities and inequities (Centers for Disease Control and Prevention (CDC), 2020; Poteat et al., 2020) and the disproportionate effect COVID-19 has had on Black and Hispanic populations in the US who make up a large proportion of the essential workforce.

Relative to the group least impacted by COVID-19 in the first wave, counties with a greater proportion of the adult population with a high-school education (Fig. 3 & Supplemental Table 4) and college education (Supplemental Fig. 1) had lower odds of being in counties with an April start to the epidemic, and had lower odds of being in the worst infection rate counties. Enrollment in high-school (Fig. 3) and college (Supplemental Fig. 2) were generally not associated with the different COVID-19 case trajectories; however, enrollment in grades 1–8 was associated with higher odds of being in counties with the early epidemics in April (OR = 1.03 95% CI 1.00, 1.06) and in May (OR = 1.04 95% CI 1.01, 1.07). Counties with a higher proportion living in non-family households had lower odds of being in the June surge designation, lower odds of being in the early surge counties that brought down the curve, and lower odds of being in the worst rate counties. Counties with a greater percentage of the population below the poverty level had lower odds of being in each group membership. Counties with a higher proportion without health insurance had higher odds of being in counties with an April start, a May start, and a June surge. Lastly, employment in the private service-producing and goods-producing industries was associated with higher odds of being in the June surge counties and the worst rate counties, while employment in the public sector was not associated with case trajectory groups (Fig. 3 & Supplemental
3.2. **COVID-19 death trajectories and their correlates**

The average number of COVID-19 deaths per 100,000 across counties in each of the five COVID-19 death groups are presented in Table 1. The majority of US counties (n = 1394) were classified as “Least impacted by COVID-19 deaths,” with nearly 0 deaths per 100,000 through May that increased to <1 death per 100,000 through the June time periods. 245 counties were labeled as having a “Smaller death peak in May” because there were nearly 0 deaths per 100,000 through April, that peaked to 5.1 deaths per 100,000 in late May and fell back down to 2.7 deaths per 100,000 in late June. 503 counties were classified as having a “Smaller death peak in April and late June” because the deaths per 100,000 were on average 0.5 in January–March that peaked to ~3 deaths per 100,000 in early and late April, decreased in May and began to rise again in June to an average of 1.3 deaths per 100,000. 311 counties were classified as “Larger death peak in May” because deaths per 100,000 started low in January through early April with <1 death per 100,000, but then steadily increased to a peak of 11.0 deaths per 100,000 in late May. 688 counties were described as “Most impacted by COVID-19 deaths with an April death peak” because deaths per 100,000 started off in January–March with the highest death rates compared to other groups and increased sharply through April and early May to an average of 8.1–11.4 deaths per 100,000. Histograms showing the distribution of daily COVID-19 deaths among US counties classified in five epidemic trajectories are presented in Fig. 4.

Fig. 5 shows the spatial distribution of the results of death trajectory classifications. Small clusters of counties that were classified as the “Most impacted by COVID-19 deaths with an April death peak” or “Larger death peak in May” were mainly observed in the eastern part of the US and the southwestern border. While the death trajectory groups were heterogeneous distributed across the US, the vast majority of the inland areas were predominately characterized as counties least impacted by COVID-19 deaths during the first wave.

Descriptive statistics for demographic, social, and economic characteristics for each COVID-19 death trajectory group are presented in Supplemental Table 5. Adjusted ORs estimates comparing odds of group membership (“Smaller death peak in May,” “Smaller death peak in April and in late June,” “Larger death peak in May,” and “Most impacted by COVID-19 deaths with an April death peak”) to the odds of being in counties “Least impacted by COVID-19 deaths” for every 1% increase in a covariate are presented in Fig. 6 and Supplemental Table 6. Several covariates were inversely associated with death trajectory groups: median age, poverty, enrollment in grades 1–8, non-family households, lack of health insurance, and White race. Female sex, Black race, Hispanic ethnicity, college enrollment, employment in the private sector industries and smoking had generally positive associations with death trajectory groups. Meanwhile, covariates that were generally not associated with death trajectory groups included high school enrollment, adult education levels, employment in the public sector and obesity.

More specifically, counties that had an older median age had lower odds of being in the group with the worst death trajectories (OR = 0.75 per 5 year increase in age 95% CI 0.64, 0.88) and had lower odds of being in the group with the largest death peak in May (OR = 0.73 per 5 year increase in age 95% CI 0.60, 0.89) (Fig. 6 and Supplemental Table 6), again highlighting the role of younger counties having more widespread infections and impacting larger lethal outbreaks. Counties with a 1% higher proportion of females had higher odds of being in the small death peak in May group (OR = 1.06 95% CI 1.00, 1.13), the small death peak in June group (OR = 1.09 95% CI 1.03, 1.15), the worst counties (OR = 1.21 95% CI 1.13, 1.30), and the large death peak in May group (OR = 1.11 95% CI 1.04, 1.19). Counties with a 1% higher proportion of Black populations also had higher odds of being in the small death peak in May group (OR = 1.06 95% CI 1.00, 1.13), the small death peak in June group (OR = 1.09 95% CI 1.03, 1.15), the worst counties (OR = 1.21 95% CI 1.13, 1.30), and the large death peak in May group (OR = 1.11 95% CI 1.04, 1.19). Counties with a 1% higher proportion of Black populations also had higher odds of being in the small death peak in May group (OR = 1.06 95% CI 1.00, 1.13), the small death peak in June group (OR = 1.09 95% CI 1.03, 1.06), the worst counties (OR = 1.07 95% CI 1.05, 1.09), and the large death peak in May group (OR = 1.07 95% CI 1.04, 1.10). Counties with a higher percentage of Hispanic populations had higher odds of being in the trajectory group most impacted by COVID-19 mortality (OR = 1.03 95% CI 1.02, 1.05) and higher odds of a small death peak in May. Counties with
Fig. 4. COVID-19 daily deaths per 100,000 for every county within each of the five identified trajectory groups.
a higher percentage of the population employed in the private service sector and the goods-producing industry had higher odds of being in the group with a small death peak in May and in June, and were more markedly associated with the worst death rate counties (OR for service industry = 1.15 95% CI 1.12, 1.18; OR for goods-producing industry = 1.09 95% CI 1.06, 1.12), and the large death peak in May counties (OR for service industry = 1.09 95% CI 1.06, 1.12; OR for goods-producing industry = 1.08 95% CI 1.06, 1.11). While county-level adult obesity was not associated with COVID-19 death trajectories, counties with a higher percentage of the adult population who smoked had higher odds of being in the COVID-19 death trajectories with a small death peak in May and in June, the worst death rate counties (OR = 1.11 95% CI 1.04, 1.17) and the large death peak in May counties (OR = 1.07 95% CI 1.00, 1.15) (Fig. 6 and Supplemental Table 6).

**Fig. 5.** Spatial distribution of the five identified trajectory groups for COVID-19 death rates.

**Fig. 6.** Mutually adjusted odds ratios (ORs) of COVID-19 death trajectories and 95% confidence intervals (CI) in relation to demographic and socioeconomic county-level covariates in 3141 US counties. ORs are comparing odds relative to that of the trajectory group that was “Least impacted by COVID-19 deaths.” ORs are for a 1% increase, except for median age (per 5 year increase).
County-level high-school education was generally not significantly associated with COVID-19 death trajectories; however, we did observe that counties with higher percentage of high-school educated adult populations had lower odds of being in the trajectory group of a large death peak in May (OR = 0.96 95% CI 0.92, 1.00). Results were similar when including the percentage of college-educated (Supplemental Fig. 3). County enrollment in high school was not associated with COVID-19 death trajectories, but enrollment in Grades 1–8 was inversely associated with COVID-19 death trajectories (Fig. 6) and college enrollment was positively associated with COVID-19 death trajectories (Supplemental Fig. 4). County-level poverty, lack of health insurance and living in non-family households were inversely associated with COVID-19 death trajectories (Fig. 6). Excluding poverty from the models did not change the results (Supplemental Fig. 5). While it is not clear, these counterintuitive findings may reflect reduced mobility patterns and/or social contacts in populations with lower incomes, lack of health insurance and those living in single non-family households. Adjustment for population size did not materially change the results (Supplemental Fig. 7).

4. Discussion

In summary, there were several place-based correlates of COVID-19 case and death trajectories that are in line with the current state of knowledge of population transmission dynamics and offer additional insight. The place-based characteristics that were observed to be associated with worse death trajectories included counties that were younger, had higher proportion of female and Black populations, more workers in the private sector, and adult smokers. Unexpectedly, we observed that counties with greater poverty and more of the population without health insurance were less likely to be the worst COVID-19 death trajectory groups, which has also been observed in other place-based studies (Andersen et al., 2021; Karaye and Horney, 2020; Sun et al., 2020a; Zhang and Schwartz, 2020; Mollalo et al., 2020). Generally, the estimated ecological associations between these social, economic and demographic factors and death trajectories were stronger for counties with the largest death peaks and those that were worst off from the beginning as opposed to counties with smaller death peaks.

Age has been inversely associated with COVID-19 cases and death in other ecological studies (Andersen et al., 2021; Sun et al., 2020a). Younger age populations are less likely to show COVID-19 symptoms and can insidiously contribute largely to spread that is thought to be a major contributor to this pandemic (Kasper et al., 2020; Sun et al., 2020b; Goldstein et al., 2020). Sun et al. estimated that more than half of the spread in Hunan, China was transmitted by pre-symptomatic individuals, and there was no evidence of differences in COVID-19 infectivity by age (Sun et al., 2020b). Similarly, while younger people are less susceptible to becoming severely symptomatic, serological studies suggest high incidence of the virus in people less than 35 years old (Goldstein et al., 2020). Interestingly, we observed that younger counties were more likely to be in the worst death and case trajectory groups, but the age association washed out for counties with case and death trajectories that peaked in May (OR of May start to the epidemic for a 5 year increase in age = 1.06 95% CI 0.88, 1.27; OR of small death peak in May for a 5 year increase in age = 0.97 95% CI 0.80, 1.17). This wash-out of the age-association is likely reflecting the 2–4 week delay in observing the benefits of the stay-at-home orders that were in place in April 2020 that prevented substantial mobility for all ages preventing cases and deaths in May. Additionally, we observed that school enrollment in grades 1–8 was associated with higher odds of having an early start to the epidemic in April and May, which may reflect the asymptomatic spread early on in the epidemic. Initiatives to vaccinate youth populations will likely contribute to a decline in cases and deaths caused by asymptomatic spread.

Counties with a greater proportion of females, Black and Hispanic populations were more likely to be in areas with the worst COVID-19 case and death trajectories. Prior ecological studies also have observed correlations between COVID-19 case and death counts with higher percentage of Black populations (Andersen et al., 2021; Karaye and Horney, 2020; Sun et al., 2020a; Mollalo et al., 2020), females (Mollalo et al., 2020), and percentage nurse practitioners (Mollalo et al., 2020). The observations in the current study appear to reflect the demographic profile of essential workers with high-risk of exposure to COVID-19 (Hawkins, 2020; Van Houtven et al., 2020), may reflect differences in social and mobility patterns during a time when personal protective equipment was difficult to obtain, and the disproportionate burden of COVID-19 mortality in these populations. Compared to Whites, Black and Hispanic populations are more likely to be employed in high-COVID-19 risk industries and be employed in occupations with more frequent close-proximity encounters such as essential industries in healthcare and transportation (Hawkins, 2020) and long-term care workers (Van Houtven et al., 2020). Furthermore, employment in private sector industry like service-producing and good-producing industries were associated with the worst COVID-19 death trajectories during the first wave; on the contrary, employment in the public sector was not associated with COVID-19 trajectories. This may be reflecting differences in precautionary measures in occupational settings and the differences in occupations that were able to work from home. Communities of color and those with many in healthcare, the service and goods industries will benefit from increasing access to vaccination to reduce health inequities in viral exposure and will benefit from communication on appropriate guidelines from city, state and federal public health authorities for proper implementation of non-pharmaceutical interventions (ex. social distancing, mask wearing, crowd size, ventilation, etc.) particularly in workplace settings that could benefit from federal financial assistance to implement the guidelines.

County-level poverty, lack of health insurance and living in non-family households were associated with lower odds of being the worse COVID-19 case and death trajectories groups. This was consistent with other COVID-19 ecological studies of income and percent uninsured (Sun et al., 2020a; Mollalo et al., 2020). While it is not clear, this finding may reflect reduced mobility patterns in those with lower incomes and those without health insurance. Workplace mobility decreased dramatically across the US in April 2020 (Google LLC, 2020), and COVID-19 related unemployment particularly affected lower wage-earning jobs. Interestingly, we observed that counties with more non-family homes were less likely to be in the worst case and death trajectory groups. Conversely, counties with more family homes were more likely to be in the worst case and death trajectory groups, suggesting that household transmission was also contributing to the spread and rise of COVID-19 even during the first wave of the pandemic. In Hunan, China, Sun et al. observed that household contacts posed the greatest risk of transmission, particularly during lock-down periods (Sun et al., 2020b), that supports these ecological observations of household structure.

Lastly, adult smoking percentages were strongly associated with worse COVID-19 death trajectories. Smoking causes numerous disease outcomes including diabetes, cardiovascular disease, respiratory and cancer outcomes that are also risk factors for having COVID-19 symptoms, being hospitalized and dying from COVID-19 (CDC, 2020c). While cigarette smoking has decreased over time, still an estimated 20% of US adults smoke (Cramer, 2018), and electronic cigarette use was increasing prior to the pandemic, particularly in younger age groups with an estimated 19.6% of high school students and 4.7% of middle school students reported current use of e-cigarettes (Wang, 2020). Whether smoking itself directly impacts COVID-19 remains to be evaluated prospectively (Samet, 2020; Alqahtani et al., 2020), but smoking has aggravated disease outcomes that are in turn contributing to COVID-19 hospitalizations and mortality. Smoking continues to be a major public health threat and additional resources to promote tobacco cessation and prevent tobacco use remain vital.

There are several limitations and strengths of the current study. Because this study is ecological, the findings can assist in hypothesis
Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. NCD received support from the University of Louisville Center for Integrative Environmental Health Sciences (CIEHS) P30 ES030283. This manuscript is the responsibility of the authors and does not represent the official views of the National Institutes of Health.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2021.147495.

References

Alqhtani, J.S., Oyelade, T., Alkhalid, A.M., Alghandi, S.M., Almehmadi, M., Alqhtani, A.S., et al., 2020. Prevalence, severity and mortality associated with COPD and smoking in patients with COVID-19: a rapid systematic review and meta-analysis. PloS One 15, e0233147.
Andersen, L.M., Harden, S.R., Sugg, M.M., Rinkle, J.D., Lundquist, T.E., 2021. Analyzing the spatial determinants of local COVID-19 transmission in the United States. Sci. Total Environ. 754, 142396.
APM Research Lab. (2021). Color of coronavirus. https://www.apmresearchlab.org/covid-deaths-by-race.
Bassett, M.T., Chen, J.T., Krieger, N., 2020. Variation in racial/ethnic disparities in COVID-19 mortality by age in the United States: a cross-sectional study. PLoS Med. 17, e1003462.
CDC. (2020a). Coronavirus disease 2019 (COVID-19) in the U.S. [Internet]. Cent. Dis. Control Prev. Available from: https://www.cdc.gov/covid-data-tracker.
CDC. (2020b). COVID data tracker: daily and total trends [Internet]. Cent. Dis. Control Prev. Available from: https://covid19.cdc.gov/covid-data-tracker/
CDC. (2020c). COVID-19 and your health [internet]. Cent. Dis. Control Prev. [cited 2021 Feb 4]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html.
Centers for Disease Control and Prevention (CDC). (2020). Coronavirus disease 2019 (COVID-19) in racial and ethnic minority groups. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/racial-ethnic-minorities.html.
County Health Rankings & Roadmaps, 2021. 2020 County health rankings key findings report [internet]. Available from: https://www.countyhealthrankings.org/reports/2020-county-health-rankings-key-findings-report.
Cramer, M.R., 2014. Tobacco product use and cessation indicators among adults — United States, MMWR Morb Mortal Wkly Rep [Internet]. 2019 [cited 2020 Dec 17]; 68. Available from: https://www.cdc.gov/mmwr/volumes/68/wr/mm6814a2.htm.
Databases, 2021. Tables & calculators by subject [internet]. Available from: https://www. bls.gov/data.
van, Dorn A., Cooney, R.E., Sabin, M.L., 2020. COVID-19 exacerbating inequalities in the US. Lancet Lond. Engl. 395, 1243–1244.
Goldstein, E., Lipsitch, M., Cevik, M., 2020. On the effect of age on the transmission of SARS-CoV-2 in households, schools and the community. J. Infect. Dis. [internet]. [cited 2020 Dec 2]; Available from: https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jjaa619/5943164.
Google LLC “Google COVID-19 community mobility reports”. https://www.google.com/ covid19/mobility/ Accessed: December 20, 2020.
Hamidi, S., Saurou, S., Ewng, R., 2020. Does density aggravate the COVID-19 pandemic? J. Am. Plann. Assoc. Routledge 86, 495–509.
Hawkins, D., 2020. Differential occupational risk for COVID-19 and other infection exposure according to race and ethnicity. Am. J. Ind. Med. 63, 817–820.
Jones, B.L., Nagington, D.S., 2007. Advances in group-based trajectory modeling and an SAS procedure for estimating them. Sociol. Methods Res. 35, 542–571.
Jones, B.L., Nagington, D.S., Rueter, K., 2001. A SAS procedure based on mixture models for estimating developmental trajectories. Sociol. Methods Res. 29, 374–393. SAGE Publications Inc.
Karaye, I.M., Horney, J.A., 2020. The impact of social vulnerability on COVID-19 in the U.S.: an analysis of spatially varying relationships. Am. J. Prev. Med. 59, 317–325.
Kaiser, M.R., Geibe, J.R., Sears, C.L., Riegspeldos, A.J., Lise, T., Van Thun, A.M., et al., 2020. An outbreak of Covid-19 on an aircraft carrier. N. Engl. J. Med. 382;820.
Lefondrè, K., Abrahamowicz, M., Regeasse, A., Hawker, G.A., Badley, E.M., McCusker, J., et al., 2004. Statistical measures were proposed for identifying longitudinal patterns of change in quantitative health indicators. J. Clin. Epidemiol. 57, 1049–1062.
Marvel, S.W., House, J.S., Wheeler, M., Song, K., Zhou, Y.-H., Wright, F.A., et al., 2021. The COVID-19 pandemic vulnerability index (PVI) dashboard: monitoring county-level
vulnerability using visualization, statistical modeling, and machine learning. Environ. Health Perspect. 129, 17701.

Mollalo, A., Vahedi, B., Rivera, K.M., 2020. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Sci. Total Environ. 728, 138884.

Potera, T., Millett, G.A., Nelson, L.E., Beyrer, C., 2020. Understanding COVID-19 risks and vulnerabilities among black communities in America: the lethal force of syndemics. Ann. Epidemiol. 47, 1–3.

Samet, J.M., 2020. Tobacco products and the risks of SARS-CoV-2 infection and COVID-19. Nicotine Tob. Res. 22, S93–S95.

Song, M., Hu, F.B., Wu, K., Must, A., Chan, A.T., Willett, W.C., et al., 2016a. Trajectory of body shape in early and middle life and all cause and cause specific mortality: results from two prospective US cohort studies. BMJ. 353, i2195.

Song, M., Willett, W.C., Hu, F.B., Spiegelman, D., Must, A., Wu, K., et al., 2016b. Trajectory of body shape across the lifespan and cancer risk. Int. J. Cancer 138, 2383–2395.

Sun, F., Matthews, S.A., Yang, T.-C., Hu, M.-H., 2020a. A spatial analysis of the COVID-19 period prevalence in U.S. counties through June 28, 2020: where geography matters? Ann. Epidemiol. 52, 54–59.e1.

Sun, K., Wang, W., Gao, L., Wang, Y., Luo, K., Ren, L., et al., 2020b. Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. Science [internet]. Available from: https://science.sciencemag.org/content/early/2020/11/23/science.abe2424.

Susser, M., 1994. The logic in ecological: I. The logic of analysis. Am. J. Public Health 84, 825–829.

Sylvestre, M.-P., McCusker, J., Cole, M., Regesae, A., Belzile, E., Abrahamowicz, M., 2006. Classification of patterns of delirium severity scores over time in an elderly population. Int. Psychogeriatr. 18, 667–680.

The New York Times, 2021. nytimes/covid-19-data [Internet]. Available from: https://github.com/nytimes/covid-19-data.

U.S. Census Bureau, 2021. County population totals: 2010–2019 [internet]. Available from: https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-total.html.

U.S. Census Bureau, 2018. American Community Survey, 2018 American community survey 5-year estimates, tables DP02, DP03, DP04, DP05; using data.census.gov. https://data.census.gov/cedsci/

U.S. Census Bureau, 2021. North American Industry Classification System (NAICS) [internet]. Available from: https://www.census.gov/naics/

Van Houtven, C.H., DePasquale, N., Coe, N.B., 2020. Essential long-term care workers commonly hold second jobs and double- or triple-duty caregiving roles. J Am Geriatr Soc [Internet]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7267626/

Wang, T.W., 2020. E-cigarette use among middle and high school students—United States. MMWR Morb Mortal Wkly Rep [Internet]. Available from: https://www.cdc.gov/mmwr/volumes/69/wr/mm6937e1.htm.

WHO, 2020. Coronavirus disease (COVID-19) dashboard. [Internet]. Available from: https://covid19.who.int.

Yuan, Z., Yang, Y., Wang, C., Liu, J., Sun, X., Liu, Y., et al., 2018. Trajectories of long-term normal fasting plasma glucose and risk of coronary heart disease: a prospective cohort study. J. Am. Heart Assoc. 7.

Zhang, C.H., Schwartz, G.G., 2020. Spatial disparities in coronavirus incidence and mortality in the United States: an ecological analysis as of May 2020. J. Rural. Health 36, 433–445.