




Disparities in case frequency and mortality of coronavirus disease 2019 (COVID-19) among various states in the United States

Rohit S. Loomba^{a,b} , Gaurav Aggarwal^c, Saurabh Aggarwal^d, Saul Flores^{e,f} , Enrique G. Villarreal^g , Juan S. Farias^g and Carl J. Lavie^h

^aAdvocate Children's Hospital, Chicago, IL, USA; ^bChicago Medical School, Rosalind Franklin University of Medicine and Science, Chicago, IL, USA; ^cJersey City Medical Center, Jersey City, NJ, USA; ^dUnityPoint Health, Des Moines, IA, USA; ^eTexas Children's Hospital, Houston, TX, USA; ^fBaylor College of Medicine, Houston, TX, USA; ^gTecnologico de Monterrey, Escuela de Medicina y Ciencias de la Salud, Monterrey, Mexico; ^hOchsner Medical Center, New Orleans, LA, USA

ABSTRACT

Objective: To utilize publicly reported, state-level data to identify factors associated with the frequency of cases, tests, and mortality in the USA.

Materials and methods: Retrospective study using publicly reported data collected included the number of COVID-19 cases, tests and mortality from March 14th through April 30th. Publicly available state-level data was collected which included: demographics comorbidities, state characteristics and environmental factors. Univariate and multivariate regression analyses were performed to identify the significantly associated factors with percent mortality, case and testing frequency. All analyses were state-level analyses and not patient-level analyses.

Results: A total of 1,090,500 COVID-19 cases were reported during the study period. The calculated case and testing frequency were 3332 and 19,193 per 1,000,000 patients. There were 63,642 deaths during this period which resulted in a mortality of 5.8%. Factors including to but not limited to population density (beta coefficient 7.5, $p < .01$), transportation volume (beta coefficient 0.1, $p < .01$), tourism index (beta coefficient -0.1 , $p = .02$) and older age (beta coefficient 0.2, $p = .01$) are associated with case frequency and percent mortality.

Conclusions: There were wide variations in testing and case frequencies of COVID-19 among different states in the US. States with higher population density had a higher case and testing rate. States with larger population of elderly and higher tourism had a higher mortality.

KEY MESSAGES

- There were wide variations in testing and case frequencies of COVID-19 among different states in the USA.
- States with higher population density had a higher case and testing rate.
- States with larger population of elderly and higher tourism had a higher mortality.

ARTICLE HISTORY

Received 17 June 2020

Revised 2 September 2020

Accepted 9 September 2020

KEYWORDS

COVID-19; coronavirus; climate; risk factors; epidemiology



Introduction

The coronavirus disease 2019 (COVID-19) has spread worldwide after its onset in Wuhan, China, reaching over three million cases [1]. In the United States (US), over seven million tests have been performed with over one million tests resulting positive. This disease has resulted in over 70,000 deaths in the US and a case fatality rate at approximately 6.2%. Recent reports have shown differences in case and fatality rates between nearby localities and thought to be likely due to differences in demographic and socioeconomic factors [2]. Furthermore, differences in climate, access

to healthcare and adherence to social distancing have also been hypothesized to affect these rates [3–5]. However, factors mediating the differences in case frequency and testing frequency between the states in the USA have not been formally assessed. The aim of this study is to use publicly available, state-level data to determine the factors associated with COVID-19 percent mortality and case and testing frequency.

Materials and methods

This study utilized publicly available, deidentified, state-level data and so no institutional review board

CONTACT Carl J. Lavie  clavie@ochsner.org  John Ochsner Heart and Vascular Institute, Ochsner Medical Center, 1514 Jefferson Hwy, New Orleans 70121, LA, USA

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

approval was required or sought. The study was performed in accordance with Declaration of Helsinki.

Variable identification and data collection

Absolute counts for COVID-19 cases, tests and mortality were obtained from Worldometer [6]. Data were collected from March 13th, 2020 through April 30th, 2020. These dates were selected as March 13th was the first day when these data were publicly reported, and April 30th was the last full day prior to data collection. The number of cases and tests were then normalized to the specific state's population to develop a frequency per 1,000,000 population. Any reference in this manuscript to "case frequency" or "testing frequency" refers to the normalized values in this manner, unless explicitly stated otherwise.

Next, data were collected at a state-wide level to help characterize the population, environment, and infrastructure in the state. The data sources are listed in Supplementary file 1. The selected variables were identified by a literature review of the factors that impact the frequency and severity of viral illnesses, including COVID-19. The following variables were collected: age, gender, underinsured population, ethnicity, influenza vaccination status, population density (persons per square mile), urban air quality rank (lower number signifying better air quality), drinking water quality rank (lower number signifying better drinking water quality), ultraviolet index, precipitation (in inches), temperature (in degree Fahrenheit), average household income (in US\$), per capita spending on healthcare (in US\$) and high school graduation rate. To capture comorbid conditions, the prevalence of obesity, prevalence of smoking, prevalence of cocaine abuse, prevalence of marijuana abuse, alcohol consumption (gallons per person per year), prevalence of asthma, prevalence of diabetes, prevalence of chronic obstructive pulmonary disease, prevalence of myocardial infarction, prevalence of coronary artery disease, prevalence of hypertension, prevalence of hyperlipidaemia and presence of inactivity were collected. To capture immunosuppressed states, the annual incidence of new cancer and HIV cases per 100,000 population were collected. Finally, the social distancing score by global positioning satellite data, public transportation volume, number of incarcerated inmates, number of nursing home residents, and tourism rank (lower number implies more tourists) were collected. Of these, ultraviolet index, temperature and precipitation were averages for March and April of 2020, whereas the remainder of data were collected from

the most recent iteration for each state. Much of the data was collected from government sources, such as the Centres for Disease Control and the complete list of sources is provided

The collected data represents state-level and not patient-level data. The endpoints were divided amongst the authors and collected by everyone. The data for each endpoint was then verified by another author who did not primarily collect the data. Finally, values in the top and bottom 10th percentile were identified and verified by a third author.

Statistical analyses

As the data was collected for each state and intended for state-level analyses, the absolute number of COVID-19 cases, tests, and mortality were converted to a frequency using the state population. The frequencies for all the endpoints were calculated per 1,000,000 population. The case frequencies were then used as the dependent variables in a series of single-independent variable linear regressions to determine the univariate association between case frequency and the other variables previously defined and served as the univariate analyses. Next, a stepwise multivariate regression was conducted with p -value of .05 or less required for inclusion into the final model. Of the resulting models, the one with the highest R-squared value was selected as the final model. The same process was repeated for testing frequency and percent mortality as the dependent variable. Collinearity analyses were run with all multivariate regressions.

All statistical analyses were done using the user-coded, syntax-based interface of SPSS Version 23.0. A p -value of .05 or less was considered statistically significant. The use of the word significant throughout the manuscript refers to "statistically significant" unless explicitly specified otherwise. All statistical analyses were done at the state-level with state-level data. Analyses were not conducted at a patient-level with patient-level data. The subjects here were the 50 states. The age, gender and comorbidity prevalence are not based on patient-specific data but rather the state prevalence.

Results

COVID-19 cases, testing and mortality

A total of 1,090,500 COVID-19 cases were reported in the study period. This resulted in a case frequency of approximately 3332 per 1,000,000 patients (3.3%). In the same period, 6,299,143 tests were done, resulting in a testing frequency 19,193 per 1,000,000 patients

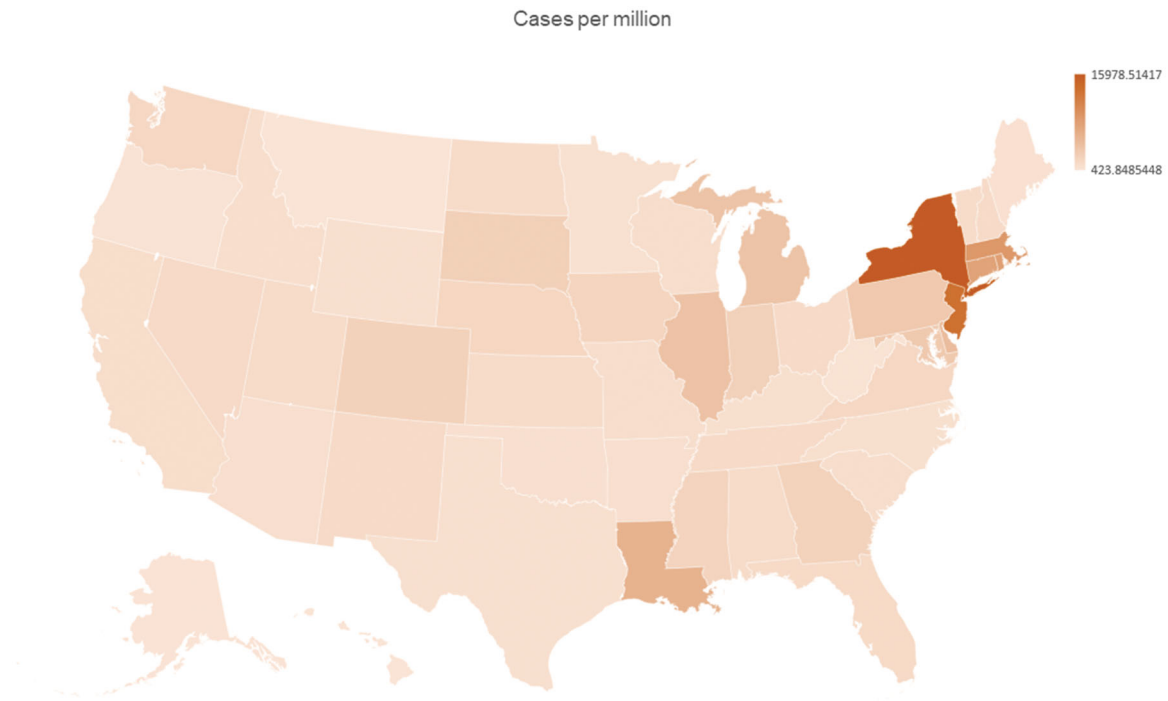


Figure 1. Cases per million in the United States divides by state.

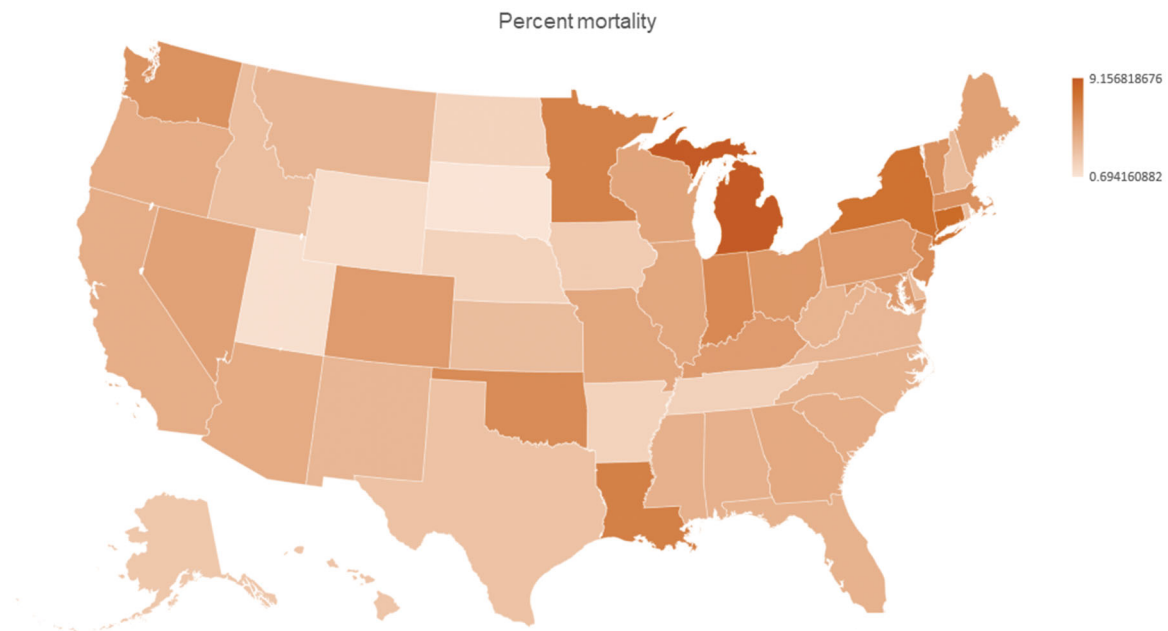


Figure 2. Percent mortality in the United States divides by state.

out of which 17.3% were reported positive. There were 63,642 deaths during this period which resulted in a mortality of 5.8%. **Figures 1** and **2** demonstrate the case frequency and the percent mortality by state.

COVID-19 case frequency, univariate analyses

The following factors were associated with greater case frequency on univariate linear regression

analyses: female gender (beta-coefficient 1,095.8, $p = .03$), higher population density (beta-coefficient 8.9, $p < .01$), lower ultraviolet index (beta coefficient -825.3 , $p = .03$), lower prevalence of obesity (beta coefficient -248.3 , $p = .03$), lower prevalence of uninsured (beta coefficient -298.6 , $p = .01$), higher frequency of other race (beta coefficient 38,638.7, $p = .01$), lower prevalence of current smokers (beta coefficient -294.5 , $p = .02$), higher per capita health

Table 1. Univariate analyses for factors associated with COVID-19 illness in the United States.

	Cases per million	Tests per million	Mortality percent
Age	194.0 (−185.8 to 573.9, $p = .31$)	−61.7 (−1,292.3 to 1,168.9, $p = .92$)	0.2 (0.1 to 0.4, $p = .01$)*
Gender	1095.8 (57.4 to 2,134.3, $p = .03$)*	1065.2 (−2,401.5 to 4,532.1, $p = .54$)	0.5 (−0.1 to 1.1, $p = .05$)
Race, White	−4,715.4 (−11,765.1 to 2,334.2, $p = .18$)	−8,401.8 (−31,287.3 to 14,483.6, $p = .46$)	−1.6 (−5.8 to 2.4, $p = .41$)
Race, Black	7,312.7 (−2,051.3 to 16,676.9, $p = .12$)	−1,715.4 (−32,483.3 to 29,052.3, $p = .91$)	4.3 (−1.0 to 9.7, $p = .11$)
Race, Native American	−23,515 (−54,374.7 to 7,342.9, $p = .13$)	28,602.2 (−72,344.9 to 129,549.3, $p = .57$)	−16.6 (−34.2 to 0.9, $p = .06$)
Race, Asian	6,237.8 (−10,201.4 to 22,677.1, $p = .44$)	15,722.4 (−37,084.8 to 68,529.6, $p = .55$)	1.0 (−8.5 to 10.6, $p = .82$)
Race, Islander	−30,422.1 (−94,071.5 to 33,227.3, $p = .34$)	20,730.6 (−185,121.7 to 226,583.0, $p = .84$)	−20.5 (−57.2 to 16.1, $p = .26$)
Race, other race	38,638.7 (8,221.4 to 69,056.1, $p = .01$)*	87,436.9 (−13,313.1 to 188,186.9, $p = .08$)	9.4 (−9.0 to 28.0, $p = .30$)
Race, multiple race	−13,133.5 (−41,037.2 to 14,770.1, $p = .34$)	5,794.8 (−84,447.0 to 96,036.7, $p = .89$)	−4.6 (−20.8 to 11.6, $p = .57$)
Uninsured	−298.6 (−544.2 to −52.9, $p = .01$)*	−900.9 (−1,693.8 to −107.9, $p = .02$)*	−0.1 (−0.3 to −0.1, $p = .02$)*
Obesity	−248.3 (−477.1 to −19.6, $p = .03$)*	−577.5 (−1,327.6 to 172.5, $p = .40$)	−0.1 (−0.2 to 0.1, $p = .39$)
Asthma	94.2 (−665.3 to 853.8, $p = .80$)	1,845.9 (−530.2 to 4,222.2, $p = .12$)	0.4 (−0.1 to 0.8, $p = .04$)
Diabetes	−143.6 (−623.5 to 336.2, $p = .55$)	−394.9 (−1,934.5 to 1,144.6, $p = .60$)	0.1 (−0.2 to 0.3, $p = .69$)
Chronic obstructive pulmonary disease (COPD)	−254.7 (−665.7 to 156.2, $p = .21$)	−557.3 (−1,885.9 to 771.2, $p = .40$)	0.1 (−0.1 to 0.3, $p = .60$)
Myocardial infarction (MI)	−586.2 (−1,427.6 to 255.1, $p = .16$)	−827.7 (−3,568.2 to 1,912.8, $p = .54$)	−0.1 (−0.5 to 0.4, $p = .80$)
Coronary artery disease (CAD)	−517.2 (−1,336.5 to 302.0, $p = .21$)	−1,566.3 (−4,196.7 to 1,064.1, $p = .54$)	0.1 (−0.3 to 0.6, $p = .80$)
Hypertension	−77.1 (−296.6 to 142.4, $p = .48$)	−223.8 (−928.2 to 480.4, $p = .52$)	0.1 (−0.1 to 0.2, $p = .82$)
Hyperlipidaemia	−1.9 (−347.0 to 343.1, $p = .99$)	−372.6 (−1,473.3 to 728.1, $p = .49$)	0.1 (−0.1 to 0.3, $p = .24$)
Cancer	23.1 (−3.5 to 49.8, $p = .08$)	7.3 (−80.8 to 95.4, $p = .86$)	0.1 (−0.1 to 0.2, $p = .16$)
Stroke	−846.5 (−1,921.7 to 228.6, $p = .12$)	−2,146.2 (−5,625.6 to 1,333.1, $p = .22$)	−0.1 (−0.6 to 0.6, $p = .94$)
Human immunodeficiency virus (HIV)	114.5 (−24.2 to 253.2, $p = .10$)	19.4 (−437.7 to 476.6, $p = .93$)	0.1 (−0.1 to 0.2, $p = .11$)
Physical inactivity	−58.6 (−410.0 to 292.8, $p = .73$)	−102.1 (−1,229.3 to 1,025.0, $p = .85$)	−0.1 (−0.2 to 0.1, $p = .80$)
Received influenza vaccination	−56.7, (−229.0 to 115.4, $p = .51$)	−397.0 (−939.4 to 145.2, $p = .14$)	0.1 (−0.1 to 0.2, $p = .70$)
State population density	8.9 (6.5 to 11.2, $p < .01$)*	19.2 (9.5 to 29.0, $p < .01$)*	0.1 (0.1 to 0.2, $p = .02$)*
Urban air quality rank (Lower number is better quality)	22.4 (−42.5 to 87.3, $p = .49$)	5.1 (−204.7 to 214.9, $p = .96$)	0.1 (−0.1 to 0.1, $p = .97$)
Drinking water quality rank (lower number is better quality)	−50.4 (−114.9 to 13.9, $p = .12$)	−50.8 (−262.9 to 161.2, $p = .63$)	−0.1 (−0.1 to 0.1, $p = .57$)
UV index	−825.3 (−1,580.7 to −70.0, $p = .03$)*	−1,429.2 (−3,951.7 to 1,093.3, $p = .26$)	−0.3 (−0.7 to 0.1, $p = .09$)
Precipitation	103.7 (−444.8 to 652.2, $p = .70$)	−199.6 (−1,975.1 to 1,575.8, $p = .82$)	0.1 (−0.1 to 0.4, $p = .46$)
Average temperature	−12.3 (−96.0 to 71.3, $p = .76$)	−159.7 (−425.4 to 106.9, $p = .23$)	−0.1 (−0.1 to 0.1, $p = .88$)
Average household income	0.1 (0.1 to 0.2, $p < .01$)*	0.1 (−0.1 to 0.2, $p = .08$)	0.1 (−0.1 to 0.1, $p = .09$)
High school grad percent	−82.5 (−381.6 to 216.5, $p = .58$)	−75.1 (−1,036.5 to 886.2, $p = .87$)	−0.1 (−0.2 to 0.1, $p = .52$)
Social distancing score	442.3 (−1,146.3 to 2,030.9, $p = .57$)	1,986.2 (−3,104.0 to 7,076.4, $p = .43$)	0.1 (−0.7 to 1.0, $p = .77$)
Alcohol consumption	−189.2 (−1,924.8 to 1,446.3, $p = .81$)	251.7 (−4,992.5 to 5,496.0, $p = .92$)	−0.1 (−1.1 to 0.7, $p = .68$)
Current cigarette smoker	−294.5 (−556.9 to −32.2, $p = .02$)*	−683.0 (−1,544.8 to 178.7, $p = .11$)	−0.1 (−0.2 to 0.1, $p = .50$)
Cocaine	754.2 (−763.9 to 2,272.5, $p = .32$)	2,852.1 (−1,993.8 to 7,698.0, $p = .24$)	1.0 (0.2 to 1.9), $p = .01$)*
Marijuana	−9.1 (−209.5 to 191.2, $p = .92$)	262.9 (−374.8 to 900.7, $p = .41$)	0.1 (−0.1 to 0.2, $p = .06$)
Per capita health care spending	1.0 (0.3 to 1.7, $p < .01$)*	3.68 (1.3 to 5.9, $p < .01$)*	0.1 (0.1 to 0.1, $p = .18$)
Public transportation volume	0.1 (0.1 to 0.2, $p < .01$)*	0.1 (0.1 to 0.2, $p = .01$)*	0.1 (0.1 to 0.2, $p = .02$)*
Number of residents in nursing home facilities	0.1 (0.1 to 0.2, $p < .01$)*	−0.1 (−0.2 to 0.1, $p = .88$)	0.1 (0.1 to 0.2, $p = .01$)*
Number of inmates in prisons	0.1 ($p = -0.1$ to 0.1, $p = .98$)	−0.1 (−0.2 to 0.1, $p = .11$)	0.1 (−0.1 to 0.1, $p = .37$)
Tourism ranking 2018 (lower ranking means more tourists)	−79.8 ($p = -146.0$ to −13.7, $p = .01$)*	−50.9 (−275.1 to 173.2, $p = .65$)	−0.1 (−0.2 to −0.1, $p = .02$)*

*Statistically significant.

care spending (beta coefficient 1.0, $p < .01$), higher public transportation volume (beta coefficient 0.1, $p < .01$), number of residents in nursing home facilities (beta coefficient 0.1, $p < .01$) and lower number for tourism ranking and thus more tourists (beta coefficient −79.8, $p = .01$) (Table 1 shows full univariate data and Table 2 shows univariate results).

COVID-19 case frequency, multivariate analyses

The following factors were associated with greater case frequency on multivariable analyses: higher population density (beta coefficient 7.5, $p < .01$) and increased public transportation volume (beta coefficient 0.1, $p < .01$). The R-square for this model was 0.78. Collinearity analyses did not demonstrate any significant collinearity (Table 2 shows multivariate analyses).

COVID-19 testing frequency, univariate analyses

The following factors were associated with greater testing frequency on univariate linear regression analyses: higher population density (beta coefficient 19.2, $p < .01$), lower prevalence of uninsured (beta coefficient −900.9, $p = .02$), higher per capita health-care spending (beta coefficient 3.68, $p < .01$), and higher public transportation volume (beta coefficient 0.1, $p = .01$).

COVID-19 testing frequency, multivariate analyses

The following factors were associated greater testing frequency on multivariable analyses: higher population density (beta coefficient 19.9, $p < .01$). The R-square for this model was 0.27. Collinearity analyses did not demonstrate any significant collinearity.

Table 2. Associations between demographic factors, comorbidities, and environmental factors on case numbers, test numbers and percent mortality in univariate analysis.

	Cases per million	Tests per million	Mortality percent
Age	No	No	Yes (more mortality with increased age)
Gender	Yes (higher frequency with more females)	No	No
Race, White	No	No	No
Race, Black	No	No	No
Race, Native American	No	No	No
Race, Asian	No	No	No
Race, Islander	No	No	No
Race, Other race	Yes (higher frequency with other race)	No	No
Race, multiple race	No	No	No
Uninsured	Yes (higher frequency with less uninsured)	Yes (lower frequency with more uninsured)	Yes (lower mortality with more uninsured)
Obesity	Yes (higher frequency with lower obesity)	No	No
Asthma	No	No	Yes (higher mortality with more asthma)
Diabetes	No	No	No
Chronic obstructive pulmonary disease (COPD)	No	No	No
Myocardial infarction	No	No	No
Coronary artery disease	No	No	No
Hypertension	No	No	No
Hyperlipidaemia	No	No	No
Cancer	No	No	No
Stroke	No	No	No
Human immunodeficiency virus (HIV)	No	No	No
Physical inactivity	No	No	No
Received influenza vaccination	No	No	No
State population density	Yes (higher frequency with more density)	Yes (higher frequency with more density)	Yes (more mortality with greater density)
Urban air quality rank (lower number is better quality)	No	No	No
Drinking water quality rank (lower number is better quality)	No	No	No
Ultraviolet (UV) index	Yes (higher frequency with lower UV index)	No	No
Precipitation	No	No	No
Average temperature	No	No	No
Average household income	Yes (higher frequency with higher average income)	No	No
High school grad percent	No	No	No
Social distancing score	No	No	No
Alcohol consumption	No	No	No
Current cigarette smoker	Yes (higher frequency with lower smoking)	No	No
Cocaine	No	No	Yes
Marijuana	No	No	No
Per capita health care spending	Yes (higher frequency with higher spending)	Yes (higher frequency with higher spending)	No
Public transportation volume	Yes (higher frequency with higher volume)	Yes (higher frequency with higher volume)	Yes (higher mortality with higher volume)
Number of residents in nursing home facilities	Yes (higher frequency with higher number)	No	Yes (higher mortality with higher number)
Number of inmates in prisons	No	No	No
Tourism ranking 2018 (lower number means more tourists)	Yes (higher frequency with more tourists)	Yes (higher frequency with more tourists)	Yes (higher mortality with more tourists)

COVID-19 percent mortality, univariate analyses

The following factors were associated with greater percent mortality on univariate linear regression analyses: older age in years (beta coefficient 0.2, $p = .01$), population density (beta coefficient 0.1, $p = .02$), higher prevalence of asthma (beta coefficient 0.4, $p = .04$), lower prevalence of uninsured (beta

coefficient -0.1 , $p = .02$), higher prevalence of cocaine use (beta coefficient 1.0, $p = .01$), higher public transportation volume (beta coefficient 0.1, $p = .02$), higher number of residents in nursing home facilities (beta coefficient 0.1, $p = .01$), and lower number for tourism ranking and thus more tourists (beta coefficient -0.1 , $p = .02$).

Table 3. Associations between demographic factors, comorbidities, and environmental factors on case numbers, test numbers and percent mortality in multivariable analysis.

	Cases per million	Tests per million	Mortality percent
Age	-0.10 ($p = .09$)	-0.16 ($p = .22$)	0.33 (0.12–0.55, $p = .01$)*
Gender	-0.01 ($p = .97$)	-0.08 ($p = .55$)	0.16 ($p = .22$)
State population density	0.64 (0.20–1.01), $p < .01$ *)	0.52 (0.30–0.74, $p < .01$ *)	0.24 ($p = .07$)
Urban air quality rank (Lower number is better quality)	0.02 ($p = .65$)	0.04 ($p = .72$)	0.04 ($p = .73$)
Drinking water quality rank (lower number is better quality)	0.04 ($p = .54$)	0.07 ($p = .56$)	-0.07 ($p = .61$)
UV index	0.05 ($p = .47$)	-0.06 ($p = .62$)	-0.11 ($p = .42$)
Precipitation	0.01 ($p = .99$)	-0.10 ($p = .44$)	0.18 ($p = .19$)
Average temperature	0.07 ($p = .28$)	-0.17 ($p = .18$)	0.01 ($p = .98$)
Average household income	-0.08 ($p = .28$)	-0.15 ($p = .34$)	0.20 ($p = .16$)
High school grad percent	-0.05 ($p = .44$)	-0.02 ($p = .84$)	-0.03 ($p = .81$)
Obesity	0.07 ($p = .28$)	-0.02 ($p = .87$)	-0.07 ($p = .63$)
Asthma	-0.10 ($p = .11$)	0.20 ($p = .11$)	0.22 ($p = .10$)
Diabetes	0.02 ($p = .74$)	-0.03 ($p = .79$)	0.09 ($p = .53$)
COPD	0.01 ($p = .92$)	-0.02 ($p = .85$)	0.12 ($p = .37$)
MI	-0.01 ($p = .91$)	0.01 ($p = .90$)	0.03 ($p = .83$)
CAD	0.01 ($p = .92$)	-0.05 ($p = .65$)	0.10 ($p = .48$)
Hypertension	0.04 ($p = .47$)	-0.06 ($p = .62$)	0.09 ($p = .50$)
Hyperlipidaemia	0.04 ($p = .44$)	-0.10 ($p = .41$)	0.15 ($p = .26$)
Cancer	0.04 ($p = .47$)	-0.03 ($p = .77$)	0.11 ($p = .42$)
Stroke	0.03 ($p = .61$)	-0.08 ($p = .53$)	0.04 ($p = .77$)
HIV	0.12 ($p = .09$)	-0.10 ($p = .41$)	0.19 ($p = .15$)
Physical inactivity	0.06 ($p = .27$)	0.07 ($p = .57$)	-0.01 ($p = .96$)
Received influenza vaccination	-0.11 ($p = .07$)	-0.25 ($p = .05$)	-0.01 ($p = .98$)
Uninsured	0.11 ($p = .13$)	-0.18 ($p = .18$)	-0.20 ($p = .17$)
Race, White	-0.05 ($p = .44$)	0.08 ($p = .55$)	-0.18 ($p = .18$)
Race, Black	-0.08 ($p = .16$)	-0.12 ($p = .33$)	0.21 ($p = .13$)
Race, Native American	0.02 ($p = .73$)	0.21 ($p = .11$)	-0.22 ($p = .09$)
Race, Asian	-0.11 ($p = .14$)	-0.12 ($p = .37$)	0.16 ($p = .28$)
Race, Islander	-0.07 ($p = .24$)	0.15 ($p = .24$)	-0.23 ($p = .09$)
Race, Other race	-0.06 ($p = .39$)	0.11 ($p = .38$)	-0.02 ($p = .84$)
Race, multiple race	-0.08 ($p = .16$)	-0.05 ($p = .68$)	0.10 ($p = .45$)
Social distancing score	0.05 ($p = 0.40$)	0.15 ($p = .23$)	-0.03 ($p = .78$)
Alcohol consumption	-0.01 ($p = .89$)	0.01 ($p = .94$)	-0.18 ($p = .17$)
Current cigarette smoker	0.06 ($p = .36$)	-0.05 ($p = .70$)	0.10 ($p = .48$)
Cocaine	-0.12 ($p = .05$)	0.08 ($p = .53$)	0.06 ($p = .63$)
Marijuana	-0.08 ($p = .19$)	0.08 ($p = .51$)	0.05 ($p = .70$)
Per capita health care spending	-0.04 ($p = .58$)	0.24 ($p = .09$)	0.14 ($p = .37$)
Public transportation volume	0.57 (0.23–0.84, $p < .01$)	0.24 ($p = .06$)	0.11 ($p = .43$)
Number of residents in nursing home facilities	0.02 ($p = .86$)	-0.16 ($p = .21$)	0.11 ($p = .49$)
Number of inmates in prisons	-0.25 (-0.41 to -0.04, $p < .01$ *)	-0.24 ($p = .05$)	-0.09 ($p = .61$)
Tourism ranking 2018 (lower ranking means more tourists)	-0.05 ($p = 0.52$)	0.10 ($p = .45$)	-0.30 (-0.53 to -0.83, $p = .02$ *)

*Statistically significant.

COVID-19 percent mortality, multivariate analyses

The following factors were associated with greater percent mortality using multivariable analyses: median age in years (beta coefficient 0.2, $p = .01$) and tourism ranking (beta coefficient -0.1, $p = .02$) (Table 3). The R-square for this model was 0.29. Collinearity analyses did not demonstrate significant collinearity.

Power analyses

For multivariate regression analyses for case frequency, for which there is a relatively adequate effect size, and two predictors in the model, 31 subjects would be required to achieve 80% power. With 50 states and thus 50 subjects in these analyses, the multivariate analyses for case frequency are adequately powered.

For multivariable regression analyses for testing frequency, for which there is a relatively low effect size, and one predictor in the model, 385 subjects would be required to achieve 80% power. With 50 states and thus 50 subjects in these analyses, the multivariable analyses for testing frequency are not adequately powered.

For multivariable regression analyses for percent mortality, for which there is relatively low effect size, and two predictors in the model, 478 patients would be required to achieve 80% power. With 50 states and thus 50 subjects in these analyses, the multivariable analyses for percent mortality are not adequately powered.

Discussion

In these analyses, wide variations in case frequency, testing frequency, and percent mortality were

observed. These results also identified factors, such as population density, transportation volume, tourism index and older age to be some of the factors affecting the above outcomes. To the authors' knowledge, this is the first study to incorporate large number of variables to study the differences in COVID-19 transmission among different states in the US.

Several laboratory and clinical risk factors have been postulated to predispose patients to symptomatic infection with COVID-19 and related mortality [7–9]. However, non-clinical factors that affect such outcomes are unknown. After multivariate analysis in this study, a higher population density was found to be associated with higher case and testing frequency. The results also identified increased public transportation to be associated with higher case frequency. Finally, the results identified older age and increased tourism independently related to higher mortality.

Since these analyses were based on state-level data, the power for multivariate regression was reduced compared to if the analyses were completed with patient-level data. Hence, the findings of the univariate analyses deserve attention as well here. We found direct relationships between case frequency in a state and female gender, underinsured status, average household income and per capita healthcare spending and inverse relationship with obesity, smoking and UV light exposure. Public transportation was significantly associated with frequency of cases, testing and mortality rates on univariate analysis. However, on multivariate analysis, only the frequency of cases remained significantly associated with public transportation. Tourism ranking was also a significant predictor of all three endpoints on univariate analysis and remained significantly associated with mortality even after multivariate regression.

There is limited data on the nature of healthcare disparities during the COVID-19 pandemic. The cumulative COVID-19 incidence has been reported to be significantly variable among jurisdictions, ranging from 20.6 per 100,000 cases in Minnesota to 915.3 per 100,000 cases in New York City [10]. The timing of the introduction of COVID-19 in the state and the extent of mitigation measures may mediate some of this variation. The age of patients has been shown to be a significant predictor of COVID-19 infection and worse outcomes [9,11]. The race and gender of patients have also been reported to be associated with a higher case frequency and worse outcomes in patients with COVID-19 [12,13]. It continues to be shown that population density is associated with an increase in

transmission and infection for the high-risk population [14,15].

The results of this study showed that public transport volume was also linked to a higher case rate. Prior studies have demonstrated similar findings with influenza-like illnesses and that its use increases the individual's risk for acquiring an acute respiratory infection [16,17]. A simulation model indicated that the high level of subway usage in New York can influence disease spread in an influenza epidemic and that between 4 and 5% of total infections would occur on subways [18]. This information is particularly important as recommendations to maintaining strict disinfecting guidelines for public transport along with shelter-in-place whenever possible are established [19,20].

The results of this study identified that higher tourism is associated with increased mortality. A possible explanation of this identified phenomenon has been published and thought to be related to an influx of infected patients presenting late in the disease course [21,22]. Furthermore, prior studies have shown that air transportation accelerates viral spreading mainly related to high passenger traffic and risk of surface contamination in airports [23,24]. Similar findings have been reported in trains and other types of commercial vessels [25,26]. Hence, the widespread implementation of travel restrictions, social distancing and lockdowns has become the main preventative intervention to decrease viral spreading during this pandemic. However, in this analysis, social distancing score was not associated with COVID-19 case frequency or mortality. The result from prior studies showed that social distancing is an effective measure at decreasing viral spreading when comprised of quarantines, school closure and workplace distancing [27,28]. However, our current analysis showed a lack of association between social distancing and COVID-19 case frequency, which may also represent a limitation of the process of measuring the social distancing score. This is particularly important as some countries like Sweden have encountered higher COVID-19 case frequencies after adopting more lenient social distancing measures [29]. This analysis also showed a lack of impact of climate-related factors on case and testing frequency. These findings require further validation as conflicting reports have been published [5,30,31]. Surprisingly, in our analysis states with a higher number of uninsured patients were found to have lower mortality which could possibly be related to underreporting in such populations as both case numbers and testing numbers were also lower in such states.

Finally, most of the comorbidities analyzed were not found to be independently associated with the case mortality in this analysis. This may suggest the complex interplay between demographics, environmental factors and diseases processes [32,33]. The findings from this analysis also highlight some of the potential limitations of state-level data rather than patient-level data, as previous studies have found a few comorbidities to be associated with increases in case frequency and percent mortality.

These analyses offer some early assessment of the factors that may be mediating COVID-19 case frequency, testing frequency, and percent mortality. These analyses present associations using state-level data and not patient-level data. While these analyses offer novel data regarding case frequency, testing frequency and percent mortality in the US, these analyses are not without their limitations. First, all the study data was captured from publicly available sources which only had data until 2018. The use of state-level data reduced the power of analyses, as we used for the multivariate regression models the number of states as the subjects. Although the data collection carries a risk of bias, this was minimized by utilizing multiple investigators for the accuracy of data captured. The ecologic design of the paper and use of various data sources with varying methods are other limitations of our study. Finally, the lack of granularity to county or city level data further limits our interpretations.

With these limitations in mind, it is important to frame the intentions of this study appropriately. These analyses are by no means intended to be definitive data but are intended to be exploratory data to help identify variables that should be accounted for in larger, multicenter studies that utilize patient-level data. Factors such as the environmental and local infrastructural characteristics appear to modulate the case frequency and percent mortality and thus could be beneficial to capture in future studies. The data from those variables may assist with the understanding of viral spreading and the pandemic evolution. For instance, the identification of the association between higher tourism volume with higher case frequency and percent mortality may help implement faster travel restrictions for future pandemics. Similarly, the association between public transportation volume and its association with increased case frequency and percent mortality may assist in developing a future public response.

Conclusion

This observational analysis of publicly reported state-level data identified factors associated with increased

case frequency, testing frequency, and percent mortality for COVID-19. These data can guide future study design and develop risk prediction models.

Disclosure statement

No potential conflict of interest was reported by the author(s).

In memoriam

We dedicate this paper to the memory of Luis Carlos Gamboa Chavez (1974–2020), devoted father, son and friend.

ORCID

Rohit S. Loomba  <http://orcid.org/0000-0002-9116-2835>

Saul Flores  <http://orcid.org/0000-0002-6274-5775>

Enrique G. Villarreal  <http://orcid.org/0000-0001-8993-5192>

References

- [1] World Health Organization. Coronavirus Disease 2019 (COVID-19) Pandemic. 2020; <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Accessed May 3, 2020.
- [2] Wadhera RK, Wadhera P, Gaba P, et al. Variation in COVID-19 hospitalizations and deaths across New York City boroughs. *JAMA*. 2020;323(21):2192.
- [3] Gómez-Ríos D, Ramirez-Malule D, Ramirez-Malule H. The effect of uncontrolled travelers and social distancing on the spread of novel coronavirus disease (COVID-19) in Colombia. *Travel Med Infect Dis*. 2020; 35:101699.
- [4] Matrajt L, Leung T. Evaluating the effectiveness of social distancing interventions to delay or flatten the epidemic curve of coronavirus disease. *Emerg Infect Dis*. 2020;26(8):1740–1748.
- [5] Sobral MFF, Duarte GB, da Penha Sobral AIG, et al. Association between climate variables and global transmission of SARS-CoV-2. *Sci Total Environ*. 2020; 729:138997.
- [6] Worldometer. COVID-19 coronavirus pandemic. 2020. Available from: <https://www.worldometers.info/coronavirus/>. Accessed May 2, 2020.
- [7] Aggarwal G, Cheruiyot I, Aggarwal S, et al. Association of cardiovascular disease with coronavirus disease 2019 (COVID-19) severity: a meta-analysis. *Curr Probl Cardiol*. 2020;45(8):100617.
- [8] Aggarwal G, Lippi G, Michael Henry B. Cerebrovascular disease is associated with an increased disease severity in patients with Coronavirus Disease 2019 (COVID-19): a pooled analysis of published literature. *Int J Stroke*. 2020;15(4): 385–389.
- [9] Aggarwal S, Garcia-Telles N, Aggarwal G, et al. Clinical features, laboratory characteristics, and outcomes of patients hospitalized with coronavirus disease 2019

- (COVID-19): Early report from the United States. *Diagnosis (Berl)*. 2020;7(2):91–96.
- [10] CDC COVID-19 Response team. Geographic differences in COVID-19 cases, deaths, and incidence - United States, February 12-April 7, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(15):465–471.
- [11] Adams ML, Katz DL, Grandpre J. Population-based estimates of chronic conditions affecting risk for complications from coronavirus disease, United States. *Emerg Infect Dis*. 2020;26(8):1831–1833.
- [12] Dorn AV, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. *Lancet*. 2020;395(10232):1243–1244.
- [13] Gausman J, Langer A. Sex and gender disparities in the COVID-19 pandemic. *J Womens Health (Larchmt)*. 2020;29(4):465–466.
- [14] Tarwater PM, Martin CF. Effects of population density on the spread of disease. *Complexity*. 2001;6(6):29–36.
- [15] Dalziel BD, Kissler S, Gog JR, et al. Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities. *Science*. 2018;362(6410):75–79.
- [16] Goscé L, Johansson A. Analysing the link between public transport use and airborne transmission: mobility and contagion in the London underground. *Environ Health*. 2018;17(1):84.
- [17] Troko J, Myles P, Gibson J, et al. Is public transport a risk factor for acute respiratory infection? *BMC Infect Dis*. 2011;11(1):16.
- [18] Cooley P, Brown S, Cajka J, et al. The role of subway travel in an influenza epidemic: a New York City simulation. *J Urban Health*. 2011;88(5):982–995.
- [19] Anderson EL, Turnham P, Griffin JR, et al. Consideration of the aerosol transmission for COVID-19 and public health. *Risk Anal*. 2020;40(5):902–907.
- [20] Park J. Changes in subway ridership in response to COVID-19 in Seoul, South Korea: implications for social distancing. *Cureus*. 2020;12(4):e7668.
- [21] Carletti F, Lalle E, Messina F, et al. About the origin of the first two Sars-CoV-2 infections in Italy: inference not supported by appropriate sequence analysis. *J Med Virol*. 2020;92(9):1404–1405.
- [22] Lo IL, Lio CF, Cheong HH, et al. Evaluation of SARS-CoV-2 RNA shedding in clinical specimens and clinical characteristics of 10 patients with COVID-19 in Macau. *Int. J. Biol. Sci*. 2020;16(10):1698–1707.
- [23] Browne A, Ahmad SS, Beck CR, et al. The roles of transportation and transportation hubs in the propagation of influenza and coronaviruses: a systematic review. *J Travel Med*. 2016;23(1):tav002.
- [24] Ikonen N, Savolainen-Kopra C, Enstone JE, et al.; for the PANDHUB consortium. Deposition of respiratory virus pathogens on frequently touched surfaces at airports. *BMC Infect Dis*. 2018;18(1):437.
- [25] Cui F, Luo H, Zhou L, et al. Transmission of pandemic influenza A (H1N1) virus in a train in China. *J Epidemiol*. 2011;21(4):271–277.
- [26] Brotherton JM, Delpech VC, Gilbert GL, et al.; Cruise Ship Outbreak Investigation Team. A large outbreak of influenza A and B on a cruise ship causing widespread morbidity. *Epidemiol Infect*. 2003;130(2):263–271.
- [27] Chao DL, Halloran ME, Obenchain VJ, et al. FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Comput Biol*. 2010;6(1):e1000656.
- [28] Fong MW, Gao H, Wong JY, et al. Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings-social distancing measures. *Emerg Infect Dis*. 2020;26(5):976–984.
- [29] Erbrink CAC. 'Life has to go on': how Sweden has faced the virus without a lockdown. 2020. <https://www.nytimes.com/2020/04/28/world/europe/sweden-coronavirus-herd-immunity.html>
- [30] Jahangiri M, Jahangiri M, Najafgholipour M. The sensitivity and specificity analyses of ambient temperature and population size on the transmission rate of the novel coronavirus (COVID-19) in different provinces of Iran. *Sci Total Environ*. 2020;728:138872.
- [31] Yao Y, Pan J, Liu Z, et al. No Association of COVID-19 transmission with temperature or UV radiation in Chinese cities. *Eur Respir J*. 2020;55(5):2000517.
- [32] Bookman EB, McAllister K, Gillanders E, et al.; NIH GxE Interplay Workshop participants. Gene-environment interplay in common complex diseases: forging an integrative model—recommendations from an NIH workshop. *Genet Epidemiol*. 2011;35(4):217–225.
- [33] McMichael AJ. Population, environment, disease, and survival: past patterns, uncertain futures. *The Lancet*. 2002;359(9312):1145–1148.