

## RESEARCH ARTICLE

# Contextual factors and the COVID-19 outbreak rate across U.S. counties in its initial phase

Wolfgang Messner  | Sarah E. Payson

Darla Moore School of Business, University of South Carolina, Columbia, South Carolina

**Correspondence**

Wolfgang Messner, Darla Moore School of Business, University of South Carolina, 1014 Greene Street, Columbia, SC 29208.  
Email: wolfgang.messner@moore.sc.edu

**Funding information**

Center for International Business Education and Research (CIBER) at the University of South Carolina

**Abstract**

**Background:** This study examines the association of contextual factors with the COVID-19 outbreak rate across U.S. counties in its initial phase.

**Methods:** Contextual factors are simultaneously tested at the county- and state-level with a multilevel linear model using full maximum likelihood.

**Results:** The variation between states is substantial and significant ( $ICC = 0.532$ ,  $u_0 = 8.20E-04$ ,  $P < .001$ ). At the state level, the cultural value of collectivism and the contextual factor of government spending are positively associated with the outbreak rate. At the county level, the racial and ethnic composition contributes to outbreak differences, disproportionately affecting black/African, native, Asian, and Hispanic Americans as well as native Hawaiians. Counties with a higher median age and a higher household income have a stronger outbreak. Better education and personal health are generally associated with a lower outbreak. Obesity and smoking are negatively related to the outbreak, in agreement with the value expectancy concepts of the health belief model. Air pollution is another significant contributor to the outbreak.

**Conclusions:** Because of a high variation in contextual factors, policy makers need to target pandemic responses to the smallest subdivision possible, so that countermeasures can be implemented effectively.

**KEYWORDS**

COVID-19, novel coronavirus, outbreak, pandemic, regional differences

## 1 | INTRODUCTION

First reports of a pneumonia of unknown etiology emerged in Wuhan, China, on 31 December 2019. The extremely contagious virus was identified as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and spread quickly beyond Wuhan. In the United States, the first case of COVID-19, the disease caused by SARS-CoV-2, was reported on 22 January 2020. Despite unprecedented government action, the number of cases in the United States crossed one million on April 28.<sup>1</sup>

The local press and epidemiological research alike have reported regional differences in the outbreak.<sup>2</sup> A community's susceptibility to any virus is determined by a variety of factors, inter alia, biological determinants, demographic profiles, and socioeconomic characteristics.<sup>3</sup> These factors vary significantly across the United States; for instance, COVID-19 fatalities in New York, an epicenter of the initial outbreak in the United States, disproportionately affected males and people belonging to older age groups, from black/African and Hispanic ethnicities, and with certain comorbidities.<sup>4</sup> However, as of 9 May 2020, more than half of COVID-19 data reported by the

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

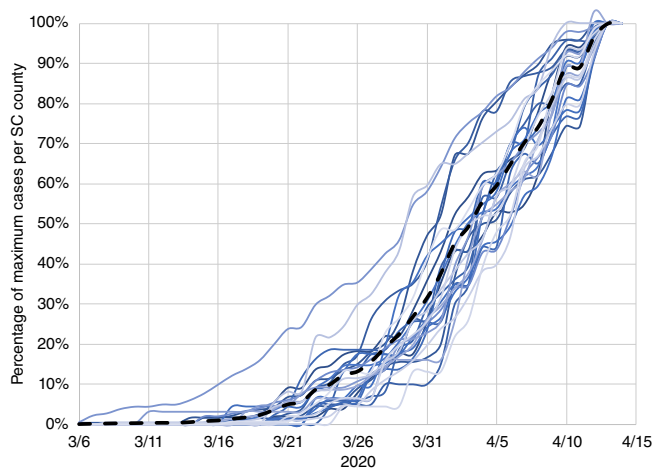
© 2021 The Authors. *Health Science Reports* published by Wiley Periodicals LLC.

Centers for Disease Control and Prevention (CDC) were missing race and ethnicity disaggregation; other individual variables were lacking as well. To understand local differences in the outbreak rate and risk of contracting COVID-19, we therefore deploy an ecological analysis using contextual factors. A two-level hierarchical linear model with full maximum likelihood allows us to simultaneously test and disentangle county- and state-level effects.

Our study contributes to various strands of current COVID-19 research. First, we note that contextual factors influence the COVID-19 outbreak. Because significant variations in the outbreak exist between states and counties within a state (Figures 1 and 2),<sup>2</sup> we recommend policy makers to look at pandemics from the smallest subdivision possible for effective implementation of countermeasures and provision of critical resources. Second, we develop an understanding of how regional cultural differences relate to outbreak variations, driven by specific psychological functioning of individuals and the enduring effects of such differences on political processes, governmental institutions, and public policies.<sup>5,6</sup> Third, we cannot support rumors propagated by the popular press that a state's leadership, as expressed by the political party in control or the gender of its governor, has a statistically significant influence on the outbreak.<sup>7</sup> Fourth, we identify how the virus affects counties differently, depending on their demographic profile. Fifth, while good personal health is generally associated with a lower risk, we identify the prevalence of obesity and smoking in counties to be negatively related with the outbreak. Sixth, while previous studies link air pollution to the death rate, we show that it also contributes to the case load.

## 2 | METHODS

We now explain the estimation of the outbreak rate and the reasons for including certain contextual factors; Table 1 summarizes the data sources.



**FIGURE 1** Epidemic days at county level (South Carolina). The spaghetti lines trace the COVID-19 outbreak in South Carolina (black dashed line) and the counties (blue straight lines) as a percentage of the cases reported on 14 April, 2020. Cases unallocated to a county due to lack of information are included in the state line; counties with less than 20 reported cases are not shown in the diagram

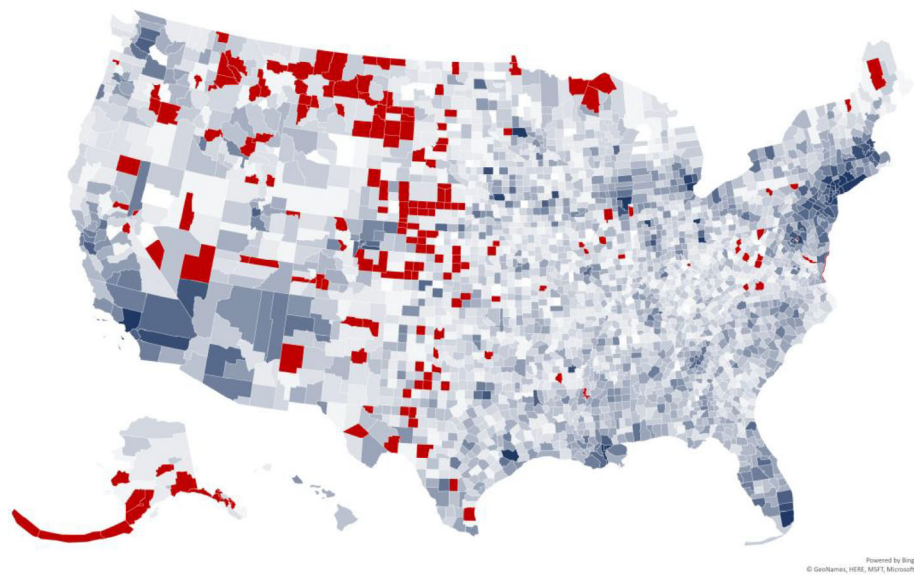
### 2.1 | Outbreak rate

We obtain COVID-19 outbreak data from USA Facts.<sup>1</sup> Since January 22, this database has aggregated data from the Centers for Disease Control and Prevention (CDC) and other public health agencies. We discard cases only allocated at the state level due to lack of information. As of April 14, these are only 308 cases per state on average, but a few states have as many as 4866 (New Jersey), 1300 (both Rhode Island and Georgia), or 1216 (Washington State). Also, the 21 cases on the Grand Princess cruise ship are not attributed to any counties in California. We determine the average outbreak start in the United States with a minimum of 10 reported cases to be 127.753 days after 31 December 2019. Because we are interested in the initial outbreak, we disregard counties after this date plus one standard deviation of 48.770. Thus, our sample consists of 2958 out of 3142 counties across the 50 U.S. states. Baseline transmission characteristics of specific pathogens in their social contexts are captured by mathematical models,<sup>8</sup> which use time-series data to estimate the force of infection.<sup>9</sup> We use the initial outbreak data at county level for the first 30 days after a minimum of 10 cases was reached. Most epidemics grow approximately exponentially during their initial phase.<sup>10</sup> A relaxation of the assumption of exponential growth is not necessary as the COVID-19 outbreak is mainly airborne.<sup>9</sup> Following approaches by the Institute for Health Metrics and Evaluation at the University of Washington<sup>11</sup> and the COVID-19 Modeling Consortium at the University of Texas at Austin,<sup>12</sup> we model the outbreak using the exponential growth equation  $\frac{dy}{dt} = by$ , where  $b$  is a positive constant called the relative growth rate with units of inverse time. Going forward, we simply refer to  $b$  as the outbreak rate. The shape of the trends in case counts enables us to see differences between counties.<sup>13</sup> Solutions to this differential equation have the form  $y = a e^{bt}$ , where  $a$  is the initial value of cases  $y$ . The doubling time  $T_d$  can be calculated as  $T_d = \frac{\ln(2)}{b}$ . Similarly,  $b$  is also related to the basic reproduction number  $R_0$ , as derived from classic SIR-type (susceptible-infected-removed) compartmental transmission models:  $R_0 = 1 + \frac{b}{\gamma}$ , where  $\frac{1}{\gamma}$  is the mean infectious period.<sup>9,10,14</sup> Taken together, our model is a statistical, but not an epidemiological model, that is, we are neither trying to model infection transmission nor estimate epidemiological parameters, such as the pathogen's reproductive or attack rate. Instead, we are fitting curves to observed outbreak data at the county level. A change-point analysis using the Fisher discriminant ratio as a kernel function does not show any significant change points in the outbreak and therefore justifies modeling the COVID-19 outbreak as a phenomenon of unrestricted population growth.<sup>15</sup> We cannot forecast outbreak dynamics with this statistical approach, though we do not require extrapolated data in our work.

### 2.2 | Cultural values

Culture can be defined as a set of values that are shared in a given social group. While cultural values are often used to distinguish countries,<sup>16</sup> more than 80% of cultural variation resides within

**FIGURE 2** Variation in outbreak rates at U.S. county level. This geo map reveals a large variation in outbreak rates at U.S. county level. Lighter blue colors signify that the pandemic has a slower relative growth rate, and darker blue colors point to a faster growth rate. Counties colored in red are excluded from the analysis because of a late start of the outbreak



countries.<sup>17</sup> The original North American colonies were settled by people hailing from various countries, who have spread their influence across mutually exclusive areas. Their distinct cultures are still with us today.<sup>6</sup> Although today's U.S. states are not strictly synonymous with these cultural areas, there is abundant evidence that political boundaries can serve as useful proxies for culture.<sup>18</sup>

One of the most useful constructs to emerge from cultural social psychology is the individualism-collectivism bipolarity. It has proven useful in describing cultural variations in behaviors, attitudes, and values. Briefly, individualism is a preference for a loosely knit social framework, whereas collectivism represents a preference for a tightly knit framework, in which its members are interdependent and expected to look after each other in exchange for unquestioning loyalty. While the majority of research on collectivism involves comparing countries,<sup>16</sup> we use an index developed at state level solely within the United States.<sup>5</sup> Previous studies have shown that the regional prevalence of pathogens and international differences in the COVID-19 outbreak are positively associated with collectivism.<sup>18,19</sup>

### 2.3 | Institutional confounders

In addition to culture, we include various institutional confounders at the state level, such as the political affiliation of a state's governor, the gender of the governor, and government spending per capita. Government plays a critical role in policy development and implementation, and so state-level differences could influence the outbreak rate.<sup>20</sup>

### 2.4 | Racial composition

While first systematic reviews about COVID-19 incidences from China relied on ethnically homogenous cohorts,<sup>21,22</sup> ethnically diverse

populations, such as in the United Kingdom and United States, may exhibit different susceptibility or response to infection because of socioeconomic, cultural or lifestyle factors, genetic predisposition, and pathophysiological differences. Certain vitamin or mineral deficiencies, differences in insulin resistance, or vaccination policies in countries of birth may also be contributing factors.<sup>22</sup> We include variables measuring the composition of U.S. counties regarding racial and ethnic groups.

### 2.5 | Income and education

Poverty is arguably the greatest risk factor for acquiring and succumbing to disease worldwide but has historically received less attention from the medical community than genetic or environmental factors. The global HIV crisis brought into sharp relief the vulnerability of financially strapped health systems and revealed disparities in health outcomes along economic fault lines.<sup>23</sup> We include the median household income to quantify potential economic disparities between U.S. counties. In addition, we measure nonproficiency in English and math performance of students. Lower educational levels may result in a lower aptitude as it relates to understanding and effectively responding to the pandemic.

### 2.6 | Other demographics

Age and gender also play a potential role in a population's susceptibility. During the aging process, immune functions decline, rendering the host more vulnerable to certain viruses.<sup>24</sup> We use the percentage of population below 18 years of age and their median age to determine potential effects of differences in mobility, response, and lifestyle factors. We also control for the percentage of the population that is female, as one COVID-19 study in Italy showed that about 82% of critically ill people admitted into intensive care were men.<sup>25</sup>

**TABLE 1** Variables and descriptive statistics

Variable	Primary source	Secondary source	N	Year(s)	Median	Minimum; maximum	Standard deviation
<i>State institutions</i>							
Party control	WIK		50	2020			
Gender of governor	WIK		50	2020			
Government spending	SIP	Census Bureau	51	2015	10.059	16.553	3.186
<i>People cultural values</i>							
Collectivism	VAN		50	1997	49.500	[31; 91]	11.336
<i>Racial composition</i>							
Black and African American	CHR	Census population est.	3130	2018	2.251	[0.512; 85.414]	14.370
Native American	CHR	Census population est.	3141	2018	0.640	[0.000; 92.515]	7.600
Asian American	CHR	Census population est.	3141	2018	0.736	[0.000; 43.357]	2.953
Native Hawaiian	CHR	Census population est.	3141	2018	0.063	[0.000; 48.900]	1.081
Hispanic American	CHR	Census population est.	3141	2018	4.405	[0.610; 96.360]	14.273
<i>Income and education</i>							
Household income	CHR	Small area income and poverty est.	3140	2018	50 547.500	[15.229; 140 382]	14 124.747
Nonproficiency in English	CHR	Census population est.	3141	2014-18	0.748	[0.000; 51.77]	3.720
Math grade	CHR	Stanford education data archive	2467	2016	3.013	[1.654; 68 943]	3107.118
<i>Other demographics</i>							
Persons under 18 years	CHR	Census population est.	3140	2018	22.063	[7.069; 41.991]	3.461
Median age	SCP	American community survey	3142	2012-16	41.000	[21.500; 66.000]	5.355
Female persons	CHR	Census population est.	3138	2018	50.301	[0.192; 76.208]	2.659
<i>Personal health</i>							
Social associations	CHR	County business patterns	3141	2017	11.096	[0.000; 52.314]	5.912
Sleep deprivation	CHR	Behavioral risk factor surveillance system	3141	2016	32.949	[8.937; 46.708]	4.282
Preventable hospitalization	CHR	Mapping Medicare disparities tool	3098	2017	4705	[34; 16 851]	1856.793
Obesity	CHR	United States Diabetes Surveillance System	3072	2016	31.300	[11.800; 47.600]	4.510
Smoking	CHR	Behavioral Risk Factor Surveillance System	3072	2020	17.409	[6.546; 41.389]	11.600
<i>External health</i>							
Air pollution	CHR	Environmental public health tracking network	3107	2014	9.400	[2.300; 19.700]	1.985

**TABLE 1** (Continued)

Variable	Primary source	Secondary source	N	Year(s)	Median	Minimum; maximum	Standard deviation
Rural area	CHR	Census population est.	3134	2010	59.517	[0.000; 100.000]	31.437
Food environment	CHR	USDA food environment atlas; map the meal gap from Feeding America		2015-17	7.700	[0.000; 34.5000]	1.512
<i>Other confounders</i>							
Density	SCP	American community survey	3142	2012-16	44.967	[0.384; 71 615.813]	1787.612
Temperature <sup>a</sup>	NCDC		3141	2017	43.000	[-14.200; 73.500]	11.650

Note: This table lists the independent variables at both levels of analysis and their provenance. CHR: County health rankings, [www.countyhealthrankings.org](http://www.countyhealthrankings.org); SCP: Social Capital Project, <https://www.jec.senate.gov/public/index.cfm/republicans/2018/4/the-geography-of-social-capital-in-america>; NCDC: National Centers for Environment Information, <https://www.ncdc.noaa.gov/cag/county/mapping/1/tavg/202003/2/value>; VAN: Collectivism index proposed by Vandello and Cohen<sup>5</sup>; WIK: [https://en.wikipedia.org/wiki/List\\_of\\_United\\_States\\_governors](https://en.wikipedia.org/wiki/List_of_United_States_governors).

<sup>a</sup>The NCDCs do not provide temperature mapping for Hawaii; all Hawaiian counties replaced by historical average data from <http://holiday-weather.com/hawaii/averages>. All websites accessed in May 2020.

## 2.7 | Personal health

Good overall personal health is a general indicator for disease resistance. Additionally, the health belief model suggests that a person's belief in a personal threat of a disease, together with faith in the effectiveness of behavioral recommendations, predicts the likelihood of the person adopting the recommendation.<sup>26</sup> We use the percentage of the population that reports insufficient amount of sleep, is obese (as defined by a body mass index above 30), and smokes daily. Given the latter two are publicized risk factors for COVID-19, there is a potential for greater caution following the value-expectancy concepts of the health belief model. Yet, medicinal nicotine has been identified as a potential protective factor against infection by SARS-CoV-2.<sup>27</sup> We also measure the preventable hospitalization rate (ie, the rate of hospital stays for ambulatory-care sensitive conditions) as a potential indicator of poor personal health and the social association rate (ie, the average number of membership associations), which is generally connected with positive mental health and happiness.

## 2.8 | External health

Previous studies suggest that exposure to pollution can suppress immune responses and proliferate the transmission of infectious diseases,<sup>28</sup> and that the COVID-19 mortality rate is associated with air pollution.<sup>29</sup> However, the impact of air pollution on the spread of COVID-19 is not yet known.<sup>28</sup> We use the 2014 average daily density of fine particulate matter PM<sub>2.5</sub> to measure air pollution across U.S. counties, and the percentage of population living in rural areas to account for physical distancing being more prevalent in rural areas. In addition, the food environment index reflects access to grocery stores and healthy foods.

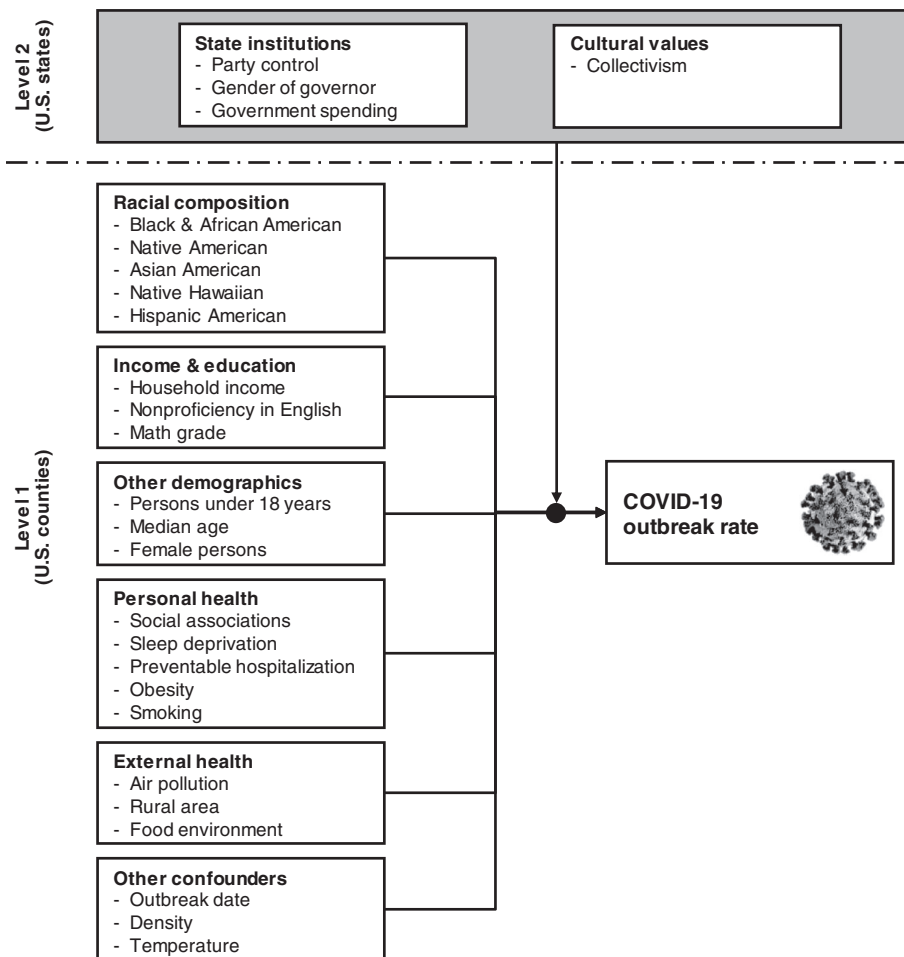
## 2.9 | Other confounders

Population density and overcrowding are significant when considering public health crises, facilitating the spread of diseases in developing and developed countries alike.<sup>30</sup> As the climate is another highly publicized confounder potentially influencing the COVID-19 transmission rate,<sup>31</sup> we also include each county's average temperature during February and March 2020. To control for the temporality of the outbreak with respect to heterogeneous contact patterns on the spreading dynamics between geographic regions,<sup>32</sup> we bring in a variable representing the number of days between January 1 and the 10th confirmed case reported.

## 3 | STATISTICAL RESULTS

To simultaneously test county- and state-level effects of contextual factors on the outbreak rate with cross-level interactions, we estimate a two-level linear model using full maximum likelihood in HLM 7.03 (Figure 3). This accounts for potential similarities in counties within the same state.<sup>28</sup> The data files for both levels are available as Appendix S1 and S2 in the supporting information. We center all predictors around the group mean at level 1 and grand mean at level 2. We first estimate a one-way random effects ANOVA (unconditional model), which has an intraclass correlation coefficient (ICC) of 0.532. That is, more than 53% of the variation in the outbreak rate is between states, and about 47% is within the states and between their counties. The variation between states is statistically significant ( $u_0 = 8.20E-04$ ,  $P < .001$ ). We thus deem it prudent to proceed with a multilevel model as follows:

Level 1 (counties): Outbreak rate<sub>ij</sub> =  $\beta_{0j}$  +  $\beta_{1j}$  [Black & African American] +  $\beta_{2j}$  [Native American] +  $\beta_{3j}$  [Asian American] +  $\beta_{4j}$  [Native Hawaiian] +  $\beta_{5j}$  [Hispanic American] +  $\beta_{6j}$  [Household income] +  $\beta_{7j}$



**FIGURE 3** Multilevel research model. This figure details the multilevel research model and the variables used at state- and county-level

[Nonproficiency in English] +  $\beta_{8j}$  [Math grade] +  $\beta_{9j}$  [Persons under 18 years] +  $\beta_{10j}$  [Median age] +  $\beta_{11j}$  [Female persons] +  $\beta_{12j}$  [Social associations] +  $\beta_{13j}$  [Sleep deprivation] +  $\beta_{14j}$  [Preventable hospitalization] +  $\beta_{15j}$  [Obesity] +  $\beta_{16j}$  [Smoking] +  $\beta_{17j}$  [Air pollution] +  $\beta_{18j}$  [Rural area] +  $\beta_{19j}$  [Food environment] +  $\beta_{20j}$  [Outbreak date] +  $\beta_{21j}$  [Density] +  $\beta_{22j}$  [Temperature] +  $r_{ij}$ .

Level 2 (states):  $\beta_{0j} = \gamma_{00} + \gamma_{01}$  [Party control] +  $\gamma_{02}$  [Gender of governor] +  $\gamma_{03}$  [Government spending] +  $\gamma_{04}$  [Collectivism] +  $u_{0j}$ ;  $\beta_{1j} = \gamma_{10} + u_{1j}$ ;  $\beta_{2j} = \gamma_{20}$ ;  $\beta_{3j} = \gamma_{30}$ ;  $\beta_{4j} = \gamma_{40}$ ;  $\beta_{5j} = \gamma_{50}$ ;  $\beta_{6j} = \gamma_{60}$ ;  $\beta_{7j} = \gamma_{70}$ ;  $\beta_{8j} = \gamma_{80}$ ;  $\beta_{9j} = \gamma_{90}$ ;  $\beta_{10j} = \gamma_{100}$ ;  $\beta_{11j} = \gamma_{110}$ ;  $\beta_{12j} = \gamma_{120}$ ;  $\beta_{13j} = \gamma_{130}$ ;  $\beta_{14j} = \gamma_{140}$ ;  $\beta_{15j} = \gamma_{150}$ ;  $\beta_{16j} = \gamma_{160}$ ;  $\beta_{17j} = \gamma_{170}$ ;  $\beta_{18j} = \gamma_{180}$ ;  $\beta_{19j} = \gamma_{190}$ ;  $\beta_{20j} = \gamma_{200}$ ;  $\beta_{21j} = \gamma_{210}$ ;  $\beta_{22j} = \gamma_{220}$

We provide the interitem correlation matrix in Table 2 and the results of the multilevel model in Table 3. Additionally, we perform several checks and robustness tests to inform our results.

First, because outbreak rates change over time and their estimation is somewhat sensitive to the starting figure, we alternatively calculate the rate after 25 (instead of 10) cases for a time series of 30 days, finding a high correlation of 0.837,  $P < .001$ . Similarly, we reduce the time series from 30 to 20 days and again find a high correlation of .963,  $P < .001$  between the rates. Even more importantly, the results of the multilevel model are stable when using these alternative calculations of the outbreak rate.

Second, we iteratively include several other contextual variables and logged versions to assess the robustness of the results. But because it is nearly impossible to establish a complete list of confounding variables, we quantify the potential impact of unobserved confounds (Table 3; impact threshold).<sup>33</sup> For instance, the necessary impact of such a confound for air pollution would be 0.101, that is, to invalidate the variable's inference on the outbreak rate, a confounding variable would have to be correlated with both the outbreak rate and air pollution at  $\sqrt{0.101} = 0.317$ . Next, we ask how many counties would have to be replaced with unobserved cases for which the null hypothesis is true (ie, a contextual variable has no influence on the outbreak rate) in order to invalidate the inference.<sup>28,33</sup> As Table 3 (confound threshold) shows, 73.162% of the counties would have to be replaced with counties for which the effect is zero in order to invalidate the influence of air pollution. In summary, it can be claimed that the influence of the identified contextual variables on the pandemic is reasonably robust.

Third, a potential omission of relevant variables can lead to multicollinearity issues, which are generally a serious problem in epidemiological studies.<sup>34</sup> Even though HLM 7.03 checks for multicollinearity, we conduct several additional diagnostics to eliminate any potential issues. In the interitem correlation matrix (Table 2), the average (absolute) correlation is 0.172, and the highest correlation

**TABLE 2** Interitem correlation matrix

	b	c	d	e	f	g	h	i	j	k	l	m	n
a	-0.034	-0.251	0.035	0.083	0.048	-0.095	0.011	0.121	-0.106	0.046	-0.007	0.206	-0.088
b		0.052	-0.368	-0.102	0.049	-0.027	-0.024	-0.104	0.003	-0.066	-0.023	0.001	0.079
c			-0.389	-0.200	0.004	0.134	0.007	-0.043	0.269	-0.013	0.004	-0.030	0.054
d				0.539	-0.209	0.045	-0.044	0.051	-0.165	0.013	-0.057	-0.015	-0.094
e					-0.104	0.013	-0.011	-0.127	-0.321	-0.035	-0.002	-0.074	-0.141
f						-0.047	0.015	0.009	-0.095	-0.002	0.004	0.232	-0.110
g							0.060	0.195	0.483	0.221	-0.030	0.049	-0.269
h								0.250	-0.139	0.630	0.755	-0.025	-0.341
i									0.009	0.749	0.254	0.324	-0.398
j										-0.083	-0.201	0.160	-0.031
k											0.699	0.164	-0.470
l												-0.098	-0.339
m													-0.447
	o	p	q	r	s	t	u	v	w	x	y	z	
a	-0.006	-0.083	-0.002	-0.007	0.115	0.028	0.057	0.025	-0.156	0.014	-0.095	0.258	
b	0.014	0.173	-0.178	-0.113	0.108	-0.054	-0.089	0.066	0.017	0.042	-0.062	-0.237	
c	-0.039	0.198	-0.296	-0.133	-0.153	-0.217	-0.292	-0.071	0.214	-0.024	0.129	-0.477	
d	0.128	-0.272	0.557	0.301	0.203	0.138	0.407	-0.030	-0.241	-0.116	0.045	0.715	
e	0.154	-0.123	0.568	0.278	0.404	0.360	0.230	-0.045	-0.442	-0.128	0.119	0.533	
f	-0.034	-0.049	-0.064	0.047	0.076	0.270	-0.219	0.091	-0.105	0.064	-0.047	-0.108	
g	0.058	-0.169	-0.046	-0.137	-0.362	-0.315	0.097	-0.501	0.087	-0.448	0.553	0.033	
h	-0.338	-0.070	-0.228	-0.150	-0.014	0.014	-0.130	-0.033	0.516	-0.064	0.012	-0.018	
i	-0.187	-0.217	-0.159	-0.093	-0.239	-0.234	-0.169	-0.290	0.191	-0.104	0.092	0.237	
j	0.064	-0.051	-0.327	-0.322	-0.518	-0.666	-0.031	-0.384	0.309	-0.353	0.246	-0.276	
k	-0.188	-0.171	-0.235	-0.143	-0.140	-0.109	-0.167	-0.203	0.494	-0.131	0.153	0.116	
l	-0.228	-0.030	-0.283	-0.150	0.004	0.039	-0.166	0.040	0.656	-0.017	-0.015	-0.049	
m	0.138	-0.117	-0.020	0.103	0.144	0.004	0.055	-0.190	-0.063	-0.028	-0.062	0.021	
n	0.079	0.276	-0.046	-0.007	-0.014	-0.118	-0.039	0.440	-0.107	0.239	-0.145	-0.108	
o		0.064	0.118	0.045	0.069	0.037	0.176	-0.156	-0.105	-0.119	0.097	0.080	
p			-0.259	-0.031	0.041	-0.076	-0.092	0.142	0.063	0.206	-0.065	-0.288	
q				0.436	0.456	0.568	0.501	-0.018	-0.510	-0.046	0.098	0.514	
r					0.386	0.433	0.228	0.115	-0.323	0.098	-0.027	0.325	
s						0.611	0.289	0.283	-0.296	0.246	-0.228	0.213	
t							0.227	0.248	-0.372	0.203	-0.118	0.222	
u								-0.114	-0.176	-0.136	0.083	0.340	
v									0.001	0.518	-0.369	-0.064	
w										-0.070	0.029	-0.360	
x											-0.342	-0.107	
y												0.043	
z													

Note: This table shows the interitem correlations between the variables at both levels of analysis. Variables: a: Party control; b: gender of governor; c: government spending; d: collectivism; e: black and African American; f: native American; g: Asian American; h: native Hawaiian; i: Hispanic American; j: household income; k: nonproficiency in English; l: Math grade; m: persons under 18 years; n: median age; o: female persons; p: social associations; q: sleep deprivation; r: preventable hospitalization; s: obesity; t: smoking; u: air pollution; v: rural area; w: food environment; x: outbreak date; y: density; z: temperature.

**TABLE 3** HLM contextual model

Fixed effect	Coefficients <sup>a</sup>	Standard error	Confidence interval	P	Effect size <sup>b</sup>	Reliability	Impact threshold	Confound threshold
Outbreak rate	64.085	0.004	[56.574; 71.596]	<.001		0.958		
<i>State institutions</i>								
Party control <sup>c</sup>	-7.695	0.007	[-21.687; 6.297]	.287	-15.390			
Gender of governor <sup>d</sup>	-3.197	0.006	[-14.884; 8.490]	.595	-8.197			
Government spending	5.948	0.002	[1.963; 9.933]	.005	2.279		0.935	90.134%
<i>People cultural values</i>								
Collectivism	1.330	< 0.001	[0.760; 1.900]	<.001	0.117		0.401	55.875%
<i>Racial composition</i>								
Black and African American	0.485	< 0.001	[0.265; 0.705]	<.001	0.101		0.045	54.720%
Native American	0.909	< 0.001	[0.515; 1.303]	<.001	0.119		0.049	56.643%
Asian American	2.629	0.001	[1.420; 3.838]	<.001	0.888		0.044	53.983%
Native Hawaiian	5.924	< 0.001	[5.530; 6.318]	.010	5.485		0.049	56.643%
Hispanic American	0.269	< 0.001	[0.048; 0.490]	.017	0.019		0.027	77.765%
<i>Income and education</i>								
Household income <sup>e</sup>	2.418	0.001	[0.044; 4.792]	.046	1.715		0.001	1.799%
Nonproficiency in English	1.884	0.001	[0.320; 3.448]	.018	0.505		0.008	16.948%
Math grade <sup>f</sup>	-0.002	< 0.001	[-0.004; 0.000]	.010	0.000		0.001	1.952%
<i>Other demographics</i>								
Persons under 18 years	0.029	< 0.001	[-0.418; 0.476]	.898	0.008			
Median age	0.619	< 0.001	[0.207; 1.031]	.003	0.101		0.019	33.479%
Female persons	0.286	< 0.001	[-0.316; 0.888]	.353	0.093			
<i>Personal health</i>								
Social associations	-0.552	< 0.001	[-0.966; -0.138]	.009	-0.093		0.013	25.050%
Sleep deprivation	1.638	< 0.001	[0.829; 2.447]	<.001	0.382		0.038	50.562%
Preventable hospitalization	0.001	< 0.001	[0.001; 0.001]	.002	0.000		1.000	100.000%
Obesity	-0.896	< 0.001	[-1.361; -0.431]	<.001	-0.199		0.035	48.136%
Smoking	-1.789	0.001	[-3.077; -0.501]	.007	-0.504		0.015	27.992%
<i>External health</i>								
Air pollution	6.159	0.001	[4.507; 7.811]	<.001	3.158		0.101	73.162%
Rural area	-0.322	< 0.001	[-0.406; -0.238]	<.001	-0.010		0.104	73.816%
Food environment	-0.901	0.001	[-3.478; 1.676]	.493	-0.593			
<i>Other confounders</i>								
Outbreak date	-0.060	< 0.001	[-0.162; 0.042]	.247	-0.006			
Density	0.173	< 0.001	[0.065; 0.281]	.002	0.000		0.022	37.663%
Temperature	-0.263	< 0.001	[-0.780; 0.254]	.319	-0.023			
<b>Random effects</b>			<b>Variance</b>	<b>df</b>	<b><math>\chi^2</math></b>	<b>P</b>		
Variance between state intercepts ( $\tau_{00}$ )			5.70E-04	43	350.334	<.001		
Variance within states ( $\sigma^2$ )			7.20E-03					

Note: This table provides the detailed results for the multilevel linear model. Run-time deletion reduced the number of level-1 records from 3118 to 2958 and level-2 from 50 to 48.

<sup>a</sup>The coefficients are multiplied with 1000 for more intuitive figures. Ditto for the confidence interval.

<sup>b</sup>The effect size is calculated as coefficient/standard deviation, again multiplied with 1000.

<sup>c</sup>Party control: 0 = Democratic, 1 = Republican.

<sup>d</sup>Gender of governor: 0 = male, 1 = female.

<sup>e</sup>The variable for household income is divided by 10 000.

<sup>f</sup>All effects for Math grade are from a separately calculated model because the variable is unavailable for 675 counties across the United States. Consequently, run-time deletion reduced the number of level-1 records to 2403 and level-2 to 43. This updates some P-values but does not affect the sign of the coefficients.

is 0.754, which is below the typical cutoff of 0.8. Most high correlations exist between racial composition and income and education. Additionally, we conduct a linear regression analysis at level 1 in IBM SPSS 27 ( $R^2 = .696$ ; without variable math grade), and find that the variable inflation factor (VIF) never exceeds the threshold of 5 (average 2.466; highest at 4.787 for nonproficiency in English). The variance-decomposition matrix also does not show any groups of predictors with high values. The results of the multilevel model are directionally confirmed, with the following observations: The effects for Native Hawaiian, Hispanic American populations, and obesity are no longer statistically significant. Conversely, the effect of the outbreak date and temperature are both significant (Beta = -0.060,  $P < .001$  and -0.098,  $P < .001$ , respectively).

Fourth, we rerun our model excluding the 23 counties of the New York metropolitan area. As a COVID-19 hotspot, they could unduly influence our analysis. All coefficients keep their sign and significance, with the exception of household income (1.983, [-0.106; 4.072],  $P = .063$ ).

Fifth, because there is no statistically correct choice for centering decisions in multilevel models,<sup>35</sup> we retest our model with raw values. With the exception of the variables at level 2 losing statistical significance, the results are fully consistent with the group- and grand-mean centered predictors in Table 3.

Sixth, we consider the assumption of multivariate normality in the multilevel model. We use a probability plot of the Mahalanobis



distance and the expected values of the order statistic to gauge the extent of normality at level 2 and find that points are not substantially distanced from the reference line. While the Kolmogorov-Smirnov test suggests a nonnormal distribution of the residuals at level 1 (0.045,  $df = 2958$ ,  $P < .001$ ), the histograms show only some non-normality in the left tail. Moreover, even severe nonnormality in multi-level models does not cause the regression coefficients and associated standard errors to have a substantial bias.<sup>36</sup>

Lastly, we are aware that an accurate estimation and comparison of the outbreak rate across units depend on similar testing strategies, test sensitivities, specificities, and reporting of tests performed vs individuals tested.<sup>13,37</sup> Even within the United States, some states report tests performed and others individuals tested.<sup>37</sup> The number of tests administered and the number of confirmed cases therefore correlate to varying extents across states.<sup>38</sup> By using a multilevel model and an exponential growth coefficient, we aim to accommodate such differences between states.

## 4 | DISCUSSION OF RESULTS

In the absence of national-level data controlled for location and disaggregated by race and ethnicity, demographics, information about comorbidity, and other personal health variables, an ecological analysis provides an alternative way of measuring the disproportionate impact of COVID-19 across the United States and among segments of Americans. It may be contrary to expectations that the outbreak rate of a new pathogen, which is able to infect virtually anyone, manifests contextual disparities. But for other conditions, such as HIV and cancer, regional health disparities have been reported before,<sup>39,40</sup> and with the current study, we show that contextual factors in the United States are also associated with a variation in COVID-19 cases.

Our analysis indicates that higher outbreak rates can be found in U.S. states characterized by a higher cultural value of collectivism (coefficient 1.330, confidence interval [0.760; 1.900],  $P < .001$ ). As Table 2 shows, collectivistic values are more prevalent in counties that are warmer (correlation with temperature 0.715,  $P < .001$ ) and have a higher percentage of people with a black/African background (with Black/African American 0.539,  $P < .001$ ). This mirrors findings from international cultural research.<sup>16</sup> Government spending is also positively linked to the outbreak (5.948, [1.693; 9.333],  $P = .005$ ), likely because the expansionary economic effect of public spending leads to more social interactions.<sup>41</sup> Conversely, we cannot find any statistical evidence that the gender of the governor or the party in control would be in any way linked to the outbreak; this certainly does not support reporting by some popular media.<sup>7</sup>

A disproportionately stronger outbreak of COVID-19 cases can be found in counties with a higher percentage of Black/African (0.485, [0.265; 0.705],  $P < .001$ ) and Asian Americans (2.629, [1.420; 3.838],  $P < .001$ ), which support prior infection and mortality studies in the United States and United Kingdom.<sup>22,42</sup> The former counties are also characterized by a higher rate of sleep deprivation (0.568,  $P < .001$ ) and warmer temperatures (0.533,  $P < .001$ ). The latter are

typically not rural ( $-0.501$ ,  $P < .001$ ) and have a higher population density (0.553,  $P < .001$ ). Native American communities also witnessed a higher initial outbreak rate (0.909, [0.515; 1.303],  $P < .001$ ). This also holds true for native Hawaiian (5.924, [0.515; 1.303],  $P = .010$ ) and Hispanic American populations (0.269, [0.048; 0.490],  $P = .017$ ), which are both characterized with lower proficiency in English (0.630 and 0.749, respectively,  $P < .001$ ).

The model also unveils a positive influence of population density on the outbreak rate (0.173, [0.065; 0.281],  $P = .002$ ). A negative association of higher average temperatures with the outbreak is only directionally informative but not statistically significant ( $-0.263$ , [ $-0.780$ ; 0.254],  $P = .319$ ), which could potentially be explained by people spending less time indoors. We see that better language fluency and higher education levels are associated with a less aggressive outbreak (nonproficiency in English: 1.884; [0.320; 3.448];  $P = .018$ ; math grade:  $-0.002$ , [ $-0.004$ ; 0.000];  $P = .010$ ), but higher income levels show a positive association (2.418, [0.044; 4.782];  $P = .046$ ). In counties with a higher household income, the obesity rate and the percentage of smokers tend to be lower ( $-0.518$ ,  $P < .001$  and  $-0.666$ ,  $P < .001$ , respectively), which are both negatively associated with the outbreak rate ( $-0.896$ , [ $-1.361$ ;  $-0.431$ ],  $P < .001$  and  $-1.789$ , [ $-3.077$ ;  $-0.501$ ],  $P = .007$ , respectively). Studies report that people with obesity are at increased risk of developing severe COVID-19 symptoms,<sup>43</sup> but, to the best of our knowledge, a link to the infection rate has not yet been established. A potential explanation of this is that people with obesity heed the warnings issued by the CDC and are extra careful in avoiding social contact, in line with the value expectancy concepts of the health belief model.<sup>26</sup> Other studies report that smoking or medicinal nicotine might be a protective factor against infection by SARS-CoV-2.<sup>27</sup> Many other variables related to good personal health are associated with a slower outbreak (social associations:  $-0.552$ , [ $-0.966$ ;  $-0.138$ ],  $P = .009$ ; sleep deprivation: 1.638, [0.829; 2.447],  $P < .001$ ; preventable hospitalization: 0.001, [0.001; 0.001],  $P = .002$ ). A better food environment is not significantly associated with the outbreak rate ( $-0.901$ , [ $-3.478$ ; 1.676],  $P = .493$ ). While the food environment index is usually associated with a healthier lifestyle, better access to grocery stores and supermarkets in the vicinity also means more interaction with other people and thus an increased likelihood of transmission.

Regarding age-related demographics, we confirm that counties with an older population are more affected by the outbreak (median age: 0.619, [0.207; 1.031],  $P = .003$ ), but the percentage of persons under 18 years is not significantly associated with the outbreak rate (0.029, [ $-0.418$ ; 0.476],  $P = .898$ ). Also, we find no effect of differences in gender (0.286, [ $-0.316$ ; 0.888],  $P = .353$ ). None of these demographic variables are strongly correlated with any other variable.

Air pollution is a significant contributor to the outbreak (6.159, [4.507; 7.811],  $P < .001$ ), and, concurrently, counties with a rural environment experience a slower outbreak ( $-0.322$ , [ $-0.406$ ;  $-0.238$ ],  $P < .001$ ). This calls for studies linking air pollution to the lethality of COVID-19<sup>28,29</sup> to include the outbreak rate as a potential confounding variable.

As a final point, we want to note that we have presented associations between contextual factors and the COVID-19 outbreak which are consistent with the deliberations leading to our research model. However, these associations, even when statistically significant, are not an inference of causality. Establishing causal inference is, of course, critical for our understanding of and fight against COVID-19 but this represents a direction for further research using more detailed data at the level of individual patients.

## ACKNOWLEDGMENTS

We thank Dr Xi Wen, editor of Health Science Reports, for his guidance and suggestions; we also greatly appreciate the anonymous reviewers for their insightful comments and suggestions on earlier versions of this paper. In parallel, reviewers and participants of the AIB US Southeast 2020 Annual Conference provided further valuable suggestions.

## CONFLICT OF INTEREST

The authors declare no conflicts of interest.

## AUTHOR CONTRIBUTIONS

Conceptualization: Wolfgang Messner

Data curation: Wolfgang Messner, Sarah E. Payson

Formal analysis: Wolfgang Messner, Sarah E. Payson

Methodology: Wolfgang Messner

Project Administration: Wolfgang Messner

Writing—Original Draft Preparation: Wolfgang Messner, Sarah E. Payson

Writing—Review & Editing: Wolfgang Messner, Sarah E. Payson

All authors have read and approved the final version of the manuscript.

Wolfgang Messner had full access to all of the data in this study and takes complete responsibility for the integrity of the data and the accuracy of the data analysis.

## TRANSPARENCY STATEMENT

Wolfgang Messner affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

## HUMAN STUDIES AND SUBJECTS

No humans or animals participated in this study.

## DATA AVAILABILITY STATEMENT

The authors confirm that the data supporting the findings of this study are available within the supplementary materials of the article. These data were derived from the following resources available in the public domain: Outbreak rate: USA Facts. Coronavirus locations: COVID-19 map by county and state. Available at <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/> (accessed 28 October 2020); Cultural values: Vandello JA,

Cohen D. Patterns of individualism and collectivism across the United States. *J Pers Soc Psychol* 1999;77:279-292. County health rankings: <https://www.countyhealthrankings.org/>; Social Capital Project: <https://www.jec.senate.gov/public/index.cfm/republicans/2018/4/the-geography-of-social-capital-in-america>; Information about state institutions in the U.S.: [https://en.wikipedia.org/wiki/List\\_of\\_United\\_States\\_governors](https://en.wikipedia.org/wiki/List_of_United_States_governors); National Centers for Environment Information: <https://www.ncdc.noaa.gov/cag/county/mapping/1/tavg/202003/2/value>; Temperature mapping for Hawaii: <http://holiday-weather.com/hawaii/averages> (all accessed in May 2020).

## ORCID

Wolfgang Messner  <https://orcid.org/0000-0003-3229-6363>

## REFERENCES

- USA Facts. Coronavirus locations: COVID-19 map by county and state. <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>. Accessed October 28, 2020.
- Messner W, Payson SE. Variation in COVID-19 outbreaks at U.S. state and county levels. *Public Health*. 2020;187:15-18.
- Chen JT, Kahn R, Li R, et al. U.S. county-level characteristics to inform equitable COVID-19 response. *medRxiv*. 2020;2020.04.08.20058248:1-38.
- New York State Department of Health. COVID-19 fatalities. <https://covid19tracker.health.ny.gov/views/NYS-COVID19-Tracker/NYSDOHCOVID-19Tracker-Fatalities?%3Aembed=yes&%3Atoolbar=no&%3Atabs=n#/views/NYS%25dCOVID19%25>. Accessed May 9, 2020.
- Vandello JA, Cohen D. Patterns of individualism and collectivism across the United States. *J Pers Soc Psychol*. 1999;77:279-292.
- Woodard C. *American nations*. London: Penguin; 2011.
- Wittenberg-Cox A. What do countries with the best coronavirus responses have in common? Women leaders. *Forbes*. <https://www.forbes.com/sites/avivahwittenbergcox/2020/04/13/what-do-countries-with-the-best-coronavirus-reponses-have-in-common-women-leaders>. Accessed May 13, 2020.
- Chowell G, Sattenspiel L, Bansal S, Viboud C. Mathematical models to characterize early epidemic growth: a review. *Phys Life Rev*. 2016;18:66-97.
- Viboud C, Simonsen L, Chowell G. A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. *Epidemics*. 2016;15:27-37.
- Ma J. Estimating epidemic exponential growth rate and basic reproduction number. *Infect Dis Model*. 2020;5:129-141.
- Murray CJ. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. *medRxiv*. 2020;2020.03.27.20043752:1-26.
- Woody S, Tec M, Dahan M, et al. Projections for first-wave COVID-19 deaths across the U.S. using social-distancing measures derived from mobile phones measures. [https://www.tacc.utexas.edu/ut\\_covid-19\\_mortality\\_forecasting\\_model\\_report](https://www.tacc.utexas.edu/ut_covid-19_mortality_forecasting_model_report). Accessed April 19, 2020.
- Pearce N, Vandenbroucke JP, VanderWeele TJ, Greenland S. Accurate statistics on COVID-19 are essential for policy guidance and decisions. *Am J Public Heal*. 2020;110(7):949-951.
- Anderson RM, May RM. *Infectious Diseases of Humans. Dynamics and Control*. Oxford: Oxford University Press; 1991.
- Texier G, Farouh M, Pellegrin L, et al. Outbreak definition by change point analysis: a tool for public health decision? *BMC Med Inform Decis Mak*. 2016;16:1-12.

16. Hofstede G. *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations*. 2nd ed. Thousand Oaks, CA: Sage; 2001.
17. Kirkman BL, Lowe KB, Gibson CB. A retrospective on culture's consequences: the 35-year journey. *J Int Bus Stud*. 2017;48:12-29.
18. Fincher CL, Thornhill R, Murray DR, Schaller M. Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism. *Proc R Soc B Biol Sci*. 2008;275:1279-1285.
19. Messner W. The institutional and cultural context of cross-national variation in COVID-19 outbreaks. *Int Public Heal J*. 2021;13: (forthcoming).
20. Adolph C, Amano K, Bang-Jensen B, Fullman N, Wilkerson J. Pandemic politics: Timing state-level social distancing responses to COVID-19. *J Health Polit Policy Law*. 2020; 1-19.
21. Li B, Yang J, Zhao F, et al. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. *Clin Res Cardiol*. 2020;109: 531-538.
22. Khunti K, Singh AK, Pareek M, Hanif W. Is ethnicity linked to incidence or outcomes of COVID-19? *BMJ*. 2020;369:14-15.
23. Alsan MM, Westerhaus M, Herce M, Nakashima K, Farmer PE. Poverty, global health, and infectious disease: lessons from Haiti and Rwanda. *Infect Dis Clin North Am*. 2011;25:611-622.
24. Schoggins JW. A phospholipase linkAGE to SARS susceptibility. *J Exp Med*. 2015;212:1755.
25. Grasselli G, Zangrillo A, Zanella A, et al. Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy Region, Italy. *JAMA*. 2020;323:1574-1581.
26. Strecher VJ, Rosenstock IM. The health belief model. In: Baum A, Newman S, Weinman J, McManus C, West R, eds. *Cambridge Handbook of Psychology, Health and Medicine*. Cambridge: Cambridge University Press; 1997:113-117.
27. Tindle HA, Newhouse PA, Freiberg MS. Beyond smoking cessation: investigating medicinal nicotine to prevent and treat COVID-19. *Nicotine Tob Res*. 2020;22(9):1669-1670.
28. Wu X, Nethery RC, Sabath MB, Braun D, Dominici F. Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study. *Sci Adv*. 2020;6(45):1-6.
29. Travaglio M, Yu Y, Popovic R, Selley L, Leal NS, Martins LM. Links between air pollution and COVID-19 in England. *Environ Pollut*. 2021; 268:1-10.
30. Kaneda T, Greenbaum C. How Demographic Changes Make us more Vulnerable to Pandemics like the Coronavirus <https://www.prb.org/how-demographic-changes-make-us-more-vulnerable-to-pandemics-like-the-coronavirus/>. Accessed May 13, 2020.
31. Chiyomaru K, Takemoto K. Global COVID-19 transmission rate is influenced by precipitation seasonality and the speed of climate temperature warming. *medRxiv*. 2020;2020.04.10.20060459:1-13.
32. Payen A, Tabourier L, Latapy M. Spreading dynamics in a cattle trade network: size, speed, typical profile and consequences on epidemic control strategies. *PLoS One*. 2019;14:1-24.
33. Frank KA. Impact of a confounding variable on a regression coefficient. *Soc Method Res*. 2000;29:147-194.
34. Vatcheva P, Lee M. Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology*. 2016;6:1-9.
35. Kreft IGG, de Leeuw J, Aiken LS. The effect of different forms of centering in hierarchical linear models. *Multivariate Behav Res*. 1995;30: 1-21.
36. Maas CJM, Hox JJ. Robustness issues in multilevel regression analysis. *Stat Neerl*. 2004;58:127-137.
37. Roser M, Ritchie H, Ortiz-Ospina E. Coronavirus disease (COVID-19)—Statistics and research. <https://ourworldindata.org/coronavirus>. Accessed March 21, 2020.
38. Kaashoek J, Santillana M. COVID-19 positive cases, evidence on the time evolution of the epidemic or an indicator of local testing capabilities? A case study in the United States. *SSRN*. 2020;3574849:1-14.
39. Ransome Y, Kawachi I, Braunstein S, Nash D. Structural inequalities drive late HIV diagnosis: the role of black racial concentration, income inequality, socioeconomic deprivation, and HIV testing. *Heal Place*. 2016;42:148-158.
40. Fang CY, Tseng M. Ethnic density and cancer: a review of the evidence. *Cancer*. 2018;124:1877-1903.
41. Galí J, López-Salido JD, Vallés J. Understanding the effects of government spending on consumption. *J Eur Econ Assoc*. 2007;5: 227-270.
42. Millett GA, Jones AT, Benkeser D, et al. Assessing differential impacts of COVID-19 on black communities. *Ann Epidemiol*. 2020; 47:37-44.
43. Stefan N, Birkenfeld AL, Schulze MB, Ludwig DS. Obesity and impaired metabolic health in patients with COVID-19. *Nat Rev Endocrinol*. 2020;16:341-342.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Messner W, Payson SE. Contextual factors and the COVID-19 outbreak rate across U.S. counties in its initial phase. *Health Sci Rep*. 2021;4:e242. <https://doi.org/10.1002/hsr2.242>