Preventive Medicine Reports 24 (2021) 101624

Contents lists available at ScienceDirect

Preventive Medicine Reports

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

Predicting coronavirus disease (COVID-19) outcomes in the United States early in the epidemic

Eduardo J. Simoes [a,b,*], Chester L. Schmaltz [a,c], Jeannette Jackson-Thompson [a,b,c]

[a] University of Missouri School of Medicine, Department of Health Management and Informatics, CE707 CS&E Bldg., DC006.00 Columbia, MO 65212, USA
[b] MU Institute for Data Science and Informatics, USA
[c] Missouri Cancer Registry and Research Center (MCR-ARC), USA

ARTICLE INFO

Keywords:
COVID-19
Associated factors
Prediction
NB model
Poisson model

ABSTRACT

By 21 October 2020, the coronavirus disease (COVID-19) epidemic in the United States (US) had infected 8.3 million people, resulting in 61,364 laboratory-confirmed hospitalizations and 222,157 deaths. Currently, policymakers are trying to better understand this epidemic, especially the human-to-human transmissibility of the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in relation to social, populational, air travel related and environmental exposure factors. Our study used 50 US states' public health surveillance datasets (January 1-April 1, 2020) to measure associations of confirmed COVID-19 cases, hospitalizations and deaths with these variables. Using the resulting associations and multivariate regression (Negative Binomial and Poisson), predicted cases, hospitalizations and deaths were generated for each US state early in the epidemic. Factors associated with a significantly increased risk of COVID-19 disease, hospitalization and death included: population density, enplanement, Black race and increased sun exposure; in addition, COVID-19 disease and hospitalization were also associated with morning humidity. Although predictions of the number of cases, hospitalizations and deaths due to COVID-19 were not accurate for every state, those states with a combination of large number of enplanements, high population density, high proportion of Black residents, high humidity or low sun exposure may expect more rapid than expected growth in the number of COVID-19 events early in the epidemic.

1. Introduction

By 19 October 2020, coronavirus diseases (COVID-19) caused 40.11 million reported human infections due to the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and 1.11 million deaths worldwide (Coronavirus, 2020). In the United States (US) by 21 October 2020, 8.3 million people had developed COVID-19, with 61,364 laboratory-confirmed COVID-19-associated hospitalizations and at least 222,157 related deaths (Covid in the U.S., 2020). Fourteen months after first reported in November 2019 (China), the COVID-19 pandemic is still unabated in many parts of the world.

Policymakers are pouring over data to better understand the human-to-human transmissibility of SARS-CoV-2. As with other infectious diseases, the probability of infection with SARS-CoV-2 is not random. Factors influencing SARS-CoV-2 transmission and preventive actions are linked to host biology (human) (Coronavirus COVID-19 (SARS-CoV-2), 2020) and agent (Hu et al., 2021), host behavior and the environment (Poirier et al., 2020).

Human biological factors may include cell-mediated and antibody-mediated immunity (Shah et al., 2020). Both are affected by a host's age, gender, prior infections, nutrition, genetics and other factors.

Human behaviors affecting transmission of respiratory viruses like SARS-CoV-2 include personal hygiene (e.g., wearing mask), socialization (e.g., social distancing), employment status (e.g., essential vs. not essential) (CDC, 2019). Social distancing, an effective way to lower transmissibility, is difficult in a socio-ecological environment where participating in large gatherings or where keeping close human contact are common (CDC, 2019). Examples of such activities include watching a sporting event (e.g., baseball); sharing office space with coworkers; traveling by airplane, with close proximity of passengers in flight and at

* Corresponding author at: University of Missouri School of Medicine, Department of Health Management and Informatics, CE707 CS&E Bldg., DC006.00 Columbia, MO 65212, USA.

E-mail addresses: simoese@health.missouri.edu (E.J. Simoes), schmaltzc@health.missouri.edu (C.L. Schmaltz), jacksonthompsonj@health.missouri.edu (J. Jackson-Thompson).

https://doi.org/10.1016/j.pmedr.2021.101624
Received 3 May 2021; Received in revised form 3 September 2021; Accepted 23 October 2021
Available online 25 October 2021

Table 1
Bivariate analysis[†] of COVID-19 outcomes with enplanements, sociodemographic factors and environmental factors.

	Cases	Hospitalizations (Note: N = 32 states)	Deaths
Population density	0.3348 (0.0006***)	0.3128 (0.01194*)	0.3054 (0.0018**)
Enplanements	0.1984 (0.0421*)	0.2419 (0.0531˜)	0.1722 (0.0776˜)
Enplanements (excluding Delaware)	0.2211 (0.0250*)	N/A	0.2058 (0.0373*)
Enplanements (DE + PA)	0.2211 (0.0250*)	N/A	0.2058 (0.0373*)
Elderly (%)	−0.0297 (0.7630)	−0.0877 (0.4848)	−0.0545 (0.5804)
Poverty rate (%)	−0.096 (0.3275)	0.0631 (0.6147)	0.009 (0.9267)
Black (%)	0.1918 (0.0493*)	0.2661 (0.0328*)	0.2310 (0.0179*)
Hispanic (%)	0.1057 (0.2804)	0.0942 (0.3359)	0.2105 (0.0915˜)
Average winter temperature	−0.0450 (0.6454)	0.1091 (0.3811)	0.0156 (0.8737)
Sun (%)	−0.2473 (0.0128*)	−0.0738 (0.5584)	−0.274 (0.0058**)
Afternoon humidity	0.1960 (0.04873*)	0.0798 (0.5262)	0.2461 (0.0134**)
Morning humidity	0.0673 (0.5015)	0.0309 (0.8071)	0.074 (0.4597)

[†]Kendall's τ (p-value) used for bivariate analyses; ***p-value ≤ 0.001; **p-value ≤ 0.01; *p-value ≤ 0.05; ˜p-value ≤ 0.1.

Table 2
Fully adjusted NB model for cases.

	Estimate (log-scale)	*p*-value	Estimate (RR[‡])	RR[‡], LCLˆ (95%)	RR[‡], UCLˆ (95%)
Population density	0.0011	0.0106*	1.001	1.000	1.002
Enplanements (in millions)	0.0038	0.4125	1.004	0.994	1.015
Elderly (%)	0.0089	0.8682	1.009	0.904	1.125
Poverty rate (%)	−0.0028	0.9481	0.997	0.906	1.101
Black (%)	0.0396	0.0060**	1.040	1.005	1.077
Hispanic (%)	0.0149	0.3465	1.015	0.980	1.053
Sun (%)	−0.0714	0.0000***	0.931	0.903	0.959
Morning humidity	−0.0472	0.0109*	0.954	0.918	0.990

Note: The NB model with a dispersion parameter fit better than the simplified Poisson model (p-value ≈ 0).
[‡] rate ratio; ˆlower confidence limit; ˆupper confidence limit; ***p < 0.001; **p < 0.01; *p < 0.05; ˜p < 0.1.

the airport; and being shoulder-to-shoulder during public transportation. These activities all involve high concentrations of people in a small area (high population density) (Hamidi et al., 2020).

The socioecological environment appears to impact on human behaviors associated with transmission of SARS-CoV-2 virus. Researchers reported higher rates of COVID-19 hospitalizations and deaths among the poor (Lewis et al., 2020), people of Black race and those of Hispanic ethinicity (Webb Hooper et al., 2020).

Physical environmental factors affecting transmissibility of respiratory viruses include direct sunlight exposure, humidity (Shaman et al., 2011) and temperature (Barreca and Shimshack, 2012). Cold temperatures and lack of sunlight prolong viability of SARS-CoV-2 virus in the environment and increase its transmissibility from an environmental source to a human host. Conversely, higher temperatures and low humidity are associated with doubling time (time needed to duplicate number of COVID-19 infected subjects) (Oliveiros et al., 2020).

Table 3
Final reduced NB model for cases.[†]

	Estimate (log-scale)	*p*-value	Estimate (RR[‡])	RR[‡], LCLˆ (95%)	RR[‡], UCLˆ (95%)
Population density	0.0012	0.0004***	1.001	1.001	1.002
Enplanements	0.0068	0.0438*	1.007	0.999	1.015
Black (%)	0.0383	0.0008***	1.039	1.016	1.064
Sun (%)	−0.0658	0.0000***	0.936	0.912	0.961
Morning humidity	−0.0525	0.0022**	0.949	0.916	0.983

Note: Proportion elderly and then percent living in poverty dropped from model; adding percent Hispanic to the final model made it worse. The NB model with a dispersion parameter fit better than the simplified Poisson model (p-value ≈ 0). Random forest regression on the residuals did not indicate that non-linear or interaction terms were necessary.
[‡] rate ratio; ˆlower confidence limit; ˆupper confidence limit; ***p < 0.001; **p < 0.01; *p < 0.05; ˜p < 0.1.

Table 4
Fully adjusted NB model for hospitalizations (N = 32 states).

	Estimate (log-scale)	*p*-value	Estimate (RR[‡])	RR[‡], LCLˆ (95%)	RR[‡], UCLˆ (95%)
Population density	0.0002	0.7883	1.000	0.999	1.002
Enplanements (in millions)	0.0000	0.9979	1.000	0.978	1.024
Elderly (%)	0.0247	0.7239	1.025	0.887	1.182
Poverty rate (%)	−0.0054	0.9250	0.995	0.868	1.133
Black (%)	0.0396	0.0484*	1.040	0.997	1.088
Hispanic (%)	0.1001	0.0049**	1.105	1.025	1.193
Sun (%)	−0.0737	0.0003***	0.929	0.890	0.969
Morning humidity	−0.0119	0.6465	0.988	0.936	1.044

Note: The NB model with a dispersion parameter fit better than the simplified Poisson model (p-value ≈ 0).
[‡] rate ratio; ˆlower confidence limit; ˆupper confidence limit; ***p < 0.001; **p < 0.01; *p < 0.05; ˜p < 0.1.

Table 5
Final reduced NB model for hospitalizations (N = 32 states).

	Estimate (log-scale)	*p*-value	Estimate (RR[‡])	RR[‡], LCLˆ (95%)	RR[‡], UCLˆ (95%)
Black (%)	0.0337	0.0035**	1.034	1.013	1.058
Hispanic (%)	0.1081	0.0000***	1.114	1.074	1.160
Sun (%)	−0.0733	0.0000***	0.929	0.900	0.957

Note: Percent elderly, population density and then percent living in poverty were dropped; adding percent Hispanic to the final model additionally causes enplanements and morning humidity to be dropped. The NB model with a dispersion parameter fit better than the simplified Poisson model (p-value ≈ 0). Random forest regression on the residuals did not indicate that non-linear or interaction terms were necessary.
[‡] rate ratio; ˆlower confidence limit; ˆupper confidence limit; ***p < 0.001; **p < 0.01; *p < 0.05; ˜p < 0.1.

However, evidence for temperature impact from different studies is contradictory (Xie and Zhu, 2020).

Several observational studies evaluated the associations of COVID-19 occurrence with multiple factors (Zakeri et al., 2020). Other studies used rate of transmission and factors potentially affecting SARS-Cov-2 transmissibility to generate expectations of future COVID-19 events (IHME, 2020; Ferguson et al., 2020). These studies used complex mathematical models, transmission rate simulations and other parameters to generate expectations with wide range of results (Ioannidis

Table 6
Fully adjusted NB model for deaths.

	Estimate (log-scale)	*p*-value	Estimate (RR‡)	RR‡, LCLˇ (95%)	RR‡, UCLˆ (95%)
Population density	0.0008	0.2036	1.001	1.000	1.002
Enplanements (in millions)	−0.0012	0.8565	0.999	0.984	1.016
Elderly (%)	0.0071	0.9290	1.007	0.846	1.201
Poverty rate (%)	0.0166	0.7929	1.017	0.876	1.182
Black (%)	0.0551	0.0099**	1.057	1.004	1.113
Hispanic (%)	0.0381	0.1001	1.039	0.983	1.100
Sun (%)	−0.0931	0.0001***	0.911	0.871	0.951
Morning humidity	−0.0488	0.0790˜	0.952	0.898	1.007

Note: The NB model with a dispersion parameter fit better than the simplified Poisson model (p-value ≈ 0).
‡ rate ratio; ˇlower confidence limit; ˆupper confidence limit; ***p < 0.001; **p < 0.01; *p < 0.05; ˜p < 0.1.

Table 7
Final reduced NB model for deaths.

	Estimate (log-scale)	*p*-value	Estimate (RR‡)	RR‡, LCLˇ (95%)	RR‡, UCLˆ (95%)
Population density	0.0009	0.0562˜	1.001	1.000	1.002
Enplanements (in millions)	0.0513	0.0010***	1.053	1.023	1.085
Enplanements^2 (in millions^2)	−0.0005	0.0019**	1.000	0.999	1.000
Black (%)	0.0294	0.0239*	1.030	1.006	1.057
Sun (%)	−0.0494	0.0013**	0.952	0.927	0.977

Note: From the full model with a quadratic term for enplanements, we dropped proportion elderly, then poverty rate and then morning humidity; adding percent Hispanic to the final model made it worse. The NB model with a dispersion parameter fit better than the simplified Poisson model (p-value ≈ 0). Random forest regression on the residuals did not indicate any additional non-linear or interaction terms were necessary.
‡ rate ratio; ˇlower confidence limit; ˆupper confidence limit; ***p < 0.001; **p < 0.01; *p < 0.05; ˜p < 0.1.

Table 8
Final reduced NB model for cases with Delaware and Pennsylvania combined.

	Estimate (log-scale)	*p*-value	Estimate (RR‡)	RR‡, LCLˇ (95%)	RR‡, UCLˆ (95%)
Population density	0.0012	0.0003***	1.001	1.001	1.002
Enplanements (in millions)	0.0062	0.0673˜	1.006	0.999	1.015
Black (%)	0.0401	0.0005***	1.041	1.017	1.066
Sun (%)	−0.0661	0.0000***	0.936	0.912	0.961
Morning humidity	−0.0540	0.0017**	0.947	0.914	0.981

‡ rate ratio; ˇlower confidence limit; ˆupper confidence limit; ***p < 0.001; **p < 0.01; *p < 0.05; ˜p < 0.1.

et al., 2020). Our study aimed to use readily available public health surveillance and census data and model-based predictions using established methods (i.e., Poisson and NB) to identify "communities" at higher risk for rapid growth of the COVID-19 early in the epidemic.

Our study uses the information available in 50 US states' public health surveillance datasets for two purposes. First, it permits measurement of the association of COVID-19 incidence, hospitalization and mortality with available information about selected behavioral-environmental factors (population density and enplanements (i.e., the number of passengers boarding at an airport), demographic factors (race, advanced age) and physical environmental factors (e.g., average air temperature, amount of sunlight and humidity). Secondly, results were used to predict the expected number of COVID-19 events (cases, hospitalizations and deaths) occurring over a specified period for each of the 50 US states.

Table 9
Final reduced NB model for cases with Delaware dropped.

	Estimate (log-scale)	*p*-value	Estimate (RR‡)	RR‡, LCLˇ (95%)	RR‡, UCLˆ (95%)
Population density	0.0012	0.0003***	1.001	1.001	1.002
Enplanements (in millions)	0.0062	0.0673˜	1.006	0.999	1.015
Black (%)	0.0401	0.0005***	1.041	1.017	1.066
Sun (%)	−0.0661	0.0000***	0.936	0.911	0.961
Morning humidity	−0.0541	0.0017**	0.947	0.914	0.981

‡ rate ratio; ˇlower confidence limit; ˆupper confidence limit; ***p < 0.001; **p < 0.01; *p < 0.05; ˜p < 0.1.

Table 10
Final reduced NB model for deaths with Delaware and Pennsylvania combined.

	Estimate (log-scale)	*p*-value	Estimate (RR‡)	RR‡, LCLˇ (95%)	RR‡, UCLˆ (95%)
Population density	0.0009	0.0726˜	1.001	1.000	1.002
Enplanements (in millions)	0.0526	0.0010***	1.054	1.024	1.087
Enplanements^2 (in millions^2)	−0.0005	0.0019**	1.000	0.999	1.000
Black (%)	0.0285	0.0314*	1.029	1.005	1.056
Sun (%)	−0.0500	0.0012**	0.951	0.926	0.977

‡ rate ratio; ˇlower confidence limit; ˆupper confidence limit; ***p < 0.001; **p < 0.01; *p < 0.05; ˜p < 0.1.

Table 11
Final reduced NB model for deaths with Delaware dropped.

	Estimate (log-scale)	*p*-value	Estimate (RR‡)	RR‡, LCLˇ (95%)	RR‡, UCLˆ (95%)
Population density	0.0009	0.0721˜	1.001	1.000	1.002
Enplanements (in millions)	0.0525	0.0010**	1.054	1.024	1.087
Enplanements^2 (in millions^2)	−0.0005	0.0020**	1.000	0.999	1.000
Black (%)	0.0286	0.0309*	1.029	1.005	1.057
Sun (%)	−0.0499	0.0013**	0.951	0.926	0.977

‡ rate ratio; ˇlower confidence limit; ˆupper confidence limit; ***p < 0.001; **p < 0.01; *p < 0.05; ˜p < 0.1.

2. Methods

2.1. Data sources

The state-level COVID-19 new cases, hospitalizations and deaths as of 1 April 2020 were obtained from the COVID Tracking Project (The COVID Tracking Project). According to Google's COVID dashboard (Coronavirus (COVID-19), 2020), 1 April 2020 is near the peak of the initial case surge in the US, the low point of workplace and transit station mobility changes and at the height of the residential mobility change. Data for the 50 states were retained, but only 32 states were reporting cumulative hospitalizations as of 1 April 2020 (The COVID Tracking Project, 2021). Our study was based on publicly available

Table 12
State (N = 50) rankings of cases and observed-to-expected ratio.

	Observed		Expected (from final reduced model)		Observed-to-expected ratio			
State	Rate (per 1 million)	Rank	Rate (per 1 million)	Rank	Ratio	Ratio 95% LCLˆ	Ratio 95% UCLˆ	Rank
AL	219.7	22	481.9	34	0.46	0.31	0.66	7
AK	181.8	15	827.1	43	0.22	0.12	0.39	2
AZ	194.1	18	197.2	9	0.98	0.52	1.86	31
AR	193.5	17	237.6	13	0.81	0.61	1.09	23
CA	206.4	20	435.3	32	0.47	0.26	0.88	8
CO	515.0	40	355.5	25	1.45	0.90	2.33	44
CT	997.7	45	841.1	44	1.19	0.79	1.77	39
DE	377.9	34	714.1	41	0.53	0.38	0.74	10
FL	323.8	31	402.2	29	0.81	0.47	1.39	22
GA	436.8	38	483.6	35	0.90	0.57	1.44	27
HI	146.9	6	212.8	11	0.69	0.50	0.95	18
ID	293.8	27	258.9	16	1.13	0.83	1.55	36
IL	550.8	41	893.2	45	0.62	0.33	1.15	14
IN	381.0	35	362.8	26	1.05	0.84	1.31	35
IA	174.0	12	237.0	12	0.73	0.57	0.94	20
KS	165.4	9	159.0	4	1.04	0.77	1.41	33
KY	132.3	3	371.6	27	0.36	0.29	0.44	6
LA	1381.9	47	577.6	39	2.39	1.51	3.80	47
ME	255.9	25	201.6	10	1.27	0.92	1.74	42
MD	328.3	32	1743.6	50	0.19	0.11	0.32	1
MA	1099.2	46	1104.0	47	1.00	0.61	1.62	32
MI	2027.3	48	593.6	40	3.42	2.62	4.46	50
MN	172.5	10	309.8	21	0.56	0.45	0.69	12
MS	360.5	33	395.5	28	0.91	0.52	1.60	28
MO	257.6	26	282.2	18	0.91	0.74	1.12	29
MT	194.6	19	298.7	20	0.65	0.48	0.89	17
NE	108.6	2	174.5	5	0.62	0.46	0.84	15
NV	415.2	37	132.0	2	3.15	2.04	4.86	49
NH	305.2	28	253.7	14	1.20	0.86	1.69	41
NJ	2505.6	49	1590.7	49	1.58	0.81	3.08	45
NM	150.2	7	192.8	8	0.78	0.49	1.24	21
NY	4303.2	50	1578.9	48	2.73	1.70	4.37	48
NC	151.0	8	513.6	36	0.29	0.23	0.38	4
ND	182.4	16	188.9	7	0.97	0.72	1.30	30
OH	217.9	21	767.7	42	0.28	0.21	0.38	3
OK	181.7	14	153.8	3	1.18	0.86	1.63	38
OR	174.5	13	344.1	23	0.51	0.35	0.73	9
PA	453.4	39	549.0	38	0.83	0.68	1.01	24
RI	630.6	43	978.9	46	0.64	0.35	1.17	16
SC	251.1	24	351.3	24	0.71	0.48	1.07	19
SD	145.8	5	124.2	1	1.17	0.78	1.77	37
TN	392.6	36	435.3	31	0.90	0.72	1.13	26
TX	137.8	4	434.9	30	0.32	0.20	0.50	5
UT	315.7	30	263.4	17	1.20	0.87	1.64	40
VT	575.3	42	452.2	33	1.27	0.88	1.85	43
VA	173.9	11	316.7	22	0.55	0.41	0.74	11
WA	971.6	44	523.1	37	1.86	1.29	2.67	46
WV	106.6	1	178.7	6	0.60	0.43	0.83	13
WI	312.8	29	298.2	19	1.05	0.80	1.37	34
WY	224.6	23	254.6	15	0.88	0.59	1.32	25

ˆlower confidence limit; ˆupper confidence limit.

anonymized databases, and thus exempt from ethical compliance.

These case, hospitalization and death data were merged with the following covariate data items by state:

- Census population as of 1 July 2019 (Bureau, 2020)
- 2019 population density calculated by dividing the 1 July 2019 population by the 2008 Census estimates of land area (Land and Water Area of States and Other Entities, 2021)
- Number of enplanements by city (Passenger Boarding (Enplanement), 2020) which were then summarized to the state level
- Percent of the US population age 65 and older (Bureau, 2020)
- Percent of persons living in poverty as defined by the Census Bureau (Bureau UC. American Community Survey (ACS), 2020)
- Percent of the population that is Black alone as of 2010 (Bureau, 2020)
- Percent of the population that is Hispanic (Bureau, 2020)

and

- Weather/environmental data: average winter temperature; percent of days with sun; total hours of sun, clear days; and morning humidity and afternoon humidity.

2.2. Analyses

Analyzing the factors associated with COVID-19 cases, hospitalizations and deaths involved bivariate and multivariate analysis using GNU R (R: The R Project for Statistical Computing, 2020).

2.3. Bivariate analysis

The rates (per 1 July 2019 Census population) of the outcomes of interest (cases, hospitalizations and deaths) were bivariately compared against the covariates: enplanements, population density, proportion of

Table 13
State rankings of hospitalizations for the states reporting hospitalizations (N = 32 as of 1 April 2020) and observed-to-expected ratio.

State	Observed		Expected (from final reduced model)		Observed-to-expected ratio			
	Rate (per 1 million)	Rank	Rate (per 1 million)	Rank	Ratio	Ratio 95% LCLˆ	Ratio 95% UCLˆ	Rank
AK	12.3	3	125.5	29	0.10	0.05	0.18	2
AZ	80.0	26	67.3	25	1.19	0.50	2.82	20
AR	29.8	11	46.5	17	0.64	0.48	0.85	8
CO	88.4	28	71.2	27	1.24	0.74	2.07	21
FL	44.2	19	186.7	31	0.24	0.13	0.42	3
GA	89.7	29	70.7	26	1.27	0.72	2.23	22
HI	9.2	2	18.5	2	0.50	0.32	0.77	6
ID	25.7	8	38.5	12	0.67	0.48	0.93	11
IA	31.4	14	29.8	9	1.05	0.78	1.43	18
KS	39.1	17	39.5	13	0.99	0.75	1.31	17
ME	46.9	20	21.9	5	2.14	1.45	3.17	31
MD	86.3	27	122.7	28	0.70	0.41	1.19	13
MA	98.1	30	64.0	22	1.53	1.19	1.98	27
MN	21.6	7	33.1	11	0.65	0.49	0.87	9
MS	111.6	31	65.1	24	1.71	0.85	3.46	28
MT	15.9	5	22.2	6	0.72	0.49	1.04	14
NH	41.2	18	32.3	10	1.28	0.89	1.83	23
NY	1024.4	32	464.4	32	2.21	1.01	4.82	32
ND	30.2	12	20.4	4	1.48	1.00	2.18	26
OH	58.1	23	64.3	23	0.90	0.63	1.30	16
OK	55.3	22	28.1	8	1.97	1.41	2.76	30
OR	36.5	16	136.1	30	0.27	0.15	0.47	4
PA	48.4	21	45.4	16	1.07	0.83	1.37	19
SC	19.8	6	49.8	18	0.40	0.24	0.67	5
SD	13.6	4	16.6	1	0.82	0.54	1.24	15
TN	29.3	10	57.2	21	0.51	0.37	0.70	7
UT	28.4	9	40.4	15	0.70	0.49	1.00	12
VT	72.1	25	40.0	14	1.80	1.17	2.79	29
VA	35.7	15	54.3	20	0.66	0.47	0.92	10
WV	0.6	1	18.9	3	0.03	0.02	0.04	1
WI	68.4	24	53.1	19	1.29	0.96	1.73	24
WY	31.1	13	22.4	7	1.39	0.94	2.04	25

ˆlower confidence limit; ˆupper confidence limit

elderly, percent living in poverty, percent Black, percent Hispanic, average winter temperature, percent sun and morning humidity. The correlation was measured in terms of Kendall's τ and the corresponding p-value was performed using the "cor.test" function in GNU R (R: The R Project for Statistical Computing, 2020).

During the analysis, the state of Delaware appeared to be an extreme low outlier for enplanements. Therefore, two additional *post hoc* analyses were performed for cases and deaths; one dropped Delaware and one combined Delaware with Pennsylvania (Tables 8–11). Delaware was excluded from hospitalization analyses since it did not report cumulative hospitalizations as of 1 April 2020.

2.4. Multivariate analysis

The three response variables were separately modelled for their associations with covariates, initially with a negative binomial (NB) regression using the "glm.nb" function within the "MASS" package (Venables and Ripley, 2002) of GNU R (R: The R Project for Statistical Computing, 2020); the log-link function was used with eight covariates: population density, enplanements (in millions), percent elderly, percent living in poverty, percent Black, percent Hispanic, percent sun, and morning humidity along with the (log) population size as an offset. Next, the simplified Poisson model that drops the NB dispersion parameter was compared to the NB model via a likelihood ratio test. The number of potential covariates is large relative to the number of observations, and some are correlated with each other.

A stepwise procedure was used to reduce models to avoid overfitting and keep them parsimonious and interpretable. The better of the full NB and Poisson models was reduced by sequentially dropping the covariate with the largest p-value if larger than 5% to arrive at a reduced model. If the full NB model was reduced, then the reduced NB model was compared to the corresponding Poisson model to determine if the NB dispersion parameter is still necessary.

To determine if any important non-linear or interaction terms were missing, a nonparametric random forest regression of the reduced model's residuals was run on the full set of the original covariates using the "ranger" package (Wright and Ziegler, 2017). The necessity of additional modelling was assessed using the "importance_pvalues" function on the "Altmann" method's permutation p-values.

Final results are given in terms of the model estimates on the log-scale (and p-values), along with their effects and 95% confidence intervals (CIs) transformed to the original scale (case, death and hospitalization rate ratios). Predictions of the three outcomes are based on the final reduced models. Observed and predicted rates, cumulative cases, hospitalizations and deaths and the ratio of observed to expected counts with 95% confidence interval are also presented for all 50 states with their predicted values in Appendix B, Tables 12–14. Two *ad hoc* sensitivity analyses regarding Delaware were also performed in the multivariate analyses.

We performed a *post-hoc* power study based on re-fitting the final reduced models (selected with statistically significant terms at the 5% level). The observed power was generally above 80% for most predictors in the three final models (cases, hospitalizations and deaths) except those with p-values relatively near the cut-off of 5%: enplanement in the case model and population density and percent of the population Black in the death model. Fitting the full models indicated inflated Type I error rates and lower observed powers. An appendix with results of power and Type I error rate simulations is available upon request.

3. Results

3.1. Bivariate analysis

Bivariate analyses revealed statistically significant associations (p-

Table 14
State (N = 50) rankings of deaths and observed-to-expected ratio.

State	Observed		Expected (from final reduced model)		Observed-to-expected ratio			
	Rate (per 1 million)	Rank	Rate (per 1 million)	Rank	Ratio	Ratio 95% LCLˋ	Ratio 95% UCLˆ	Rank
AL	5.3	30	10.1	30	0.53	0.31	0.90	22
AK	4.1	22	12.4	34	0.33	0.17	0.65	11
AZ	4.0	20	3.1	3	1.30	0.56	2.99	38
AR	3.3	14	5.9	19	0.56	0.38	0.81	24
CA	4.3	25	3.5	7	1.22	0.44	3.40	36
CO	12.0	39	7.1	21	1.68	0.93	3.04	43
CT	23.8	43	12.7	36	1.87	1.04	3.37	45
DE	13.4	41	9.9	29	1.35	0.79	2.29	39
FL	4.1	21	15.2	39	0.27	0.14	0.50	6
GA	13.1	40	29.3	47	0.45	0.21	0.95	15
HI	0.7	3	5.5	17	0.13	0.08	0.21	3
ID	5.0	27	3.3	5	1.55	1.02	2.34	41
IL	11.1	38	7.2	22	1.54	0.53	4.48	40
IN	9.7	37	8.7	26	1.11	0.82	1.50	31
IA	2.9	10	4.5	13	0.63	0.44	0.92	25
KS	3.4	15	3.5	6	0.99	0.67	1.47	29
KY	3.8	17	7.7	23	0.49	0.37	0.67	18
LA	58.7	48	15.6	40	3.76	1.99	7.11	49
ME	5.2	29	4.5	14	1.15	0.77	1.74	33
MD	8.6	35	30.7	48	0.28	0.15	0.51	7
MA	26.9	45	22.9	45	1.18	0.59	2.33	34
MI	67.7	49	21.4	44	3.16	2.02	4.93	48
MN	3.0	13	10.1	31	0.30	0.20	0.45	9
MS	7.4	33	10.7	32	0.69	0.31	1.54	26
MO	2.9	11	9.2	27	0.32	0.24	0.42	10
MT	4.7	26	4.1	10	1.15	0.77	1.73	32
NE	2.1	6	4.3	12	0.48	0.34	0.69	17
NV	14.6	42	4.7	16	3.09	1.56	6.09	47
NH	2.9	12	5.8	18	0.51	0.33	0.78	20
NJ	53.5	47	44.6	49	1.20	0.48	3.01	35
NM	2.4	9	1.9	1	1.23	0.67	2.25	37
NY	99.8	50	60.1	50	1.66	0.79	3.51	42
NC	1.0	4	23.2	46	0.04	0.03	0.07	2
ND	3.9	18	4.0	9	1.00	0.66	1.50	30
OH	5.6	31	16.2	41	0.34	0.24	0.50	12
OK	7.6	34	3.6	8	2.11	1.41	3.15	46
OR	4.3	24	11.0	33	0.39	0.23	0.66	14
PA	5.8	32	16.2	42	0.36	0.25	0.51	13
RI	9.4	36	13.0	37	0.73	0.31	1.72	28
SC	5.0	28	9.4	28	0.54	0.31	0.95	23
SD	2.3	8	3.2	4	0.70	0.46	1.07	27
TN	3.5	16	12.5	35	0.28	0.20	0.38	8
TX	2.0	5	13.9	38	0.14	0.08	0.27	5
UT	2.2	7	4.6	15	0.48	0.32	0.71	16
VT	25.6	44	6.6	20	3.87	2.31	6.48	50
VA	4.0	19	7.9	25	0.50	0.34	0.75	19
WA	37.0	46	21.2	43	1.75	0.93	3.31	44
WV	0.6	2	4.1	11	0.13	0.09	0.20	4
WI	4.1	23	7.9	24	0.52	0.37	0.73	21
WY	0.0	1	2.5	2	0.00	0.00	0.00	1

ˋlower confidence limit; ˆupper confidence limit.

value < 0.05) of all three outcomes with population density and percent of population that was Black (Table 1). Enplanements (including or excluding Delaware) was statistically associated with cases, marginally so with hospitalizations and with mixed results for deaths, depending on how the outlier—Delaware—is handled. Percent sun was significantly associated with cases and deaths. Afternoon—but not morning—humidity was statistically associated with cases and deaths. Percent of the population that was Hispanic was insignificant statistically but kept in multivariate models because of its potential confounding effect.

3.2. Multivariate analysis

The statistically significant factors associated with any of the three outcomes in bivariate analyses were included in all initial multivariate regression models of the three outcomes. In addition, we evaluated if excluded factors improved residual prediction and stability of full models. The percent of population Hispanic was retained for improving model fitness. Afternoon humidity was correlated with percent sun (Pearson's correlation ρ = -0.7964411 [p-value ≈ 0], Kendall's τ = -0.5608519 [p-value < 0.001]) and was later replaced with morning humidity due to better model stability and interpretability.

3.2.1. Cases

Regressing case counts on the eight predictor variables revealed four statistically significant associations at the 5% level: population density, percent of population Black, percent sun and morning humidity (Table 2). After sequentially dropping statistically insignificant variables from the model, the final reduced model included population density, number of enplanements, percent Black, percent sun and morning humidity (Table 3). The reduced NB model with a dispersion parameter was retained. Analysis of the residuals did not indicate additional modelling was necessary. Incidence rate of COVID-19 increased between 1% and 4% per every unit increase of population density, enplanement and percent Black. Rate decreased 5% and 6% for morning humidity and percent sun, respectively. Our predicted number of cases early in the epidemic was 56% accurate, with the 95% CIs of the observed to expected ratios including the value "1″ in 28 out of 50 states (Appendix B, Table 12).

3.2.2. Hospitalizations

Regressing the 32 hospitalization counts on the selected predictor variables revealed statistically significant associations with percent population Black, percent of population Hispanic and percent sun (Table 4). After sequentially dropping statistically insignificant predictors from model, percent of population Black, percent of population Hispanic and percent sun remained in the final reduced model (Table 5). The NB model with a dispersion parameter was retained. Analysis of the residuals did not indicate additional modelling was necessary. Hospitalization rate of COVID-19 increased 3% and 11 % per every unit increase of percent population Black and percent population Hispanic, respectively. Rate decreased 7% per unit increase of percent sun. Our predicted number of hospitalizations early in the epidemic was 44% accurate, with the 95% CIs of the observed to expected ratios including the value "1″ in 14 out of 32 states (Appendix B, Table 12).

3.2.3. Deaths

Regressing death counts on the predictor variables resulted in percentage of population Black and percent sun being statistically significant (Table 6). After sequentially dropping statistically insignificant factors, the NB model with a dispersion parameter was retained. Analysis of the residuals indicated that—despite enplanements being dropped from the model—this was an important variable for explaining the residuals in this initial reduced model. Enplanements was re-added to the model along with a quadratic term. After adding the first- and second-order terms on enplanements, population density was retained in final model because it was marginally statistically significant (p-value 0.0562) and improved model fit. The model reduction process was replicated using a quadratic term to start with, and it resulted in the same final reduced model. The reduced model had statistically significant terms for enplanements (first- and second order), percent of population Black and percent sun, with population density marginally statistically significant (Table 7). Death rate of COVID-19 increased < 1% for every unit increase in population density. Death rate increased 3% per every unit increase of percent of population Black, while decreasing 5% per every unit increase in percent sun. Our predicted number of deaths early in the epidemic was 36% accurate, with the 95% CIs of the observed to expected ratios including the value "1″ in 18 out of 50 states (Appendix B, Table 12).

4. Discussion

This study aimed to identify factors associated with COVID-19 epidemic growth in the US early in the epidemic, and to use this information to generate predictions of COVID-19 occurrence. Differently from other studies focusing on forecasting number of cases, hospitalizations or deaths (IHME, 2020; Ferguson et al., 2020), our study focused on factors that could indicate target subpopulations for preventive actions to mitigate early growth of the COVID-19 epidemic. Our study confirmed associations of COVID-19 cases included the variables population density, enplanement, percent of population Black, percent sun exposure and morning humidity. COVID-19 hospitalization was significantly associated with enplanement, percent of population Black, percent of population Hispanic and percent sun. Finally, COVID-19 death was found to be significantly associated with population density, enplanements (quadratically), percent of population Black and percent sun. Across all states, the percent accuracy of COVID-19 predictions was 56%, 44% and 36% for cases, hospitalizations and deaths, respectively.

Enplanement: This study's findings of positive associations of COVID-19 cases and hospitalizations with enplanement are consistent with the Centers for Disease Control and Prevention (CDC) reports of increased risk of COVID-19 clusters due to on-board transmission of SARS-CoV-2 during long flights (Khanh et al., 2020). Other studies reported increased risk of SARS-CoV-2 infections associated with air travel (Mouton et al., 2020; Hanson et al., 2020). However, another study reported that the risk of contracting COVID-19 during air travel was lower than the risk in an office building, classroom, supermarket or commuter train (Pombal et al., 2020). Also, two separate publications indicated a low risk of in-flight transmissibility (Hoehl et al., 2020; Schwartz et al., 2020). Simulation results for COVID-19 transmission during flight suggests that rapid mixing, dilution and removal of airplane air limit exposure risk for aerosol contaminants (transcom-report-final.pdf, 2021; Guidelines for Environmental Infection Control in Health-Care Facilities, 2003; ANSI/ASHRAE/ASHE).

Population density: This study finding of the positive relationships of population density with COVID-19 cases and deaths has been inconsistently reported. While some studies found such associations (Pekmezaris et al., 2021; Kadi and Khelfaoui, 2020; Olusola et al.), others could not confirm when adjusted for other factors such as population size (Hamidi et al., 2020). Such finding could possibly be due to superior healthcare systems.

Environmental factors–Humidity, Percent Sun and Temperature: In this study, humidity and percent sun were inversely associated with increases in the observed number of COVID-19 cases, hospitalizations and deaths. Research suggests that absolute humidity conditions has a role in controlling the timing of influenza A (H1N1) epidemics (Shaman et al., 2011); humidity levels were associated with increases in influenza mortality in another study (Barreca et al., 2012). The association between humidity and number of COVID-19 cases has not been consistent. A recent study found that doubling time for confirmed COVID-19 cases correlates positively with temperature and inversely with humidity, suggesting a decrease in the rate of progression of COVID-19 with arrival of spring and summer in the northern hemisphere (Oliveiros et al., 2020). Another reported no clear association between humidity and number of COVID-19 cases (Poirier et al., 2020).

This study found no associations of the three COVID-19 outcomes with average winter temperature. Research suggests that temperature increase is associated with decreased transmission of COVID-19 cases (Poirier et al., 2020), while in winter (temperature below 3 °C) every 1 °C increase in temperature was associated with increased numbers of COVID-19 cases (Xie and Zhu, 2020).

Sociodemographic–Racial factor: This study findings of percent of the population Black associations with COVID-19 infections, hospitalizations and deaths and percent of population Hispanic association with hospitalization have been reported elsewhere (CDC, 2019). Much of the disproportional burden of COVID-19 among Black and Hispanic populations is attributable to lower socioeconomic conditions, comorbid chronic conditions and structural inequities. In the US, factors that increase COVID-19 transmissibility, such as living in a multi-generational or multifamily household, are more common among Hispanic and Black populations (NW, 2021); also, factors that negatively impact early treatment of COVID-19, such as lacking health insurance, are more common among Black and other minority groups (Mar 05 ADP, 2020). Healthcare resources impacting the management of COVID-19, such as diagnostic testing, are scarce in many predominantly minority communities (Bilal et al., 2020; Bilal et al., 2020).

Socioeconomic–Poverty: This study could not confirm associations of poverty and three COVID-19 outcomes (Lewis et al., 2020; Finch and Hernández Finch, 2020). Another study, however, found no evidence of poverty association with COVID-19 cases and deaths (Ettensperger, 2020). Differently from previous research that used a community-based index sensitive to relative poverty and disadvantage (IDD-Technical-documentation-1.pdf, 2021), this study used data on percent of population living in poverty, which is less sensitive to detect association of COVID-19 with poverty.

5. Limitations

Our study had important limitations and strengths. First, the study suffers from low *a priori* statistical power due to the relatively small number of states used in the analyses. A *post-hoc* simulation study of the observed power indicating inflated Type I errors and some terms with low observed power is available upon request. Secondly, identified associations are ecological in nature, thus precluding one from drawing individual-level inferences on developing COVID-19. Thirdly, as a state-level analysis, environmental factors (e.g., temperature) as well as sociodemographic factors (e.g., percent Black) ignore within-state variations. However, the very nature of this ecological association can be useful for large-scale public health action oriented towards a community or a subpopulation.

6. Conclusions

In summary, our study results indicate that US states with a combination of large number of enplanements, high population density, higher percentage of Black residents, high humidity and low sun exposure may expect a more rapid increase in the number of cases, hospitalizations or deaths due to COVID-19 in the early phase of a COVID-19 epidemic. Predictions for the three COVID-19 outcomes based on study multivariate models were accurate in approximately half of the states.

7. Future directions

If the study model-based analyses could be repeated using granular data at the city, county or Public Use Microdata Area level, the accuracy of predictive equations could be improved. This type of study can help public health prioritize and take a more effective and proactive stance on the prevention of a future respiratory-disease pandemic similar to COVID-19 by: 1) Identifying and targeting higher-risk subpopulations or geographical areas with preventive action early in the epidemic; and 2) Providing estimates of future healthcare and prevention needs based on new cases, hospitalizations and deaths expected to accumulate over time.

CRediT authorship contribution statement

Eduardo J. Simoes: Conceptualization, Methodology, Supervision, Writing – original draft, Formal analysis. **Chester L. Schmaltz:** Data curation, Formal analysis, Software, Writing – review & editing, Validation. **Jeannette Jackson-Thompson:** Writing – review & editing, Validation.

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.

Acknowledgement

We would like to thank collaborators David Katz, Ali Mokdad and Ross C. Brownson, for their help with ideas about the causes of the COVID-19 epidemic during 2020 discussions.

Appendices

Appendix A: Sensitivity analyses

Appendix B: State ranking of cases, hospitalizations and deaths

References

ANSI/ASHRAE/ASHE Addendum a to ANSI/ASHRAE/ASHE Standard 170-2017, 18.
Ai B, Jp S. Absolute humidity, temperature, and influenza mortality: 30 years of county-level evidence from the United States. Am. J. Epidemiol. doi:10.1093/aje/kws259.
Barreca AI, Shimshack JP. Absolute Humidity, Temperature, and Influenza Mortality: 30 Years of County-Level Evidence from the United States. *Am J Epidemiol*. 2012;176 (suppl_7):S114-S122. doi:10.1093/aje/kws259.
Bilal, U., Barber, S., Diez-Roux, A., 2020. Early Evidence of Disparities in COVID-19 Testing in US Cities. Epidemiology. https://doi.org/10.1101/2020.05.01.20087833.
Bilal U, Barber S, Tabb L, Diez-Roux AV. Spatial Inequities in COVID-19 Testing, Positivity, Incidence and Mortality in 3 US Cities: a Longitudinal Ecological Study. medRxiv. Published online October 23, 2020:2020.05.01.20087833. doi:10.1101/2020.05.01.20087833.
Bureau UC. American Community Survey (ACS). The United States Census Bureau. Accessed November 17, 2020. https://www.census.gov/programs-surveys/acs.
Bureau UC. Population and Housing Unit Estimates. The United States Census Bureau. Accessed November 17, 2020. https://www.census.gov/popest.
CDC. Coronavirus Disease 2019 (COVID-19). Centers for Disease Control and Prevention. Published February 11, 2020. Accessed January 4, 2021. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html.
CDC. Coronavirus Disease 2019 (COVID-19). Centers for Disease Control and Prevention. Published February 11, 2020. Accessed November 17, 2020. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html.
CDC. Coronavirus Disease 2019 (COVID-19) - Transmission. Centers for Disease Control and Prevention. Published October 28, 2020. Accessed November 17, 2020. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html.
Coronavirus (COVID-19). Google News. Accessed November 17, 2020. https://news.google.com/covid19/map?hl=en-US&gl=US&ceid=US:en.
Coronavirus COVID-19 (SARS-CoV-2) | Johns Hopkins ABX Guide. Accessed November 17, 2020. https://www.hopkinsguides.com/hopkins/view/Johns_Hopkins_ABX_Guide/540747/all/Coronavirus_COVID_19__SARS_CoV_2_?refer=true.
Coronavirus (COVID-19) Deaths - Statistics and Research. Our World in Data. Accessed October 16, 2020. https://ourworldindata.org/covid-deaths.
Covid in the U.S.: Latest Map and Case Count. Published March 3, 2020. Accessed November 12, 2020. https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html.
Ettensperger F. The Link between Poverty and COVID-19 Case and Mortality Rates in Germany. medRxiv. Published online August 13, 2020:2020.08.09.20171207. doi:10.1101/2020.08.09.20171207.
Neil M. Ferguson, Daniel Laydon, Gemma Dejati-Gilani et al. Report 9 - Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imp. Coll. Lond. Published online March 16, 2020. doi:https://doi.org/10.25561/77482.
Finch, W.H., Hernández Finch, M.E., 2020. Poverty and Covid-19: Rates of Incidence and Deaths in the United States During the First 10 Weeks of the Pandemic. Front. Sociol. 5 https://doi.org/10.3389/fsoc.2020.00047.
Guidelines for Environmental Infection Control in Health-Care Facilities, 2003. 545922006–001. Published online. https://doi.org/10.1037/e545922006-001.
Hamidi S, Sabouri S, Ewing R. Does Density Aggravate the COVID-19 Pandemic? J Am Plann Assoc. Published online June 18, 2020. Accessed November 17, 2020. https://www.tandfonline.com/doi/abs/10.1080/01944363.2020.1777891.
Hanson, R., Mouton, C., Grissom, A., Godges, J. (Eds.), 2020. COVID-19 Air Traffic Visualization: A New Tool Helps Analyze Commercial Air Travel Involving Infected Passengers. RAND Corporation.
Hoehl, S., Karaca, O., Kohmer, N., Westhaus, S., Graf, J., Goetsch, U., Ciesek, S., 2020. Assessment of SARS-CoV-2 Transmission on an International Flight and Among a Tourist Group. JAMA Netw. Open 3 (8), e2018044. https://doi.org/10.1001/jamanetworkopen.2020.18044.
Hu B, Guo H, Zhou P, Shi Z-L. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol. Published online October 6, 2020:1-14. doi:10.1038/s41579-020-00459-7.
IDD-Technical-documentation-1.pdf. Accessed January 4, 2021. https://poverty.umich.edu/files/2020/01/IDD-Technical-documentation-1.pdf.
IHME, 2020. COVID-19 health service utilization forecasting team MCJL. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. medRxiv Published online. https://doi.org/10.1101/2020.03.27.20043752.
Ioannidis JPA, Cripps S, Tanner MA. Forecasting for COVID-19 has failed. Int J Forecast. doi:10.1016/j.ijforecast.2020.08.004.
Kadi, N., Khelfaoui, M., 2020. Population density, a factor in the spread of COVID-19 in Algeria: statistic study. Bull. Natl. Res. Cent. 44 (1) https://doi.org/10.1186/s42269-020-00393-x.
Khanh NC, Thai PQ, Quach H-L, et al. Transmission of SARS-CoV 2 During Long-Haul Flight - Volume 26, Number 11—November 2020 - Emerging Infectious Diseases journal - CDC. doi:10.3201/eid2611.203299.
Land and Water Area of States and Other Entities: Table 12s0358.xls. In: Index of /library/publications/2011/compendia/statab/131ed/tables. Accessed March 18,

2021. https://www2.census.gov/library/publications/2011/compendia/statab/131ed/tables/.

Lewis NM. Disparities in COVID-19 Incidence, Hospitalizations, and Testing, by Area-Level Deprivation — Utah, March 3–July 9, 2020. MMWR Morb Mortal Wkly Rep. 2020;69. doi:10.15585/mmwr.mm6938a4.

Mar 05 ADP, 2020. Changes in Health Coverage by Race and Ethnicity since the ACA, 2010-2018. KFF. Published March 5, 2020. Accessed February 7, 2021. https://www.kff.org/racial-equity-and-health-policy/issue-brief/changes-in-health-coverage-by-race-and-ethnicity-since-the-aca-2010-2018/.

Mouton, C., Hanson, R., Grissom, A., Godges, J. (Eds.), 2020. COVID-19 Air Traffic Visualization: By January 31, 2020, at Least 1.5 Daily Infected Passengers Were Originating in China. RAND Corporation.

NW 1615 L. St, Suite 800Washington, Inquiries D 20036USA202-419-4300 | M-857-8562 | F-419-4372 | M. A majority of young adults in the U.S. live with their parents for the first time since the Great Depression. Pew Research Center. Accessed February 7, 2021. https://www.pewresearch.org/fact-tank/2020/09/04/a-majority-of-young-adults-in-the-u-s-live-with-their-parents-for-the-first-time-since-the-great-depression/.

Oliveiros B, Caramelo L, Ferreira NC, Caramelo F. Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases. medRxiv. Published online March 8, 2020:2020.03.05.20031872. doi:10.1101/2020.03.05.20031872.

A. Olusola, B. Olusola, O. Onafeso, F. Ajiola, S. Adelabu, Early geography of the coronavirus disease outbreak in Nigeria. GeoJournal. doi:10.1007/s10708-020-10278-1.

Passenger Boarding (Enplanement) and All-Cargo Data for U.S. Airports – Airports. Accessed November 17, 2020. https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/.

Pekmezaris R, Zhu X, Hentz R, Lesser ML, Wang JJ, Jelavic M. Sociodemographic predictors and transportation patterns of COVID-19 infection and mortality. J Public Health. doi:10.1093/pubmed/fdab191.

Poirier, C., Luo, W., Majumder, M.S., Liu, D., Mandl, K.D., Mooring, T.A., Santillana, M., 2020. The role of environmental factors on transmission rates of the COVID-19 outbreak: an initial assessment in two spatial scales. Sci. Rep. 10 (1) https://doi.org/10.1038/s41598-020-74089-7.

Pombal, R., Hosegood, I., Powell, D., 2020. Risk of COVID-19 During Air Travel. JAMA 324 (17), 1798. https://doi.org/10.1001/jama.2020.19108.

R: The R Project for Statistical Computing. Accessed November 17, 2020. https://www.r-project.org/.

Schwartz KL, Murti M, Finkelstein M, et al. Lack of COVID-19 transmission on an international flight. *CMAJ*. 2020;192(15):E410-E410. doi:10.1503/cmaj.75015.

Shah, V.K., Firmal, P., Alam, A., Ganguly, D., Chattopadhyay, S., 2020. Overview of Immune Response During SARS-CoV-2 Infection: Lessons From the Past. Front. Immunol. 11 https://doi.org/10.3389/fimmu.2020.01949.

Shaman, J., Goldstein, E., Lipsitch, M., 2011. Absolute Humidity and Pandemic Versus Epidemic Influenza. Am. J. Epidemiol. 173 (2), 127–135. https://doi.org/10.1093/aje/kwq347.

The COVID Tracking Project. The COVID Tracking Project. Accessed November 17, 20. https://covidtracking.com/.

The COVID Tracking Project. The COVID Tracking Project. Accessed February 17, 2021. https://covidtracking.com/.

transcom-report-final.pdf. Accessed February 5, 2021. https://assets.documentcloud.org/documents/20397280/transcom-report-final.pdf.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, 4th ed. Springer.

Webb Hooper, M., Nápoles, A.M., Pérez-Stable, E.J., 2020. COVID-19 and Racial/Ethnic Disparities. JAMA 323 (24), 2466. https://doi.org/10.1001/jama.2020.8598.

Wright, M.N., Ziegler, A., 2017. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J. Stat. Softw. 77 (1), 1–17. https://doi.org/10.18637/jss.v077.i01.

J X, Y Z. Association between ambient temperature and COVID-19 infection in 122 cities from China. The Science of the total environment. doi:10.1016/j.scitotenv.2020.138201.

Zakeri, R., Bendayan, R., Ashworth, M., Bean, D.M., Dodhia, H., Durbaba, S., O'Gallagher, K., Palmer, C., Curcin, V., Aitken, E., Bernal, W., Barker, R.D., Norton, S., Gulliford, M., Teo, J.T.H., Galloway, J., Dobson, R.J.B., Shah, A.M., 2020. A case-control and cohort study to determine the relationship between ethnic background and severe COVID-19. EClinicalMedicine 28, 100574. https://doi.org/10.1016/j.eclinm.2020.100574.