



OPEN

Country-level predictors of COVID-19 mortality

Paul A. Brown

This study aimed to identify country-level predictors of COVID-19 mortality, after controlling for diverse potential factors, and utilizing current worldwide mortality data. COVID-19 deaths, as well as geographic, demographic, socioeconomic, healthcare, population health, and pandemic-related variables, were obtained for 152 countries. Continuous variables were examined with Spearman's correlation, categorical variables with ANOVA or Welch's Heteroscedastic F Test, and country-level independent predictors of COVID-19 mortality identified by weighted generalized additive models. This study identified independent mortality predictors in six limited models, comprising groups of related variables. However, in the full model, only WHO region, percent of population ≥ 65 years, Corruption Perception Index, hospital beds/100,000 population, and COVID-19 cases/100,000 population were predictive of mortality, with model accounting for 80.7% of variance. These findings suggest areas for focused intervention in the event of similar future public health emergencies, including prioritization of the elderly, optimizing healthcare capacity, and improving deficient health sector-related governance.

Coronavirus disease 2019 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2, an enveloped single-strand-RNA β -coronavirus, from the family *Coronaviridae*^{1,2}. The virus was initially named 2019 novel coronavirus, after isolation from patients with viral pneumonia in Wuhan China in late December 2019³. Poor outcomes have been associated with multiple host factors, including age, sex, comorbidities, laboratory markers and lack of vaccination^{4,5}. Higher mortality was also previously associated with the delta variant^{6–8}. This variant originated in October 2021 in India and has three clades, 21A, 21I and 21J, with latter dominating across continents⁹. More recent work suggests delta variant was not more fatal than pre-delta variants⁴, but more fatal than omicron variant⁵.

However, World Health Organization describes a wide range of determinants of health, encompassing aspects of our social, economic, and physical environments, in addition to personal characteristics and behaviors¹⁰. It is therefore plausible that in addition to individual risks, country-level factors, including demographic-, socio-economic-, and environment-related health parameters, may play an important role in COVID-19 incidence and subsequent mortality.

Previous authors have reported on COVID-19 mortality risk factors within geographic regions^{11–14}, or during defined pandemic phases e.g., the first wave^{12,14}. However, other authors have reported worldwide country-level COVID-19 mortality risk factors using data close to time of publication. Implicated risk factors include general pre-pandemic variables comprising country-level demographics (population ≥ 60 years¹⁵); socioeconomic and governance factors e.g., higher GDP per capita^{16,17}, higher income disparity¹⁶, higher transport infrastructure quality and lower government effectiveness¹⁸; national healthcare metrics including lower general health expenditure, lower infectious disease system responsiveness and greater accountability¹⁹; and population health characteristics like higher prevalence of obesity^{15,16}, chronic obstructive pulmonary disease, Alzheimer's disease, and depression¹⁷. Reported risk factors also comprise pandemic-specific variables. These include pandemic features e.g., duration¹⁵ and case load¹⁹; pandemic-associated public health policies inclusive of delayed international travel restrictions¹⁵ and lower testing¹⁸; and pandemic-associated public behaviours, e.g., shorter duration of mask wearing¹⁵. In addition, prior work also explored worldwide country-level COVID-19 mortality with a focus on select predictors, for example health system parameters^{19,20}.

However, to the best of my knowledge, no report has tested potential country-level predictors of COVID-19 mortality using a wide range of potential predictors, unrestricted geographic scope, and late 2022 COVID-19 mortality data. This study therefore aimed to identify country-level independent predictors of COVID-19 mortality after controlling for a diverse range of potential factors utilizing current worldwide mortality data.

Department of Basic Medical Sciences, Faculty of Medical Sciences Teaching and Research Complex, The University of the West Indies, Mona, Kingston 7, Jamaica. email: paul.brown02@uwimona.edu.jm

Results

Study data. There were no duplicates or invalid entries. Number of countries with missing data ranged from 0 (Country) to 46 (Delta21Jp100K). One hundred and fifty-two countries had complete data for all variables. The study data therefore comprised 152 cases (countries) and 41 variables, used for all analyses.

Descriptive statistics are presented in Table 1. The dataset comprised 38, 26, 18, 48, 8, and 14 countries within the Africa, Americas, Eastern Mediterranean, Europe, South-East Asia, and Western Pacific UN-regions, respectively. COVID-19 deaths/100,000 population ranged widely between 0.13 and 660.38. Cases/100,000 population also ranged widely from 39.12 to 70,445.77. Mean per capita deaths/cases were 26.7/2300, 204/13,531, 73.3/9334,

Variable	Minimum	Mean	Median	Maximum	SD	IQR	Skewness	Kurtosis
CovDp100K	0.13	135.96	92.15	660.38	138.21	194.97	1.18	1.03
AvgTemp	- 3.71	19.01	21.74	30.01	8.24	14.09	- 0.63	- 0.69
pPopFem_Yr	2021.00	2021.00	2021.00	2021.00				
pPopFem	25.00	49.77	50.30	53.90	3.71	1.20	- 4.23	20.78
pPop65_Yr	2021.00	2021.00	2021.00	2021.00				
pPop65	1.00	9.95	7.50	29.00	6.96	12.00	0.61	- 0.99
pPopUrb_Yr	2021.00	2021.00	2021.00	2021.00				
pPopUrb	13.00	62.40	64.00	100.00	21.83	35.00	- 0.35	- 0.76
AdLit_Yr	2021.00	2021.00	2021.00	2021.00				
AdLit	0.00	82.70	93.65	100.00	23.28	25.08	- 1.59	1.69
GDPpC_Yr	2014.00	2020.91	2021.00	2021.00				
GDPpC	236.80	16,628.54	6312.80	135,682.80	23,124.36	18,265.03	2.27	5.90
IGSpGDP_Yr	2004.00	2020.39	2021.00	2021.00				
IGSpGDP	1.90	45.28	40.15	176.70	25.44	27.97	2.00	6.47
pWfUem_Yr	2021.00	2021.00	2021.00	2021.00				
pWfUem	0.30	8.36	6.40	33.60	6.07	6.35	1.52	2.45
CPI_Yr	2021.00	2021.00	2021.00	2021.00				
CPI	14.00	44.34	40.00	88.00	18.63	26.25	0.68	- 0.40
HEpGDP_Yr	2011.00	2018.89	2019.00	2019.00				
HEpGDP	1.80	6.37	6.20	16.77	2.58	3.88	0.61	0.54
GHS.In_Yr	2021.00	2021.00	2021.00	2021.00				
GHS.In	16.70	41.79	39.35	75.90	13.65	22.50	0.43	- 0.78
LEB_Yr	2020.00	2020.00	2020.00	2020.00				
LEB	54.00	72.93	74.00	85.00	7.43	10.25	- 0.63	- 0.38
HBp100K_Yr	2021.00	2021.00	2021.00	2021.00				
HBp100K	0.00	19.93	14.60	94.00	17.30	21.42	1.61	3.35
MDp100K_Yr	2021.00	2021.00	2021.00	2021.00				
MDp100K	0.10	22.91	22.40	84.50	18.96	30.97	0.62	- 0.26
NCDMortp100K_Yr	2021.00	2021.00	2021.00	2021.00				
NCDMortp100K	13.80	71.50	69.60	100.00	16.48	22.25	- 0.42	0.26
pPopH20_Yr	2007.00	2019.86	2020.00	2020.00				
pPopH20	37.00	89.01	97.00	100.00	15.62	15.00	- 1.55	1.31
PoeMgt_Yr	2021.00	2021.00	2021.00	2021.00				
CovCp100K	39.12	16,634.37	9672.71	70,445.77	18,106.17	25,029.41	1.07	0.05
Delta21Jp100K	0.00	44.73	13.85	927.08	93.53	34.53	6.18	52.87
VacFullp100	0.23	53.56	57.04	99.01	25.06	41.14	- 0.33	- 1.03

Table 1. Descriptive statistics. AdLit = adult literacy rate (percent of 15+ years); AvgTemp = average temperature (2021); CovCp100K = COVID-19 Cases/100,000 population; CovDp100K = COVID-19 Deaths/100,000 population; CPI = Corruption Perception Index; Delta21Jp100K = COVID-19 Delta 21J sequence count/100,000 tests; GDP = gross domestic product; GDPpC = GDP per capita (US\$); GHS.In = Global Health Security Index; HBp100K = hospital beds/100,000 population; HEpGDP = health expenditure (% of GDP); IGSpGDP = imports of goods and services (% of GDP); IQR = interquartile range; LEB = life expectancy at birth; MDp100K = doctors/100,000 population; NCDMortp100K = non-communicable disease (age-standardized) mortality rate/100,000 population; pPop65 = ≥ 65 years of age (percent of population); pPopFem = females (percent of population); pPopH20 = using at least basic drinking water services (percent of population); pPopUrb = urban population (percent of population); pWfUem = unemployment (percent of workforce); VacFullp100 = number of persons fully vaccinated/100 population; SD = standard deviation; Yr = year of latest available data for the associated variable.

251/33,367, 47/6971 and 44.2/18,842, within the Africa, Americas, Eastern Mediterranean, Europe, South-East Asia, and Western Pacific UN-regions, respectively, with Africa reporting the lowest per capita cases and deaths.

Data quality checks revealed most variables contained recent data (2018–2022), as seen with GDPpC_Yr, IGSpGDP_Yr, HEpGDP_Yr, and pPopH20_Yr, with minimum year of 2014, 2004, 2011, and 2007 respectively. Datapoints earlier than 2018 occurred in 1, 5, 2, and 3 cases for GDPpC, IGSpGDP, HEpGDP, and pPopH20, respectively. Thirteen variables contained outliers, but these were relatively few for most variables. Quality checks also identified a few additional datapoints of potential concern, including within the AdLit and HBp100K variables, with a solitary minimal value of 0, for Chad and Mali, respectively. However, these were not outliers as confirmed with the *boxplot* function (*graphics* package²¹), and therefore not removed. Also, no outliers were removed as they were judged to represent valid data after visualization with *plot* function (*base* package²²).

Correlation analysis. Correlations between COVID-19 mortality and continuous variables of interest ranged between -0.58 and 0.78 , with all 19 correlations being statistically significant, and the strongest correlation being with COVID-19 cases per capita (Table 2). Among the 19 associations, there were 1, 7, and 10 small, medium, and large positive correlations, as well as 1 large negative correlation (Table 2, Supplementary Fig. S1).

ANOVA analysis/Welch's Heteroscedastic F Test. Summary of the ANOVA and Welch's Heteroscedastic F Test results are presented in Table 3. Mean COVID-19 deaths did not differ across PoeMgt groups. However, there was a statistically significant difference across WHO_Region groups. Pairwise comparisons demonstrated significant differences (two-sided) between several pairs (Supplementary Table S1).

Multivariate analysis. The GAMs analyzed included 6 limited models (geographic, demographic, socioeconomic, health metric, population health, and pandemic-related) and one full model. As shown in Table 4, WHO_Region and $s(\text{AvgTemp})$, $s(\text{pPop65})$ and $s(\text{AdLit})$, $s(\text{pWfUem})$ and $s(\text{CPI})$, $s(\text{HEpGDP})$ and $s(\text{MDp100K})$, $s(\text{pPopH20})$, and $s(\text{CovCp100K})$, were identified as independent predictors of COVID-19 mortality for the gam.MODEL_1Geo, gam.MODEL_2Dem, gam.MODEL_3SoEc_mod, gam.MODEL_4Heal_mod, gam.MODEL_5PopH, and gam.MODEL_6Pand models respectively. For the gam.MODEL_Full_mod model,

Dependent	Independent	rho ^a	95% CI	Effect size ^b	p (2-sided)
CovDp100K	AvgTemp	-0.58	[-0.68, -0.46]	Large	<0.001
CovDp100K	pPopFem	0.42	[0.27, 0.54]	Medium	<0.001
CovDp100K	pPop65	0.77	[0.69, 0.83]	Large	<0.001
CovDp100K	pPopUrb	0.43	[0.28, 0.55]	Medium	<0.001
CovDp100K	AdLit	0.67	[0.57, 0.75]	Large	<0.001
CovDp100K	GDPpC	0.63	[0.52, 0.72]	Large	<0.001
CovDp100K	IGSpGDP	0.32	[0.17, 0.46]	Medium	<0.001
CovDp100K	pWfUem	0.30	[0.15, 0.45]	Medium	<0.001
CovDp100K	CPI	0.49	[0.36, 0.61]	Medium	<0.001
CovDp100K	HEpGDP	0.58	[0.47, 0.68]	Large	<0.001
CovDp100K	GHS.In	0.63	[0.51, 0.72]	Large	<0.001
CovDp100K	LEB	0.60	[0.49, 0.70]	Large	<0.001
CovDp100K	HBp100K	0.63	[0.53, 0.72]	Large	<0.001
CovDp100K	MDp100K	0.73	[0.65, 0.80]	Large	<0.001
CovDp100K	NCDMortp100K	0.44	[0.30, 0.56]	Medium	<0.001
CovDp100K	pPopH20	0.66	[0.56, 0.75]	Large	<0.001
CovDp100K	CovCp100K	0.78	[0.70, 0.84]	Large	<0.001
CovDp100K	Delta21Jp100K	0.25	[0.09, 0.40]	Small	0.002
CovDp100K	VacFullp100	0.37	[0.22, 0.50]	Medium	<0.001

Table 2. Summary of correlation analysis. AdLit = adult literacy rate (percent of 15+ years); AvgTemp = average temperature (2021); CI = confidence interval; CovCp100K = COVID-19 Cases/100,000 population; CovDp100K = COVID-19 Deaths/100,000 population; CPI = Corruption Perception Index; Delta21Jp100K = COVID-19 Delta 21J sequence count/100,000 tests; GDP = gross domestic product; GDPpC = GDP per capita (US\$); GHS.In = Global Health Security Index; HBp100K = hospital beds/100,000 population; HEpGDP = health expenditure (% of GDP); IGSpGDP = imports of goods and services (% of GDP); LEB = life expectancy at birth; MDp100K = doctors/100,000 population; NCDMortp100K = non-communicable disease (age-standardized) mortality rate/100,000 population; pPop65 = ≥ 65 years of age (percent of population); pPopFem = females (percent of population); pPopH20 = using at least basic drinking water services (percent of population); pPopUrb = urban population (percent of population); pWfUem = unemployment (percent of workforce); VacFullp100 = number of persons fully vaccinated/100 population. ^aSpearman's correlation coefficient (two-sided), degrees of freedom = 150. ^bBased on Cohen's criteria²³.

ANOVA						
	df	Sum Sq	Mean Sq	F	p (1-sided)	R ²
PoeMgt	2	61,628	30,814	1.626	0.2	0.021
Residuals	149	2,822,852	18,945			
Welch's Heteroscedastic F Test						
	df _{num}	df _{denom}	Statistic	p (1-sided)		
WHO_Region	5	51.26133	29.86029	<0.001		

Table 3. Summary of ANOVA/Welch's Heteroscedastic F Test. df = degrees of freedom; denom = denominator; F = F statistic; Mean Sq = mean square; num = numerator; PoeMgt = point of entry management; Sum Sq = sum of squares; WHO_Region = WHO region.

independent predictors were WHO_Region, s(pPop65), s(CPI), s(HBp100K), and s(CovCp100K), accounting for 80.7% of variance. All models demonstrated practical significance with at least medium effect size: $R^2 > 13\%$ or Cohen's $f^2 > 0.15^{23}$. Models fitted with standardized independent variables produced similar results. Backward elimination using the R^2 criterion, identified gam.MODEL_Full_mod as most parsimonious, without further modification (not shown).

Mortality varied across regions with highest mean deaths in Europe and the Americas. Initial analysis revealed that mortality was significantly greater in the Americas compared with Africa, with an average increase of 66.1 deaths/100,000 population. This analysis also suggested that the Americas and to a lesser extent Eastern Mediterranean and Europe appeared to stand out from the other regions. Additional models were therefore run with these WHO regions as reference. With the Americas as reference group, in addition to Africa, mortality was significantly greater in the Americas compared with South-East Asia and Western Pacific, with an average increase of 60.4 and 80.4 deaths/100,000 population respectively. With Eastern Mediterranean as reference group, mortality was significantly greater in this region compared with Western Pacific, with an average increase of 50.8 deaths/100,000 population. Finally, with Europe as reference group, mortality was significantly greater in Europe compared with Western Pacific, with an average increase of 49.6 deaths/100,000 population.

Partial effects. Partial effects plots assessed impact of individual predictors. The gam.MODEL_1Geo model demonstrated greatest COVID-19 mortality at average temperature between approximately 10 °C to 15 °C (Supplementary Fig. S2). In the gam.MODEL_2Dem model, mortality increased with percent of population ≥ 65 up to approximately 20% then tapered off, but was initially unchanged with increasingly percent of adult literacy, before increasing above approximately 70% (Supplementary Fig. S3). Among socioeconomic variables, deaths increased logarithmically with Corruption Perception Index (CPI), but increased with greater percent unemployment up to approximately 15% before leveling off (Supplementary Fig. S4). COVID-19 death increased progressively with health expenditure relative to GDP and doctors/100,000, as shown in gam.MODEL_4Heal_mod model (Supplementary Fig. S5). In the gam.MODEL_5PopH and gam.MODEL_6Pand models, deaths increased exponentially with percent of population using at least basic drinking water services (Supplementary Fig. S6), but increased logarithmically with COVID-19 Cases/100,000 population (Supplementary Fig. S7), respectively.

Including all eligible variables in the gam.MODEL_Full_mod model, demonstrated that COVID-19 deaths increased progressively with greater percent population ≥ 65 and up to approximately 30,000 cases/100,000, but decreased progressively above a CPI of approximately 50 and above approximately 50 hospital beds/100,000 population. Assessment of the trend changes for these independent predictors comparing the modified full model with the corresponding smooth term versus the linear term, using the F-test²⁵, demonstrated significant differences (1-sided) for CPI ($F = 3.3735$, $p = 0.03838$) and CovCp100K ($F = 5.7823$, $p = 0.003924$). The partial effects plots also suggested deaths increased progressively with increasing COVID-19 Delta 21J sequence count/100,000 tests, and decreased progressively above life expectancy at birth (LEB) of approximately 75. However, these latter trends were not statistically significant (Fig. 1).

Differences in continuous independent predictors across who regions. As with COVID-19 deaths/100,000 population, WHO regions also differed significantly in percent of population ≥ 65 years of age (Welch's Heteroscedastic F Test statistic = 71.86161, $df_{num} = 5$, $df_{denom} = 38.79369$, $p < 0.001$), CPI (Welch's Heteroscedastic F Test statistic = 11.31819, $df_{num} = 5$, $df_{denom} = 45.55676$, $p < 0.001$), HBp100K (Welch's Heteroscedastic F Test statistic = 26.15499, $df_{num} = 5$, $df_{denom} = 40.21732$, $p < 0.001$), and COVID-19 cases/100,000 population (Welch's Heteroscedastic F Test statistic = 28.97315, $df_{num} = 5$, $df_{denom} = 37.22894$, $p < 0.001$). For all five variables tested across WHO regions, Africa had the lowest mean, while Europe had the highest. Pairwise comparisons demonstrated significant differences (two-sided) between several WHO region pairs for these variables: always including Africa vs. Europe and South-East Asia vs. Europe, as well as Africa vs. Americas for cases, deaths, and percent of population ≥ 65 years (Supplementary Tables S1–S5).

Linear modelling. Multiple linear regression identified similar independent predictors for the limited models, except the MODEL_1Geo model where only WHO_Region was predictive. However, in the corresponding full linear model (MODEL_Full_mod), COVID-19 mortality independent predictors were WHO_Region, pPop65, CPI, and AdLit, accounting for 63.2% of variance. Neither HBp100K nor CovCp100K were predictive (not shown).

Model/variable ^a	Coefficient ^c	EDF	t/F	p-value ^d	R ² (%) ^e	f ^{2f}	Power ^f
gam.MODEL_1Geo					58.7	0.956	1.00
Intercept	33.906	1.000	4.355	0.000			
WHO_Region: Americas	162.379	1.000	7.719	0.000			
WHO_Region: Eastern Mediterranean	43.162	1.000	3.287	0.001			
WHO_Region: Europe	168.861	1.000	7.599	0.000			
WHO_Region: South-East Asia	27.732	1.000	1.698	0.092			
WHO_Region: Western Pacific	22.890	1.000	1.964	0.051			
s(AvgTemp)	34.646	2.142	5.033	0.004			
gam.MODEL_2Dem					60.4	0.926	1.00
Intercept	113.384	1.000	17.250	0.000			
s(pPopFem)	- 10.148	1.810	1.371	0.289			
s(pPop65)	120.969	2.064	14.395	0.000			
s(pPopUrb)	18.034	2.286	1.860	0.243			
s(AdLit)	- 4.664	2.794	6.911	0.000			
gam.MODEL_3SoEc_mod ^b					30.7	0.325	1.00
Intercept	116.803	1.000	12.500	0.000			
s(pWfUem)	45.540	1.415	3.338	0.024			
s(CPI)	79.084	1.492	20.556	0.000			
gam.MODEL_4Heal_mod ^b					53.6	0.935	1.00
Intercept	119.299	1.000	16.980	0.000			
s(HEpGDP)	30.374	1.892	8.218	0.000			
s(LEB)	65.371	2.164	3.775	0.057			
s(HBp100K)	79.553	1.911	2.253	0.063			
s(MDp100K)	53.354	1.448	5.084	0.014			
gam.MODEL_5PopH					45.5	0.373	1.00
Intercept	108.766	1.000	12.370	0.000			
s(NCDMortp100K)	16.963	2.408	0.477	0.699			
s(pPopH20)	- 25.640	2.419	23.591	0.000			
gam.MODEL_6Pand					61.3	0.603	1.00
Intercept	103.031	1.000	11.180	0.000			
PoeMgt: future emergency plan	21.546	1.000	1.267	0.207			
PoeMgt: none	4.902	1.000	0.643	0.521			
s(CovCp100K)	215.521	1.526	70.287	0.000			
s(Delta21Jp100K)	6.918	1.104	0.726	0.434			
s(VacFullp100)	17.292	2.315	2.091	0.097			
gam.MODEL_Full_mod ^b					80.7	2.706	1.00
Intercept	101.832	1.000	5.902	0.000			
WHO_Region: Americas	66.142	1.000	2.849	0.005			
WHO_Region: Eastern Mediterranean	36.547	1.000	1.809	0.073			
WHO_Region: Europe	35.363	1.000	1.229	0.221			
WHO_Region: South-East Asia	5.706	1.000	0.207	0.836			
WHO_Region: Western Pacific	- 14.271	1.000	- 0.619	0.537			
PoeMgt: Future Emergency Plan	- 8.947	1.000	- 0.511	0.610			
PoeMgt: None	- 3.372	1.000	- 0.287	0.774			
s(AvgTemp)	27.531	2.569	1.141	0.303			
s(pPop65)	45.536	2.349	8.569	0.000			
s(pPopUrb)	8.429	2.543	0.538	0.675			
s(AdLit)	- 44.298	2.900	1.191	0.290			
s(pWfUem)	- 2.551	2.137	0.480	0.750			
s(CPI)	41.194	2.269	5.307	0.002			
s(HEpGDP)	- 8.210	2.514	0.544	0.642			
s(LEB)	63.659	2.535	2.618	0.056			
s(HBp100K)	11.532	2.324	3.344	0.015			
s(CovCp100K)	188.652	2.026	10.701	0.000			
s(Delta21Jp100K)	14.773	1.342	3.644	0.059			
s(VacFullp100)	21.700	2.479	0.432	0.725			

Table 4. Independent predictors of COVID-19 mortality. ^aAll models with CovDp100K as dependent variable. ^bModel modified by removing variable(s) with high partial concavity. ^cCoefficient for smooth terms = sum of basis function coefficients. ^dp-values are approximations computed from relevant statistic and effective (estimated²⁴) degrees of freedom²⁵. ^eReported as “Deviance explained”²⁵. ^f f^2 and power (rounded) computed with `pwr.f2.test` function (`pwr` package²⁶). 1Geo = geographic; 2Dem = demographic; 3SoEc = socioeconomic; 4Heal = health metric; 5PopH = population health; 6Pand = pandemic-related; AdLit = adult literacy rate (percent ≥ 15 years²⁷); AvgTemp = average temperature (2021); CovCp100K = COVID-19 Cases/100,000 population; CPI = Corruption Perception Index (0–100, where 100 = best²⁷); Delta21Jp100K = COVID-19 Delta 21J sequence count/100,000 tests; EDF = effective (estimated²⁴) degrees of freedom, a measure of smooth complexity, with 1 implying linearity²⁸; F = F statistic (numeric variables); gam.MODEL = generalized additive model; GDP = gross domestic product; HBp100K = hospital beds/100,000 population; HEpGDP = health expenditure (% of GDP); LEB = life expectancy at birth; MDp100K = doctors/100,000 population; mod = modified; NCDMortp100K = non-communicable disease (age-standardized) mortality rate/100,000 population; PoeMgt = point of entry management; pPop65 = ≥ 65 years of age (percent of population); pPopFem = females (percent of population); pPopH₂O = using at least basic drinking water services (percent of population); pPopUrb = urban population (percent of population); pWfUem = unemployment (percent of workforce); t = t statistic (intercept and categorical variables); VacFullp100 = number of persons fully vaccinated/100 population; WHO = World Health Organization; WHO_Region = WHO region. Significant values are in bold.

Discussion

This study identified WHO region, percent of population ≥ 65 years, CPI, hospital beds/100,000 population, and COVID-19 Cases/100,000 population as independent predictors of COVID-19 mortality, accounting for 80.7% of variance. Mortality varied across regions with highest mean deaths in Europe and the Americas. The partial effects plots further demonstrated that COVID-19 deaths increased progressively with greater percent of population ≥ 65 and up to approximately 30,000 cases/100,000, but decreased progressively above a CPI of approximately 50 and above approximately 50 hospital beds/100,000 population.

Mortality was significantly greater in the Americas, compared with Africa. Caseload was lowest for Africa in the dataset and consistent with WHO’s confirmation of under-reporting in this region²⁹. However, the low African mortality was determined after controlling for eligible variables including caseload. Potential contributors to the low African mortality include younger age structure with associated reduced comorbidities, genetic factors including decreased response to angiotensin-converting enzyme inhibitors, natural selection conferring protection, trained immunity-based herd immunity, lower life expectancy, and low seeding rate due to lower air traffic to the continent^{30–33}. Additional analyses revealed that mortality was also significantly greater in the Americas compared with South-East Asia and Western Pacific, in Eastern Mediterranean compared with Western Pacific, and in Europe compared with Western Pacific region. Various factors have been suggested for the low Western Pacific mortality including prior investment in pandemic preparation, as well as rapid and stringent public health responses, such as aggressive testing and early case management^{34,35}. The causes of the lower Africa, South-East Asia, and Western Pacific mortality warrant further study.

Progressively greater COVID-19 mortality with increasing percent of population ≥ 65 is not surprising given previous similar findings for proportion of population ≥ 60 ¹⁵, and can be explained by multiple factors. These include atypical presentation of respiratory infections with associated delayed intervention and polypharmacy, age-related altered immune response, increased presence of multiple comorbidities, and polypharmacy-associated enhanced susceptibility to viral infections³⁶. This finding is also consistent with the younger age-structure in Africa, contrasting that in Europe and the Americas. The initial enhanced mortality with increasing caseload also appears logical. It is less clear why mortality levels off beyond approximately 30,000 cases/100,000. Possible explanations include increasing competence^{37,38}, or resource allocation³⁹, in settings with high caseloads, offsetting the heightened burden.

There was progressively decreasing mortality above a CPI of approximately 50. CPI is a composite index derived from studies and expert surveys, published annually by Transparency International, and measures perceived public sector corruption. The index ranges from 1 to 100, with 100 indicating the lowest level of perceived corruption⁴⁰, and is strongly correlated with other measures of corruption⁴¹. It is possible that higher levels of corruption could negatively impact reporting, and recent work has shown that high CPI was associated with increased daily reported COVID-19 cases and deaths. However, this analysis was restricted to data for the initial 120 days from first confirmed case⁴², and could be reflective of the early pandemic response. In contrast, the present CPI finding is consistent with findings of a significant negative association between CPI and poor health outcomes, and a positive association between health-sector corruption specifically and chronic disease, using data covering the period 2004–2015⁴³. It is also consistent with evidence corruption undermines various aspects of healthcare system performance, including efficiency^{44,45}. However, efficiency is not guaranteed by abundance of healthcare system inputs including health expenditure⁴⁶, which may help explain why countries with comparable health expenditure differ with respect to important health outcomes including LEB and infant mortality rate⁴⁷. This lack of congruence is consistent with the present finding that health expenditure was not an independent predictor of COVID-19 mortality in the modified full model.

The decreased mortality above approximately 50 hospital beds/100,000 population is also not consistent with some prior reports. An early study, using global October 2020 data, found no significant association between

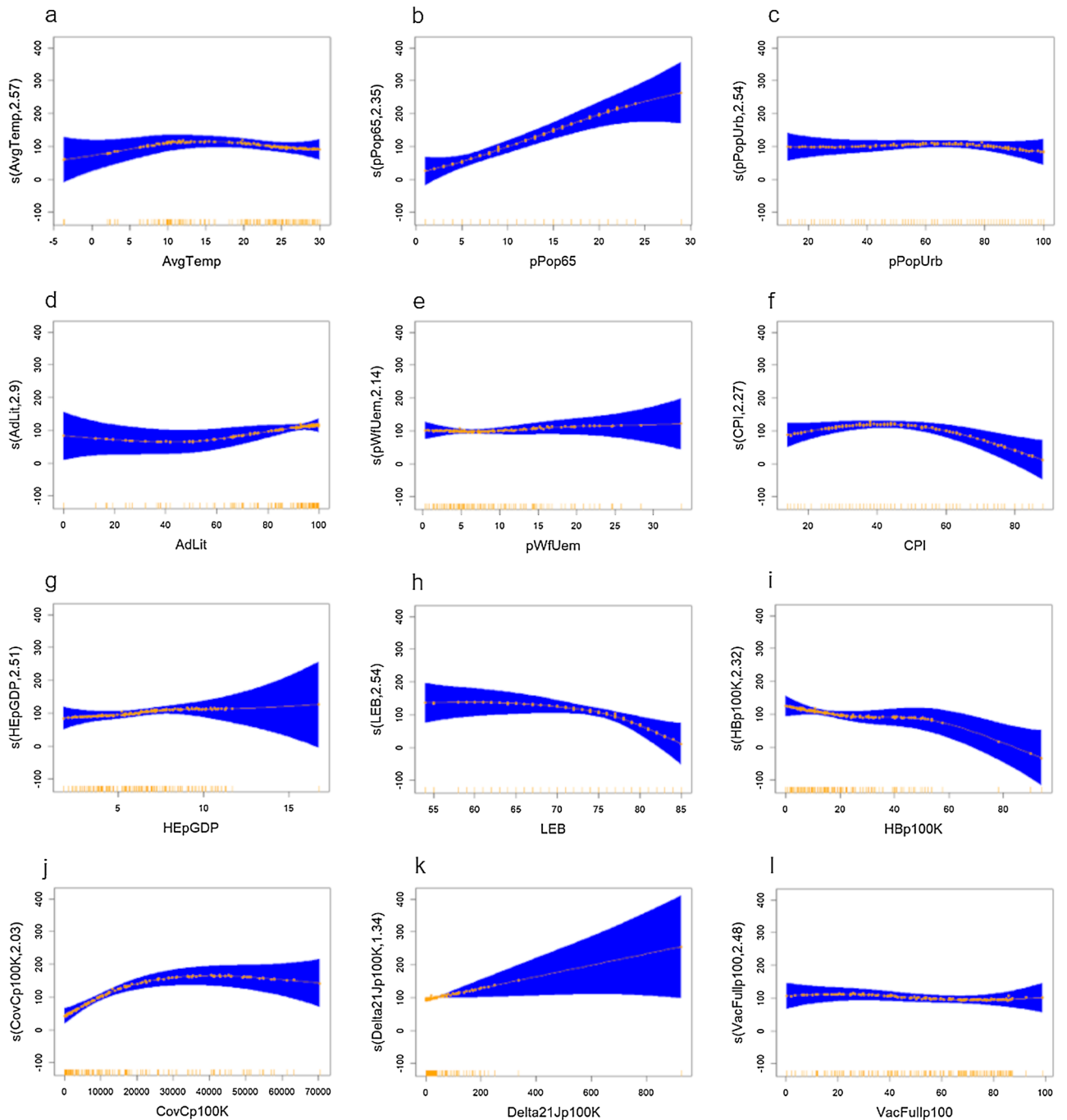


Figure 1. Partial effects plots for gam.MODEL_Full_mod model. (a) AvgTemp, (b) pPop65, (c) pPopUrb, (d) AdLit, (e) pWfUem, (f) CPI, (g) HEpGDP, (h) LEB, (i) HBp100K, (j) CovCp100K, (k) Delta21Jp100K, and (l) VacFullp100. Plotted with the *shift* argument to shift the scale based on the intercept value, for a more natural interpretation²⁸. The smooth function on y-axis therefore represents the partial effect of the independent variable on COVID-19 mortality (presented as the independent variable, with effective degrees of freedom). Shading reflects 95% confidence interval for the mean shape of the effect²⁸. AvgTemp = average temperature (2021); pPop65 = ≥ 65 years of age (percent of population); pPopUrb = urban population (percent of population); AdLit = adult literacy rate (percent ≥ 15 years); pWfUem = unemployment (percent of workforce); CPI = Corruption Perception Index; HEpGDP = health expenditure (% of GDP); LEB = life expectancy at birth; HBp100K = hospital beds/100,000 population; CovCp100K = COVID-19 Cases/100,000 population; Delta21Jp100K = COVID-19 Delta 21J sequence count/100,000 tests; VacFullp100 = number of persons fully vaccinated/100 population.

beds/100,000 and COVID-19 deaths²⁰, a result that could at least partly be due to the early data. A more recent study found increased mortality in Italian regions with higher beds per capita, after adjusting for percentage of population ≥ 65 /LEB/aging index, health expenditure per capita, general practitioners per 1000, and number of long-term care facilities⁴⁸. However, as these authors suggest, regions with higher beds per capita are more centralized, and will likely attract higher caseloads and hence mortality. This is supported by the observation that hospital beds/100,000 population was lost as an independent predictor if percentage of population ≥ 65 and cases per capita were removed from the current paper's modified full model. The current findings are also consistent with a USA report that regions with more general medicine/surgical beds per COVID-19 case had significantly lower COVID-19 mortality³⁹. Populations served by less than 50 hospital beds/100,000 may therefore be at risk.

The study results suggest some important implications. They highlight the complex nature of the relationships under investigation. For example, even though average temperature, adult literacy rate, health expenditure, doctors per capita, and percent of population using at least basic drinking water services, were all highly correlated with mortality and identified as independent predictors in their respective limited models, these relationships were lost after controlling for all eligible variables in the full model. Likewise, the present study found that Africa had the lowest cases, deaths, and CPI among WHO regions. Low African caseload was consistent with the low mortality seen in both the relevant limited model and the full model. However, although there was a positive medium correlation between CPI and mortality and progressively greater mortality with increasing CPI in the relevant limited model, after adjusting for other variables in the full model, low CPI was associated with high COVID-19 mortality. This implies the low African CPI does not adequately explain the low COVID-19 mortality seen in Africa. Also, full vaccination per capita, moderately correlated with COVID-19 deaths, was not a predictor of mortality in the full model. Previous USA⁴⁹ and global⁵⁰ analyses suggested that vaccination reduces mortality. However, the USA analysis considered any level of vaccination, controlled for county population size, social vulnerability index, and mobility changes, and assessed data from the alpha/delta phase of the pandemic⁴⁹, during which vaccinations may have been more impactful, considering the lower post-delta mortality risk⁵. Similarly, although the global analysis included a diverse range of covariates, the data was from late 2021/early 2022⁵⁰, in proximity to the delta wave⁴. Therefore, the timing of the current dataset may partly explain why Delta 21J sequence count/100,000 tests was not identified as an independent predictor of COVID-19 mortality. These findings suggest the need for future work to determine the temporal relationships between COVID-19 mortality and potential predictors. The results also imply that a substantial portion of COVID-19 mortality risk originates from factors beyond the control of individuals. Accordingly, the WHO's Sustainable Development Goal 3, "Good Health and Well-Being"⁵¹, arguably represents a justifiable mandate for countries to assume substantial responsibility for the welfare of their citizens. However, further work also seems prudent to assess the impact of other potential predictors relevant to personal responsibility.

Among the study's strengths, the dataset comprised complete data on 152 countries, with representation from all UN-defined regions. A comprehensive list of variables was also included in the models. This was important, as reported COVID-19 deaths may also depend on extraneous factors including population demographics, governance, and health system capacity. Including such variables therefore facilitated controlling for diverse factors. Additionally, the methodology utilized generalized additive models, allowing for analysis and visualization of complex, non-linear relationships. Regarding analysis, use of GAMs probably explains why HBp100K and CovCp100K were identified as independent predictors in the gam.MODEL_Full_mod model, but not the linear MODEL_Full_mod model, based on their clearly non-linear partial effects plots. Regarding visualization, comparing gam.MODEL_Full_mod with corresponding smooth term versus linear term for CPI and CovCp100K, demonstrated significant differences, implying non-linear trends, supporting utility of visualizing trend changes with GAM-based partial effects plots.

There were also some limitations. Firstly, the cross-sectional design prevents causal assumptions. In addition, the database used was dependent on available sources. It is possible that some sources under-reported COVID-19 cases and deaths, as suggested for Africa²⁹, as well as other variables. The results must therefore be interpreted accordingly. Further, data quality could have varied between different sources. However, the overall data quality was generally good, with most data being recent (2018–2022), with only a solitary datapoint in two variables appearing potentially questionable.

In conclusion, COVID-19 mortality varied across regions with highest mean deaths in Europe and the Americas. Mortality increased progressively with increasing population ≥ 65 , as well as with caseload up to $\sim 30,000/100,000$ population. Finally, mortality decreased progressively at high CPI and high hospital beds per capita. These findings suggest areas for focused intervention in the event of similar future public health emergencies, including prioritization of the elderly, optimizing healthcare capacity, and improving deficient health sector-related governance.

Methods

Study design and sample size calculation. This was a cross-sectional study. The required sample size was calculated based on the Raosoft online sample size calculator at <http://www.raosoft.com/samplesize.html>. Assuming a population of approximately 200 United Nation-defined countries, the minimum required sample size to achieve a 5% margin of error at the 95% confidence level was 132 countries.

Raw data. Data was obtained from several sources (Table 5), including two datasets from the World Health Organization (WHO); one from Trading Economics; nine from the World Bank; and one each from the Nuclear Threat Initiative/Johns Hopkins Center for Health Security/Economist Impact (NTI/JHCHS/EI), CoVariants, and Worldometer. Country, WHO region, as well as COVID-19 cases and deaths were from 16th December 2022. COVID-19 Delta 21J sequence count per country (based on data made available by GISAID: the Global

Variable ^{a,b}	Dataset	Source	Access	Ref
Identifier				
Country	WHO-COVID-19-global-table-data	WHO	Public	29
Dependent				
CovDp100K	WHO-COVID-19-global-table-data	WHO	Public	29
Independent				
Geographic				
WHO_Region	WHO-COVID-19-global-table-data	WHO	Public	29
AvgTemp	Average temperature by country	Trading Economics	With permission	56
Demographic				
pPopFem ^d	Population, female (% of total population)	The World Bank	CC-BY 4.0	57
pPop65	Population ages 65 and above (% of total population)	The World Bank	CC-BY 4.0	58
pPopUrb	Urban population (% of total population)	The World Bank	CC-BY 4.0	59
AdLit	2022/04/2021-GHS-Index-April-2022	NTI/JHCHS/EI	Public	60
Socioeconomic				
GDPpC ^c	GDP per capita (current US\$)	The World Bank	CC-BY 4.0	61
IGSpGDP ^c	Imports of goods and services (% of GDP)	The World Bank	CC-BY 4.0	62
pWfUem	Unemployment, total (% of total labor force) (modeled ILO estimate)	The World Bank	CC-BY 4.0	63
CPI	2022/04/2021-GHS-Index-April-2022	NTI/JHCHS/EI	Public	60
Healthcare				
HEpGDP	Current health expenditure (% of GDP)	The World Bank	CC-BY 4.0	64
GHS.In ^c	2022/04/2021-GHS-Index-April-2022	NTI/JHCHS/EI	Public	60
LEB	Life expectancy at birth, total (years)	The World Bank	CC-BY 4.0	65
HBp100K	2022/04/2021-GHS-Index-April-2022	NTI/JHCHS/EI	Public	60
MDp100K ^d	2022/04/2021-GHS-Index-April-2022	NTI/JHCHS/EI	Public	60
Population health				
NCDMortp100K ^d	2022/04/2021-GHS-Index-April-2022	NTI/JHCHS/EI	Public	60
pPopH ₂ ^d	People using at least basic drinking water services (% of population)	The World Bank	CC-BY 4.0	66
Pandemic				
PoeMgt	2022/04/2021-GHS-Index-April-2022	NTI/JHCHS/EI	Public	60
CovCp100K	WHO-COVID-19-global-table-data	WHO	Public	29
Delta21Jp100K	21J.Delta_table Reported cases and deaths by country or territory (total tests)	CoVariants Worldometer	CC-BY 4.0 Public	54,55
VacFullp100	who-data/vaccination-data	WHO	Public	67

Table 5. Data description, sources, and access. AdLit = adult literacy rate (percent ≥ 15 years²⁷); AvgTemp = average temperature (2021); CC-BY 4.0 = Creative Commons Attribution 4.0 International license; CovCp100K = COVID-19 Cases/100,000 population; CovDp100K = COVID-19 Deaths/100,000 population; CPI = Corruption Perception Index (0–100, where 100 = best²⁷); Delta21Jp100K = COVID-19 Delta 21J sequence count/100,000 tests; EI = Economist Impact; GDP = gross domestic product; GDPpC = GDP per capita (US\$); GHS = Global Health Security; GHS.In = Global Health Security Index (0–100, where 100 = best²⁷); HBp100K = hospital beds/100,000 population; HEpGDP = health expenditure (% of GDP); IGSpGDP = imports of goods and services (% of GDP); JHCHS = Johns Hopkins Center for Health Security; LEB = life expectancy at birth; MDp100K = doctors/100,000 population; NCDMortp100K = non-communicable disease (age-standardized) mortality rate/100,000 population; NTI = Nuclear Threat Initiative; PoeMgt = point of entry management; pPop65 = ≥ 65 years of age (percent of population); pPopFem = females (percent of population); pPopH₂ = using at least basic drinking water services (percent of population); pPopUrb = urban population (percent of population); Public = publicly available link/data; pWfUem = unemployment (percent of workforce); VacFullp100 = number of persons fully vaccinated/100 population; WHO = World Health Organization; WHO_Region = WHO region. ^aYear of data also extracted with associated variable of interest values. ^bRelated independent variables grouped and analyzed as limited models. ^cVariable removed due to high partial concurrency in limited model. ^dVariable removed due to high partial concurrency in full model.

initiative on sharing all influenza data^{52,53}) was from 15th December 2022. The per capita vaccination data was from 13th December 2022, and average temperature data was for 2021. Other variables were obtained for the most recent year for each country. As a result, for each of these variables, the year of collection varied between countries, which was collected as an associated “_Yr” variable (Table 1). All variables were numeric, except for WHO region and point of entry management (PoeMgt). There were six WHO regions: Africa, Americas, Eastern Mediterranean, Europe, South-East Asia, and Western Pacific. PoeMgt was measured with 3 groups: no plan; plan between public health system and border control authorities to identify international cases, and trace and

quarantine contacts in response to active public health emergencies; and plan between public health system and border control authorities to identify international cases, and trace and quarantine contacts to prepare for future public health emergencies²⁷. All variables were used as obtained from source, except for Delta 21J sequence count/100,000 tests, which was computed based on sequence counts per country from CoVariants⁵⁴ and total tests per country from Worldometer⁵⁵. Variables of interest were extracted from the original datasets and manually merged based on the “Country” identifier, in Microsoft Excel comma separated values (.csv) format.

Preprocessing. Raw data was read into R software with the *base*²² and *haven*⁶⁸ packages, and comprised 221 cases (countries) with 41 variables, including country name, COVID-19 deaths, WHO region, average temperature, COVID-19 cases, Delta 21J sequence counts per country, total tests per country, Delta 21J sequence count/100,000 tests, and per capita vaccination, in addition to 16 other potential independent variables and their associated year (Table 5). The categorical variable *PoeMgt*, originally coded as 0, 1, and 2, but normalized to 0, 50, and 100 respectively, to make them directly comparable with other indicators²⁷, was re-coded (*mutate* function from *dplyr* package⁶⁹) to reflect the underlying categories as defined in the source documentation²⁷. Data was screened for duplicates (*distinct* function from *dplyr* package⁶⁹), invalid entries (empty rows or columns), and missing data (*is.na* function from R base package²²). Countries with incomplete data were removed with the *complete.cases* function from the *stats* package⁷⁰, and lack of missing values subsequently confirmed. Outliers, defined as datapoints $> Q3 + 3 \times IQR$ or $< Q1 - 3 \times IQR$, were then detected with the *rstatix* package⁷¹. Finally, the *e1071* package⁷² was used to compute descriptive statistics.

Statistical analysis. Due to non-linearity and outliers among continuous bivariate models, Spearman’s correlation (*correlation* package⁷³) was used to examine the associations between COVID-19 mortality and continuous variables. The relationship with *PoeMgt* was tested with ANOVA (*stats* package⁷⁰), but with WHO_Region using Welch’s Heteroscedastic F Test (*onewaytests* package⁷⁴), because of heteroscedasticity between variable groups. Based on the presence of non-linear heteroscedastic multivariate models, country-level independent predictors of COVID-19 mortality were identified by weighted generalized additive models (GAMs) using the *gam* function (*mgcv* package⁷⁵). Due to cone-shaped residual plots, inverse error variances were applied as weights, as previously described for weighted least squares^{76,77}. Briefly, model residuals were regressed on model fitted values, and weights estimated as the inverse of the squared extracted fitted values. GAMs apply smooth functions to continuous independent variables that capture non-linear aspects of non-linear relationships, with each flexible smooth comprised of smaller basis functions that model a portion of the relationship⁷⁸. Adequacy of basis functions (model complexity) and concavity were tested for all models. GAMs were fitted without adjusting model complexity (k-value), using a smoothing parameter of 0.0001 to minimize risk of overfitting⁷⁸, and after removal of model variables with high partial concavity (Table 5) defined as > 0.8 between a variable pair²⁸. Non-significant or less significant high-concavity variables, were removed first. Limited models were initially fitted for groups of related variables (Table 5), and then a final (full) model for all variables that did not violate the concavity limit. Each model was evaluated with Cohen’s R^2 and f^2 to identify practically significant models defined as at least medium effect size: Cohen’s $R^2 > 13\%$ or $f^2 > 0.15$ ²³. The impact of individual variables was then assessed with partial effects plots. All seven models were fitted with multiple linear regression, fitted with standardized independent variables, and the final model was run with various WHO regions as reference, for comparison. Trend changes for identified independent predictors were further assessed, by comparing the full model with the corresponding smooth term versus the linear term, using the F-test²⁵. Differences in continuous independent predictors across WHO regions was also tested with Welch’s Heteroscedastic F Test (*onewaytests* package⁷⁴). Finally, the full model was tested by backward elimination (R^2 criterion) to determine the most parsimonious model. All data preprocessing, statistical analysis, and data visualization were performed with R, version 4.2.1 (The R Foundation for Statistical Computing, 2022). A p value of < 0.05 was considered statistically significant. However, for multiple comparisons, p-values were adjusted by Bonferroni correction.

Data availability

The dataset used in this paper was compiled from publicly available sources, each with or without a specified licence, as outlined in Table 5. The sources include two datasets from the World Health Organization; one from Trading Economics; nine from The World Bank; and one each from the Nuclear Threat Initiative/Johns Hopkins Center for Health Security/Economist Impact, CoVariants, and Worldometer, as follows: (1) WHO-COVID-19-global-table-data. WHO [accessed December 16th 2022]: <https://covid19.who.int/WHO-COVID-19-global-table-data.csv>. (2) WHO-data/vaccination-data. WHO [accessed December 13th 2022]: <https://covid19.who.int/who-data/vaccination-data.csv>. (3) Average Temperature by Country. Trading Economics [accessed December 5th 2022]: <https://tradingeconomics.com/country-list/temperature>. (4) Population, female (% of total population). The World Bank [accessed November 7th 2022]: <https://data.worldbank.org/indicator/SP.POP.TOTL.FE.ZS>. (5) Population ages 65 and above (% of total population). The World Bank [accessed November 7th 2022]: <https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS>. (6) Urban population (% of total population). The World Bank [accessed November 7th 2022]: <https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS>. (7) GDP per capita (current US\$). The World Bank [accessed October 15th 2022]: <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>. (8) Imports of goods and services (% of GDP). The World Bank [accessed November 7th 2022]: <https://data.worldbank.org/indicator/NE.IMP.GNFS.ZS>. (9) Unemployment, total (% of total labor force) (modeled ILO estimate). The World Bank [accessed November 7th 2022]: <https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS>. (10) Current health expenditure (% of GDP). The World Bank [accessed October 15th 2022]: <https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS>. (11) Life expectancy at birth, total (years). The World Bank [accessed November 7th 2022]: <https://data.worldbank.org/indicator/SP.DYN.LE00.IN>. (12)

People using at least basic drinking water services (% of population). The World Bank [accessed November 7th 2022]: <https://data.worldbank.org/indicator/SH.H2O.BASW.ZS>. (13) 2022/04/2021-GHS-Index-April-2022. Nuclear Threat Initiative/Johns Hopkins Center for Health Security/Economist Impact [accessed October 31st 2022]: <https://www.ghsindex.org/wp-content/uploads/2022/04/2021-GHS-Index-April-2022.csv>. (14) CoVariants: SARS-CoV-2 Mutations and Variants of Interest. CoVariants [accessed December 15th 2022]: https://github.com/hodcroflab/covariants/blob/83d7fdf5a782193ef64d82d8ddd93cdbfa889539/cluster_tables/21J.Delta_table.tsv. (15) Reported Cases and Deaths by Country or Territory. Worldometer [accessed December 16th 2022]: <https://www.worldometers.info/coronavirus/>. The compiled raw data is available as a supplementary file.

Code availability

R scripts used for data preprocessing, statistical analysis, and data visualization are available as supplementary files.

Received: 22 January 2023; Accepted: 3 June 2023

Published online: 07 June 2023

References

- Cevik, M., Kuppalli, K., Kindrachuk, J. & Peiris, M. Virology, transmission, and pathogenesis of SARS-CoV-2. *BMJ* **371**, m3862. <https://doi.org/10.1136/bmj.m3862> (2020).
- Samudrala, P. K. *et al.* Virology, pathogenesis, diagnosis and in-line treatment of COVID-19. *Eur. J. Pharmacol.* **883**, 173375. <https://doi.org/10.1016/j.ejphar.2020.173375> (2020).
- Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet* **395**, 565–574. [https://doi.org/10.1016/s0140-6736\(20\)30251-8](https://doi.org/10.1016/s0140-6736(20)30251-8) (2020).
- Stepanova, M. *et al.* The impact of variants and vaccination on the mortality and resource utilization of hospitalized patients with COVID-19. *BMC Infect. Dis.* **22**, 702. <https://doi.org/10.1186/s12879-022-07657-z> (2022).
- Wang, C. *et al.* Differences in incidence and fatality of COVID-19 by SARS-CoV-2 Omicron variant versus Delta variant in relation to vaccine coverage: A world-wide review. *J. Med. Virol.* <https://doi.org/10.1002/jmv.28118> (2022).
- Lin, L., Liu, Y., Tang, X. & He, D. The disease severity and clinical outcomes of the SARS-CoV-2 variants of concern. *Front. Public Health* <https://doi.org/10.3389/fpubh.2021.775224> (2021).
- Bast, E., Tang, F., Dahn, J. & Palacio, A. Increased risk of hospitalisation and death with the delta variant in the USA. *Lancet. Infect. Dis* **21**, 1629–1630. [https://doi.org/10.1016/S1473-3099\(21\)00685-X](https://doi.org/10.1016/S1473-3099(21)00685-X) (2021).
- Dhar, M. S. *et al.* Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. *Science* **374**, 995–999. <https://doi.org/10.1126/science.abj9932> (2021).
- Chakraborty, C., Bhattacharya, M., Sharma, A. R., Dhama, K. & Lee, S.-S. Continent-wide evolutionary trends of emerging SARS-CoV-2 variants: Dynamic profiles from Alpha to Omicron. *GeroScience* <https://doi.org/10.1007/s11357-022-00619-y> (2022).
- WHO. Determinants of health. (2017). <https://www.who.int/news-room/questions-and-answers/item/determinants-of-health>.
- Zhang, F. *et al.* Predictors of COVID-19 epidemics in countries of the World Health Organization African Region. *Nat. Med.* **27**, 2041–2047. <https://doi.org/10.1038/s41591-021-01491-7> (2021).
- Pana, T. A. *et al.* Country-level determinants of the severity of the first global wave of the COVID-19 pandemic: An ecological study. *BMJ Open* **11**, e042034. <https://doi.org/10.1136/bmjopen-2020-042034> (2021).
- Mulchandani, R., Babu, G. R., Kaur, A., Singh, R. & Lyngdoh, T. Factors associated with differential COVID-19 mortality rates in the SEAR nations: A narrative review. *IJID Regions* **3**, 54–67. <https://doi.org/10.1016/j.ijregi.2022.02.010> (2022).
- Klement, R. J. & Walach, H. Identifying factors associated with COVID-19 related deaths during the first wave of the pandemic in Europe. *Front. Public Health* <https://doi.org/10.3389/fpubh.2022.922230> (2022).
- Leffler, C. T. *et al.* Association of country-wide coronavirus mortality with demographics, testing, lockdowns, and public wearing of masks. *Am. J. Trop. Med. Hyg.* **103**, 2400–2411. <https://doi.org/10.4269/ajtmh.20-1015> (2020).
- Chaudhry, R., Dranitsaris, G., Mubashir, T., Bartoszko, J. & Riazi, S. A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes. *EClinicalMedicine* <https://doi.org/10.1016/j.eclinm.2020.100464> (2020).
- Hashim, M. J., Alsuwaidi, A. R. & Khan, G. Population risk factors for COVID-19 mortality in 93 countries. *J. Epidemiol. Glob. Health* **10**, 204–208. <https://doi.org/10.2991/jegh.k.200721.001> (2020).
- Rojas, D., Saavedra, J., Petrova, M., Pan, Y. & Szapocznik, J. Predictors of COVID-19 fatality: A worldwide analysis of the pandemic over time and in Latin America. *J. Epidemiol. Glob. Health* **12**, 150–159. <https://doi.org/10.1007/s44197-022-00031-x> (2022).
- Neogi, S. B., Pandey, S., Preetha, G. S. & Swain, S. The predictors of COVID-19 mortality among health systems parameters: An ecological study across 203 countries. *Health Res. Policy Syst.* **20**, 75. <https://doi.org/10.1186/s12961-022-00878-3> (2022).
- Sen-Crowe, B., Sutherland, M., McKenney, M. & Elkbuli, A. A closer look into global hospital beds capacity and resource shortages during the COVID-19 pandemic. *J. Surg. Res.* **260**, 56–63. <https://doi.org/10.1016/j.jss.2020.11.062> (2021).
- R-core. *graphics* (version 3.6.2). <https://www.rdocumentation.org/packages/graphics/versions/3.6.2> (Accessed 13 Jan 2023) (1969).
- The R Base Package: Documentation for package 'base' version 4.3.0.*, <https://stat.ethz.ch/R-manual/R-devel/library/base/html/00Index.html> (2022).
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* 2nd edn. (Erlbaum, 1988).
- Wood, S. N. *Generalized Additive Models: An Introduction with R* (Chapman & Hall, 2006).
- Faraway, J. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models* 2nd edn. (CRC Press, 2016).
- De Rosario, H. *Functions in pwr (1.3-0) RDocumentation*, <https://www.rdocumentation.org/packages/pwr/versions/1.3-0> (2020).
- Bell, J. A. & Nuzzo, J. B. Global Health Security Index: GHS Index Methodology. <https://www.GHSIndex.org> (2021).
- Ross, N., Miller, D., Simpson, G. L. & Pedersen, E. J. *Generalized Additive Models in R. Chapter 2*, <https://noamross.github.io/gams-in-r-course/chapter2> (Accessed 03 Apr 2022) (2019).
- WHO. WHO-COVID-19-global-table-data. (2022). <https://covid19.who.int/WHO-COVID-19-global-table-data.csv>.
- Ghosh, D., Bernstein, J. A. & Mersha, T. B. COVID-19 pandemic: The African paradox. *J. Glob. Health* **10**, 020348. <https://doi.org/10.7189/jogh.10.020348> (2020).
- Njenga, M. K. *et al.* Why is there low morbidity and mortality of COVID-19 in Africa?. *Am. J. Trop. Med. Hyg.* **103**, 564–569. <https://doi.org/10.4269/ajtmh.20-0474> (2020).
- Osayomi, T. *et al.* A geographical analysis of the African COVID-19 paradox: Putting the poverty-as-a-vaccine hypothesis to the test. *Earth Syst. Environ.* **5**, 799–810. <https://doi.org/10.1007/s41748-021-00234-5> (2021).
- Lawal, Y. Africa's low COVID-19 mortality rate: A paradox?. *Int. J. Infect. Dis.* **102**, 118–122. <https://doi.org/10.1016/j.ijid.2020.10.038> (2021).

34. Zheng, X.-Y., Guan, W.-J. & Zhong, N.-S. Clinical characteristics of COVID-19 in developing countries of western pacific: Low case-fatality rate unraveled. *Lancet Reg. Health Western Pac.* <https://doi.org/10.1016/j.lanwpc.2020.100073> (2021).
35. Kasai, T. *From COVID-19 containment to suppression in the Western Pacific Region: 2020 Lessons for 2021*, <https://www.who.int/westernpacific/news-room/commentaries/detail-hq/from-covid-19-containment-to-suppression-in-the-western-pacific-region-2020-lessons-for-2021> (2021).
36. Watson, A. & Wilkinson, T. M. A. Respiratory viral infections in the elderly. *Ther. Adv. Respir. Dis.* **15**, 1753466621995050. <https://doi.org/10.1177/1753466621995050> (2021).
37. Chen, C.-H., Chen, Y.-H., Lin, H.-C. & Lin, H.-C. Association between physician caseload and patient outcome for sepsis treatment. *Infect. Control Hosp. Epidemiol.* **30**, 556–562. <https://doi.org/10.1086/597509> (2015).
38. Hogg, R. S. *et al.* Relation between hospital HIV/AIDS caseload and mortality among persons with HIV/AIDS in Canada. *Clin. Investig. Med.* **21**, 27–32 (1998).
39. Janke, A. T. *et al.* Analysis of hospital resource availability and COVID-19 mortality across the United States. *J. Hosp. Med.* **16**, 211–214. <https://doi.org/10.12788/jhm.3539> (2021).
40. Domashova, J. & Politova, A. The Corruption Perception Index: Analysis of dependence on socio-economic indicators. *Procedia Comput. Sci.* **190**, 193–203. <https://doi.org/10.1016/j.procs.2021.06.024> (2021).
41. Wilhelm, P. G. International validation of the corruption perceptions index: Implications for business ethics and entrepreneurship education. *J. Bus. Ethics* **35**, 177–189. <https://doi.org/10.1023/A:1013882225402> (2002).
42. Khan, A. R., Abedin, S., Rahman, M. M. & Khan, S. Effects of corruption and income inequality on the reported number of COVID-19 cases and deaths: Evidence from a time series cross-sectional data analysis. *PLOS Glob. Public Health* **2**, e0001157. <https://doi.org/10.1371/journal.pgph.0001157> (2022).
43. Ferrarri, L. & Salustri, F. The relationship between corruption and chronic diseases: Evidence from Europeans aged 50 years and older. *Int. J. Public Health* **65**, 345–355. <https://doi.org/10.1007/s00038-020-01347-w> (2020).
44. Tormusa, D. & Mogom, A. The impediments of corruption on the efficiency of healthcare service delivery in Nigeria. *J. Health Ethics* <https://doi.org/10.18785/ojhe.1201.03> (2016).
45. Gonzalez-Aquines, A., Mohamed, B. & Kowalska-Bobko, I. Corruption in the health care sector: A persistent threat to European health systems. *Zdrowie Publiczne i Zarządzanie* <https://doi.org/10.4467/20842627OZ.21.007.15761> (2022).
46. Asandului, L., Roman, M. & Fatulescu, P. The efficiency of healthcare systems in Europe: A data envelopment analysis approach. *Procedia Econ. Finance* **10**, 261–268. [https://doi.org/10.1016/S2212-5671\(14\)00301-3](https://doi.org/10.1016/S2212-5671(14)00301-3) (2014).
47. Zarulli, V., Sopina, E., Toffolutti, V. & Lenart, A. Health care system efficiency and life expectancy: A 140-country study. *PLoS ONE* **16**, e0253450. <https://doi.org/10.1371/journal.pone.0253450> (2021).
48. Ferrara, N. *et al.* Relationship between COVID-19 mortality, hospital beds, and primary care by Italian regions: A lesson for the future. *J. Clin. Med.* <https://doi.org/10.3390/jcm11144196> (2022).
49. Suthar, A. B. *et al.* Public health impact of covid-19 vaccines in the US: Observational study. *BMJ* **377**, e069317. <https://doi.org/10.1136/bmj-2021-069317> (2022).
50. Watson, O. J. *et al.* Global impact of the first year of COVID-19 vaccination: A mathematical modelling study. *Lancet. Infect. Dis* **22**, 1293–1302. [https://doi.org/10.1016/S1473-3099\(22\)00320-6](https://doi.org/10.1016/S1473-3099(22)00320-6) (2022).
51. WHO. Targets of Sustainable Development Goal 3. (2022). <https://www.who.int/europe/about-us/our-work/sustainable-development-goals/targets-of-sustainable-development-goal-3>.
52. Shu, Y. & McCauley, J. GISAIID: Global initiative on sharing all influenza data—from vision to reality. *Eurosurveillance* <https://doi.org/10.2807/1560-7917.es.2017.22.13.30494> (2017).
53. Khare, S. *et al.* GISAIID's role in pandemic response. *China CDC Wkly.* **3**, 1049–1051. <https://doi.org/10.46234/ccdcw2021.255> (2021).
54. Hodcroft, E. B. CoVariants: SARS-CoV-2 Mutations and Variants of Interest. (2021). <https://github.com/hodcroftlab/covariants/commit/83d7fdf5a782193ef64d82d8dd93c8bfa889539>.
55. Worldometers. Reported Cases and Deaths by Country or Territory. (2022). <https://www.worldometers.info/coronavirus/>.
56. ECONOMICS, T. Average Temperature by Country. (2022). <https://tradingeconomics.com/country-list/temperature>.
57. Bank, T. W. Population, female (% of total population). (2022). <https://data.worldbank.org/indicator/SP.POP.TOTL.FE.ZS>.
58. Bank, T. W. Population ages 65 and above (% of total population). (2022). <https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS>.
59. Bank, T. W. Urban population (% of total population). (2022). <https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS>.
60. Initiative, N. T. & Security, J. H. C. f. H. 2022/04/2021-GHS-Index-April-2022. (2022). <https://www.ghsindex.org/wp-content/uploads/2022/04/2021-GHS-Index-April-2022.csv>.
61. Bank, T. W. GDP per capita (current US\$). (2022). <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>.
62. Bank, T. W. Imports of goods and services (% of GDP). (2022). <https://data.worldbank.org/indicator/NE.IMP.GNFS.ZS>.
63. Bank, T. W. Unemployment, total (% of total labor force) (modeled ILO estimate). (2022). <https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS>.
64. Bank, T. W. Current health expenditure (% of GDP). (2022). <https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS>.
65. Bank, T. W. Life expectancy at birth, total (years). (2022). <https://data.worldbank.org/indicator/SP.DYN.LE00.IN>.
66. Bank, T. W. People using at least basic drinking water services (% of population). (2022). <https://data.worldbank.org/indicator/SH.H2O.BASW.ZS>.
67. WHO. WHO-data/vaccination-data. (2022). <https://covid19.who.int/who-data/vaccination-data.csv>.
68. Wickham, H. *haven*, <https://www.rdocumentation.org/packages/haven/versions/2.4.3> (Accessed 11 Nov 2022) (2021).
69. Wickham, H. *Functions in dplyr (0.7.8)*, <https://www.rdocumentation.org/packages/dplyr/versions/0.7.8> (Accessed 11 Nov 2022) (2018).
70. R-core. *Functions in stats (3.6.2)*, <https://rdocumentation.org/packages/stats/versions/3.6.2> (Accessed 11 Nov 2022) (1969).
71. Kassambara, A. *rstatix*, <https://www.rdocumentation.org/packages/rstatix/versions/0.7.0> (Accessed 03 Apr 2022) (2021).
72. Meyer, D. *e1071 (version 1.7-12)*, <https://www.rdocumentation.org/packages/e1071/versions/1.7-12> (Accessed 29 Nov 2022) (2022).
73. Wiernik, B. M. *correlation*, <https://www.rdocumentation.org/packages/correlation/versions/0.8.3> (Accessed 29 Nov 2022) (2022).
74. Dag, O. *onewaytests (version 2.6)*, <https://www.rdocumentation.org/packages/onewaytests/versions/2.6> (Accessed 29 Nov 2022) (2021).
75. Wood, S. *Functions in mgcv (1.8-39)*, <https://www.rdocumentation.org/packages/mgcv/versions/1.8-39> (Accessed 03 Apr 2022) (2022).
76. Foley, M. *Weighted Least Squares: How to address heteroscedasticity in linear regression with R*, <https://rpubs.com/mpfoley73/500818> (2019).
77. Pardoe, I., Simon, L. & Young, D. *13.1 - Weighted Least Squares*, <https://online.stat.psu.edu/stat501/lesson/13/13.1> (Accessed 03 Apr 2022) (2022).
78. Ross, N., Miller, D., Simpson, G. L. & Pedersen, E. J. *Generalized Additive Models in R. Chapter 1.*, <https://noamross.github.io/gams-in-r-course/chapter1> (Accessed 03 Apr 2022) (2019).

Acknowledgements

PAB gratefully acknowledges all sources of the original datasets used in this secondary analysis. These include the World Health Organization, Trading Economics, the World Bank, the Nuclear Threat Initiative/Johns Hopkins Center for Health Security/Economist Impact collaboration, CoVariants, and Worldometer. In addition, regarding COVID-19 21J sequence count data from CoVariants, PAB gratefully acknowledges all data contributors, i.e., the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata, and sharing via the GISAID Initiative.

Author contributions

P.A.B.: conceptualization, methodology, formal analysis, writing—original draft, writing—review and editing, visualization, final approval.

Competing interests

The author declares no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-36449-x>.

Correspondence and requests for materials should be addressed to P.A.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023