**Original Article**

# Global variation of COVID-19 mortality rates in the initial phase

**Saman Hasan Siddiqui**[ID]**, Azza Sarfraz**[ID]**, Arjumand Rizvi**[ID]**, Fariha Shaheen**[ID]**, Mohammad Tahir Yousafzai**[ID]**, Syed Asad Ali**[ID]

Department of Pediatrics and Child Health, Aga Khan University, Karachi, Pakistan

## ABSTRACT

**Objectives:** Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused devastation in over 200 countries. Italy, Spain, and the United States (US) were most severely affected by the first wave of the pandemic. The reasons why some countries were more strongly affected than others remain unknown. We identified the most-affected and less-affected countries and states and explored environmental, host, and infrastructure risk factors that may explain differences in the SARS-CoV-2 mortality burden.

**Methods:** We identified the top 10 countries/US states with the highest deaths per population until May 2020. For each of these 10 case countries/states, we identified 6 control countries/states with a similar population size and at least 3 times fewer deaths per population. We extracted data for 30 risk factors from publicly available, trusted sources. We compared case and control countries/states using the non-parametric Wilcoxon rank-sum test, and conducted a secondary cluster analysis to explore the relationship between the number of cases per population and the number of deaths per population using a scalable EM (expectation–maximization) clustering algorithm.

**Results:** Statistically significant differences were found in 16 of 30 investigated risk factors, the most important of which were temperature, neonatal and under-5 mortality rates, the percentage of under-5 deaths due to acute respiratory infections (ARIs) and diarrhea, and tuberculosis incidence ($p < 0.05$)

**Conclusion:** Countries with a higher burden of baseline pediatric mortality rates, higher pediatric mortality from preventable diseases like diarrhea and ARI, and higher tuberculosis incidence had lower rates of coronavirus disease 2019-associated mortality, supporting the hygiene hypothesis.

**Keywords:** Coronavirus; Environment; Mortality; Public health

## Introduction

By April 1, 2020, over 810,000 confirmed cases and 40,000 confirmed deaths were reported due

to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) globally [1]. In the initial stages of the pandemic of coronavirus disease 2019 (COVID-19), which is caused by SARS-CoV-2, the incidence and mortality rates varied widely in different countries. Among the countries with the highest incidence and mortality rates due to COVID-19, certain populous countries such as the United States (US), Brazil, and Mexico were particularly affected [1]. The trajectories of COVID-19-specific mortality rates highlight the discrepancies of the COVID-19 impact worldwide. Within the US, New York City became the epicenter of the US and was particularly hard-hit early. In the first month following the confirmation of COVID-19, 130,689 cases were identified and the mortality rate was 244 deaths per million. In comparison, Pennsylvania, which has a comparable total population size to New York City, had 28,685 confirmed cases in the first month and a significantly lower mortality rate, at 59 deaths per million [1]. Similar discrepancies were noted across different countries such as Italy and France, which have comparable population sizes; however, Italy had 69,176 confirmed cases and 126 deaths per million by the end of the first month, as reported by the World Health Organization (WHO) [1], whereas France had 43,977 confirmed cases in the first month and a lower mortality rate, at 45 deaths per million, within the same period [1].

Based on variations at the state, regional, and national levels, as well as the beliefs of leaders, different policies and practices were implemented to contain the pandemic with varying success. The extent of the contribution of natural and environmental factors to variation in the COVID-19 incidence and mortality rates across different countries is unclear, as is the contribution of various policies implemented by countries and regions. Consequently, we estimated the mortality rates in the first month of the pandemic for all countries, as the total number of cases of COVID-19 is not yet known given gaps in testing for COVID-19 as reported by the WHO [1]. All countries were assessed in their first month of the pandemic, and factors associated with high morbidity and mortality were explored. This has implications because it can provide insights into what makes a society vulnerable to the rapid spread of COVID-19, with relevance for the future. In particular, identifying the risk factors that contributed to variation in mortality rates has important implications for the second and third waves of the pandemic. A careful case-control analysis of contributors (including environmental, natural, and policy differences) to the severity of disease and mortality in COVID-19 was conducted to shed light on discrepancies in the first wave of the pandemic globally.

## Materials and Methods

Cases and controls were selected by matching countries with the highest COVID-19 mortality rates compared to those with the lowest. Matching was done based on the population size and WHO region (where possible). This study covered the period between March and May 2020 to represent the first wave of the COVID-19 pandemic. We shortlisted all countries that had at least 100 cases by March 31 (106 countries), as identified by Ritchie et al. [2]. We referred to the day when a country reported its 100th case as day X. For each country, we took the time of 1 month (30 days) after the day of the first 100 cases (day X). We referred to this as day Y, whereby Y = X + 30 days. All further data were obtained for day Y of each country, as identified by Ritchie et al. [3]. All countries were arranged in descending order by the total deaths/population in those 30 days, and the top 10 countries were selected. For countries with an area greater than 5,000,000 km$^2$, the unit of comparison was by state/province. These countries included Russia, Canada, the US, Brazil, and Australia. The top 10 countries in terms of COVID-19 deaths per population were selected for analysis as cases. To select controls, countries and US states were listed by decreasing order of population.

For each case, 6 controls were matched according to the closest population size. Controls were selected based on 3 criteria: (1) the control countries were closest in population size to the case country, (2) 2 controls were from the same WHO region and 4 from different WHO regions, and (3) the COVID-19 death rates per population of the controls were at least 3 times lower than that of the case (Table 1, Figure 1).

### Data Sources, Collection Process, and Tool

Based on 3 different contributors (the environment, infrastructure, and host-related factors), potential risk factors that may determine the severity of COVID-19 burden in these 3 categories were shortlisted (Table 2). For each variable, reliable sources of publicly available information for all countries were utilized. Data were manually entered in Microsoft Excel and a tool was utilized to process the data. A complete list of sources is found in Table S1.

After piloting the data, 2 control countries (Taiwan and Hong Kong) had many missing variables due to the non-availability of data, and variables were treated as missing. For 23 out of 30 variables, comparable data were not found for states within the US (Table S1). Therefore, data available for the entire country were used instead. For infrastructure factors, data for public transportation were missing; given high intercity variability, only the top metropolitan city in the country was considered. Population size was

**Table 1.** List of case and control countries/states

| Case country/state | Same WHO region controls | | | Different WHO region controls | | | |
|---|---|---|---|---|---|---|---|
| New Jersey | Tennessee | Indiana | Austria | United Arab Emirates | Hong Kong | Belarus |
| Connecticut | Oklahoma | Utah | Kuwait | Bosnia and Herzegovina | Moldova | Armenia |
| New York | Florida | Texas | Australia | Niger | Sri Lanka | Taiwan |
| Michigan | Washington | Arizona | Jordan | Sweden | Portugal | Kyrgyzstan |
| Louisiana | Alabama | Kentucky | Georgia | Slovakia | New Zealand | Croatia |
| Spain | Uzbekistan | Poland | Argentina | Iraq | Afghanistan | Algeria |
| Belgium | Czech Republic | Greece | Tunisia | Cuba | Dominican Republic | Rwanda |
| Italy | Turkey | Ukraine | South Africa | Colombia | South Korea | Kenya |
| Netherlands | Romania | Kazakhstan | Ecuador | Guatemala | Cambodia | Senegal |
| United Kingdom | Germany | Russia | Thailand | Egypt | Philippines | Vietnam |

WHO, World Health Organization.
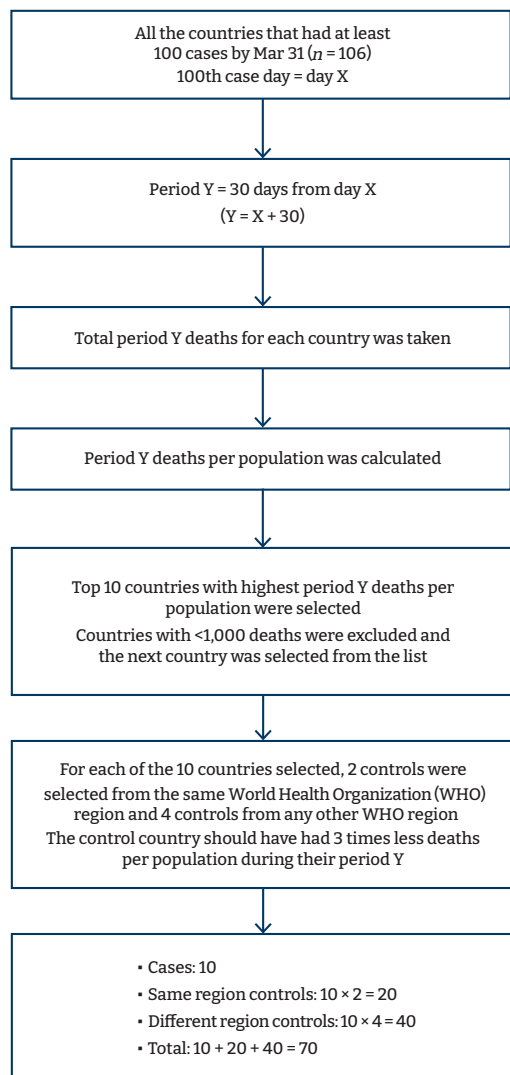


**Figure 1.** Algorithm used to select case and control countries/states.

used to ascertain the top metropolitan city, and the public transportation system of the selected cities was analyzed for further analysis.

### Data Analysis

Temperatures in March and April were found separately, using the arithmetic mean to obtain the average temperature of March and April. Public transportation was divided into 4 categories: (1) none, (2) over-ground transport system only (buses/trams), (3) underground mass public transport only, and (4) over-ground and underground transport systems. Data analysis was carried out in Stata 15.1 (StataCorp., College Station, TX, USA), and Power BI (Microsoft Corp., Redmond, WA, USA) was used for the graphics. Due to the skewed distribution of variables, data were reported as median and interquartile range. The non-parametric Wilcoxon rank-sum test was used to compare factors between the 2 groups. The cluster analysis was performed using a scalable EM clustering algorithm, which iteratively refines an initial cluster model to fit the data and determines the probability that a data point exists in a cluster. The algorithm ends the process when the probabilistic model fits the data. The function used to determine the fit was the log-likelihood of the data given the model.

### Ethical Approval

Ethical approval was not required, as all information was freely available in the public domain.

## Results

Geographically, all cases were in the northern hemisphere; 5 were states within the US, whereas the other 5 were

countries in Europe (Figure 2). Controls were located across 6 continents (Figure 2). The risk factors that were significantly different in cases versus control countries/states are given in Figures 3 and 4; of the 30 risk factors that were investigated, 16 were found to be significant. Host-related factors were also observed to contribute to the difference in mortality rates (Figures 5, 6). A secondary analysis was conducted to compare the number of cases and the number of deaths in period Y, and the results are shown in Figure 7. Italy and Spain had a much higher proportion of deaths per reported cases than the US. There were limited data for the case countries/states on

**Table 2.** Comparison of risk factors in case and control countries and states (significant factors only)

| Risk factor | Control | Case | *p* |
|---|---|---|---|
| Environmental factor | | | |
| March temperature (°C) | 12.3 (4.1–20.8) | 5 (2.5–8.6) | 0.035 |
| April temperature (°C) | 15.3 (9.6–21.6) | 8.1 (7.6–11.7) | 0.013 |
| Average temperature (°C) | 13.5 (6.8–20.6) | 6.6 (5.5–10.2) | 0.023 |
| UV light index (J/m²) | 3,230 (2,138–4,862) | 1,662 (1,645–2,444) | 0.008 |
| Host factor | | | |
| Life expectancy at birth (y) | 75.2 (71.5–77.6) | 81.4 (81.2–82.8) | 0.040 |
| Under-5 mortality rate (/1,000 live births) | 10 (6–22) | 4 (3–4) | 0.003 |
| Neonatal mortality rate (/1,000 live births) | 6 (3–12) | 2 (2–2) | 0.008 |
| Tuberculosis incidence (/100,000 population) | 26.5 (5.5–69.0) | 4.6 (2.3–8.0) | 0.004 |
| % of under-5 deaths due to diarrhea | 1.7 (0.4–5.4) | 0.3 (0.2–0.4) | 0.016 |
| % of under-5 deaths due to acute respiratory infection | 1.0 (5.3–15.1) | 1.8 (1.4–2.4) | 0.002 |
| Presence of national BCG program | 42 (70) | 0 | < 0.001 |
| Infrastructure factor | | | |
| Education index | 0.69 (0.60–0.78) | 0.81 (0.79–0.86) | 0.012 |
| Population density (people/km²) | 90 (57–136) | 275 (205–378) | 0.017 |
| Human Development Index | 0.77 (0.70–0.84) | 0.92 (0.89–0.92) | 0.003 |
| % of GDP on healthcare | 6.8 (5.3–8.2) | 9.6 (8.9–10.1) | 0.014 |
| Population of age ≥ 65 (% of total population) | 12 (5–16) | 18 (16–19) | 0.001 |

Data are presented as medians (interquartile range) or *n* (%).
IQR, interquartile range; UV, ultraviolet; BCG, Bacillus Calmette-Guérin; GDP, gross domestic product.
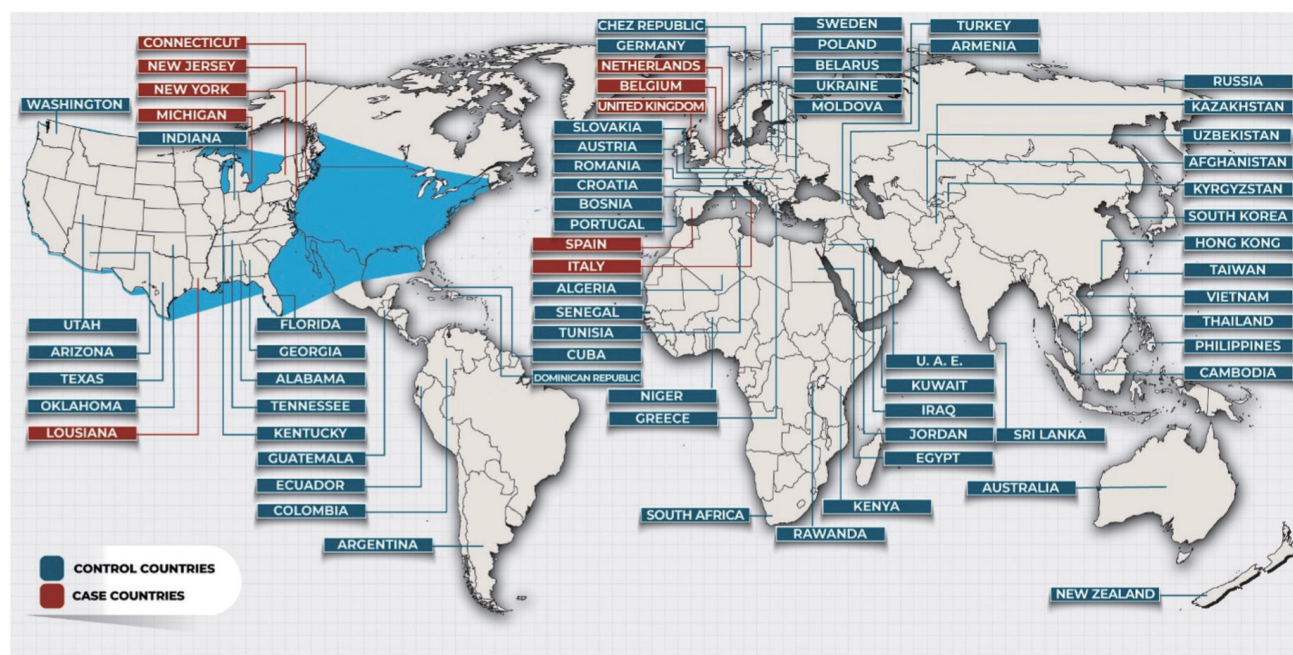


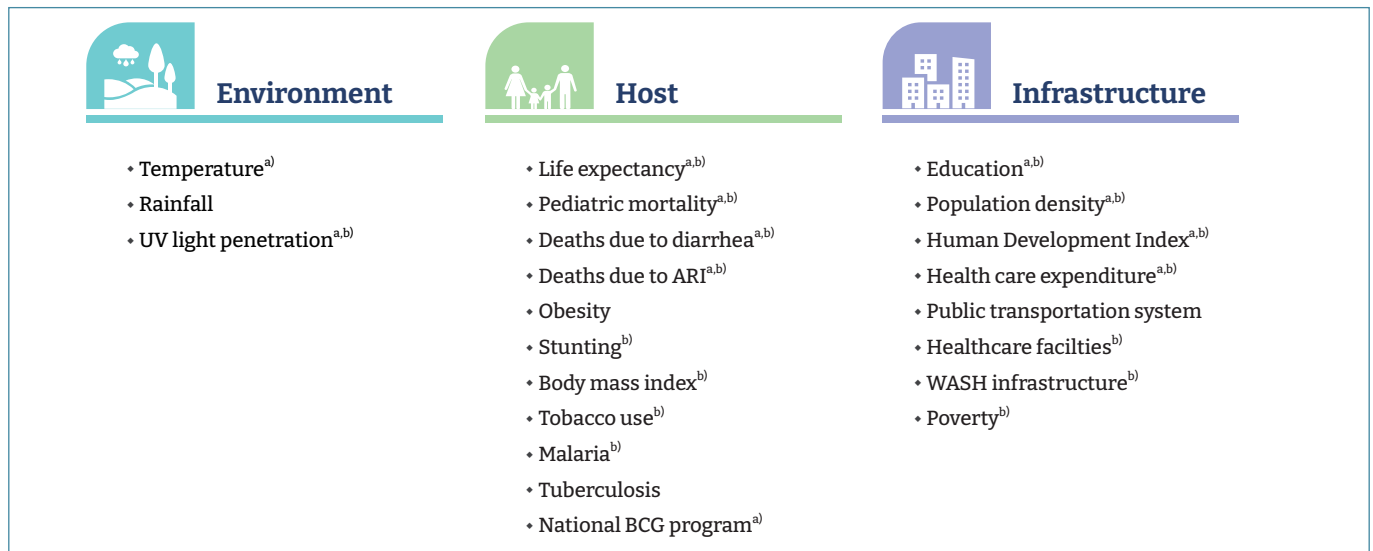**Figure 2.** Map showing case and control countries/states.

**Figure 3.** Three domains of risk factors studied.
[a)]Statistically significant differences noted between cases and controls. [b)]Individual State data was not available; therefore, data was taken for whole of United States. UV, ultraviolet; ARI, acute respiratory infection; BCG, Bacillus Calmette-Guérin; WASH, water, sanitation, and hygiene.
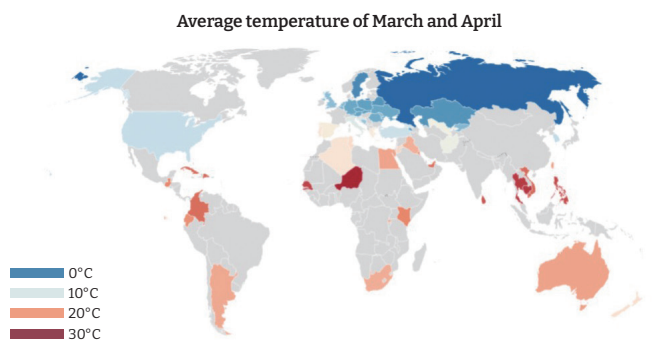


**Figure 4.** Map showing temperatures in included countries/states.



**Figure 5.** Comparison of pediatric mortality rates in case and control countries/states.

stunting, malaria incidence, and poverty.

## Environmental Factors
The case countries/states recorded lower average temperatures in March and April (6.6°C) than the controls within that period (13.5°C, $p = 0.023$) (Table 2). The ultraviolet (UV) light index was lower in the case countries/states (1,662.0 J/m$^2$) than in the controls (3,229.5 J/m$^2$, $p = 0.008$) (Table 2).

## Host-Related Factors
Among the host-related factors, pediatric and neonatal mortality rates, causes of under-5 deaths, adult life expectancy, and others were significantly different among the cases and controls. The neonatal mortality rate was 2 per 1,000 live births in the cases versus 6 per 1,000 live births in the controls ($p = 0.008$)
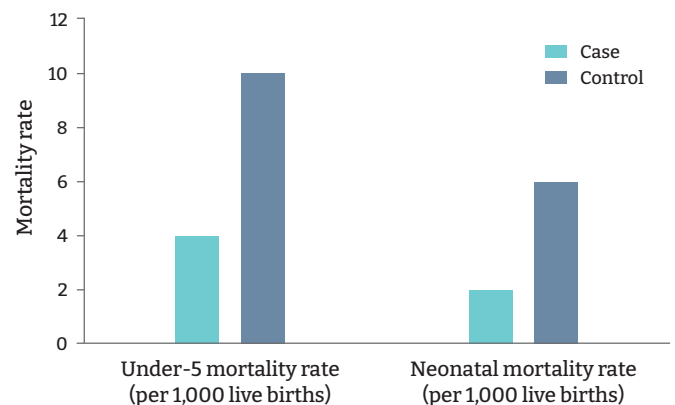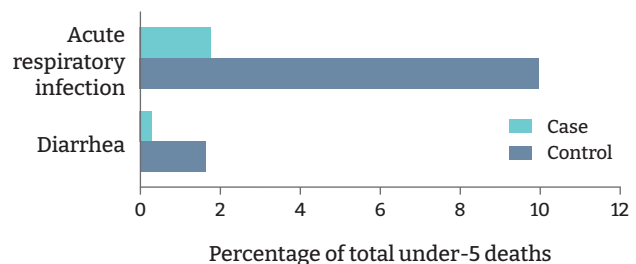


**Figure 6.** Comparison of top causes of under-5 deaths in case and control countries/states.

(Table 2). The under-5 mortality rates were 4 and 10 per 1,000 live births in the cases and controls, respectively ($p = 0.003$) (Table 2, Figure 5). In the case countries/states, 0.3% of under-5
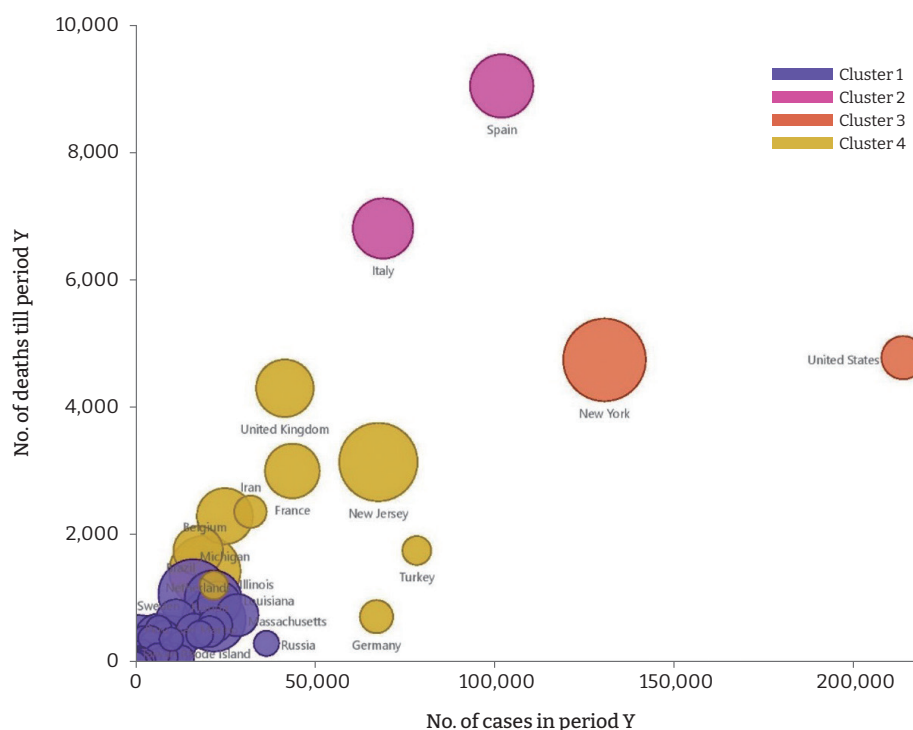
**Figure 7.** Comparison of number of deaths and number of cases in period Y in all included countries/states.

deaths were due to diarrhea, versus 1.7% in the controls ($p = 0.016$) (Table 2). Furthermore, 1.8% of all under-5 deaths in the case countries/states resulted from acute respiratory infections, versus 10.0% in the controls ($p = 0.002$) (Table 2, Figure 6). In the cases, the tuberculosis incidence was found to be 4.6 per 100,000 population, versus 26.5 in the controls ($p = 0.004$) (Table 2). National Bacillus Calmette-Guérin (BCG) vaccination programs were not present in the cases, whereas 42 control countries/states (70%) had these programs ($p < 0.001$) (Table 2).

### Infrastructure

The education index was 0.81 in the case countries/states and 0.69 in the controls ($p = 0.012$), whereas the Human Development Index was 0.92 versus 0.77 in the case and control countries/states, respectively ($p = 0.003$) (Table 2). The case countries/states spent 2.87% of gross domestic product more than the control countries on healthcare, and this was a significant difference ($p = 0.014$). The population density was 3 times higher in the case countries/states than in the controls (275 vs. 90 people/km$^2$ of land area, $p = 0.017$) (Table 2).

## Discussion

The analysis of our data showed many interesting results. The cases were all high-income countries in North America and Europe. There was a significant difference in living standards between the case and control countries, with the controls belonging to the low-income group (overall) with poorer healthcare and education and higher mortality from diarrhea and respiratory infections in childhood. Our findings suggest that countries with poorer economies and health systems did not experience worse outcomes in the early stage of the COVID-19 pandemic.

The baseline mortality due to acute respiratory infections was more than 5 times higher in the controls than in the cases. Our findings suggest that young children, including infants, who have a higher susceptibility and burden of other respiratory infections, may have milder COVID-19 due to the long-term boost of the innate immune response, known as trained immunity, as noted by Netea et al. [4]. The case countries all had lower incidence rates of tuberculosis, as well as no national BCG vaccination programs. An association of anti-tuberculosis antibodies with decreased severity of COVID-19 has been reported by Kovacic [5] and Miller et al. [6]. Countries with a long-standing BCG program had some degree of protection against COVID-19, corroborating the finding of Escobar et al. [7] that every 10% increase in the BCG index was associated with a 10.4% reduction in COVID-19 mortality. Taken together, the higher burden of tuberculosis and pediatric mortality in the control countries can be explained by the "hygiene hypothesis,"

which relates to the decreasing incidence of infections in Western countries, as identified by Stiemsma et al. [8].

Welliver [9] noted that temperature and the UV light index were significant indicators, whereby colder temperatures seemed to favor a high burden of COVID-19, similar to the associations of temperature with influenza and other respiratory viruses. However, further studies are needed to confirm this finding. Population density seems to have been a vital factor that differentiated the cases and controls. The case countries/states had more than 3 times the number of people living per square kilometer of land. This is a logical result, since crowding leads to higher rates of transmission of droplet infections. In countries with lower population densities, Corburn et al. [10] pointed out that most of the population lives in rural settings with automatic distancing, whereas heavily urbanized areas pose difficulties in implementing social distancing due to the living environment.

It is of note that several of our indicators, including obesity, stunting, and tobacco use, proved not to be statistically significant. We suspected that the case countries/states would primarily have cities with underground mass public transit available, but our findings were insignificant. Seventy percent (7 out of 10) of our case countries/states and 51.7% (31 out of 60) of our controls had underground mass public transit available, and this difference was not statistically significant. Tobacco use and obesity, which are risk factors for many diseases, were also not found to be significant between the case and control countries/states. Stunting, malaria incidence, and poverty index data were not available for many developed countries. As malaria and stunting are not major public health concerns in that part of the world, these parameters are not rigorously documented. Moreover, these countries have stronger economies and have not been assigned a poverty index number, so this variable was treated as missing data.

### The Cases of Spain, Italy, and the United States
Our secondary analysis revealed that Italy and Spain had much higher proportions of deaths among COVID-19 cases than that of the US (Figure 7). One explanation is that the testing threshold for Spain and Italy was much higher, increasing the likelihood of high-risk patients testing positive. However, Roser et al. [11] reported that the COVID-19 positivity rates of Spain, Italy, and the USA were 3% to 5%, 5% to 10%, and 10% to 20%, respectively. Another explanation is that in Italy and Spain, a higher percentage of the population is over 65 years of age than in the US. A more detailed look at the individual country numbers indicated that 23% of the population of Italy is over 65

years of age, whereas the corresponding percentages in Spain and the US are 19% and 16%, respectively. This factor may have contributed to the high level of mortality in Italy, but other factors are also likely to have played a role. The contribution of population demographics to increased COVID-19 mortality has been explored by Kontis et al. [12], who showed a correlation with the elderly population across different countries. Another possible explanation posited by Islam et al. [13] is that the COVID-19 virus strains present in Europe were deadlier than the North American strains. A genome-wide association study of severe COVID-19 patients conducted by Severe Covid-19 GWAS Group et al. [14] identified a gene cluster on chromosome 3 as a risk locus for respiratory failure. The COVID-19 Host Genetics Initiative [15] reported that nearly 50% of individuals in South Asia are carriers of this suspicous region, while Zeberg and Paabo [16] identified that 16% of people in Europe are carriers. Although the specific genes contributing to the severity of COVID-19 remain elusive, other contributors, including socioeconomic changes and preexisitng conditions, may have contributed to the high disparisitng in the death toll.

### Strengths
We systematically explored 30 variables related to the life sciences, including climate, population demographics, public health indices, and infrastructure. The study employed a holistic approach aiming to account for many risk factors. By using mortality per population as the principal measure of severity of the disease, 2 goals were achieved: first, the study eliminated patients with mild disease and therefore focused only on those who were a burden on the health system, and second, a fair comparison was made between countries with larger and smaller populations. The control countries were matched based on population size for all dependent variables. Lastly, all data were obtained from publicly available, verifiable sources.

### Limitations
The study aimed to identify the top 10 countries and states that were worst affected by COVID-19 in the early stage of the pandemic. Many countries/states had a limited number of laboratory polymerase chain reaction testing kits and were unable to test a sufficient number of suspected individuals. The case and control countries/states were identified based on the COVID-19 mortality data from only March and April 2020; analyzing the spread of the virus in the following months may yield different results regarding the top 10 countries/states. For 23 of the 30 variables, data were not available for individual states of the USA and combined

data of the USA were used, leaving no differences across the states. Furthermore, population-level data (not individual data) were employed, which may have caused this study to be affected by the ecological fallacy, which is an important bias in studies involving population-level data collection.

## Conclusion

The death toll following the first wave of the COVID-19 pandemic was affected by all 3 types of determinants (environmental, host-related, and infrastucture-related). We investigated various factors that potentially contributed to the global discrepancy in mortality rates. Countries implemented various control measures such as suspending public transport and promoting physical isolation. Even as the national emergency responses varied across countries, significant impacts of key environmental, host-related, and infrastructure-related determinants were identified. Our results shine light on the epidemiological contributors to variation in COVID-19 mortality. Our findings may guide future researchers to investigate these parameters concerning COVID-19 and help deduce factors related to COVID-19 preparedness for the second and third waves of the pandemic. Actions to mitagate future global health threats require identifying deficiencies in the existing health systems, as well as improving coordination and implementation.

## Supplementary Material

Table S1. All indicators except public transportation. Supplementary data is available at https://doi.org/10.24171/j.phrp.2021.12.2.03.

## Notes

## References

1. World Health Organization (WHO). Coronavirus disease (COVID-19) weekly epidemiological update and weekly operational update [Internet]. Geneva: WHO; 2019 [cited 2020 Jun 10]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.

2. Ritchie H, Ortiz-Ospina E, Beltekian D, et al. Coronavirus (COVID-19) cases [Internet]. Our World in Data; 2020 [cited 2020 Jun 15]. Available from: https://ourworldindata.org/covid-cases.

3. Ritchie H, Ortiz-Ospina E, Beltekian D, et al. Coronavirus (COVID-19) deaths [Internet]. Our World in Data; 2020 [cited 2020 Jun 17]. Available from: https://ourworldindata.org/covid-deaths.

4. Netea MG, Giamarellos-Bourboulis EJ, Domínguez-Andrés J, et al. Trained immunity: a tool for reducing susceptibility to and the severity of SARS-CoV-2 infection. Cell 2020;181:969–77.

5. Kovacic D. Immunological principles by which the tuberculosis BCG vaccine may possibly reduce SARS-CoV2-related morbidity and mortality in vaccinated patients: a case for BCG vaccination during the COVID-19 pandemic [Preprint]. Posted 2020 Apr. Research Gate. https://www.researchgate.net/publication/340484800.

6. Miller A, Reandelar MJ, Fasciglione K, et al. Correlation between universal BCG vaccination policy and reduced morbidity and mortality for COVID-19: an epidemiological study [Preprint]. Posted 2020 Mar 28. medRxiv 2020.03.24.20042937v1. https://doi.org/10.1101/2020.03.24.20042937.

7. Escobar LE, Molina-Cruz A, Barillas-Mury C. BCG vaccine protection from severe coronavirus disease 2019 (COVID-19). Proc Natl Acad Sci U S A 2020;117:17720–6.

8. Stiemsma LT, Reynolds LA, Turvey SE, et al. The hygiene hypothesis: current perspectives and future therapies. Immunotargets Ther 2015;4:143–57.

9. Welliver RC Sr. Temperature, humidity, and ultraviolet B radiation predict community respiratory syncytial virus activity. Pediatr Infect Dis J 2007;26(11 Suppl):S29–35.

10. Corburn J, Vlahov D, Mberu B, et al. Slum health: arresting COVID-19 and improving well-being in urban informal settlements. J Urban Health 2020;97:348–57.

11. Roser M, Ritchie H, Ortiz-Ospina E, et al. Coronavirus pandemic (COVID-19) [Internet]. Our World in Data; 2020 [cited 2020 Jun 25]. Available from: https://ourworldindata.org/coronavirus.

12. Kontis V, Bennett JE, Rashid T, et al. Magnitude, demographics and dynamics of the effect of the first wave of the COVID-19 pandemic on all-cause mortality in 21 industrialized countries. Nat Med 2020;26:1919–28.

13. Islam MR, Hoque MN, Rahman MS, et al. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates hetero-

geneity. Sci Rep 2020;10:14004.

14. Severe Covid-19 GWAS Group, Ellinghaus D, Degenhardt F, et al. Genomewide association study of severe Covid-19 with respiratory failure. N Engl J Med 2020;383:1522–34.

15. COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. Eur J Hum Genet 2020;28:715–8.

16. Zeberg H, Paabo S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. Nature 2020;587:610–2.