**Descriptive Analysis of COVID-19 cases in Africa, Asia, and Europe**

**Emma Landry1, Javier Magallanes Mendoza2, Mounica Lakshmi Dugginapeddi3, Seamus Gallivan4**

1PhD Biostatistics, 2MPH Biostatistics, 3MPH Community Health Sciences, 4MS Biostatistics

**I. INTRODUCTION**

On December 8th, 2019, several cases of pneumonia were reported in Wuhan, China, most of which were associated with the local seafood market of wild animals.1 On December 31st, 2019, China informed about these pneumonia cases to the World Health Organization (WHO).1 The WHO identified the cause of those pneumonia cases as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) on 7th January 2020 and the disease was termed as coronavirus disease 2019 (COVID-19).1 Since then, the virus has quickly spread throughout the world with the WHO declaring COVID-19 a pandemic on 11 March 2020.1 As of October 29, 2021, the WHO reported a total of 245,373,039 confirmed cases of COVID-19 and 4,979,421 deaths due to COVID-19 throughout the world.2

Due to the proximity to China, Asia has been badly hit early on with COVID-19 along with several Pacific islands countries.3 However, countries reported extremely different impacts, ranging from a huge number of deaths in India to a very few number of losses in New Zealand.3 Variation in demographic characteristics, population density, rural-urban composition, and degree of international visitors, among others, could help explain these observed differences in death rates.3 While most people who are infected with COVID-19 recover, the death toll in Asia is considerably high overall.3

In Europe, the pandemic resulted in large scale government interventions, often constraining basic social and political rights.4 These developments created an economic shock which is likely to surpass that of the financial crisis of the last decade.4 The health threat of the pandemic combined with the economic shock led many countries to issue far-reaching lockdown measures and border closures.4 Because of these actions, social integration, and cohesion have been threatened both within and across European countries leading to deepening of disparities in the society.4

Africa, on the other hand, reported fewer COVID-19 cases and deaths than other continents which has been attributed to poor access to healthcare services and fragile healthcare systems. Many countries have been facing a deadlier third wave of cases since mid-2021.5 The pandemic could have devastating impacts on the already strained health systems of the continent.6 This could rapidly turn into a social and economic crisis throughout the continent.6 In addition, asymptomatic transmission of the virus, without effective testing and reporting systems, allows for potential virus mutation.6

We have conducted a descriptive study to elaborate and compare the COVID-19 outcome measures like confirmed cases, recovered cases, and mortalities among various countries of Asia, Africa, and Europe. Given the diversity of demographic features and COVID-19 outcomes among these continents,3,4,5,6 we sought to compare the countries of these continents in our study to further understand the differences in COVID-19 outcome measures both within and among continents.

**II. METHODS & MATERIALS**

*A. Data Set*

The source of our datasets is the worldometer website which provides publicly available COVID-19 related datasets of various countries. The data is an amalgamation of various prospectively collected health data from worldwide government agencies and local health ministries.

The datasets for Africa, Asia, and Europe were imported to SAS Studio. A new variable continent was added. Total cases per 1,000,000 and total deaths per 1,000,000 were renamed in the Asia file to match the naming convention of the other datasets. The data was merged vertically to obtain a dataset of 145 observations and 11 variables.

A fourth dataset, found on a github repository collecting COVID-19 data from Our World in Data, was queried to extract GDP and HDI for all countries in the world. Using a PROC SQL statement in SAS, a left join merging by Country was used to populate GDP and HDI for all 145 countries in our dataset. Finally, GDP and HDI were standardized before running a generalized linear model.

*B. Study Population*

For our study, we used three datasets corresponding to Asia, Africa, and Europe respectively. Each data set mentions the total number of confirmed cases, recovered cases, active cases, and deaths of each country in a continent. In addition, total cases per million and total deaths per million population are also mentioned. With regards to testing, the total number of COVID-19 tests done and tests per million population are mentioned.

The data includes 145 observations, representing countries in the 3 continents of study. Total cases per 1,000,000 population has a minimum of 67, maximum of 220,244, and mean of 49,152. Total deaths per 1,000,000 population has a minimum of 3, maximum of 3,414, and mean of 744. GDP has a minimum of 661.2, maximum of 116,935.6, and mean of 19,511.4. HDI is a scale from 0 to 1 and has a minimum of 0.394, maximum of 0.957, and mean of 0.714.

*C. Statistical Methods*

We used SAS statistical software version 9.4 to discover descriptive statistics and conduct regression analysis. We created a histogram to observe the distribution of deaths per million and boxplots to view the correlation between deaths per million and human development index. Two general linear models were performed to analyze the relationship between global economic development indicators and COVID-19 deaths and cases in each continent.

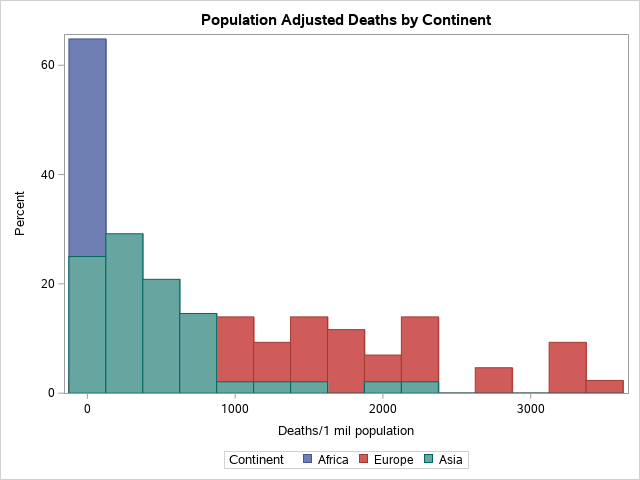


Figure 1: COVID-19 Deaths by Continent

Figure 1 and Table 1 indicate that there are differences in deaths per 1,000,000 population and cases per 100,000. To assess why these differences exist, gross domestic product (GDP) per capita, a measure of the economic production value attributed to each individual citizen, and human development index (HDI), a country's average measure of health, knowledge, and standard of living, were analyzed.

The relationships between COVID-19 deaths 1,000,000 and the continent, HDI, and GDP are analyzed formally using a linear regression model, in order to determine whether these differences are statistically significant. The same is done for COVID-19 cases per 100,000 population. A linear model was selected due to the ease of understanding and interpretation.

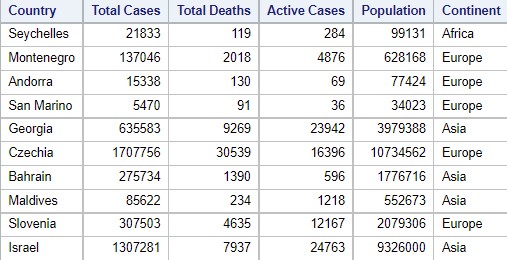
Letting Yi denote the response of our model (deaths in the first, and cases in the second) the linear model can be formulated as :

.

The denote the predictors. These two numerical predictors are the standardized values of HDI and GDP. We also add a categorical predictor for the continent, which is introduced in the model through two dummy binary variables for Africa and Asia. The ’s are the coefficients we wish to estimate, and express the contribution of each predictor to the outcome.

**III. RESULTS**

Table 1: Countries with most Total Cases (per 100k)



Boxplots were used to explore the relationship between deaths per 1,000,000 population against GDP and HDI and determine whether there is a relationship that could be analyzed further using statistical analysis. Figure 2 shows that in general, countries with higher incomes determined by their GDP value have more variability in total deaths reported when stratified by continent. Figure 3 shows a similar trend, but there is a noticeable increase in deaths in higher HDI countries, with the high HDI countries experiencing the most deaths in their respective continents.

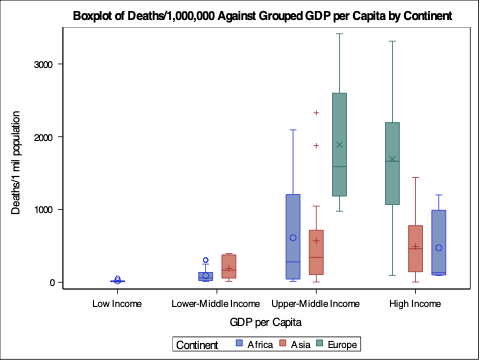


Figure 2: COVID-19 Deaths against GDP

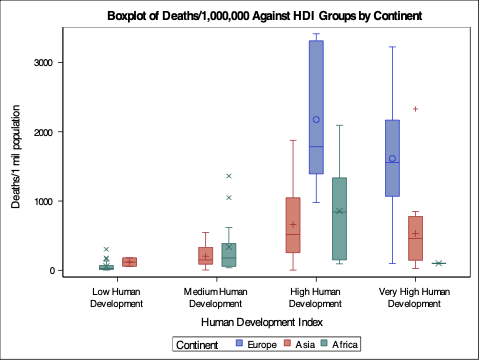


Figure 3: COVID-19 Deaths against HDI

We now proceed to the analysis of our linear regression results, obtained using the GLM procedure in SAS, which is compatible with categorical variables. In tables 3 and 4, cases and deaths per 1,000,000 population are the response. The estimate column gives the estimates of the parameters, . The column of most interest is the last one, Pr > |t|. This the p-value associated with the hypothesis test with null hypothesis H0: the parameter is zero, performed for each parameter. A small p-value thus indicates that we may reject the null hypothesis that the parameter should be zero, and thus that the predictor has a meaningful contribution to the model.

Table 2: Regression Results for COVID-19 Deaths

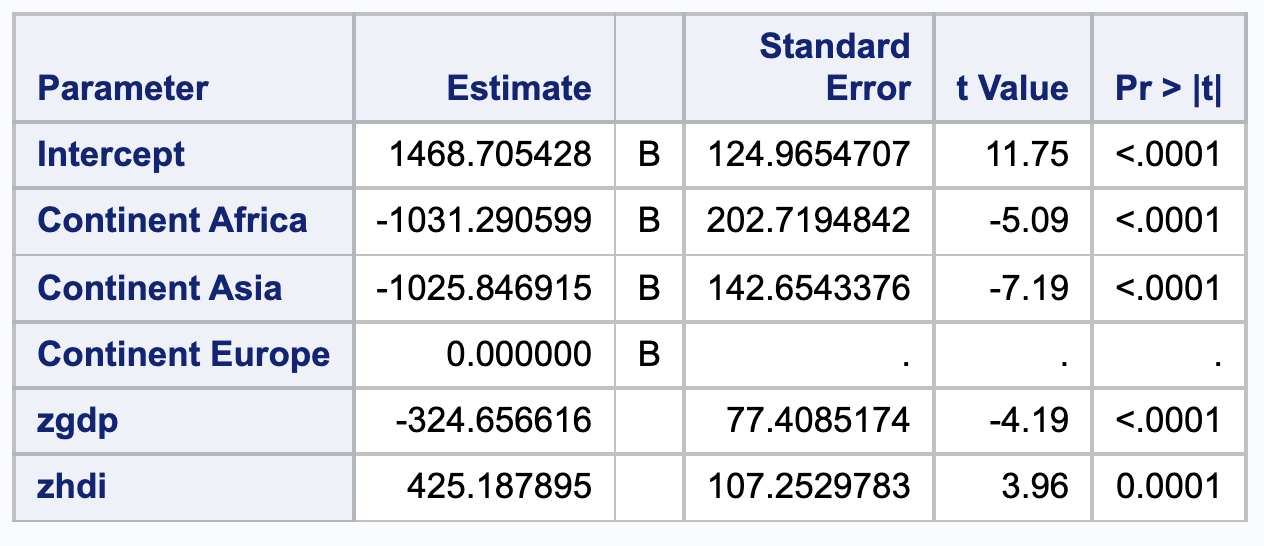


Table 2 provides the regression results for the COVID-19 deaths per 1,000,000 population response variable. We note that all the p-values are 0.0001 or less, hence all the considered predictors appear to have a meaningful relationship with COVID-19 cases. It is also interesting to note that the parameters for Africa, Asia, and GDP have negative values, and hence these predictors are negatively correlated with the number of deaths. These quantitative results are consistent with the qualitative observations made regarding the boxplots.

Table 3: Regression Results for COVID-19 Cases

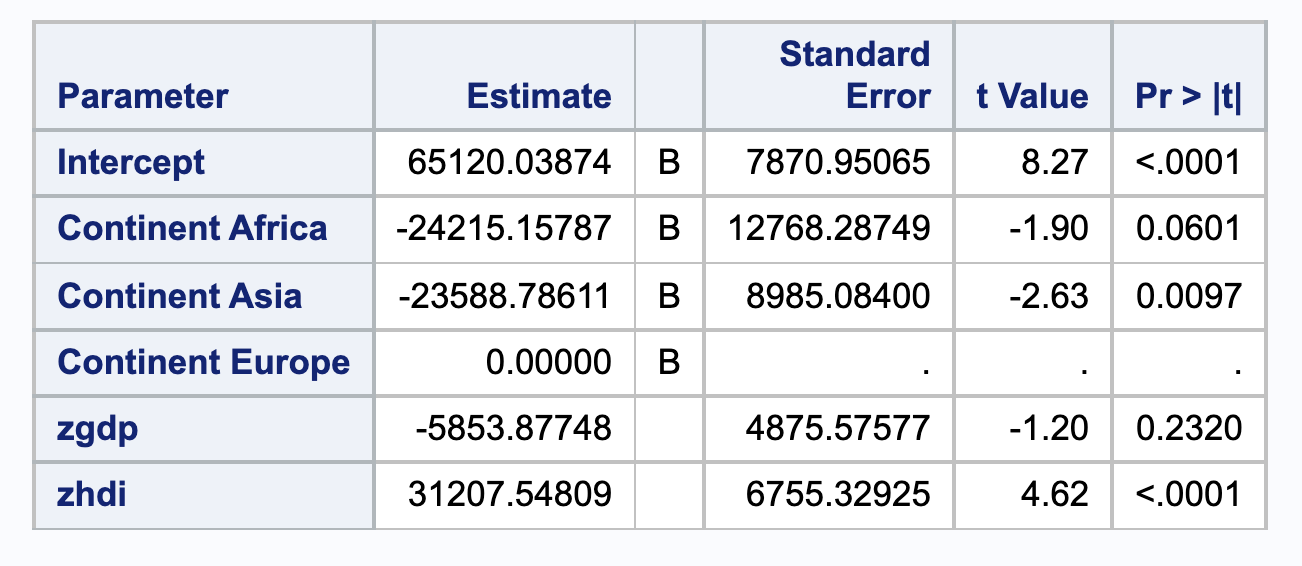


Table 3 shows the results for COVID-19 cases per 100,000 population. We now notice that some of the predictors are less meaningful than they were for predicting deaths. Considering the usual 0.05 level for the test, we see that the p-values for Africa and GDP are greater than that threshold, and are thus not particularly meaningful. Considering the values of the parameter estimates, we make the same remark that Africa, Asia, and GDP are negatively correlated with the number of deaths.

**IV. DISCUSSION**

There are some limitations to our analysis. First, we addressed data from Asia, Africa and Europe exclusively. Due to the tremendous impact of geography on infectious disease transmission, we are unable to estimate any impacts of COVID-19 on the Americas and Australia. Additionally, it is impossible to collect exact numbers of COVID-19 cases and deaths by any health data collection agency or method. Thus our data are imperfect and do not necessarily represent the true toll of this pandemic.

**V. CONCLUSIONS**

COVID-19 had an enormous impact across Asia, Africa, and Europe, challenging the resilience of economies and health systems of the countries.3,4,5,6 A complete assessment of the impact will only be possible after the pandemic ends, but these early findings can help to guide the public health protocols of various regions.

**ACKNOWLEDGEMENTS**

Thank you to our professor, Dr. Kim and teaching assistant, Elvis.

**REFERENCES**

1. Zahid MN, Perna S. Continent-wide analysis of Covid 19: Total cases, deaths, tests, socio-economic, and morbidity factors associated to the mortality rate, and forecasting analysis in 2020–2021. *International Journal of Environmental Research and Public Health.* 2021;18(10):5350. doi:10.3390/ijerph18105350
2. World Health Organization, “WHO Coronavirus (COVID-19) Dashboard”.
3. OECD and World Health Organization, “Health at a Glance: Asia/Pacific 2020: Measuring Progress Towards Universal Health Coverage,” 2020.
4. Maria Grasso, Martina Klicperová-Baker, Sebastian Koos, Yuliya Kosyakova, Antonello Petrillo & Ionela Vlase (2021) The impact of the coronavirus crisis on European societies. What have we learnt and where do we go from here? – Introduction to the COVID volume, European Societies, 23:sup1, S2-S32, DOI: [10.1080/14616696.2020.1869283](https://doi.org/10.1080/14616696.2020.1869283)
5. Okonji, E.F., Okonji, O.C., Mukumbang, F.C. and Van Wyk, B. (2021), Understanding varying COVID-19 mortality rates reported in Africa compared to Europe, Americas and Asia. Trop Med Int Health, 26: 716-719. <https://doi.org/10.1111/tmi.13575>
6. OECD, “COVID-19 and Africa: Socio-economic implications and policy responses,” 2020.

**Appendix**

1. **Source Data File**

The data comes from worldometer open, public data. A Kaggle user, Anandhu H, compiled this data into a user-friendly [CSV format](https://www.kaggle.com/anandhuh/datasets).

GDP and HDI data was found at [Data on COVID-19 by Our World in Data](https://github.com/owid/covid-19-data/tree/master/public/data)*.*