ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 13, Issue, 01, pp. 15758-15761, January, 2021

DOI: https://doi.org/10.24941/ijcr.40630.01.2021

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

VERIFYING FACTORS AFFECTING COVID-19 CASES AND DEATHS

***Juheon Rhee**

International School Manila

ARTICLE INFO

Article History:
Received 29th October, 2020
Received in revised form
18th November, 2020
Accepted 19th December, 2020
Published online 30th January, 2021

Key Words:
COVID-19, Population, GDP per capita, Regression, Correlation.

ABSTRACT

Background: The coronavirus pandemic is arguably the direst issue of 2020, impacting the greater majority of the world's population. One of the most prominent questions surrounding the COVID-19 pandemic is why there is such a wide distribution of total cases and deaths per countries, even in an adjusted population sample. This paper contains both the correlation and regression analysis showing the relationship between the coronavirus cases and deaths and their corresponding affecting factors. **Objective**: The study aims to verify several intuitions regarding the factors that impact population-adjusted total cases and total deaths with cross-sectional data of different countries. **Methods**: This was done by deducing from the correlation regression values as well as other statistical results. All the data utilized in this paper was provided by the Our World in Data database. **Results**: The statistical results showed that population-adjusted total cases are significantly positively affected by GDP per capita and COVID-19 tests per thousand, however, not as significantly affected by the population density and the population rate of elderlies. They also showed that the population-adjusted total deaths are significantly negatively influenced by population-adjusted hospital beds and GDP per capita but positively influenced by the population rate of elderlies. **Conclusion**: Through this study, it was revealed that while some results did not align with initial hypothesis in the correlation analysis, with supposedly deterring factors showing positive correlation with either population-adjusted total cases or deaths, they had negative relationships in the regression analysis when other independent factors were contained.

Citation: ***Juheon Rhee.*** "Verifying factors affecting covid-19 cases and deaths", *International Journal of Current Research*, 13, (01), 15758-15761.

INTRODUCTION

The coronavirus disease (COVID-19) was first identified in December 2019 in Wuhan, China and has been declared a global pandemic by the World Health Organization (WHO) on March 11, 2020. Cumulative worldwide cases have surged past 63.3 million and cumulative death cases past 1.5 million, as of November 30 2020. While COVID-19 has been spread out to almost all countries in the world, the degree of severity in terms of number of cases and deaths varies vastly by countries. Some researchers have already conducted to find out the factors affecting the COVID-19 cases and deaths with the data from various countries. Reference [1] highlighted the importance of wearing face masks to prevent transmission of respiratory viruses. Reference [2] compared the characteristics of the patients who have died with those of the patients who recovered, finding that those who died were on average 17 years older. Reference [3] observed how gender, pre-existing medical conditions, and different races are associated with COVID-19-related deaths. Reference [4] established that low-income countries with lower government effectiveness scores, fewer hospital beds, and a younger population demographic have a negative relationship with COVID-19 mortality cases in a cross-sectional study. Reference [5], a study based in Italy, compared the course of COVID-19-related deaths in different age groups, primarily adults younger than 65 years old and elderlies older than or are 65 years old, which revealed that individuals dying from COVID-19 presented high levels of comorbidities and that the majority whom were elderlies. This study is focused on verifying the factors affecting COVID-19 cases and deaths with sufficient data that have already been accumulated in as many countries as possible as of 30 November 2020. Following the introduction, this paper is organized as Statistical Methods in chapter II, Empirical results in chapter III, and the Conclusion in the last chapter.

***Corresponding author:* Juheon Rhee**
International School Manila

STATISTICAL METHODS

To verify whether the different factors which are expected to affect the COVID-19 cases and deaths really do affect them in the same directions as intuitively expected, correlation analysis between COVID-19 cases or deaths and the factors is conducted. Correlation is calculated in the following formula:

$$\rho_{X,Y} = c\quad(X,Y) = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y}$$

Where, σ_X and μ_X are the standard deviation and mean of variable X, respectively.

To supplement correlation analysis, a regression analysis is also conducted. Regression can show how a factor affects an explanatory variable when other influencing factors are controlled. Regression has the following formula:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_K + \varepsilon$$

Where ε is the residual value.

In the regression analysis, this study focuses on the direction of the coefficient, β, on each independent

variable, X, and the significance of each independent variable on the explanatory variable, Y. Significance of each independent variable is confirmed by the t-statistics for β, which is calculated as follows:

$$t_\beta = \frac{\beta}{s.e.(\beta)}$$

Where s.e.(β) means the standard error of the estimator of β.

The bigger the absolute value of t-statistics on β is, the smaller the p-value is, and thus, the more significant the independent variable is as a regressor. P-value is the probability of the null hypothesis, β = 0, being correct. The smaller the p-value, the more likely the null hypothesis is rejected.

EMPIRICAL RESULTS

Table 1 shows the descriptive statistics of COVID-19 for total cases and total cases per million as of 30 November 2020 of 189 different locations (locations mostly correspond to countries) and total deaths and total deaths per million of 173 different locations, all which are data from Our World in Data. Due to differences in population, with some countries being much more populated than others, a right skewed distribution is expected, as total cases and total deaths are heavily dependent and are complementary with population size. This right skewed distribution is detected from both the positive value of skewness and a mean that is greater than the median, as shown in Table 1.

Table 1. Descriptive Statistics

	Total Cases	Total Deaths	Total Cases Per Million	Total Deaths Per Million
Samples	189	173	189	173
Mean	335,088	8,489	11,995	235
Median	34,652	610	5,659	85
Max	13,592,553	268,177	87,297	1,436
Min	1	1	3	0
Standard Deviation	1,317,550	28,794	14,900	303
Skewness	7.82	6.29	1.65	1.57

Data source: Our World in Data (ourworldindata.org) with observation date of 30 November 2020

It is also observed that skewness value still remains positive even after eliminating the population effect by observing total cases and total deaths per million population. The right skewed distributions of total COVID-19 cases per million and total deaths per million are clearly shown in both Figure 1 and 2. They imply that the severity of COVID-19 cases and deaths depending on countries is neither similar nor normally distributed, but heavily skewed, indicating that some countries are much more damaged than others. Table 2 shows the factors correlated with the total COVID-19 cases per million. A positive correlation is observed between the total cases per million and total number of tests per thousand, which is in line with initial intuition; the more people in a country have taken the COVID-19 tests, the more positive cases are likely to be found.

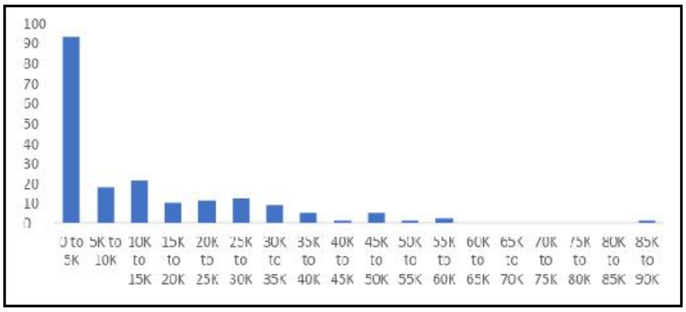

Figure 1. Histogram of Total Cases per Million of Samples

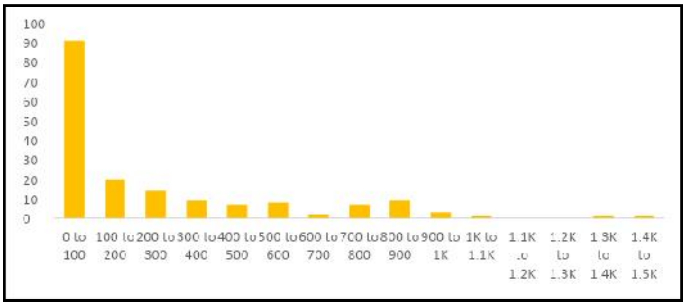

Figure 2. Histogram of Total Deaths per Million of Samples

Stringency index, based on the definition provided by Our World in Data, is the composite measure based on nine response indicators including school closures, workplace closures, and travel bans, regarding governmental response. It is rescaled into numerical values from 0 to 100, the higher the value, the more stringent. Despite the intuition that a high stringency index would deter COVID-19 spread, there is a positive correlation between total cases per million and the stringency index, as can be seen in Table 2. This result implies that greater means of stringency is taken when the Coronavirus outbreak is more serious in a country. This implication takes no regard to the effectiveness or ineffectiveness of the stringency, as the index of stringency is a resulting factor rather than an inducing factor of the COVID-19 spread. While it was originally inferred that a higher population density and total cases per million would have a positive correlation considering the manner in which COVID-19 spreads, in Table 2, it is observed that the correlation is very low. GDP per capita, which means the wealth of a country, shows a positive correlation with total cases per million in Table 2, implying that wealthier countries are more prone to the spread of COVID-19. This correlation result aligns with the current situation where COVID-19 cases have been prominently observed in wealthy countries like the United States and many European countries. Positive correlation is observed between age over 70 and total cases per million. Old people tend to be more cautious than young people, and thus, the initial expectation was for populations with a greater elderly distribution demography to have fewer total cases per million. However, the correlation result rejects this hypothesis. A possible reason for this positive correlation could be the fact that wealthier countries have higher population rates of elderlies, which seems to outweigh the possible effect from the carefulness of elderlies. As a result, a positive correlation between age over 70 and total cases per million is observed. This relationship is differently concluded in the regression result later which controls other independent variables including GDP per capita, the indicator of the wealth of a country.

B_2 × *Population density (i)* + β_3 × *Age over 70 (i)* + β_4 × *GDP per capita (i)* + ε

Where α is the intercept, β is the coefficient of each independent variable, (i) is the ith country and ε is residual value.

Table 2. Correlation with total cases per million

	Total Tests per Thousand	Stringency Index	Population Density	Age over 70	GDP per capita
Samples	88	168	186	180	182
Correlation	0.52	0.60	0.04	0.49	0.54

Data source: Our World in Data (ourworldindata.org) with observation date of 30 November 2020

Table 3. Regression results on total cases per million

Independent Variable	Coefficient	T statistics	P-value
Intercept	3,414.31	1.33	0.1872
Total tests per thousand	11.09	2.26	0.0270
Population density	-2.17	-1.38	0.1707
Age over 70	413.66	1.34	0.1842
GDP per capita	0.23	2.69	0.0088

Table 4. Correlation with total deaths per million

	Total Cases per Million	Age over 70	Hospital beds per thousand	GDP per capita
Samples	189	180	166	182
Correlation	0.78	0.49	0.18	0.29

Data source: Our World in Data (ourworldindata.org) with observation date of 30 November 2020

Table 3 shows the regression results on total cases per million. This regression analysis is conducted to show how each independent variable affects total cases per million when the other independent variables are controlled. Stringency index is excluded as an independent variable as it is considered as a resulting factor rather than an inducing factor on the number of COVID-19 cases. Below are some implications made from the regression results. First, both total tests per thousand and GDP per capita have positive coefficients and p-values that are less than 5 percent. This indicates that these two independent variables are considered significant in terms of their influence over total cases per million and have positive relationship with the aforementioned dependent variable, total cases per million. Second, the p-values from Table 3, 0.1707 and 0.1842, suggest that both the population density and age over 70 respectively, are independent variables which do not affect the total cases per million significantly. The result of population density in regression analysis is in line with the correlation analysis, implying that the population density has little influence on total cases per million. The results of age over 70 in the regression analysis differs from the correlation analysis in that the age over 70 variable becomes rather insignificant when other independent variables are controlled. Lastly, other important influencing factors may not be taken into account, as inferred from the relatively low adjusted R square, 35 percent, of the regression result. Strictness on the usage of face masks is, among other factors, considered a significant factor for total cases per million but is not included in Table 3 due to unavailability of data. According to [6], face masks have been proved effective in reducing transmission and spread of viruses. 80 countries' data from Our World in Data as of 30 November 2020 were used to reach the statistical results of the regression formula as follows. This regression shows the adjusted R square result of 0.35.

Total cases per million (i) = α + β_1 × *Total tests per thousand (i)* +

Table 4 shows the factors correlated with the total COVID-19 deaths per million. A positive correlation is observed between the total deaths per million and total cases per million, which is expected, as deaths are a form of an outcome to total cases. The more people infected with COVID-19 in a country per million, the greater number of deaths per million. Positive correlation is also observed between age over 70 and total deaths per million. This result is in line with the intuition as the older generations are more vulnerable to COVID-19. GDP per capita has a positive correlation with total deaths per million, which is due to wealthy countries having more Coronavirus cases in general. These countries have more access to hospital beds than the relatively poorer countries. This supports the positive correlation with the number of hospital beds per thousand and total of deaths per million despite the intuition that the number of hospital beds and greater wealth would contribute in deterring total deaths per million. However, these two relationships are differently concluded in the regression result which controls other independent variables.

Table 5 shows the regression results on total deaths per million. This regression analysis is conducted to show how each independent variable affects total deaths per million when the other independent variables are controlled. Below are some implications made from the regression results. First, both GDP per capita and hospital beds per thousand have negative coefficients, implying that they are negatively correlated with total deaths per million. This result is opposite to that of correlation analysis, however, is more in line with the intuition that a greater number of hospital beds per thousand and a higher GDP per capita do deter total death per million. Based on the fact that their p-values are less than 5 percent, these two independent variables can be considered significant in terms of their influence over total deaths per million. Next, age over 70 variable has a positive coefficient and p-values that are less than 5 percent, indicating that it is a significantly positively affecting factor on total deaths per million. This result is in line with the correlation result and also in line with the intuition that elderly people are more likely to die from the virus.

146 countries' data from Our World in Data as of 30 November 2020 were used to reach the statistical results of the regression formula as follows. This regression shows the adjusted R square result of 0.65.

$$\text{Total deaths per million } (i) = \alpha + \beta_1 \times \text{Total cases per million } (i) + \beta_2 \times \text{Age over 70 } (i)$$

$$+ \beta_3 \times \text{GDP per capita } (i) + \beta_4 \times \text{Hospital beds for thousand } (i) + \varepsilon$$

Where α is the intercept, β is the coefficient of each independent variable, (i) is the i^{th} country and ε is residual value.

Table 5. Regression results on total deaths per million

Independent Variable	Coefficient	Standard error	T statistics	P-value
Intercept	33.12	26.42	1.25	0.2120
Total cases per million	0.017	0.0012	13.54	0.0000
Age over 70	21.92	4.92	4.46	0.0000
GDP per capita	-0.004	0.00085	-4.66	0.0000
Hospital beds for thousand	-19.12	8.292	-2.31	0.0226

DISCUSSION AND CONCLUSION

This paper was aimed to confirm the factors affecting the population-adjusted COVID-19 cases and deaths with the cross-sectional data of as many countries as possible. Factors were selected based on initial intuition of which would most likely affect the total cases and deaths, but also considering the accessibility of data. The main findings from this paper can be summarized as follows. First, it was confirmed that GDP per capita is a significant factor affecting the total cases per million whereas population density and age over 70 are both insignificant factors. GDP per capita was positively correlated with total cases per million, which is in line with the observations that more cases have been reported in wealthy countries. Second, age over 70 positively affects total deaths per million whereas GDP per capita and hospital beds for thousand negatively affect them under the regression which is able to show the effect of an independent variable while the other variables are controlled. This result aligns with the intuition that the elderly is more likely to die from the virus and people residing in wealthy countries are more likely to be cured and provided with more hospital beds and other means for deterring and treating the virus. This paper can be improved in the future with the following factors taken into account. First is to incorporate the data on the usage of face masks for total cases per million. It is expected to achieve higher R squared value from the regression on total cases per million if a good proxy for usage of face mask is found and incorporated. Second is the time-series data analysis. Stringency may be better observed for its effect on reducing the total cases under a time-series analysis. Furthermore, it can be used as an independent variable in the regression analysis on total cases per million with the time-series results and data.

Declaration of Competing Interest: The author of this study declares that she has no conflict of competing interest.

Funding Statement: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Abboah-Offei, M., Salifu, Y., Adewale, B., Bayuo, J., Ofosu-Poku, R., & Opare-Lokko, E. (2021). A rapid review of the use of face mask in preventing the spread of COVID-19. International journal of nursing studies advances, 3, 100013. https://doi.org/10.1016/j.ijnsa.2020.100013

Cheng, V. C., Wong, S. C., Chuang, V. W., So, S. Y., Chen, J. H., Sridhar, S., To, K. K., Chan, J. F., Hung, I. F., Ho, P. L., & Yuen, K. Y. (2020). The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. The Journal of infection, 81(1), 107–114. https://doi.org/10.1016/j.jinf.2020.04.024

Jordan, R. E., Adab, P., & Cheng, K. K. (2020). Covid-19: risk factors for severe disease and death. BMJ (Clinical research ed.), 368, m1198. https://doi.org/10.1136/bmj.m1198

Liang, L. L., Tseng, C. H., Ho, H. J., & Wu, C. Y. (2020). Covid-19mortality is negatively associated with test number and government effectiveness. Scientific reports, 10(1), 12567. https://doi.org/10.1038/s41598-020-68862-x

Palmieri, L., Vanacore, N., Donfrancesco, C., Lo Noce, C., Canevelli, M., Punzo, O., Raparelli, V., Pezzotti, P., Riccardo, F., Bella, A., Fabiani, M., D'Ancona, F. P., Vaianella, L., Tiple, D., Colaizzo, E., Palmer, K., Rezza, G., Piccioli, A., Brusaferro, S., Onder, G., … Italian National Institute of Health COVID-19 Mortality Group (2020). Clinical Characteristics of Hospitalized Individuals Dying With COVID-19 by Age Group in Italy. The journals of gerontology. Series A, Biological sciences and medical sciences, 75(9), 1796–1800. https://doi.org/10.1093/gerona/glaa146

Williamson, E. J., Walker, A. J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C. E., Curtis, H. J., Mehrkar, A., Evans, D., Inglesby, P., Cockburn, J., McDonald, H. I., MacKenna, B., Tomlinson, L., Douglas, I. J., Rentsch, C. T., Mathur, R., Wong, A., Grieve, R., Harrison, D., … Goldacre, B. (2020). Factors associated with COVID-19-related death using OpenSAFELY. Nature, 584(7821), 430–436. https://doi.org/10.1038/s41586-020-2521-4

****