Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Original article

The impact of temperature, population size and median age on COVID-19 (SARS-CoV-2) outbreak

Kushan Tharuka Lulbadda a,⁎, Dhanushka Kobbekaduwa b, Malika Lakmali Guruge c

a Department of Electrical Engineering, University of Moratuwa, Bandaranayake Mawatha, Moratuwa, Sri Lanka
b Department of Civil Engineering, Xi’an Jiaotong-Liverpool University, 111 Ren Ai Road, Dushu Lake Higher Education Town, Suzhou Industrial Park, Suzhou, Jiangsu, 215123, China
c Mathematics Unit, Sri Lanka Institute of Information Technology, New Kandy Road, Malabe, Sri Lanka

A R T I C L E   I N F O

Keywords:
Coronavirus
COVID-19 outbreak
Temperature
Population size
Median age

A B S T R A C T

Objectives: This study aims to explore the association between the spread of COVID-19 and external parameters. In this regard, temperature, population size, median age, and health care facilities of 58 different countries are considered as external factors.

Methods: A negative binomial regression model was fitted to identify the associations between the factors and cases of COVID-19 during the study periods.

Results: The temperature, population size, and median age are positively associated with the spreading rate of COVID-19. There is no evidence supporting that case counts of COVID-19 could decline in countries with better health care facilities. Also, an empirical model was presented to estimate the number of cases within a country using the external parameters.

Conclusions: It is impossible to express the change in the number of cases for a unit increase in each of the variables because a change in a single variable depends on different values of other variables. However, the findings of this study provide useful implications for the authorities and decision-makers to take specific precautionary measures.

1. Introduction

The recent coronavirus (COVID-19) epidemic has had a significant impact in many countries, and it must gain the maximum degree of inspection. Therefore, it is essential to discuss the history of this virus family before considering its behavior.

Coronavirus is a large family of viruses that cause infections in humans and animals. For humans, the virus can cause illnesses extending in severity from the common cold to Severe Acute Respiratory Syndrome (SARS). Currently focusing on the development of novel therapeutic options such as antivirals and vaccines.

There are several examples as follows; one was in 2003, the severe acute respiratory syndrome coronavirus (SARS-CoV) triggered an outbreak of the disease SARS in China and Hong Kong. Middle East respiratory syndrome coronavirus (MERS-CoV) started an epidemic in Saudi Arabia and among other countries in 2012. Both outbreaks were controlled by a combination of human involvement and still unknown natural conditions. In 2020, coronaviruses became a familiar name among people.

All viruses from this family are zoonotic. They originate in animals and can then be spread to humans after mutation. The naming of more than 40 coronavirus has done by the International Committee for the Taxonomy of Viruses. Four kinds of these are community developed and have been circulating unceasingly through the human population for a very long time. Fig. 1 gives a clear idea of these special categories.

Coronaviruses are a family of enveloped viruses. The meaning is that they are coated in an oily coat called a lipid bilayer, studded with proteins that stick out like spikes of a crown, which caused them to give their name as corona which is the Latin term for the crown.

There are few external factors of the countries which have direct as well as indirect effects on the eruption of COVID-19. Among them, the average temperature, the current population of the country, the overall health system performance by taking the health care rank, as well as the median age of people are considered for this research study. Gross

⁎ Corresponding author.
E-mail addresses: kushantharuka@gmail.com, 188110B@uom.lk (K.T. Lulbadda), D.Kobbekaduwa19@student.xjtlu.edu.cn (D. Kobbekaduwa), malika.l@sliit.lk (M.L. Guruge).

https://doi.org/10.1016/j.cegh.2020.09.004
Received 26 June 2020; Received in revised form 25 August 2020; Accepted 11 September 2020
Available online 28 September 2020
2213-3984/© 2020 The Authors. Published by Elsevier, a division of RELX India, Pvt. Ltd on behalf of INDIACLEN. This is an open access article under the CC
domestic product (GDP) in Purchasing Power Parity (PPP) is added to the study to get the highest practical aspect in terms of economy. For the analysis, the top 58 countries that recorded more than 2000 confirmed cases in the initial 60 days, starting from the first case are used.

The paper is organized as follows; Section 2 is about the worldwide distribution of the virus. Section 3 discusses the impacts of external factors on the virus. Then, Section 4 describes the method of analysis. Section 5 provides the results of the investigation, and the last section presents the discussion with conclusions.

2. Worldwide distribution of the virus

The Covid-19 has spread around the world and been declared a pandemic by the World Health Organization (WHO). According to the data from WHO, 3,096,626 confirmed cases were reported worldwide at the end of April 20, 10.1 WHO has categorized the world into six regions, and the distribution of cases by the end of April 2020 is shown in Fig. 2.

The virus is spread by small droplets formed by sneezing, coughing, and talking. Rather than traveling through air for long distances, the droplets typically fall to the ground or onto surfaces. It is most transmissible during the first three days after the beginning of symptoms. Also, outspread is possible before the appearance of symptoms, and from people who do not show signs.

Suggested preventive actions include washing hands with soap for at least 20 s, maintaining a considerable distance of 1 m from each other, wearing a face mask in public places, and quarantine for suspicious or infected people. Moreover, the worldwide authorities have decided to apply travel restrictions, hazard controls for workstations, and facility closures with lockdowns to mitigate the risk.

3. Impacts of the external factors

There are many remaining unknown facts for the spreading of the virus through the environment. The high fluctuation of the behaviors and qualities of the virus can be considered as the major reason for this unpredictable outbreak. This situation is urgently required to be collaborated with environmental researchers and scientists to identify the environmental features as well as other external facts that affect the transmission of the deadly virus. There are many ongoing types of research and laboratory experiments to determine the behavior of factors such as temperature, wind, and humidity, which can cause the spreading rate. The following sections describe the external parameters that are under consideration.

3.1. Impact of temperature on COVID-19

Temperature is one of the most important fact to consider. The active period of the virus has been identified as 8–10 days on dry surfaces and 37 °C is recorded as the human body temperature where the virus survives. However, the threshold value of temperature to destroy COVID-19 is still unknown. Several kinds of research conducted by analyzing a laboratory-grown copy of the coronavirus (SARS-CoV-2) have indicated that heat has a direct impact on the behavior of the virus. A temperature of 60 °C for an hour-long period has successfully suppressed numerous deadly viruses such as Ebola. But, this temperature can be insufficient to destroy SARS-CoV-2 coronavirus.

Researchers have discovered that heating the virus to a temperature of 92 °C for 15 min can ultimately make the virus inactive. Even though numerous scholars have made valuable discoveries within laboratory conditions, the real-life situation is complicated and ungovernable.

A group of scientists at University College London has suggested that a higher number of coronavirus cases were reported in February, while the rates decreased in summer. This is due to some improvements of the immune in human bodies with increasing temperature.

3.2. Impact of population size on COVID-19

The spread of a deadly virus can be enhanced with the increasing number of human beings. Therefore, it is vital to consider the transmission of COVID-19 concerning to the population size of a geographical area. In a research study, the analysis of Receiver Operating Characteristics was used to obtain the sensitivity of population size. The results suggested that the areas with a population size greater than 1.7 million can enhance the transmission rate.

The primary concern on the spread of COVID-19 is the human to human transmission. The history of pandemics indicates that the disease in a particular population can cross international boundaries. The first outbreak of cholera in 1816, the Spanish flu pandemic in 1918, and the Asian flu pandemic in 1957 are good examples from the past which started in a single area and then transmitted to numerous parts of the world.

3.3. Impact of the health care performance on COVID-19

During a pandemic, the availability of health care facilities is a critical aspect irrespective of the disease that is being spread. In the case of COVID-19, the concern on health care services can be crucial due to the requirements of testing services, personal protective equipment for the health officials and more space to isolate the confirmed patients. A diverse kind of study has been carried out by a group of scholars to analyze the Mexican indigenous population and their capacity to

![Corona Virus](image-url)
respond to the COVID-19. The overall conclusion suggested that the considered population is more vulnerable to COVID-19 due to the identified social and health factors along with the shortage of proper health care facilities. Hand washing stations, hygienic equipment, and proper waste disposal management systems are considered as primary health care facilities. However, these necessary facilities were reported insufficient in most of the hospitals in the aforementioned countries. Therefore, in a situation of a pandemic, the available health care facilities can contribute in many ways to mitigate or control the spread of viruses.

3.4. Impact of the median age on COVID-19

A virus can infect a person regardless of his/her age. But the response from the human body to the virus can vary depending on the age of the person. The lack of immunity in most of older people is one of the identified reasons. Considering the pandemic of COVID-19, several published reports have stated that most of the non-survivors are relatively old. The presence of more than one disease in older patients such as diabetes, kidney diseases or hypertension has increased the risk due to the weakening of immunity.

A study conducted in the Philippines concluded that age has a positive influence on COVID-19 mortality. The results showed that COVID-19 patients older than 60 years of age have a chance of 8.15 times more likely to die compared to the patients below 60 years. Therefore, an in-depth analysis of age as a risk factor can be significant to safeguard senior citizens worldwide by minimizing the spread of COVID-19.

3.5. Impact of the gross domestic product (GDP) per capita on COVID-19

GDP within a country is defined as the total market or monetary value of all the finished goods and services produced within a period. This value exhibits the standard of living between countries. Also, GDP per capita based on PPP is observed to international currency using purchasing power parity rates. Moreover, GDP is a tool used to measure the economic progress of a country. Therefore, concerning the spreading rate of COVID-19, the influence of the economy can be related to the GDP. Hence, to understand the connection between the economic status of a country and the transmission rate of COVID-19, GDP per capita based on PPP values can be considered as a useful tool.

4. Method of analysis

Atmospheric factors such as rainfall, humidity, air quality index, and temperature are some dynamic factors that can modify the spread of deadly viruses. Aside from these environmental conditions, several more factors like demography of the considered area, foreign affairs of the country, technological expansion with medical development, lock down time, national net wealth, etc. could affect the outbreak situation.

Table 1 shows the list of top 58 countries that recorded more than 2000 confirmed cases for the initial 60 days from the day of the first case reported, obtained from the WHO COVID-19 dashboard which is used for the analysis. The secondary data of average temperature as given in weather to visit website and accuweather website for the considered period, population size in the country for 2020 as the worldometer website, the current health care ranking by WHO considering the overall health system performance, the median age for 2020 of the country as United Nations world population prospects and GDP of the country as PPP value from the International Monetary Fund World Economic Outlook are considered for the statistical analysis.

Various media outlets have explained the extraordinary increase in the number of incidents reported in the country. However, the delay of establishing safety rules and regulations, economy prioritized authorities, and running out of health resources and facilities are considered to be the major reasons. As the behavior of this virus outbreak situation is unpredictable, it is difficult to describe the behavior by statistical analysis.

The analysis is conducted using the negative binomial regression model. In this analysis, the number of cases recorded during the first 60 days from the first case is used, and all the independent variables are covariates. Since the dependent variable (number of cases) consists of count data, the Poisson regression model which is the basic statistical model for count data has fitted. Thus, it has tried to fit the Poisson

| Table 1 | List of the countries. |
|---------|-----------------------|
| **Turkey** | **Israel** | **Morocco** |
| **China** | **Pakistan** | **Argentina** |
| **Iran** | **Qatar** | **Republic of Moldova** |
| **Spain** | **Bangladesh** | **Luxembourg** |
| **Brazil** | **Ukraine** | **Sweden** |
| **Peru** | **Belarus** | **Algeria** |
| **Germany** | **United States of America** | **Hungary** |
| **Netherlands** | **Romania** | **Tajikistan** |
| **Switzerland** | **Indonesia** | **Uzbekistan** |
| **Portugal** | **Serbia** | **Egypt** |
| **Ecuador** | **Denmark** | **Cameroun** |
| **Saudi Arabia** | **Republic of Korea** | **Bolivia** |
| **Ireland** | **Panama** | **Bahrain** |
| **The United Kingdom** | **Czech Republic** | **Australia** |
| **Chile** | **Norway** | **Puerto Rico** |
| **France** | **Colombia** | **Kuwait** |
| **Belgium** | **Dominican Republic** | **Greece** |
| **Austria** | **South Africa** | |
| **Mexico** | **Kazakhstan** | |
regression model for all possible models of the variables. But since Pearson Chi square values were highly greater than one for all those models, it implied that there is highly overdispersion in the models, so a negative binomial regression model for data has been chosen. It is an extension for the Poisson regression model which can be used when there is an overdispersion in Poisson regression model.

Assumptions under Negative Binomial Regression Model.

- Linearity in model parameters
- Independence of individual observations
- Multiplicative effects of independent variables

4.1. Data collection

This analysis is based on secondary data, which was collected from several resources, as mentioned above. The series consists of 290 secondary data of the five factors discussed above on 58 countries.

4.2. Data analyzing by the negative binomial regression model

This analysis is done to identify the association between the number of cases and the variables. Table 2 shows the mean and standard deviations for each of those factors, which gives an idea about the dispersion of each element.

4.3. Advanced analysis

The selected model is described in Equation (1) as follows,

\[
\text{Number of Cases} = \text{exponential} \left[ 6.661 + 0.073^*X_1 - 2.337^*10^{-5}^*X_2 + 0.023^*X_3 + 3.347^*10^{-5}^*X_4 + 2.457^*10^{-9}^*X_2^*X_3 - 1.139^*10^{-12}^*X_2^*X_4 - 6.735^*10^{-11}^*X_1^*X_2^*X_3 + 6.072^*10^{-14}^*X_1^*X_2^*X_4 - 4.793^*10^{-8}^*X_1^*X_2^*X_4^2 \right] \tag{1}
\]

\[
X_1 = \text{Average temperature in } ^0\text{C} \\
X_2 = \text{Population Size} \\
X_3 = \text{Median age} \\
X_4 = \text{GDP per capita PPP}
\]

The healthcare facility rank has been removed from the above model because it showed no association with the model. Therefore, healthcare rank has a minor impact on the number of cases which is negligible. According to the model, the number of cases cannot be described from the considered factors individually due to the existence of interactions. But the interaction of the factors describes the situation perfectly. Equation (1) can be effectively used to predict the number of cases within a country. However, it is important to note that the data substituted in the formula should be in the range of the values given in this study.

5. Results

This section provides the results of the statistical analysis.

Table 2
Mean values and Standard deviations.

| Factor                  | Mean       | Standard Deviation |
|-------------------------|------------|--------------------|
| No of cases             | 18,903     | 26039.42           |
| Average Temperature in °C| 14 °C      | 8.56 °C            |
| Population Size         | 72191567   | 194857039.9        |
| Median Age              | 35.57      | 7.21               |
| GDP per Capita (ppp)    | 36576.17   | 27106.49           |
| Healthcare Rank         | 62.05      | 45.247             |

5.1. Validity of the model

The goodness of fit test (using Pearson Chi-square) has been carried out to test the validity of the model. The negative binomial regression model fits the data well for the above factors. Moreover, to check the significance of the model, the omnibus test has been carried out. According to the test result, it can be concluded that the negative binomial regression model for this data is significant at 5% level.

Further, the significance of the model parameters has been checked by comparing p values with a 0.05 significance level. According to the model effects shown in Table 3, though the average temperature, population size and median age are not significant, the interaction terms associated with those three variables are significant in the model. Therefore the main effects from those variables in the model cannot be omitted.

5.2. Checking the model adequacy

In order to check the model adequacy, residual diagnostics have been used as follows.

5.2.1. Checking outliers

To check the outliers (unusual observation), standardized Pearson residuals have used. It found that all the standardized Pearson residuals are within $-2$ and $+2$ except the last three observations which can be considered as outliers in this data set.

5.2.2. Checking whether the model has fitted for the data and linearity of the model

According to Fig. 3, it can be seen that all the observations lie within $-2$ and $+2$ of standardized deviance residuals except one observation. That means 95% of the residuals are under the absolute value of 2.0. Therefore, it is evident that the model adequately fits the data. Further, the linearity in the model parameters has satisfied since there is no pattern in the plot. So the assumed model fits the data and describes the situation correctly.

6. Discussion

The primary intension of this work is to enhance the knowledge on the spread of COVID-19 related to external parameters such as temperature, population size, median age, and health care facilities. The analyzed results were presented using a negative binomial regression model. The results suggested that the population size, average temperature and median age does not have much impact when they are individually considered within the model. However, their interactions with other variables show a significant impact on the number of cases. Therefore in the model, the effect of those variables on the number of cases...
cases cannot be ignored. In contrast, healthcare facilities within a country showed a negligible effect on the spread of the virus. GDP was considered to incorporate the economic status within a country to make the analysis more practical. Since there are interactions among factors, it is difficult to explain precisely on how the number of cases changes for a unit increase in each of the variables, because that change depends on different values taken by other variables as well. Finally, an empirical formula has been presented to estimate the number of cases reported within a country under certain conditions.

The critical analysis conducted in this study can be useful in undertaking precautionary measures to mitigate or minimize the spread throughout the world. The recommendations provided by this study will be valuable in imposing preventive measures in the future to face a similar kind of pandemic.

**Funding**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Ethical approval**

Ethical approval was not required for this work as freely available open data were used.

**Declaration of competing interest**

The authors have no competing or conflicting interests.

**Acknowledgement**

The authors wish to thank the following individuals who assisted with the medical background of this research study and proofreading of the paper: Ms. H.M.D.D. Herath of Faculty of Medicine and Allied Sciences, Rajarata University of Sri Lanka and Ms. Dilani Shashika Senanayake of Faculty of Medicine, University of Kelaniya.
coronavirus (COVID-19) in different provinces of Iran. Sci Total Environ. 2020;708:1–5. https://doi.org/10.1016/j.scitotenv.2020.138872.

19 Pandemics. How are viruses spread? Nctm.org; 2020. https://www.nctm.org/Classroom-Resources/Illuminations/Interactives/Pandemics-How-Are-Viruses-Spread/. Accessed May 21, 2020.

20 de Leon-Martínez Lorena Díaz, Sierra-de la Vega Luz de la, Palacios-Ramírez Andrés, et al. Critical review of social, environmental and health risk factors in the Mexican indigenous population and their capacity to respond to the COVID-19. Sci Total Environ. 2020;733. https://doi.org/10.1016/j.scitotenv.2020.139357.

21 Walton Brett. Healthcare Facilities in Developing Countries a High Risk for Coronavirus Transmission. New Security Beat; 2020. https://www.newsecuritybeat.org/2020/03/healthcare-facilities-developing-countries-high-risk-coronavirus-transmission/. Accessed May 21, 2020.

22 Caramelo Francisco, Ferreira Nuno, Oliveiros Barbara. Estimation of risk factors for COVID-19 mortality - preliminary results. medRxiv. 2020;19. https://doi.org/10.1101/2020.02.

23 Medina Michael Arieh P. Age as a risk factor of COVID-19 mortality in the Philippines. https://ssrn.com/abstract=3579145; 2020. Accessed May 27, 2020.

24 Chappelow Jim. Gross domestic product (GDP). Invest; 2020. https://www.investopedia.com/terms/g/gdp.asp. Accessed May 27, 2020.

25 Amadeo Kimberly. Real GDP per capita, how to calculate it, and data since 1947. https://www.thebalance.com/real-gdp-per-capita-how-to-calculate-data-since-1946-396028; 2020. Accessed May 27, 2020.

26 Smith Lisa. Does high GDP mean economic prosperity? Invest; 2020. https://www.investopedia.com/articles/economics/08/genuine-progress-indicator-gpi.asp/. Accessed May 27, 2020.

27 weather2visit. Weather to visit 2020. https://www.weather2visit.com/; 2020. Accessed May 27, 2020.

28 AccuWeather. https://www.accuweather.com/; 2020. Accessed May 27, 2020.

29 Population by Country. Worldometer; 2020. https://www.worldometers.info/world-population/population-by-country/; 2020. Accessed May 25, 2020.

30 World Population Prospects 2019. United Nations. https://population.un.org/wpp/Download/Standard/Population/; 2020. Accessed May 25, 2020.

31 List of Countries by Projected GDP per capita. International monetary fund. https://worldpopulationreview.com/countries/most-visited-countries/; 2020. Accessed May 26, 2020.

32 Månsson Kristofer. On ridge estimators for the negative binomial regression model. Econ Model. 2012;29(2):178–184. https://doi.org/10.1016/j.econmod.2011.09.009.