BRIEF REPORT

Lethality model for COVID-19 based on social determinants of health: an approximation in 67 countries [version 1; peer review: 1 approved]

Fred G. Manrique-Abril1,2, Mario J. Pacheco-López3, Cristian F. Téllez-Piñerez3, Felipe Ortíz-Rico3, Juan D. Sarmiento-Páramo3

1Facultad de Enfermería, Universidad Nacional de Colombia, Bogotá, Colombia
2Facultad de Ciencias de la Salud, Universidad Pedagógica y Tecnológica de Colombia, Tunja, Boyacá, Colombia
3Facultad de Estadística, Universidad Santo Tomas, Bogotá, Colombia

Abstract

Background: Social, geographic, economic, demographic, and health factors were analysed to identify some social determinants related to case fatality of COVID-19 in 67 countries.

Methods: A mixed generalized linear model with beta distribution with random intercept was used to estimate the effects of the explanatory variables on the lethality for COVID-19 in 67 countries.

Results: The case fatality rate (CFR) was highest in the countries with the highest percentages of people over 60 years of age, the highest number of hospital beds, the highest mortality rate from diabetes, and the highest number of COVID-19 tests. Additional increases were seen based on literacy rates, health investment, death rate from cardiovascular disease, poverty rate, ratio of men, number of air passengers mobilized, number of days from the first reported case to the start of quarantine, death rate from respiratory infections, and percentage of people living in urban areas.

Conclusions: The statistical model used to predict lethality is novel because it allows the magnitude of the CFR to be analysed over a logistic model that classifies countries considering the presence and absence of deaths. When considering a beta distribution with excess zeros, the model also allows countries without reported deaths due to COVID-19 at the analysed cut-off date to be included.

Keywords

COVID-19 Pandemic, Beta Regression, Lethality, social determinants.
This article is included in the Disease Outbreaks gateway.

This article is included in the Coronavirus collection.

Corresponding author: Fred G. Manrique-Abril (fgmanriquea@unal.edu.co)

Author roles: Manrique-Abril FG: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; Pacheco-López MJ: Data Curation, Formal Analysis, Investigation, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Téllez-Piñerez CF: Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; Ortiz-Rico F: Conceptualization, Data Curation, Funding Acquisition, Project Administration, Resources, Supervision, Writing – Original Draft Preparation; Sarmiento-Páramo JD: Data Curation, Investigation, Methodology, Resources, Software, Visualization, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2020 Manrique-Abril FG et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


Introduction

COVID-19 is an infectious disease caused by the SARS-CoV-2 virus that originated in early December 2019 from the city of Wuhan, the capital of Hubei Province, in China. Its virulence and reproducibility gave the virus a special character, and the WHO declared the disease a pandemic on March 11, 2020, due to the high number of infected people and deaths that it had caused around the world. In comparison, the 2003 outbreak of severe acute respiratory syndrome (SARS) had a fatality rate of about 10% (8,098 cases and 774 deaths), and Middle East respiratory syndrome (MERS) killed 34% of people with the disease between 2012 and 2019 (2,494 cases and 858 deaths).

However, even though its death rate is minuscule, COVID-19 so far has caused more deaths (1871) than SARS and MERS combined (1632). As such, many epidemiologists, statisticians, mathematicians, and other researchers have focused on forecasting the number of new cases at the regional or global level and on detecting possible measures to decrease the probability of contagion and death in the most vulnerable populations. Few studies can determine with sufficient statistical rigour which factors decrease the probability of death from COVID-19 with any level of statistical significance.

The SIR model is on the simplest compartmental models, and many models are derivatives of this basic form. The model consists of three compartments:

S: The number of susceptible individuals. When a susceptible and an infectious individual come into “infectious contact”, the susceptible individual contracts the disease and transitions to the infectious compartment. I: The number of infectious individuals. These are individuals who have been infected and are capable of infecting susceptible individuals. R for the number of removed (and immune) or deceased individuals. These are individuals who have been infected and have either recovered from the disease and entered the removed compartment, or died. It is assumed that the number of deaths is negligible with respect to the total population. This compartment may also be called “recovered” or “resistant”. This model is reasonably predictive for infectious diseases that are transmitted from human to human. With these models, indicators of lethality are also calculated by dividing the number of deaths by the total number of those infected. This simple calculation is called the case fatality rate (CFR).

In practice, when a pandemic enters a territory, the susceptible often do not get sick, do not infect themselves, and do not die as established in the SIR model. Thus, one assumes that other factors are included in the SIR or susceptible, exposed, infected, recovered (SEIR) models (a compartmental deterministic model) that determine the infection and mortality. This would substantially modify the CFR, which is a measure of a disease’s severity and indicates the disease’s importance in terms of its ability to produce death. Given this order of ideas, lethality is calculated as CFR:

$$CFR = \frac{Nd}{Nc} \times 100$$

Where $Nd$=Number of deaths from a disease in a given period and $Nc$=Number of cases diagnosed with the same disease in the same period.

Since each country has different characteristics in terms of the social determinants and the causes and deaths COVID-19 pandemic report is uniform, it is necessary to collect the available information to model the virus’s CFR and identify the factors that statistically increase or modify the risk of death. A model will give researchers a powerful tool to estimate the probability of death from COVID-19 and its causes will give researchers a powerful tool, alongside the existing ones for predicting the number of cases, for decision-making aimed at alleviating the consequences of deadly events. That is why, in this article, we propose to model the CFR for COVID-19 using a mixed beta model and adjust it for some social, geographic, economic, health and demographic variables as social determinants in the global context of the COVID-19 pandemic.

Methods

Based on the assumptions that the CFR by country can be affected by socioeconomic factors and by health and non-pharmaceutical measures, and that its behaviour may have varied since the start of the pandemic, a mixed beta regression model with random intercept was used to estimate the effects of the explanatory variables on the CFR for COVID-19 in 67 countries. A significance level of 10% was used to identify factors that had a statistically significant relationship with the COVID-19 CFR worldwide, see Table 1.

Data source

Data on the number of daily cases and deaths from COVID-19 were obtained from the Johns Hopkins University repository to construct the CFRs by country. The possible predictors of the CFR by country were obtained from the repositories of Population Pyramid, the World Bank, the World Health Organization and the Topographic Mission Radar Shuttle, as well as from health policy, news and macro data. The seven-day corrected fatality rate was calculated nine times: on April 8, 15, 22 and 29; May 6, 13, 20 and 27; and June 3.

Beta mixed models

The parameterized beta probability density function, denoted by $Y \sim \beta(\mu, \phi)$, in terms of its mean ($\mu$) and precision ($\phi$) parameters, is given by

$$f(y | \mu, \phi) = \frac{\gamma(\phi)}{\gamma(\mu \phi) \gamma(1 - \mu \phi)} y^{\mu \phi - 1} (1 - y)^{(1 - \mu) \phi - 1}$$

$0 < y < 1$, with $0 < \mu < 1$, $\phi > 0$, $E(Y) = \mu$, $Var(Y) = \mu(1 - \mu)/(1 + \phi)$, and $\gamma$ is as the Gamma function. For a random sample from $Y_i \sim \beta(\mu_i, \phi)$, and assuming $\phi$ is constant, the beta regression model is specified by $g(\mu_i) = x^\top_i \beta = \eta_i$, with $\beta = \beta_0^\top, \beta_2^\top, ..., \beta_p^\top$ being a vector of the $(p)$ unknown regression coefficients, $x_i = (x_{i1}, ..., x_{ip})^\top$ being a vector of $(p)$ known covariates and $\eta$ being a linear prediction. The model specification is completed by the choice of a link function, $g() : (0, 1) \rightarrow IR$. We adopt the Logit function $g(\mu) = log(\mu/(1 - \mu))$. This model does not contemplate
Table 1. Explanatory variables for COVID-19 case fatality rate.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFR (response)</td>
<td>Corrected case fatality rate: measured nine times: on April 8, 15, 22 and 29; May 6, 13, 20 and 27; and June 3</td>
</tr>
<tr>
<td>Explanatory</td>
<td></td>
</tr>
<tr>
<td>60years</td>
<td>People aged over 60 years (%)</td>
</tr>
<tr>
<td>Literacy</td>
<td>Literacy rate</td>
</tr>
<tr>
<td>Beds</td>
<td>Number of hospital beds per 10000 inhabitants</td>
</tr>
<tr>
<td>Health</td>
<td>Investment in health as a percentage of GDP</td>
</tr>
<tr>
<td>Obesity</td>
<td>Prevalence of obesity</td>
</tr>
<tr>
<td>Test</td>
<td>COVID-19 tests per million inhabitants</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Diabetes mellitus mortality rate</td>
</tr>
<tr>
<td>Respiratory diseases</td>
<td>Respiratory diseases mortality rate</td>
</tr>
<tr>
<td>HIV/AIDS</td>
<td>HIV/AIDS mortality rate</td>
</tr>
<tr>
<td>Respiratory infection</td>
<td>Respiratory infection mortality rate</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>Cardiovascular disease mortality rate</td>
</tr>
<tr>
<td>Poverty</td>
<td>Percentage of people below the poverty line</td>
</tr>
<tr>
<td>Sex</td>
<td>Sex ratio (men/women)</td>
</tr>
<tr>
<td>Doctors</td>
<td>Number of doctors per 1000 inhabitants</td>
</tr>
<tr>
<td>Passengers</td>
<td>Annual volume of air passengers (%)</td>
</tr>
<tr>
<td>Quarantine</td>
<td>Days until quarantine decree</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross domestic product</td>
</tr>
<tr>
<td>Tobacco</td>
<td>Smoker between 15 and 60 years old (%)</td>
</tr>
<tr>
<td>Urban population</td>
<td>Population living in urban areas (%)</td>
</tr>
<tr>
<td>Height</td>
<td>Height above sea level (meters above the sea level)</td>
</tr>
<tr>
<td>Density</td>
<td>Population density</td>
</tr>
<tr>
<td>Nurses</td>
<td>Number of nurses per 1000 inhabitants</td>
</tr>
</tbody>
</table>

possible dependencies such as those induced by multiple measurements on the same observational unit or time.

Including latent random effects on a grouped data structure is a parsimonious strategy, as compared to adding parameters to the fixed part of the model, while still accounting for nuisance effects.

Let $Y_i$ denote an observation $j = 1, \ldots, n_i$ within group $i = 1, \ldots, q$ and $y_i$ denote a $n_i$-dimensional vector of measurements from the $i$-th group. Let $b_i$ be a q-dimensional vector of random effects, and assume the responses $Y_i$ are conditionally independent with density

$$f_Y(y_i | b_i, \beta, \phi) = \frac{\gamma(\phi)}{\gamma(\mu_i, \phi)\gamma(1-\mu_i, \phi)} y_i^{\mu_i - 1} (1 - y_i)^{(1-\mu_i)\phi - 1}$$

and a link function $g(y_i) = x_i^T \beta + z_i^T \eta_i$, vectors of known covariates $x_i$ and $z_i$ with dimensions $(p)$ and $(q)$, respectively, a $(p)$-dimensional vector of unknown regression parameters $\beta$ and the precision parameter $\phi$. The model specification is completed by $b_i | \Sigma \sim N(0, \Sigma)$ Gaussian random effects.
Preparation of the model

A model was developed that accounted for the variables’ nature, which is bounded in the interval \([0, 1]\). A mixed beta model was then assumed to model the CFR, and a step procedure did what for the final selection. The statistical software R version 3.6.3 was used to fit the model, specifically the GAMLSS function. See 14 for more technical details.

Results

The distribution of CFR by country is shown in Figure 1. For June 3, high CFR stand out in countries like Belgium, France, Mexico, England, Italy, Hungary, Holland and Sweden, as do the low values in Russia, Australia, New Zealand, Malaysia, Chile and Israel.

According to the adjusted model (Figure 2), the variables with greater statistical significance in the model were three (60years, Health and Sex), the CFR increases inversely with health expenditure; and directly the CFR is higher where more people over 60 years of age live and there is a male population predominance.

So too does CFR increase, if the number of beds increases, if there are more diabetes mortality or greater death rate from cardiovascular diseases or death rate from respiratory infections. The CFR increase if there are a greater number of tests for SARS-CoV-2. The register lethality change with the literacy proportion in the country; the CFR increase with poverty, with a greater number of air passengers mobilised, number of days from the first case reported until the beginning of the quarantine, and percentage of people living in the urban area.

According to the model, the CFR decreased in the countries with the low death rate from respiratory diseases, low obesity rate, low death rate from HIV/AIDS, low physicians per thousand inhabitants, low GDP in health care, low population density and a high number of nurses per thousand inhabitants.

Discussion

Economic Commission for Latin America and the Caribbean (ECLAC) studied the relationship between mortality in Latin American countries with some characteristics of their social development. A set of four social indicators (hospital beds per 1000 inhabitants, protein consumption, literacy and proportion of households with drinking water) have high linear correlations with life expectancy at birth \((r = 0.94)\).[15]

This global analysis of the socioeconomic determinants of COVID-19 mortality had several limitations. Countries had clear differences in their stage of historical development, with different modes of production being found. On the other hand, mortality was highly heterogeneous mortality between countries and in different populations within each country. Finally, the review of the available information indicates that the analysis is not systematic or complete, particularly for the most important analysis categories.

Nevertheless, the available data allow us to reach some important conclusions. For example, adjustment of the presented model allows factors related to COVID-19 case fatality to be identified. The methodology can be replicated with different cut-off dates, include more related factors or be used at sub-national levels.

The calculated case fatality differs from another model’s fatality proposed by researchers at the University Hospital of Lausanne and collaborators (15), who estimated a CFR of 3.59% as of March 1, 2020, considering that an adjustment must be made for the 14 previous days before death to coincide with the date of infection spread and infection, representing the maximum
incubation period for the virus. The findings show that current figures may underestimate the potential threat of COVID-19 in symptomatic patients\textsuperscript{16}.

**Conclusions**
The statistical model used to model fatalities is novel because it allows to analyse the magnitude of the fatality rate over a logistic model that classifies the countries considering the presence and absence of fatalities. Additionally, when considering a beta distribution with excess zeros, countries without reported deaths due to COVID-19 at the analysed cut-off date can be included. Finally, the variability obtained from the partial estimates made by the countries on different dates can be analysed.

The model’s results may have a high sensitivity to the data on reported case fatality from COVID-19, due to the sub-registry inherent to the measurement of infected people, because of

**Figure 2. Results of the adjusted model.**
the small number of tests carried out in some countries. However, the modelling strategy used is optimal for the type of information considered and can be replicated with new updates to the lethality measurement\(^1\).

**Data availability**

**Source data**
The population, cases, test number and deaths number by country were obtained from the repository Johns Hopkins University (JHU) at [https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data](https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data).

**Extended data**

R analysis code for this study is available at: [https://github.com/Statistick/covid19model](https://github.com/Statistick/covid19model).


License: Creative Commons Attribution 4.0 International license.

**References**


6. Mahase E: Coronavirus: covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. BMJ. 2020; 368: m641. ISSN 1756-1833. [PubMed Abstract](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7128566/) [Publisher Full Text](https://bmj.bmj.com/content/368/bmj.m641)


World Bank data were obtained using the following program: [https://github.com/vincentarelbundock/WDI](https://github.com/vincentarelbundock/WDI).
Ahmed M. Abbas  
Department of Obstetrics & Gynecology, Faculty of Medicine, Assiut University, Assiut, Egypt  

The topic is interesting and the authors covered the aspects of the paper successfully. The research methods are appropriate and the results are presented in a clear manner. Discussion and conclusions are sufficient. Citations are updated and covering the topic. Tables are good. The article is very concise and informative.

Is the work clearly and accurately presented and does it cite the current literature?  
Yes

Is the study design appropriate and is the work technically sound?  
Yes

Are sufficient details of methods and analysis provided to allow replication by others?  
Yes

If applicable, is the statistical analysis and its interpretation appropriate?  
Yes

Are all the source data underlying the results available to ensure full reproducibility?  
Yes

Are the conclusions drawn adequately supported by the results?  
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Obstetrics and Gynecology

I confirm that I have read this submission and believe that I have an appropriate level of
expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

• Your article is published within days, with no editorial bias
• You can publish traditional articles, null/negative results, case reports, data notes and more
• The peer review process is transparent and collaborative
• Your article is indexed in PubMed after passing peer review
• Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com