Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2020;6(2): 93-100

Original Article

Country Level Socioeconomic and Health System Indicators Explain COVID-19 Mortality Worldwide

Mohammad Noorchenarboo[1], Seyed Amirreza Mousavi[2], Hamed Moehimani[2]*

[1]Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran.
[2]Students Scientific Research Center, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran.

ARTICLE INFO | ABSTRACT

Received 01.04.2020
Revised 25.04.2020
Accepted 20.05.2020
Published 20.06.2020

Key words:
COVID-19;
Health care;
Mortality;
Public health;
Socioeconomic factors

Background: COVID-19 mortality rates differ across countries. We aimed to construct a model that predicts mortality worldwide, by including only country-level socioeconomic and health system indicators and excluding variables related to short-term measures for pandemic management.
Methods: COVID-19 mortality data was collected from Johns Hopkins University resource center. Additional sources were public reports from the United Nations, the World Bank and the Heritage Foundation. We implemented multiple linear regression with backward elimination on the selected predictors.
Results: The final model constructed on seven Independent variables, significantly predicted COVID-19 mortality rate by country (F-statistic: 29.2, $p<0.001$). Regression coefficients (95% CI) in descending order of standardized effects: Annual tourist arrivals: 5.43 (4.03, 6.83); health expenditure per capita: 4.43 (2.92, 5.96); GDP (PPP): -4.60 (-6.81, -2.38); specialist surgical workforce per 100000: 2.63 (0.67, 4.59); number of physicians per 1000: -2.32 (-4.3, -0.28); economic freedom score: -1.35 (-2.60, -0.10); and total population: 1.66 (-0.19, 3.52). All VIF values were below 5, showing acceptable collinearity. R-squared (52.65%), adjusted R-squared (50.25%) and predicted R-squared (42.33%) showed strong model fit.
Conclusion: limited country-level socioeconomic and health system indicators can explain COVID-19 mortality worldwide; emphasizing the priority of attending to these fundamental structures when planning for pandemic preparedness.

Introduction

2019-coronavirus disease (COVID-19) was first reported in early 2020; World Health Organization declared it a pandemic in march and by late April 2020, it has infected more than 3 million people worldwide and total number of deaths exceed 200 thousand (1).
Health systems and country level socioeconomics affect all aspects of public healthcare and play major roles in the burden that infectious outbreaks impose on countries (2-5). To manage the recent crisis, officials have pursued different emergency approaches. Comparing number of cases and mortality rates by country, effectiveness of these plans are widely discussed (6-8); yet insufficient attention is given to the prominent effect of fundamental economic and healthcare structures that influence COVID-19 outcomes.

*.Corresponding Author Email: h-moheimani@alumnus.tums.ac.ir

Please cite this article in press as Noorchenarboo M, Mousavi SA, Moehimani H. Country Level Socioeconomic and Health System Indicators Explain COVID-19 Mortality Worldwide. *2020;6(2): 93-100*

Modeling techniques help scientists explain various individual and public features of COVID-19; they assist diagnosis, compare the effectiveness of interventions and predict disease progression in communities (9-11). In the present study, we aimed to model COVID-19 mortality rate by country, using only important country level socioeconomic and health system indicators. Even though such a model would exclude variables associated with various emergency healthcare measures that address the current pandemic, we hypothesized it would be able to accurately explain the difference in mortality rates worldwide.

Methods

***Overvie*w**

The research protocol was passed in April 2020. Due to the methodologic structure, institutional review board approval was not required.

Table 1. Selected variables and related information

Category	Variable	Median (IQR)†
Outcome	COVID-19 Mortality (Deaths per 100k population) as of April 25th	0.3 (0.09-1.9)
Population	Total Population (millions)	10.7 (4.2-37.7)
	Female/male ratio	1.01 (0.99-1.03)
	Proportion of population aged 0-20 (%)	31.4 (22.1-42.6)
	Proportion aged 20-40 (%)	29.6 (26.8-32.1)
	Proportion aged 40-60 (%)	23.6 (17.4-26.0)
	Proportion aged 60+ (%)	11.7 (5.7-22.4)
Health system indicators	Health expenditure/GDP (%)	6.5 (4.6-8.4)
	Health expenditure per capita (USD)	371 (134-1223)
	Physicians per 1000 population	1.8 (0.4-3.0)
	Nurses and midwives per 1000 population	3.2 (1.1-7)
	Specialist surgical workforce per 100k population	31.2 (5.1-65.4)
Socioeconomic indicators *	Economic freedom score	61.8 (54.3-68.6)
	Property rights score	53.3 (37.2-68.5)
	Judicial effectiveness score	44.6 (32.1-56.6)
	government integrity score	36.9 (28.1-52.1)
	Government expenditure/GDP (%)	32.5 (25.5-40.4)
	GDP (PPP, billions)	158 (36-522)
	Unemployment (%)	5.8 (4.1-9.5)
Tourism	Number of yearly arrivals (millions)	3.1 (1.4-11.1)
Air pollution	Mean annual exposure to PM 2.5 (mcg/m^3)	21.4 (14.5-35.6)

† Interquartile Range
** Definitions available at heritage.org/index*

We chose COVID-19 mortality rate (defined as the number of deaths per 100000 population of countries, as of April 25th) as the outcome (dependent) measure. Countries with at least one death attributable to COVID-19 were included in the study. Independent variables associated with health system and national socioeconomic conditions were selected based on their relative importance in the published literature (2, 4, 5, 12, 13) and accessibility of adequate and reliable data (missing values for all selected variables were less than 5%). Considering the nature of this pandemic, indicators of air pollution and tourism were also selected (14, 15). Descriptive statistics and information regarding all selected variables can be viewed at table 1.

Data sources

All data were collected from publicly available recourses. COVID-19 mortality was based on the reports from Johns Hopkins coronavirus resource center (2020) (16). Indicators of health systems, air pollution and tourism were obtained from the World Bank (from 2016, 2017 and 2018 respectively) (17). Population data and socioeconomic indicators were respectively collected from the United Nations (2019) (18) and the Heritage Foundation (2020) (19).

Analysis

All analysis were performed using R, version 3.5.3 (The R Foundation for Statistical Computing, Vienna, Austria). As long as more intuitive models can explain phenomena, advanced models are better avoided (20). With this in mind, we constructed a multiple linear regression model on the standardized scores from all twenty independent variables. The final model was achieved by implementing the backwards elimination method (elimination alpha=0.1) (21, 22). Significance of the model was tested using Analysis of Variance (ANOVA). We calculated R-squared (coefficient of determination), adjusted R-squared and predicted R-squared to evaluate the goodness of fit (23, 24). Regression coefficients and standardized effects were calculated for each independent variable present in the final model and variance inflation factors (VIF) were measured to assess multi-collinearity among variables (25). For all tests, p-values less than 0.05 were considered statistically significant.

Results

After removing the countries with no reported COVID-19 deaths, we ran our analysis on 146 remaining countries. Using the backward elimination method, 13 of 20 selected variables which did not contribute significantly to model fitting were dropped and seven remained: Total population, health expenditure per capita, number of physicians per 1000, specialist surgical workforce per 100000, purchasing power parity (PPP) adjusted GDP, economic freedom score and number of annual tourist arrivals. The final model significantly predicts COVID-19 mortality by country (F-statistic: 29.2, $p<0.001$). Table 2 and figure 1 show regression coefficients and standardized effects respectively.

VIF quantifies the severity of multi-collinearity in a regression analysis; the higher the VIF value, the greater correlation between one independent variable and other variables. In this analysis, all VIF values are below 5 and therefore show acceptable amount of collinearity (25). R-squared represents the amount of variability in the outcome that can be explained by the model. Adjusted R-squared takes the number of predictors into account and penalizes the goodness of fit measure for extra predictors. Our attention to parsimony resulted in relatively similar R-squared (52.65%) and adjusted R-squared values (50.25%). Based on common interpretative criteria, these values show strong fit of the model to the data (23).

Table 2. Regression statistics for the final model

Independent variable	Coefficient (95% CI)*	VIF
Annual tourist arrivals	5.43 (4.03, 6.83)†	2.00
Health expenditure per capita	4.43 (2.92, 5.96)†	2.23
GDP (PPP)	-4.60 (-6.81, -2.38)†	4.96
Specialist surgical workforce per 100k	2.63 (0.67, 4.59)‡	3.92
Number of Physicians per 1000	-2.32 (-4.3, -0.28)‡	4.22
Economic freedom index score	-1.35 (-2.60, -0.10)‡	1.59
Total population	1.66 (-0.19, 3.52)	3.52

** To aid interpretation, coefficients and 95% confidence intervals were calculated for the standardized scores instead of raw values*

† Significant at $p < 0.001$

‡ Significant at $p<0.05$

The value for predicted R-squared depends on the degree of accuracy with which the model can predict mortality for a country, when it is excluded from the analysis and the model is built on all other countries. Here, the absence of considerable difference between predicted R-squared (42.33%) and adjusted R-squared gives us confidence that the calculated goodness of fit measures are not the result of an overly-fitted model (24).

Discussion

Success of every public health measure depends on the structure that it is laid upon. This fact is even more pronounced during times of crisis, such as 2019-2020 coronavirus pandemic (2, 26). As these foundations can only be altered in the long run, governments and health officials rely on their short term policies to alter the course of COVID-19. Consequently, most current works assess the pros and cons of varying emergency measures (6, 7, 27). Such scholarly discussions can provide decision makers with useful information at this time, but can also shift our attention from the underlying issues that play much more important roles in the mortality from both infectious and non-communicable diseases. Our study shows in the absence of variables regarding short-term policies, limited socioeconomic and health system indicators can explain COVID-19 mortality. The purpose of this work was not to diminish the value of emergency decisions that are implemented worldwide to manage this pandemic (our research methodology does not allow such conclusions), but to emphasize the fact that a stronger economy and a more resilient health system might have the most prominent effect in reducing mortality in times crisis as well as ordinary conditions (5, 28). Laws and policies that aim to improve economic and healthcare foundations today, will save the most lives when the next pandemic or health crisis arrives. Additionally, even short term and crisis-driven policies will be much more effective if they are planned ahead (29). Investigations with hierarchical type of analysis will be helpful in assessing the added value of such interventions.

In this work, regression coefficients were calculated for a model that included seven independent variables without interaction terms. While the degree of collinearity is acceptable for the whole model, presence of correlation among independent variables affects their coefficient measurements and insignificant p-values do not necessarily mean that the associated variables are unimportant (30). Even after considering such interpretative limitations, studying the standardized effects for each variable that is present in the final model can be informative (figure 1).

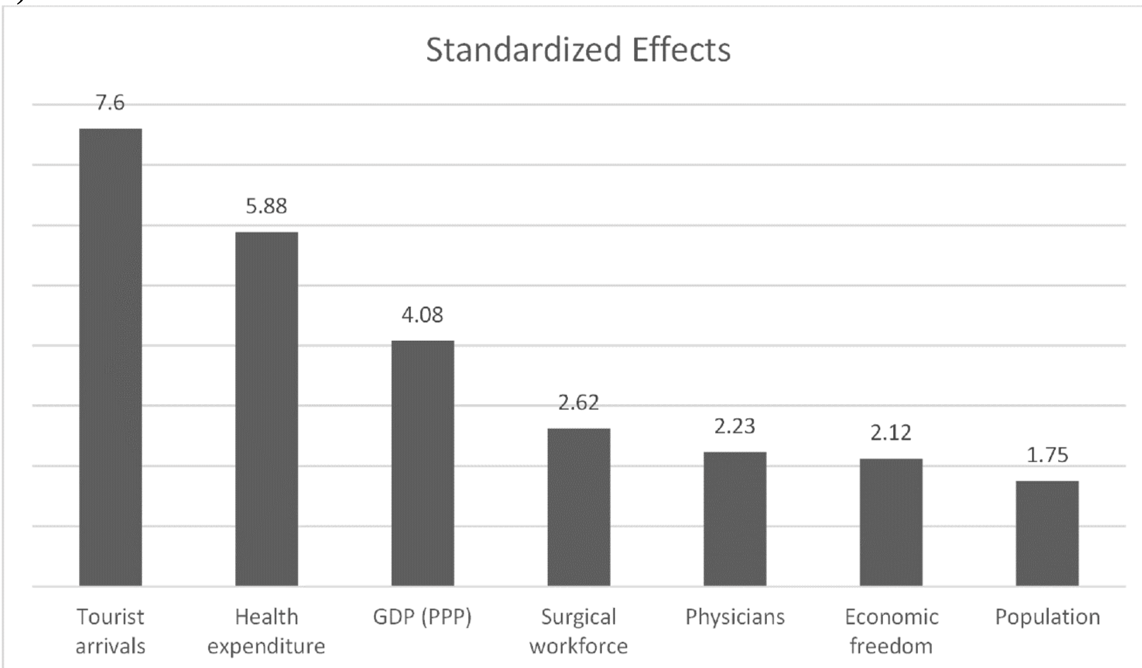

Figure 1. Standardized effects

Number of annual tourist arrivals possesses the highest effect in our model of COVID-19 mortality. If we interpret this variable as an indicator of global connectivity (31), this result is in line with known effects of tourism and globalization in facilitating propagation of COVID-19 and other infectious diseases (32, 33).

Availability of large scale screening and testing in a country is partly predicted by the amount of health expenditure (34). The positive relationship between health expenditure per capita and mortality might be the result of missed COVID-19 deaths in countries with less healthcare resources. Readers should also note the association between the outcome and health expenditure as well as surgical specialist workforce, was assessed while adjusting for the number of physicians per 1000. Future investigations can shed light on the effects of maintaining different priorities in allocating limited healthcare recourses on pandemic preparedness (25, 35).

The overall relationship of GPD (PPP), a measure of country's wealth, and COVID-19 mortality rate was strongly negative. More interestingly, economic freedom (defined as 'the ability of a society in taking economic actions' and scored by the Heritage Foundation) is independently associated with lower mortality. In this climate that the advantages and disadvantages of centralized decision-making are widely discussed (6), this result is a reminder of studies that correlate higher economic freedom with better social and health indicators (36, 37).

By and large, more developed countries have a higher proportion of aged adults and have undergone higher extent of industrialization (38). While old age and exposure to polluted air are direct risk factors for COVID-19 (13, 14), their overall relationships with mortality rates are also affected indirectly due to their association with social and environmental characteristics of developed economies. This might justify the absence of 60+ population and air pollution (as measured by annual exposure to P.M. 2.5) among significant contributors to the final model.

Although inclusion of additional country-level variables (e.g. location and climate) could potentially improve predictive power of the model, they were not relevant to our research question and their omission is not a shortcoming. However, this work bears some limitations. At the time of writing, we are still relatively early in the course of coronavirus pandemic and it is unclear if the model would yield the same results with the final mortality rates. In addition, some countries have not provided any statistics on COVID-19 and the uncertainties regarding the accuracy of officially reported numbers can limit the interpretative value of this analysis. Comparing the results and predictions from this work with future studies conducted on authenticated and finalized mortality rates can provide further insights. Lastly, this work is an observational study and subject to the associated methodological weaknesses; therefore we caution readers to only make causal interpretations after considering more comprehensive longitudinal and interventional studies, or natural experiments (39).

Conclusions

A linear model constructed on limited country level socioeconomic and health system indicators explains worldwide COVID-19 mortality. While we challenge the effectiveness of different national emergency measures, we should start planning for the future pandemics by designing and implementing policies that strengthen these fundamental structures.

Acknowledgements

The authors would like to thank the faculty members at Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, who acted as scientific advisors and supported this study.

Conflicts of interest

None declared.

References

1. WHO. Coronavirus disease 2019 (COVID-19): situation report, 100. 2020.
2. Bygbjerg IC. Double burden of noncommunicable and infectious diseases in developing countries. Science. 2012;337(6101):1499-501.
3. Loayza NV, Pennings S. Macroeconomic policy in the time of covid-19: A primer for developing countries. World Bank; 2020.
4. Dowd JB, Andriano L, Brazel DM, Rotondi V, Block P, Ding X, et al. Demographic science aids in understanding the spread and fatality rates of COVID-19. Proceedings of the National Academy of Sciences. 2020;117(18):9696-8.
5. Legido-Quigley H, Asgari N, Teo YY, Leung GM, Oshitani H, Fukuda K, et al. Are high-performing health systems resilient against the COVID-19 epidemic? The Lancet. 2020;395(10227):848-50.
6. Kupferschmidt K, Cohen J. Can China's COVID-19 strategy work elsewhere? : American Association for the Advancement of Science; 2020.
7. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? The Lancet. 2020;395(10228):931-4.
8. Gibney E. Whose coronavirus strategy worked best? Scientists hunt most effective policies. Nature. 2020.
9. Singh D, Kumar V, Kaur M. Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks. European Journal of Clinical Microbiology & Infectious Diseases. 2020:1-11.

10. Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, et al. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. The Lancet Infectious Diseases. 2020.
11. Enserink M, Kupferschmidt K. With COVID-19, modeling takes on life and death importance. American Association for the Advancement of Science; 2020.
12. Armocida B, Formenti B, Ussai S, Palestra F, Missoni E. The Italian health system and the COVID-19 challenge. The Lancet Public Health. 2020.
13. Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. Jama. 2020.
14. Martelletti L, Martelletti P. Air pollution and the novel Covid-19 disease: a putative disease risk factor. SN Comprehensive Clinical Medicine. 2020:1-5.
15. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science. 2020;368(6489):395-400.
16. Coronavirus recourse center at Johns Hopkins University (2020). Mortality analysis. Retrieved April 25, 2020 from: https://coronavirus.jhu.edu/data/mortality
17. The World Bank. Open data. https://data.worldbank.org/ Accessed April 25, 2020.
18. The United Nations. World population prospects. https://population.un.org/wpp/ Accessed April 25, 2020.
19. The Heritage Foundation. Index of economic freedom. https://www.heritage.org/index/ Accessed April 25, 2020.
20. Green KC, Armstrong JS. Simple versus complex forecasting: The evidence. Journal of Business Research. 2015;68(8):1678-85.
21. Heinze G, Wallisch C, Dunkler D. Variable selection–A review and recommendations for the practicing statistician. Biometrical Journal. 2018;60(3):431-49.
22. Derksen S, Keselman HJ. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. British Journal of Mathematical and Statistical Psychology. 1992;45(2):265-82.
23. Miles J. R squared, adjusted R squared. Wiley StatsRef: Statistics Reference Online. 2014.
24. Hawkins DM. The problem of overfitting. Journal of chemical information and computer sciences. 2004;44(1):1-12.
25. O'brien RM. A caution regarding rules of thumb for variance inflation factors. Quality & quantity. 2007;41(5):673-90.
26. Michaud C. Global Burden of Infectious Diseases. Encyclopedia of Microbiology. 2009:444.
27. Hale T, Petherick A, Phillips T, Webster S. Variation in government responses to COVID-19. Blavatnik School of Government Working Paper. 2020;31.
28. Bhutta ZA, Sommerfeld J, Lassi ZS, Salam RA, Das JK. Global burden, distribution, and interventions for infectious diseases of poverty. Infectious diseases of poverty. 2014;3(1):21.
29. Osterholm MT. Preparing for the next pandemic. New England Journal of Medicine. 2005;352(18):1839-42.
30. Gordon RA. Issues in multiple regression. American Journal of Sociology. 1968;73(5):592-616.
31. Burns PM, Novelli M. Tourism and mobilities: local-global connections: CABI; 2008.
32. Farzanegan MR, Feizi M, Gholipour HF. Globalization and outbreak of COVID-19: An empirical analysis. Joint Discussion Paper Series in Economics; 2020.
33. Peeri NC, Shrestha N, Rahman MS, Zaki R, Tan Z, Bibi S, et al. The SARS, MERS and novel coronavirus (COVID-19) epidemics,

the newest and biggest global health threats: what lessons have we learned? International journal of epidemiology. 2020.
34. Nixon J, Ulmann P. The relationship between health care expenditure and health outcomes. The European Journal of Health Economics. 2006;7(1):7-18.
35. Barrett DH, Ortmann LH, Dawson A, Saenz C, Reis A, Bolan G. Public health ethics: cases spanning the globe: Springer International Publishing; 2016.
36. Esposto AG, Zaleski PA. Economic freedom and the quality of life: an empirical analysis. Constitutional political economy. 1999;10(2):185-97.
37. Stroup MD. Economic freedom, democracy, and the quality of life. World Development. 2007;35(1):52-66.
38. Hardiess G, Meilinger T, Mallot HA. The International Encyclopedia of the Social and Behavioral Sciences. 2015.
39. Vetter TR, Schober P. Regression: the apple does not fall far from the tree. Anesthesia & Analgesia. 2018;127(1):277-83.