BMJ Open

Macro-Scale Estimators of Worldwide COVID-19 Occurrence and Mortality

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-055562
Article Type:	Original research
Date Submitted by the Author:	15-Jul-2021
Complete List of Authors:	Erdem, Sabri; Dokuz Eylül University, Department of Business Administration, İzmir/Turkey Ipek, Fulya; Hacettepe University Bars, Aybars; Dokuz Eylul Universitesi, Social Science Institute Genç, Volkan; Dokuz Eylul Universitesi, Social Science Institute Erpek, Esra; Izmir Katip Celebi University, Department of Internal Medicine Mohammadi, Shabnam; Dokuz Eylul Universitesi, Social Science Institute Altınata, Anıl; Dokuz Eylul Universitesi, Social Sciences Institute Akar, Servet; Izmir Katip Celebi University, Department of Internal Medicine
Keywords:	COVID-19, Public health < INFECTIOUS DISEASES, Health policy < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Macro-Scale Estimators of Worldwide COVID-19 Occurrence and Mortality

Running title: COVID-19 Global Analysis

Sabri Erdem, PHD, Professor

Affiliation: Dokuz Eylül University, Faculty of Business, Department of Business Administration, İzmir/Turkey

E-mail: sabri.erdem@deu.edu.tr

ORCID ID: 0000-0001-6766-3202

Fulya Ipek, PT, MSc (Corresponding Author)

Affiliation: Hacettepe University, Faculty of Physical Therapy and Rehabilitation, Ankara, Turkey

Email: fulya.ipek@hacettepe.edu.tr

ORCID ID: 0000-0001-7606-042X

Aybars Bars, MSc

Affiliation: Dokuz Eylül University, Social Sciences Institute, İzmir/Turkey

E-mail: aybars.bars@ogr.deu.edu.tr

ORCID ID: 0000-0002-8051-9165

Volkan Genç, MSc

Affiliation: Dokuz Eylül University, Social Sciences Institute, İzmir/Turkey

E-mail: volkan.genc@ogr.deu.tr

ORCID ID: 0000-0003-2184-482X

Esra Erpek, MD

Affilation: İzmir Katip Çelebi University, Department of Internal Medicine, Division of Rheumatology Atatürk Education and Research Hospital, İzmir/Turkey

E-mail: esraerpek@gmail.com

ORCID ID: 0000-0003-3540-4905

Shabnam Mohammadi, MSc

Affilation: Dokuz Eylül University, Social Sciences Institute, İzmir/Turkey

E-mail: mohammadi.shabnam@ogr.deu.edu.tr

ORCID ID: 0000-0001-6987-0301

Anıl Altınata, MSc

Affilation: Dokuz Eylül University, Social Sciences Institute, İzmir/Turkey

E-mail: altinata.anil@ogr.deu.edu.tr

ORCID ID: 0000-0001-6300-5982

Servet Akar, MD, Professor

Affilation: İzmir Katip Çelebi University, Department of Internal Medicine, Division of Rheumatology Atatürk Education and Research Hospital, İzmir/Turkey

E-mail: servet.akar@gmail.com

ORCID ID: 0000-0002-3734-1242

Fulya Ipek, PT, MSc (Corresponding Author)

Affiliation: Hacettepe University, Faculty of Physical Therapy and Rehabilitation, Ankara, Turkey

Address: Hacettepe University, Faculty of Physical Therapy and Rehabilitation, 06100, Altındağ/Ankara, Turkey

Email: fulya.ipek@hacettepe.edu.tr

fulyaipek.92@gmail.com

Phone: +90 539 272 20 22

Abstract

Objective: To investigate macro-scale estimators of the variations in COVID-19 cases and deaths among countries.

Design: Epidemiological study.

Setting: Country-based data from publicly available online databases of international organizations.

Participants: The study involved 170 countries/territories, each of which had complete COVID-19 and tuberculosis data, as well as specific health-related estimators (obesity, hypertension, diabetes, and hypercholesterolemia).

Primary and secondary outcome measures: The worldwide heterogeneity of the total number of COVID-19 cases and deaths per million on 31 December 2020 was analyzed by 17 macro-scale estimators around the health-related, socio-economic, climatic, and political factors. In 139 of 170 nations, the Best Subsets Regression was used to investigate all potential models of COVID-19 variations among countries. A multiple linear regression analysis was conducted to explore the predictive capacity of these variables. The same analysis was applied to the number of deaths per hundred thousand due to tuberculosis, a quite different infectious disease, to validate and control the differences with the proposed models for COVID-19.

Results: In the model for the COVID-19 cases (R^2=0.45); obesity (β=0.460), hypertension (β=0.214), sunshine (β=-0.157), and transparency (β=0.147), whereas in the model for COVID-19 deaths (R^2=0.41); obesity (β=0.279), hypertension (β=0.285), alcohol consumption (β=0.173), and urbanization (β=0.204) were significant factors ($p<0.05$). Unlike COVID-19, the tuberculosis model contained significant indicators like obesity, undernourishment, air pollution, age, schooling, democracy, and GINI index.

Conclusions: This study recommends the new predictors explaining the global variability of COVID-19. Thus, it might assist policy-makers in developing health policies and social strategies to deal with COVID-19.

Trial Registration: ClinicalTrials.gov Identifier: NCT04486508

Strengths and limitations of this study

- The results of this study are valid since it covers recent macro-scale data of 170 countries and Covid-19 cases and deaths for a one-year period.
- This study is isolated from the vaccine effect because Covid-19 cases and deaths data were terminated just before vaccination in the world.
- This study gives us a predictive model that provides the importance of each factor on Covid-19 cases and deaths. It incorporates social, economic, and political indicators as well as health-related ones.
- The main limitation of this study is that some macro-scale estimators were not recently published but we assume that there are no remarkable changes annually.
- Another limitation is that there might be some interruption or reliability issues in Covid-19 Cases and Deaths data flow from countries that could affect the reliability of the model.

INTRODUCTION

The novel coronavirus disease 2019 (COVID-19) caused by Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) has spread to countries worldwide. On 11 March 2020, the World Health Organization (WHO) declared the condition as a public health emergency. With new variants emerging, experts now assume that SARS-CoV-2 would not only remain an endemic virus but also would continue to circulate in communities and would make a massive burden of disease and death for years.[1]

The cumulative incidence (total number of cases per million) of COVID-19 (TotalCase-COV19) and mortality (total number of deaths per million) of COVID-19 (TotalDeath-COV19) vary significantly among the countries. Besides, old age, diabetes, high blood pressure, obesity, pregnancy, immunosuppression, cancer, and cardiovascular, respiratory (asthma and chronic obstructive pulmonary disease, etc.), and chronic kidney disease are all well-known health-related factors that contribute to COVID-19 morbidity and mortality.[2, 3] Aside from these clinical factors and comorbidities affecting infection progression, there might be additional demographic, social, economic, and environmental factors at the macro-level differentiating countries in terms of TotalCase-COV19 and TotalDeath-COV19.

The term “Macro-level data'' (e.g., Human Development Index 2020 by United Nations) could be defined as periodic (e.g., weekly, monthly, quarterly, or annual) and long-run data that gathered, summarized, and published by the government, private, public, national and international organizations about a specific subject (e.g., health, energy, social, economic, politic) in the world, continents, regions, countries, cities or territories. This kind of data shows the current position and trend of any subject in a comparable manner and guides the policymakers in countries and organizations for setting or updating their strategies and helps them take an action around that subject. Although healthcare executives and regulatory authorities have implemented regulations and recommendations such as wearing masks, social distancing, staying at home, isolation, and lockdowns to limit the spread of SARS-CoV-2, considering these macro indicators will help steer national and international health policymakers to combat the COVID-19 pandemic.

To the best of our knowledge, there has been limited information on how the nature and dynamics of SARS-CoV-2 vary across the world. [4-8]. These studies have revealed that demographic, climatic, environmental, socio-economic, and political indicators could help predict or clarify variations in TotalCase-COV19 and TotalDeath-COV19 among countries. In particular, the elderly ratio, the prevalence of comorbidities, population size, and health expenditure were shown to have been associated with the incidence and mortality of COVID-19

at various levels according to the aforementioned studies. However, these findings have been based on the early phase of the outbreak. The current study has been based on data (the cumulative number of cases and deaths) just before starting vaccinations around the world. Additionally, the findings might provide more consistent details about the progression of the disease due to covering data for a longer period. Therefore, the current study aims to identify health-related and socio-economic macro-scale indicators that could explain some of the substantial variations in TotalCase-COV19 and TotalDeath-COV19 worldwide.

METHODS

Publicly accessible data about TotalCase-COV19 and TotalDeath-COV19 were obtained from the John Hopkins University – Center for Science and Engineering (JHU-CSSE).[9] While the process of the pandemic was still ongoing mass immunization has started in several countries as of December 2020. To eliminate the effect of immunization, in this study, TotalCase-COV19 and TotalDeath-COV19 data were censored by 31 December 2020, just before vaccination has not yet become widespread. Our dataset is available in Open Science Framework. Health-related indicators from the World Health Organization and socio-economic, climatic, and political indicators from United Nations UN, World Inequality Database (WID), and Worldbank explaining variability in TotalCase-COV19 and TotalDeath-COV19 among countries were summarized in Table 1. Our dataset involves 17 macro-scale estimators of 170 countries.

[Table 1 is near here]

Statistical analyses

In this study, we conducted multiple linear regression analyses to explain variations in our target variables (i.e., TotalCase-COV19 and TotalDeath-COV19). Before the analysis, we tested all the assumptions of multiple linear regression analysis by:

- drawing “actual target variable vs predicted target value scatter” plot diagram for testing linearity,
- calculating standardized residual values for each case for detecting outlier as outside ∓ 3,
- drawing a partial regression plot diagram for each predictor vs target variable for detecting the variance of predictors,
- calculating variance inflation factor (VIF) for each predictor variable in the model for testing multicollinearity,

- calculating Durbin-Watson test statistics for detecting autocorrelation of consequent errors,
- drawing "regression predicted value vs regression standardized residual" plot diagram for detecting heteroscedasticity and
- drawing "observed cumulative probability vs expected cumulative probability (P-P)" plot diagram of standardized regression residual for testing normality of residuals.

For eliminating the variance inflation and suppressing potential interactions among predictors, we have transformed all variables into the standardized form [(x-mean(x))/standard deviation(x).], so that we could get a more acceptable variance inflation factor (VIF=$1/(1-R^2)$ <2.0) values where R^2 is obtained by regression analysis in that the inspecting variable is regarded as a dependent variable and others are regarded as its predictor. We have applied Durbin-Watson Statistics to test the autocorrelation in consequent error terms in the model and we have found all error terms are independent in all models since test statistics were between lower critical value and (4-lower critical value).[30] After having inspected the plot diagrams and histograms, it could be said that all assumptions of multiple linear regression were almost met with the help of using standardized values of all predictors and target variables in all models. The exception for slight violation was observed as variances in error terms in P-P plots and predicted values against observed values scatter plot for homoscedasticity that could be questionable if other test results were unsatisfactory.

We conducted multiple linear regression analyses for different situations/scenarios and created predictive models to determine indicators aligned with these scenarios. After having tested the dataset against the assumptions of multiple linear regression, we investigate the dataset using some considerations for each scenario/model through a best-subset approach that analyzes data in terms of a predictors subset that could create applicable models that illuminates the next step for finding the best one. We could decide the better one by inspecting each model in terms of cost-benefit ratios such as R^2, Std. Error or Mallows' Cp (min is preferred), AIC (Akaike Information Criterion: min is preferred), and BIC (Bayesian Info Criterion: min is preferred).[31]

After determining the incorporated variables in a model, we applied a hierarchical regression model with forward selection and stepwise methods. There is no constant in our models because they were all insignificant statistically and were discarded. Using the information from best-subset, literature, and best practices, we tried a variety of models by evaluating the usage of intercept, handling missing values, handling outliers, applying the

best technique for regression analysis (e.g., enter, stepwise, forward/backward selection), determining thresholds for alpha-to-enter (0,1) and remove (0,15) for accepting or rejecting the next variable to the incorporated model. We utilized the SPSS 24 and Minitab 19 for analyzing the data.

RESULTS

The descriptive statistics of all variables were provided in Table 1. We have considered possible different models, explaining the TotalCase-COV19 and TotalDeath-COV19, including both common and different estimators that could explain the variations in countries. Therefore, we have listed all possible significant models via the best subset method widely used in regression analyses for TotalCase-COV19 and TotalDeath-COV19 are demonstrated in Supplementary Table A and Table B.

Based on best-subset analysis of TotalCase-COV19 and TotalDeath-COV19 modeling, the indicators "Sunshine", "Obesity", "Hypertension", "Urbanization", "Schooling", "Alcohol", "Democracy", "Transparency", and "HDI" are commonly found significant for the most of models. Additionally, for modeling the "TotalDeath-COV19", "Age", "GINI" and "Undernourishment" were found significant; and for modeling "TotalCase-COV19", "Cholesterol" was found significant. However, it was remarkable that the best-subset model shows the results for only 139 countries with complete data for all variables at issue. That is why the recommended model could not be significant as in best-subset because of missing values and their handling strategies. In all these regression analyses, we found that regression model fit and all parameter fits are significant ($p<0.05$).

Urbanization, schooling, HDI, transparency, democracy, and the GINI index are all interrelated factors, despite the lack of their multicollinearity. Thus, countries with desirable socioeconomic conditions are likely to have high levels of urbanization, education, human development, transparency, and democracy, as well as low-income inequality.

Table 2 shows the results of the regression model indicating the determinants of the TotalCase-COV19 worldwide. Here is the regression model indicating that transparency, sunshine, and obesity, and hypertension in countries could help to explain the variability of the TotalCase-COV19 across countries. In this regression analysis R^2 value shows that the model with these four indicators explains 44.9% of all variability. As a result, the generated model is acceptable and comes out with expected roles of variables as aligned with the literature.

[Table 2 is near here]

Table 3 shows the results of the regression model indicating the determinants of the TotalDeath-COV19 worldwide. The regression model revealed that urbanization and alcohol in addition to obesity and hypertension might help to explain the variation in the TotalDeath-COV19 across countries. In this regression analysis, the R^2 value shows that the model with these four predictors explains 40.9% of all variability. Durbin-Watson value shows that there is no evidence for autocorrelation at 0,05 significance level.

[Table 3 is near here]

Finally, we analyzed the TBDeath by utilizing the same 17 macro-scale indicators for validating-controlling purposes of our model proposals for the TotalCase-COV19 and TotalDeath-COV19. As an infectious disease quite dissimilar from COVID-19; the TBDeath model was expected to be different from our proposed model. As parallel in our expectations, it was quite different in terms of their involved variables except for obesity where "Air Pollution", "Undernourishment", and "GINI" are other significant factors rather than our proposed models (Table 4). Exceptionally, VIF values of Age and Schooling variables are large, relatively and still tolerable since their VIF values are not exceeding a threshold of 5.[32]

[Table 4 is near here]

DISCUSSION

Principal findings

The present study, which aims to explain the heterogeneity in the TotalCase-COV19 and TotalDeath-COV19 worldwide, by using an analysis of the health-related, socio-economic, climatic, and political macro indicators of 170 countries showed two regression models. Well-established comorbidities, obesity and hypertension, were significant in both these models. Additionally, indicators of sunshine and transparency for the TotalCase-COV19 model, as well as indicators of alcohol and urbanization for the TotalDeath-COV19, were important.

Interpretation and comparison with previous studies

Previous studies proposed that the comorbidities (obesity, hypertension, pulmonary disease, etc.) contributed to the risk for COVID-19 and progression to severe disease by increasing the expression of angiotensin-

converting enzyme 2 (ACE2) and/or transmembrane protease serine 2 (TMPRSS2) on host lung cells and heightening the permissiveness of viral infection.[2, 33-35] In this study, the TotalCase-COV19 and TotalDeath-COV19 were greater in countries with a higher prevalence of obesity and hypertension. Global analyzes of the variability of COVID-19 mortality among countries, as in our study, revealed that TotalDeath-COV19 were positively correlated with deaths due to comorbidities, including cardiovascular, chronic respiratory, and kidney diseases, obesity, and cancer.[5, 7, 36]

The association between sunshine and the COVID-19 is arguable in the literature. Studies revealing sunshine hours positively linked to COVID-19 growth claimed that on sunny days outdoor activities increased along with a decline in adherence to lockdown rules, resulting in increased virus exposure and transmission.[37-39] Other studies, on the other hand, suggested that sunny weather could help prevent the spread of COVID-19 by minimizing air pollution and increasing vitamin D production.[40-42] Those studies revealed that UV light from the sun can inactivate SARS-CoV-2, reduce outdoor transmission or increase immune resistance through vitamin D production. These results are supported by our research, which found a high TotalCase-COV19 in countries with low sunshine.

Public corruption and GINI reflecting corruption in the flow of products, resources, and services inside a nation were found to be associated with the TotalCase-COV19 and TotalDeath-COV19.[43, 44] The present study further supports these studies by indicating that there are more TotalCase-COV19 in countries with higher transparency (i.e., lower corruption) levels. Thus, not only socio-economic indicators but also political indicators could play a critical role in the COVID-19 outbreak. Lack of access to reliable sources of information, as well as misinformation and inadequate communication, might lead to people disregarding government health alerts. Policies that promote the effective distribution of government budgets in public goods and services like healthcare and education, as well as policies that encourage transparency and information flow, could serve information accuracy about COVID-19's spread.[43, 45]

There is no country-level analysis demonstrating an association between alcohol consumption and the TotalCase-COV19 and TotalDeath-COV19 in the literature. Recent studies have emphasized that increasing alcohol consumption is linked to excessive production of pro-inflammatory cytokines by hepatic cells, resulting in higher levels of inflammatory markers (Interleukin-8 and tumor necrosis factor [TNF-a]) which has also been observed in COVID-19 patients.[46-48] As a result, it was hypothesized that acute and chronic alcohol consumption might suppress the immune system, leading to reduced resistance to acute respiratory disease and SARS-CoV-2

infection, as well as facilitating the progression of COVID-19.[49, 50] The present study supports these findings, demonstrating that TotalDeath-COV19 are higher in countries with high alcohol consumption.

Urbanization has the potential to exacerbate disease and mortality by creating a slew of challenges such as public transportation, health, and economic disparities, substandard living facilities, insufficient freshwater supplies, and ineffective sanitation and ventilation systems.[4, 6] Existing literature has shown that dense urban populations could cause less social distance, more trade, and more human mobility, leading to multiple infection routes and more rapid spread of COVID-19.[51, 52] However, no significant relation was observed between the urbanization and the TotalCase-COV19 whereas some of the best subset models involve it. One explanation for this may be the difficulty of identifying infected people in countries with low socioeconomic status due to inadequate test conditions for COVID-19. Therefore, the effect of urbanization on the TotalCase-COV19 can be hard to observe. Besides that, developed countries with a larger urban population and advanced healthcare systems have longer life expectancies and a higher percentage of the population over 65, which could lead to higher mortality rates in high-income countries.[8] Moreover, as tougher restriction policies for elderly people have been implemented, urbanization might be a significant factor in TotalDeath-COV19 rather than TotalCaseCOV-19. Our study thus showed that countries with higher urban populations have higher TotalDeath-COV19 in line with other global COVID-19 studies.[6, 53]

Best subset regression analyses show that heterogeneity in the TotalCase-COV19 and the TotalDeathCOV-19 across countries may be attributed to variations in macro indicators including obesity and hypertension, urbanization, schooling, transparency, democracy, and HDI, as well as alcohol consumption and sunshine. Additionally, while the elderly population, the GINI index, and undernourishment were related to the TotalDeathCOV-19, hypercholesterolemia was related to the TotalCase-COV19. While the elderly population was a major parameter in nearly all TotalDeathCOV-19 models in our analysis, however, it was left out of the proposed models because it suppresses other estimators.

Other worldwide and national studies suggested that socio-economic, political, climatic, environmental, and ecological factors are correlated with the TotalCase-COV19 and the TotalDeathCOV-19 at different contexts and levels.[4-8, 54-57] These studies revealed that countries with a high elderly population and level of democracy, as well as low levels of schooling and human development, had the highest mortality rate. Socioeconomic inequalities like low socioeconomic status, inadequate schooling, limited access to healthcare, and income inequality induce difficulty in accessing and affording healthy food, use of tobacco and alcohol, low

physical activity, and lower use of preventive medicine. This situation might increase the prevalence of comorbidities by altering the gut microbiome, increasing local inflammation, and compromising immunity. Hence, the high-risk population becomes more susceptible to COVID-19.[2]

To control the COVID-19 models, we perform a regression analysis to estimate the TBDeath consisting of the same macro-scale estimators. In this regression model, tuberculosis is associated with GINI, air pollution, undernourishment, age, democracy, schooling, and obesity following the literature.[58-60] There were remarkable differences between our Covid models and Tuberculosis as we expected since they are different infections and they should have their own (disease-specific) characteristics.

Limitations

There are certain limitations of the present study. Firstly, not all health-related, socio-economic, climatic, and political indicators were up to date like COVID-19 deaths and cases in 2020. At that point, it is reasonable to expect that macro-scale indicators would not change significantly over the few years. Secondly, just 139 countries out of 170 were involved in the best subset regression analyses since only these countries had complete data. Thirdly, definitions of COVID-19 cases and illness for people based on symptoms and diagnosis varied by country over a certain time lead some country data on the number of cases to be biased, and so it increases the residual in the model. Additionally, some interruptions on data flow about COVID-19 deaths and cases from countries could lead the model to be biased. To some extent, these error impacts (i.e., bias) were minimized by standardizing the data's relative position to other countries' data by considering overall mean and standard deviation as well. In deciding the cause of death, a similar condition occurred in that some deaths could be recorded such as due to respiratory problems, other infections, or multiple organ failure. Finally, the reliability of the PCR test, that is used for diagnosing COVID-19 cases as a most preferred procedure across the world, could also be regarded as an important limitation of the study since false positive and false negative results of PCR test might cause deviations both in the COVID-19 cases and models.

Conclusion and implications of findings

In conclusion, this original study reveals that health-related, as well as social, economic, climatic, and political macro indicators have been noteworthy in the COVID-19 outbreak. The findings of this study do not claim that these macro indicators cause disease or death directly because it does not conduct any clinical or laboratory research. Essentially, the present study, demonstrating that these macro indicators explain the variation in the number of COVID-19 cases and deaths across 170 countries, may serve as a basis for future clinical trials.

Additionally, if it is assumed that COVID-19 would have a lasting effect in the world for an extended period, this study might assist policymakers in developing short- and long-term health and social strategies to enhance these factors associated with COVID-19. Following the complete vaccination in nations, further research can be done to show how the vaccine affects these models.

What is already known on this topic

- On a global scale, struggling with COVID-19 is urgent, unavoidable, unworkable, and underserved in many aspects.
- Older age, obesity and hypertension are common health-related indicators for explaining both COVID-19 cumulative prevalence and mortalities.

What this study adds

- Macro-scale indicators remarkably explain the cumulative prevalence and deaths of COVID-19 for countries.
- Transparency (i.e., corruption) perception and total sunny hours of a country for COVID-19 cases whereas alcohol consumption and urbanization for COVID-19 deaths were found significant for explaining variability among countries.
- This study suggests that policy-makers should build strategies for improving these significant indicators in the long run and for similar epidemics.

Footnotes

Author Contribution: SE and SA had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: SE, SA, and FI. Acquisition, analysis, or interpretation of data: SE and FI. Drafting of the manuscript: SI and FI. Critical revision of the manuscript for important intellectual content: SE and FI. Statistical analysis: SR. Administrative, technical, or material support: FI, AB, VG, EE, SM, and AA. Supervision: SA.

Competing of interest statement: The authors declare that they have no conflict of interest.

Data sharing: The data that support the findings of this study are openly available in Open Science Framework at https://osf.io/fx3pj/?view_only=5d226a0d346a488c9492512688a85334.

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

Ethical approval: This study is registered at ClinicalTrials.gov, NCT04757870 and ethical committee approval was received from İzmir Katip Çelebi University/Non-interventional Clinical Researches Ethics Board with the 2021-GOKAE-0287 protocol number.

Public and patient involvement statement: The authors of the study plan to share the results with communities through presentations and discussions with public health leaders, some of whom have been aware of the research but were not directly involved in it.

Transparency statement: The lead authors affirm that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

REFERENCES

1 Steenhuysen J, Kelland K. 'When will it end?': How a changing virus is reshaping scientists' views on COVID-19. Reuters [Internet]. 2021, March 3. Available from: https://www.reuters.com/article/us-health-coronavirus-variants-insight-idUSKBN2AV1T1.

2 Carethers JM. Insights into disparities observed with COVID-19. J Intern Med. 2021 Apr;289(4):463-473. doi: 10.1111/joim.13199. Epub 2020 Dec 6. PMID: 33164230.

3 Wollenstein-Betech S, Cassandras CG, Paschalidis IC. Personalized predictive models for symptomatic COVID-19 patients using basic preconditions: Hospitalizations, mortality, and the need for an ICU or

ventilator. Int J Med Inform. 2020;142:104258. Epub 2020/09/15. doi: 10.1016/j.ijmedinf.2020.104258. PubMed PMID: 32927229; PubMed Central PMCID: PMCPMC7442577.

4 Ficetola GF, Rubolini D. Containment measures limit environmental effects on COVID-19 early outbreak dynamics. Sci Total Environ. 2021 Mar 20;761:144432. doi: 10.1016/j.scitotenv.2020.144432. Epub 2020 Dec 16. PMID: 33360124; PMCID: PMC7744010.

5 Arsalan M, Mubin O, Alnajjar F, Alsinglawi B. COVID-19 Global Risk: Expectation vs. Reality. Int J Environ Res Public Health. 2020 Aug 3;17(15):5592. doi: 10.3390/ijerph17155592. PMID: 32756513; PMCID: PMC7432363.

6 Su D, Chen Y, He K, Zhang T, Tan M, Zhang Y, Zhang X. Influence of socio-ecological factors on COVID-19 risk: a cross-sectional study based on 178 countries/regions worldwide. medRxiv [Preprint]. 2020 Apr 29:2020.04.23.20077545. doi: 10.1101/2020.04.23.20077545. PMID: 32511588; PMCID: PMC7276015.

7 Sorci G, Faivre B, Morand S. Explaining among-country variation in COVID-19 case fatality rate. Sci Rep. 2020 Nov 3;10(1):18909. doi: 10.1038/s41598-020-75848-2. PMID: 33144595; PMCID: PMC7609641.

8 Rahman M, Islam M, Shimanto MH, Ferdous J, Rahman AAS, Sagor PS, Chowdhury T. A global analysis on the effect of temperature, socio-economic and environmental factors on the spread and mortality rate of the COVID-19 pandemic. Environ Dev Sustain. 2020 Oct 6:1-15. doi: 10.1007/s10668-020-01028-x. Epub ahead of print. PMID: 33041644; PMCID: PMC7538192.

9 Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020;20(5):533-4. Epub 2020/02/23. doi: 10.1016/s1473-3099(20)30120-1. PubMed PMID: 32087114; PubMed Central PMCID: PMCPMC7159018.

10 WHO (World Health Organization). Prevalence of obesity among adults, BMI >= 30 (crude estimate) (%). Global Health Observatory [Internet]. 2017. Available from: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/prevalence-of-obesity-among-adults-bmi-=-30-(crude-estimate)-(-).

11 WHO (World Health Organization). Raised blood pressure (SBP ≥ 140 OR DBP ≥ 90), age-standardized (%) Estimates by country. Global Health Observatory [Internet]. 2017. Available from: https://apps.who.int/gho/data/node.main.A875STANDARD?lang=en.

12 WHO (World Health Organization). Recorded alcohol per capita consumption. Global Health Observatory [Internet]. 2020. Available from: https://apps.who.int/gho/data/node.main.A1039?lang=en.

13 WHO (World Health Organization). Raised fasting blood glucose (≥ 7.0 mmol/L or on medication) (crude estimate) Estimates by country. Global Health Observatory [Internet]. 2021. Available from: https://apps.who.int/gho/data/node.main.A869?lang=en.

14 WHO (World Health Organization). Raised total cholesterol (≥6.2 mmol/L) Data by country. Global Health Observatory [Internet]. 2021. Available from: https://apps.who.int/gho/data/view.main.2577.

15 WHO (World Health Organization). Joint effects of air pollution Data by country. Global Health Observatory [Internet]. 2018. Available from: https://apps.who.int/gho/data/node.main.ENVHEALTHJOINTAAPHAP?lang=en%0A.

16 Data.world. 2019/W44: World Cities Ranked by Annual Sunshine Hours [Internet]. 2019. Available from: https://data.world/makeovermonday/2019w44.

17 Food and Agriculture Organization of the United Nations (FAO). Vegetable supply per person kilograms per year. FAOSTAT [Internet]. 2020. Available from: https://ourworldindata.org/grapher/vegetable-consumption-per-capita?tab=table.

18 United Nations Development Programme. Global Human Development Indicators. Human Development Reports [Internet]. 2021. Available from: http://hdr.undp.org/en/content/download-data.

19 The World Bank. Urban population (% of total population). United Nations Population Division, World Urbanization Prospects [Internet]. 2021. Available from: https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS.

20 The World Bank. Population ages 65 and above (% of total population). United Nations Population Division's World Population Prospects [Internet]. 2021. Available from: https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS.

21 United Nations Development Programme. Global Human Development Indicators. Human Development Reports [Internet]. 2021. Available from: http://hdr.undp.org/en/content/download-data.

22 Food and Agriculture Organization of the United Nations (FAO). Prevalence of undernourishment (percent). FAOSTAT [Internet]. 2021. Available from: http://www.fao.org/faostat/en/#search/undernourishment.

23 Helliwell JF, Huang H, Wang S et al. World Happiness, Trust and Deaths under COVID-19. World Happiness Report 2021. Available from: https://worldhappiness.report/ed/2021/#appendices-and-data.

24 Central Intelligence Agency (CIA). Gini Index coefficient – distribution of family income. The World Factbook. 2021. Available from: https://www.cia.gov/the-world-factbook/field/gini-index-coefficient-distribution-of-family-income/.

25 The Economist Intelligence Unit (EIU). Democracy Index 2020: In sickness and in health?. 2021. Available from: https://www.eiu.com/n/campaigns/democracy-index-2020/#mktoForm_anchor.

26 Transparency International. Corruption Perceptions Index. 2021. Available from: https://www.transparency.org/en/cpi/2020/table/nzl.

27 Worldometer. Cumulative Total Deaths per Million People by 31th December 2020. Available from: https://www.worldometers.info/coronavirus/.

28 Worldometer. Cumulative Total Cases per Million People by 31th December 2020. Available from: https://www.worldometers.info/coronavirus/.

29 WHO (World Health Organization). Incidence of tuberculosis (per 100 000 population per year). Global Health Observatory [Internet]. 2020. Available from: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/incidence-of-tuberculosis-(per-100-000-population-per-year).

30 Newbold P, Carlson WL, Thorne BM. Statistics for Business and Economics. 8th ed. Pearson; 2013.

31 Graybill FA, Iyer HK. Regression Analysis: Concepts and Applications. Belmont, California. 1994.

32 Akinwande MO, Dikko HG, Samson A. Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. Open J Stat. 2015;5(07):754-67. doi: 10.4236/ojs.2015.57075.

33 Pinto BGG, Oliveira AER, Singh Y, Jimenez L, Gonçalves ANA, Ogava RLT, Creighton R, Schatzmann Peron JP, Nakaya HI. ACE2 Expression Is Increased in the Lungs of Patients With Comorbidities Associated With

Severe COVID-19. J Infect Dis. 2020 Jul 23;222(4):556-563. doi: 10.1093/infdis/jiaa332. PMID: 32526012; PMCID: PMC7377288.

34 Radzikowska U, Ding M, Tan G, Zhakparov D, Peng Y, Wawrzyniak P, Wang M, Li S, Morita H, Altunbulakli C, Reiger M, Neumann AU, Lunjani N, Traidl-Hoffmann C, Nadeau KC, O'Mahony L, Akdis C, Sokolowska M. Distribution of ACE2, CD147, CD26, and other SARS-CoV-2 associated molecules in tissues and immune cells in health and in asthma, COPD, obesity, hypertension, and COVID-19 risk factors. Allergy. 2020 Nov;75(11):2829-2845. doi: 10.1111/all.14429. Epub 2020 Aug 24. PMID: 32496587; PMCID: PMC7300910.

35 Wang B, Li R, Lu Z, Huang Y. Does comorbidity increase the risk of patients with COVID-19: evidence from meta-analysis. Aging (Albany NY). 2020 Apr 8;12(7):6049-6057. doi: 10.18632/aging.103000. Epub 2020 Apr 8. PMID: 32267833; PMCID: PMC7185114.

36 Wang J, Sato T, Sakuraba A. Coronavirus Disease 2019 (COVID-19) Meets Obesity: Strong Association between the Global Overweight Population and COVID-19 Mortality. J Nutr. 2021 Jan 4;151(1):9-10. doi: 10.1093/jn/nxaa375. PMID: 33188432; PMCID: PMC7890648.

37 Azuma K, Kagi N, Kim H, Hayashi M. Impact of climate and ambient air pollution on the epidemic growth during COVID-19 outbreak in Japan. Environ Res. 2020 Nov; 190:110042. doi: 10.1016/j.envres.2020.110042. Epub 2020 Aug 12. PMID: 32800895; PMCID: PMC7420955.

38 Menebo MM. Temperature and precipitation associate with Covid-19 new daily cases: A correlation study between weather and Covid-19 pandemic in Oslo, Norway. Sci Total Environ. 2020 Oct 1;737:139659. doi: 10.1016/j.scitotenv.2020.139659. Epub 2020 May 29. PMID: 32492607; PMCID: PMC7258804.

39 Paez A, Lopez FA, Menezes T, Cavalcanti R, Pitta MGDR. A Spatio-Temporal Analysis of the Environmental Correlates of COVID-19 Incidence in Spain. Geogr Anal. 2020 Jun 8:10.1111/gean.12241. doi: 10.1111/gean.12241. Epub ahead of print. PMID: 32836331; PMCID: PMC7300768.

40 Biasin M, Bianco A, Pareschi G, Cavalleri A, Cavatorta C, Fenizia C, Galli P, Lessio L, Lualdi M, Tombetti E, Ambrosi A, Redaelli EMA, Saulle I, Trabattoni D, Zanutta A, Clerici M. UV-C irradiation is highly effective in inactivating SARS-CoV-2 replication. Sci Rep. 2021 Mar 18;11(1):6260. doi: 10.1038/s41598-021-85425-w. PMID: 33737536; PMCID: PMC7973506.

41 Coccia M. Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. Sci Total Environ. 2020 Aug 10;729:138474. doi: 10.1016/j.scitotenv.2020.138474. Epub 2020 Apr 20. PMID: 32498152; PMCID: PMC7169901.

42 Merow C, Urban MC. Seasonality and uncertainty in global COVID-19 growth rates. Proc Natl Acad Sci U S A. 2020 Nov 3;117(44):27456-27464. doi: 10.1073/pnas.2008590117. Epub 2020 Oct 13. PMID: 33051302; PMCID: PMC7959558.

43 Chaudhry R, Dranitsaris G, Mubashir T, Bartoszko J, Riazi S. A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes. EClinicalMedicine. 2020 Aug;25:100464. doi: 10.1016/j.eclinm.2020.100464. Epub 2020 Jul 21. PMID: 32838237; PMCID: PMC7372278.

44 Farzanegan MR. The Effect of Public Corruption on COVID-19 Fatality Rate: A Cross-Country Examination. CESifo Working Paper No. 8938. 2021. Available at SSRN: https://ssrn.com/abstract=3805464.

45 Cartaxo ANS, Barbosa FIC, de Souza Bermejo PH, Moreira MF, Prata DN. The exposure risk to COVID-19 in most affected countries: A vulnerability assessment model. PLoS One. 2021 Mar 4;16(3):e0248075. doi: 10.1371/journal.pone.0248075. PMID: 33662028; PMCID: PMC7932136.

46 Bilal B, Saleem F, Fatima SS. Alcohol consumption and obesity: The hidden scare with COVID-19 severity. Med Hypotheses. 2020 Nov;144:110272. doi: 10.1016/j.mehy.2020.110272. Epub 2020 Sep 16. PMID: 33254576; PMCID: PMC7492811.

47 Pilotto A, Odolini S, Masciocchi S, Comelli A, Volonghi I, Gazzina S, Nocivelli S, Pezzini A, Focà E, Caruso A, Leonardi M, Pasolini MP, Gasparotti R, Castelli F, Ashton NJ, Blennow K, Zetterberg H, Padovani A. Steroid-Responsive Encephalitis in Coronavirus Disease 2019. Ann Neurol. 2020 Aug;88(2):423-427. doi: 10.1002/ana.25783. Epub 2020 Jun 9. PMID: 32418288; PMCID: PMC7276848.

48 Casillo GM, Mansour AA, Raucci F, Saviano A, Mascolo N, Iqbal AJ, Maione F. Could IL-17 represent a new therapeutic target for the treatment and/or management of COVID-19-related respiratory syndrome? Pharmacol Res. 2020 Jun;156:104791. doi: 10.1016/j.phrs.2020.104791. Epub 2020 Apr 14. PMID: 32302707; PMCID: PMC7194796.

49 Althobaiti YS, Alzahrani MA, Alsharif NA, Alrobaie NS, Alsaab HO, Uddin MN. The Possible Relationship between the Abuse of Tobacco, Opioid, or Alcohol with COVID-19. Healthcare (Basel). 2020 Dec 22;9(1):2. doi: 10.3390/healthcare9010002. PMID: 33375144; PMCID: PMC7822153.

50 Lassen MCH, Skaarup KG, Sengeløv M, Iversen K, Ulrik CS, Jensen JUS, Biering-Sørensen T. Alcohol Consumption and the Risk of Acute Respiratory Distress Syndrome in COVID-19. Ann Am Thorac Soc. 2020 Dec 14. doi: 10.1513/AnnalsATS.202008-988RL. Epub ahead of print. PMID: 33315543.

51 Ahmadi M, Sharifi A, Dorosti S, Jafarzadeh Ghoushchi S, Ghanbari N. Investigation of effective climatology parameters on COVID-19 outbreak in Iran. Sci Total Environ. 2020 Aug 10;729:138705. doi: 10.1016/j.scitotenv.2020.138705. Epub 2020 Apr 17. PMID: 32361432; PMCID: PMC7162759.

52 Kodera S, Rashed EA, Hirata A. Correlation between COVID-19 Morbidity and Mortality Rates in Japan and Local Population Density, Temperature, and Absolute Humidity. Int J Environ Res Public Health. 2020 Jul 29;17(15):5477. doi: 10.3390/ijerph17155477. PMID: 32751311; PMCID: PMC7432122.

53 Squalli J. Evaluating the determinants of COVID-19 mortality: A cross-country study. medRxiv 2020.05.12.20099093; doi: https://doi.org/10.1101/2020.05.12.20099093.

54 Coelho FC, Lana RM, Cruz OG, Villela DAM, Bastos LS, Pastore Y Piontti A, Davis JT, Vespignani A, Codeço CT, Gomes MFC. Assessing the spread of COVID-19 in Brazil: Mobility, morbidity and social vulnerability. PLoS One. 2020 Sep 18;15(9):e0238214. doi: 10.1371/journal.pone.0238214. PMID: 32946442; PMCID: PMC7500629.

55 Hawkins RB, Charles EJ, Mehaffey JH. Socio-economic status and COVID-19-related cases and fatalities. Public Health. 2020 Dec;189:129-134. doi: 10.1016/j.puhe.2020.09.016. Epub 2020 Oct 17. PMID: 33227595; PMCID: PMC7568122.

56 Wollenstein-Betech S, Silva AAB, Fleck JL, Cassandras CG, Paschalidis IC. Physiological and socioeconomic characteristics predict COVID-19 mortality and resource utilization in Brazil. PLoS One. 2020 Oct 14;15(10):e0240346. doi: 10.1371/journal.pone.0240346. PMID: 33052960; PMCID: PMC7556459.

57 Martins LD, da Silva I, Batista WV, Andrade MF, Freitas ED, Martins JA. How socio-economic and atmospheric variables impact COVID-19 and influenza outbreaks in tropical and subtropical regions of Brazil. Environ Res. 2020 Dec;191:110184. doi: 10.1016/j.envres.2020.110184. Epub 2020 Sep 16. PMID: 32946893; PMCID: PMC7492183.

58 Dos Santos DT, Seles Alves L, Moraes Arcoverde MA, Arroyo LH, Zamboni Berra T, Ramos ACV, Lima Dos Santos F, Arcêncio RA, Nunes C. Social risk and its association with tuberculosis mortality in a context of high inequality in South Brazil: A geo-epidemiology analysis. J Infect Public Health. 2020 Aug;13(8):1148-1155. doi: 10.1016/j.jiph.2020.03.010. Epub 2020 Apr 12. PMID: 32295755.

59 Lönnroth K, Jaramillo E, Williams BG, Dye C, Raviglione M. Drivers of tuberculosis epidemics: the role of risk factors and social determinants. Soc Sci Med. 2009 Jun;68(12):2240-6. doi: 10.1016/j.socscimed.2009.03.041. Epub 2009 Apr 23. PMID: 19394122.

60 Murray M, Oxlade O, Lin HH. Modeling social, environmental and biological determinants of tuberculosis. Int J Tuberc Lung Dis. 2011 Jun;15 Suppl 2:64-70. doi: 10.5588/ijtld.10.0535. PMID: 21740661.

Table 1. The list and the definitions of indicators and the outcome variable used in the analysis

Factors	Variables	Description	Year	N	Mean±Standard Deviation	Min	Max
Health-Related	Obesity[10]	Obesity (defined as of Body Mass Index [BMI]>= 30, crude estimate %) prevalence in ≥ 18 years old people	2016	170	10.59±5.84	2.10	37.30
	Hypertension[11]	High Blood Pressure (Systolic Blood Pressure [SBP] ≥ 140) or (Diastolic Blood Pressure [DBP] ≥ 90) Prevalence in 18 ≥ Years Old People	2015	170	23.45±5.31	12.60	41.00
	Alcohol[12]	Annual Alcohol Consumption Per Capita (Liter) in ≥ 15 Years Old People	2018	170	4.96±3.87	0.00	20.05
	Diabetes[13]	Diabetes, raised fasting blood glucose (Millimoles Per Liter [mmol/L] ≥ 7.0) Prevalence in ≥ 18 Years Old People	2014	170	7.70±3.99	1.00	22.10
	Cholesterol[14]	Hypercholesterolemia, raised total Cholesterol (Millimoles Per Liter [mmol/L] ≥ 6.2) prevalence in ≥25Years Old People	2008	170	10.59±5.84	2.50	29.10
	AirPollution[15]	Mortality attributable to joint effects of household and ambient air pollution in adults such as acute respiratory infections, cerebrovascular diseases, ischemic heart diseases, chronic obstructive pulmonary disease, and lung cancer in adults (Ambient and household air pollution attributable death rate (per 100 000 population)	2016	169	10.59±5.84	9.00	184.00
Climatic	Sunshine[16]	Total Sunshine Hours (Duration of sunlight in a year for countries)	2019	170	2519.18±576.3	1177.00	3737.00
Socio Economic	Vegetable[17]	Annual Vegetable Consumption for countries Kg/Per Capita	2017	154	91.93±69.73	6.28	377.17

	HDI[18]	Human Development Index. A measure prepared in line with the life span, literacy rate, education and life level for countries around the world (Rate between 0.800 - 1000 is very high, rate between 0.350 - 0.549 low)	2020	169	727.75±151.3	394.00	957.00
	Urbanization[19]	% of population living in urban areas for Countries around the world	2019	169	60.31±22.11	13.25	100.00
	Age[20]	Elderly people ratio whose age 65 & over for countries around the world	2019	168	9.22±6.5	1.16	28.00
	Schooling[21]	Average number of completed years of education of a countries' population aged 25 and higher around the world	2020	169	8.44±3.12	1.40	13.40
	Undernourishment[22]	The proportion of people who are malnourished as a percentage of the population that reflects the share of the population with insufficient caloric intake around the world	2019	170	0.11±0.12	0.01	0.60
	Happiness[23]	Happiness Index that shows life satisfaction in different nations, nationally representative samples of respondents are asked to imagine a ladder, (Rate between 10 representing the best possible life and a 0 representing the worst possible life.)	2019	139	5.55±1.09	2.52	7.84
	GINI[24]	Gini Inequality Index measures whether the distribution of national income in a country is equal or not. The coefficient takes values between 0 and 1, and higher values correspond to greater inequality.	2019	170	38.86±8.04	25.00	63.00
Political	Democracy[25]	Index that measures the state of democracy in 167 countries. It takes values between 0 and 10. 10 represents the perfect democracy and 0 represents the dictator regime.	2020	155	5.52±2.18	1.43	9.81

	Transparency[26]	Transparency measured by the Corruption Perception Index for each country, scored between 100 represents very clean and 0 represents highly corrupt.	2020	166	43.92±18.74	12.00	88.00
Target Variables	TotalDeath-COV19[9, 27]	Cumulative Total Deaths per Million People by 31th December 2020	2020	170	294.91±375.79	0.00	1673.00
	TotalCase-COV19[9, 28]	Cumulative Total Cases per Million People by 31th December 2020	2020	170	16874.29±20449.21	4.00	11568.00
Control Variable	TBDeath[30]	Deaths Related with Tuberculosis 100.000 population per year	2020	170	101.79±135.30	0.40	611.00

Table 2. Parameter testing results* for Linear Regression for Predicting "Total Cases Per Million People"

	Standardized Coefficients	95.0% Confidence Interval for β		Collinearity Statistics	p-value
	Beta	Lower Bound	Upper Bound	VIF	
Transparency	0.147	0.012	0.264	1.363	0.032
Obesity	0.460	0.311	0.547	1.213	0.001
Sunshine	-0.157	-0.266	-0.026	1.266	0.018
Hypertension	0.214	0.085	0.312	1.134	0.001

*R^2=0.449, Se=0.70, Durbin-Watson=2.01

*TotalCase-COV19= 0.214*Hypertension+ 0.460*Obesity-0.157*Sunshine+0.147*Transparency

Table 3. Parameter testing results* for Linear Regression for Predicting "Total Deaths Per Million People"

	Standardized Coefficients	95.0% Confidence Interval for β		Collinearity Statistics	p-value
	Beta	Lower Bound	Upper Bound	VIF	
Alcohol	0.173	0.030	0.315	1.461	0.018
Urban	0.204	0.040	0.368	1.920	0.015
Obesity	0.279	0.116	0.442	1.915	0.001
Hypertension	0.285	0.146	0.423	1.380	0.000

*R^2=0.409, Se=0.78, Durbin-Watson=2.29

*TotalDeath-COV19= 0.204*Urban+0.285*Hypertension+0.173*Alcohol+0.279*Obesity

Table 4. Parameter testing results* for Linear Regression for Predicting "Deaths Due to Tuberculosis"

	Standardized Coefficients	95.0% Confidence Interval for β		Collinearity Statistics	p-value
	Beta	Lower Bound	Upper Bound	VIF	
Obesity	-0.246	-0.427	-0.085	2.129	0.004
GINI	0.23	0.096	0.389	1.528	0.001
AirPollution	0.284	0.148	0.44	1.566	0.001
Undernourishment	0.245	0.084	0.402	2.014	0.003
Age	-0.453	-0.690	-0.224	4.193	0.001
Democracy	0.280	0.110	0.477	2.399	0.002
Schooling	0.253	0.048	0.459	3.303	0.016

*R^2=0.535, Se=0.72, Durbin-Watson=1.79

*TBDeath=-0.246*Obesity+0.230*GINI+0.284*AirPollution+0.245*Undernourishment-0.453*Age+0.280*Democracy+0.253*Schooling

Supplementary Table A. Best Subset Models for Total Cases Per Million

Number of Variables	R^2	R^2 (adjusted)	R^2 (pred)	Mallows Cp	S	Age	Happiness	Gini	Transparency	Democracy	Urbanization	Schooling	AirPollution	HDI	Vegetable	SunShine	Undernourishment	Obesity	Cholesterol	Hypertension	Alcohol	Diabet
1	38.5	38.1	36.1	16.0	0.78	X																
1	38.5	38	36.7	16.1	0.78									X								
2	46.5	45.6	43.5	-0.4	0.73	X												X				
2	44.5	43.6	41.6	4.1	0.74									X					X			
3	48.1	46.8	44.3	-2.1	0.72	X									X			X				
3	47.9	46.7	44.4	-1.7	0.72									X				X	X			
4	49.5	47.9	45.2	-3.6	0.71									X	X			X	X			
4	49.2	47.6	44.2	-2.7	0.71	X									X			X	X			
5	50.0	47.9	45.1	-2.5	0.71					X				X	X			X	X			
5	49.9	47.9	45.0	-2.4	0.71									X	X			X	X		X	
6	50.4	48.0	44.5	-1.6	0.71					X				X	X			X	X	X		
6	50.4	48.0	44.5	-1.5	0.71									X	X			X	X	X	X	
7	50.7	47.9	44.3	-0.4	0.71					X				X	X			X	X	X	X	
7	50.6	47.8	44.0	-0.1	0.71					X	X			X	X			X	X	X		
8	50.9	47.7	43.8	1.2	0.71					X	X			X	X			X	X	X	X	
8	50.9	47.6	43.8	1.4	0.71					X			X	X	X			X	X	X	X	
9	51.1	47.4	43.3	2.8	0.72					X	X		X	X	X			X	X	X	X	
9	51.0	47.3	42.6	3.0	0.72				X	X	X			X	X			X	X	X	X	
10	51.2	47.1	42.2	4.6	0.72				X	X	X		X	X	X			X	X	X	X	
10	51.1	47.1	42.7	4.7	0.72			X		X	X		X	X	X			X	X	X	X	
11	51.2	46.7	41.5	6.5	0.72		X		X	X	X		X	X	X			X	X	X	X	
11	51.2	46.7	41.3	6.5	0.72				X	X	X		X	X	X			X	X	X	X	X
12	51.3	46.3	40.7	8.4	0.72		X		X	X	X		X	X	X			X	X	X	X	X
12	51.3	46.3	40.2	8.4	0.72	X	X		X	X	X		X	X	X			X	X	X	X	
13	51.3	45.9	40.0	10.3	0.73		X	X	X	X	X		X	X	X			X	X	X	X	X
13	51.3	45.9	39.6	10.3	0.73	X	X	X	X	X	X		X	X	X			X	X	X	X	
14	51.4	45.5	38.6	12.2	0.73	X	X	X	X	X	X		X	X	X			X	X	X	X	X
14	51.4	45.5	38.6	12.2	0.73		X	X	X	X	X		X	X	X	X		X	X	X	X	X
15	51.4	45.1	37.7	14.1	0.73	X		X	X	X	X	X	X	X	X		X	X	X	X	X	X
15	51.4	45.1	37.4	14.1	0.73	X	X	X	X	X	X	X	X	X	X			X	X	X	X	X
16	51.4	44.6	37.0	16.0	0.73	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X
16	51.4	44.6	35.7	16.0	0.73	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X
17	51.4	44.1	35.3	18.0	0.74	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

139 cases used; 31 cases contain missing values.

Supplementary Table B. Best Subset Models for Total Deaths Per Million

Number of Variables	R^2	R^2 (adj)	R^2 (pred)	Mallows Cp	S	Age	Happiness	Gini	Transparency	Democracy	Urbanization	Schooling	AirPollution	HDI	Vegetable	SunShine	Undernourishment	Obesity	Cholesterol	Hypertension	Alcohol	Diabet
1	35.6	35.1	33.2	28.1	0.86	X																
1	30.5	30.0	28.6	40.3	0.89									X								
2	42.3	41.4	39.3	13.8	0.81	X												X				
2	41.4	40.5	38.9	16.0	0.82											X		X				
3	45.6	44.3	41.9	7.9	0.79											X		X		X		
3	45.4	44.1	41.8	8.4	0.79	X										X		X				
4	48.4	46.7	43.8	3.3	0.78	X			X							X		X				
4	46.9	45.2	42.6	6.8	0.79			X								X		X		X		
5	49.7	47.6	44.5	2.2	0.77	X		X	X							X		X				
5	49.1	47.1	43.9	3.4	0.77	X			X	X						X		X				
6	50.5	48.1	45.0	2.1	0.77	X		X	X							X	X	X				
6	50.3	47.9	44.1	2.6	0.77	X		X	X							X		X		X		
7	51.2	48.4	44.7	2.5	0.76	X		X	X							X	X	X		X		
7	51.1	48.3	44.8	2.7	0.76	X		X	X		X					X	X	X				
8	51.6	48.4	45.2	3.5	0.76	X		X	X			X				X	X	X			X	
8	51.5	48.3	44.4	3.7	0.76	X		X	X	X						X	X	X		X		
9	52.0	48.4	44.8	4.6	0.76	X		X	X		X	X				X	X	X			X	
9	51.9	48.3	44.5	4.8	0.76	X		X	X			X				X	X	X		X	X	
10	52.3	48.3	44.4	5.9	0.76	X		X	X	X	X	X				X	X	X			X	
10	52.2	48.2	44.1	6.1	0.76	X		X	X	X		X				X	X	X		X	X	
11	52.5	48.1	44.1	7.4	0.77	X		X	X	X	X	X		X		X	X	X			X	
11	52.4	48.0	43.8	7.5	0.77	X		X	X	X		X		X		X	X	X		X	X	
12	52.7	47.9	43.4	8.8	0.77	X		X	X	X	X	X		X		X	X	X		X	X	
12	52.6	47.8	42.8	9.0	0.77	X		X	X	X		X		X	X	X	X	X		X	X	
13	52.9	47.6	42.4	10.5	0.77	X		X	X	X	X	X		X	X	X	X	X		X	X	
13	52.8	47.6	42.5	10.5	0.77	X		X	X	X	X	X		X		X	X	X		X	X	X
14	53.0	47.3	41.4	12.2	0.77	X		X	X	X	X	X		X	X	X	X	X		X	X	X
14	52.9	47.2	41.1	12.4	0.77	X		X	X	X	X	X		X	X	X	X	X	X	X	X	
15	53.0	46.9	40.1	14.1	0.77	X		X	X	X	X	X		X	X	X	X	X	X	X	X	X
15	53.0	46.8	40.8	14.2	0.77	X	X	X	X	X	X	X		X	X	X	X	X		X	X	X
16	53.0	46.5	39.5	16.0	0.78	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X
16	53.0	46.4	39.4	16.1	0.78	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
17	53.1	46.0	38.8	18.0	0.78	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

139 cases used; 31 cases contain missing values.

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5
Bias	9	Describe any efforts to address potential sources of bias	NA
Study size	10	Explain how the study size was arrived at	5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5-7
		(*b*) Describe any methods used to examine subgroups and interactions	5-7
		(*c*) Explain how missing data were addressed	5-7
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	NA
		(*e*) Describe any sensitivity analyses	NA
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7-8
		(b) Give reasons for non-participation at each stage	7-8
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7-8
		(b) Indicate number of participants with missing data for each variable of interest	7-8
Outcome data	15*	Report numbers of outcome events or summary measures	7-8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	7-8

		(*b*) Report category boundaries when continuous variables were categorized	7-8
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	7-8
Discussion			
Key results	18	Summarise key results with reference to study objectives	8-9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	8-11
Generalisability	21	Discuss the generalisability (external validity) of the study results	8-12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	NA

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Investigating the Effect of Macro-Scale Estimators on Worldwide COVID-19 Occurrence and Mortality through Regression Analysis Using Online Country-Based Data Sources

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-055562.R1
Article Type:	Original research
Date Submitted by the Author:	27-Dec-2021
Complete List of Authors:	Erdem, Sabri; Dokuz Eylül University, Department of Business Administration, İzmir/Turkey Ipek, Fulya; Hacettepe University, Faculty of Physical Therapy and Rehabilitation Bars, Aybars; Dokuz Eylül University, Social Sciences Institute Genç, Volkan; Dokuz Eylul Universitesi, Social Science Institute Erpek, Esra; Izmir Katip Celebi University, Department of Internal Medicine, Division of Rheumatology Atatürk Education and Research Hospital Mohammadi, Shabnam; Dokuz Eylul Universitesi, Social Science Institute Altınata, Anıl; Dokuz Eylül University, Social Sciences Institute Akar, Servet; Izmir Katip Celebi University, Department of Internal Medicine, Division of Rheumatology Atatürk Education and Research Hospital
Primary Subject Heading:	Global health
Secondary Subject Heading:	Epidemiology, Health policy, Infectious diseases, Public health
Keywords:	COVID-19, Public health < INFECTIOUS DISEASES, Health policy < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Investigating the Effect of Macro-Scale Estimators on Worldwide COVID-19 Occurrence and Mortality through Regression Analysis Using Online Country-Based Data Sources

Running title: COVID-19 Global Analysis

Sabri Erdem, PHD, Professor

Affiliation: Dokuz Eylül University, Faculty of Business, Department of Business Administration, İzmir/Turkey

E-mail: sabri.erdem@deu.edu.tr

ORCID ID: 0000-0001-6766-3202

Fulya Ipek, PT, MSc (Corresponding Author)

Affiliation: Hacettepe University, Faculty of Physical Therapy and Rehabilitation, Ankara, Turkey

Email: fulya.ipek@hacettepe.edu.tr

ORCID ID: 0000-0001-7606-042X

Aybars Bars, MSc

Affiliation: Dokuz Eylül University, Social Sciences Institute, İzmir/Turkey

E-mail: aybars.bars@ogr.deu.edu.tr

ORCID ID: 0000-0002-8051-9165

Volkan Genç, MSc

Affiliation: Dokuz Eylül University, Social Sciences Institute, İzmir/Turkey

E-mail: volkan.genc@ogr.deu.tr

ORCID ID: 0000-0003-2184-482X

Esra Erpek, MD

Affilation: İzmir Katip Çelebi University, Department of Internal Medicine, Division of Rheumatology Atatürk Education and Research Hospital, İzmir/Turkey

E-mail: esraerpek@gmail.com

ORCID ID: 0000-0003-3540-4905

Shabnam Mohammadi, MSc

Affilation: Dokuz Eylül University, Social Sciences Institute, İzmir/Turkey

E-mail: mohammadi.shabnam@ogr.deu.edu.tr

ORCID ID: 0000-0001-6987-0301

Anıl Altınata, MSc

Affilation: Dokuz Eylül University, Social Sciences Institute, İzmir/Turkey

E-mail: altinata.anil@ogr.deu.edu.tr

ORCID ID: 0000-0001-6300-5982

Servet Akar, MD, Professor

Affilation: İzmir Katip Çelebi University, Department of Internal Medicine, Division of Rheumatology Atatürk Education and Research Hospital, İzmir/Turkey

E-mail: servet.akar@gmail.com

ORCID ID: 0000-0002-3734-1242

Fulya Ipek, PT, MSc (Corresponding Author)

Affiliation: Hacettepe University, Faculty of Physical Therapy and Rehabilitation, Ankara, Turkey

Address: Hacettepe University, Faculty of Physical Therapy and Rehabilitation, 06100, Altındağ/Ankara, Turkey

Email: fulya.ipek@hacettepe.edu.tr

fulyaipek.92@gmail.com

Phone: +90 539 272 20 22

Abstract

Objective: To investigate macro-scale estimators of the variations in COVID-19 cases and deaths among countries.

Design: Epidemiological study.

Setting: Country-based data from publicly available online databases of international organizations.

Participants: The study involved 170 countries/territories, each of which had complete COVID-19 and tuberculosis data, as well as specific health-related estimators (obesity, hypertension, diabetes, and hypercholesterolemia).

Primary and secondary outcome measures: The worldwide heterogeneity of the total number of COVID-19 cases and deaths per million on 31 December 2020 was analyzed by 17 macro-scale estimators around the health-related, socio-economic, climatic, and political factors. In 139 of 170 nations, the Best Subsets Regression was used to investigate all potential models of COVID-19 variations among countries. A multiple linear regression analysis was conducted to explore the predictive capacity of these variables. The same analysis was applied to the number of deaths per hundred thousand due to tuberculosis, a quite different infectious disease, to validate and control the differences with the proposed models for COVID-19.

Results: In the model for the COVID-19 cases (R^2=0.45); obesity (β=0.460), hypertension (β=0.214), sunshine (β=-0.157), and transparency (β=0.147), whereas in the model for COVID-19 deaths (R^2=0.41); obesity (β=0.279), hypertension (β=0.285), alcohol consumption (β=0.173), and urbanization (β=0.204) were significant factors ($p<0.05$). Unlike COVID-19, the tuberculosis model contained significant indicators like obesity, undernourishment, air pollution, age, schooling, democracy, and GINI index.

Conclusions: This study recommends the new predictors explaining the global variability of COVID-19. Thus, it might assist policy-makers in developing health policies and social strategies to deal with COVID-19.

Trial Registration: ClinicalTrials.gov Identifier: NCT04486508

Strengths and limitations of this study

- This study might represent recent macro-scale data of 170 countries and Covid-19 cases and deaths for a one year.
- It is free from the vaccine effect.
- It contains a reliable predictive model incorporating social, economic, and political indicators as well as health-related ones.
- Some macro-scale estimators were not recently published but we assume that there are no remarkable changes annually.
- There might be some interruption or reliability issues in Covid-19 Cases and Deaths data flow from countries.

INTRODUCTION

The novel coronavirus disease 2019 (COVID-19) caused by Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) has spread to countries worldwide. On 11 March 2020, the World Health Organization (WHO) declared the condition as a public health emergency. With new variants emerging, experts now assume that SARS-CoV-2 would not only remain an endemic virus but also would continue to circulate in communities and would make a massive burden of disease and death for years.[1]

The cumulative incidence (total number of cases per million) of COVID-19 (TotalCase-COV19) and mortality (total number of deaths per million) of COVID-19 (TotalDeath-COV19) vary significantly among the countries. Besides, old age, diabetes, high blood pressure, obesity, pregnancy, immunosuppression, cancer, and cardiovascular, respiratory (asthma and chronic obstructive pulmonary disease, etc.), and chronic kidney disease are all well-known health-related factors that contribute to COVID-19 morbidity and mortality.[2 3] Aside from these clinical factors and comorbidities affecting infection progression, there might be additional demographic, social, economic, and environmental factors at the macro-level differentiating countries in terms of TotalCase-COV19 and TotalDeath-COV19.

The term “Macro-level data'' (e.g., Human Development Index 2020 by United Nations) could be defined as periodic (e.g., weekly, monthly, quarterly, or annual) and long-run data that gathered, summarized, and published by the government, private, public, national and international organizations about a specific subject (e.g., health, energy, social, economic, politic) in the world, continents, regions, countries, cities or territories. This kind of data shows the current position and trend of any subject in a comparable manner and guides the policymakers in countries and organizations for setting or updating their strategies and helps them take an action around that subject. Although healthcare executives and regulatory authorities have implemented regulations and recommendations such as wearing masks, social distancing, staying at home, isolation, and lockdowns to limit the spread of SARS-CoV-2, considering these macro indicators will help steer national and international health policymakers to combat the COVID-19 pandemic.

To the best of our knowledge, there has been limited information on how the nature and dynamics of SARS-CoV-2 vary across the world.[4-8] These studies have revealed that demographic, climatic, environmental, socio-economic, and political indicators could help predict or clarify variations in TotalCase-COV19 and TotalDeath-COV19 among countries. In particular, the elderly ratio, the prevalence of comorbidities, population size, and health expenditure were shown to have been associated with the incidence and mortality of COVID-19

at various levels according to the aforementioned studies. However, these findings have been based on the early phase of the outbreak. The current study has been based on data (the cumulative number of cases and deaths) just before starting vaccinations around the world. Additionally, the findings might provide more consistent details about the progression of the disease due to covering data for a longer period. Therefore, the current study aims to identify health-related and socio-economic macro-scale indicators that could explain some of the substantial variations in TotalCase-COV19 and TotalDeath-COV19 worldwide.

METHODS

Publicly accessible data about TotalCase-COV19 and TotalDeath-COV19 were obtained from the John Hopkins University – Center for Science and Engineering (JHU-CSSE).[9] While the process of the pandemic was still ongoing mass immunization has started in several countries as of December 2019. To eliminate the effect of immunization, in this study, TotalCase-COV19 and TotalDeath-COV19 data were censored by 31 December 2020, just before vaccination has not yet become widespread. Our dataset is available in Open Science Framework. Health-related indicators from the World Health Organization and socio-economic, climatic, and political indicators from United Nations UN, World Inequality Database (WID), and Worldbank explaining variability in TotalCase-COV19 and TotalDeath-COV19 among countries were summarized in Table 1. Our dataset involves 17 macro-scale variables of 170 countries.

[Table 1 is near here]

Statistical analyses

In this study, we conducted multiple linear regression analyses to explain variations in our target variables (i.e., TotalCase-COV19 and TotalDeath-COV19). Before the analysis, we tested all the assumptions of multiple linear regression analysis by:

- drawing “actual target variable vs predicted target value scatter” plot diagram for testing linearity,
- calculating standardized residual values for each case for detecting outlier as outside ∓ 3,
- drawing a partial regression plot diagram for each predictor vs target variable for detecting the variance of predictors,
- calculating variance inflation factor (VIF) for each predictor variable in the model for testing multicollinearity,
- calculating Durbin-Watson test statistics for detecting autocorrelation of consequent errors,

- drawing "regression predicted value vs regression standardized residual" plot diagram for detecting heteroscedasticity and
- drawing "observed cumulative probability vs expected cumulative probability (P-P)" plot diagram of standardized regression residual for testing normality of residuals.

For eliminating the variance inflation and suppressing potential interactions among predictors, we have transformed all variables into the standardized form [(x-mean(x))/standard deviation(x).], so that we could get a more acceptable variance inflation factor (VIF=1/(1-R2) <2.0) values where R2 is obtained by regression analysis in that the inspecting variable is regarded as a dependent variable and others are regarded as its predictor. We have applied Durbin-Watson Statistics to test the autocorrelation in consequent error terms in the model and we have found all error terms are independent in all models since test statistics were between lower critical value and (4-lower critical value).[10] After having inspected the plot diagrams and histograms, it could be said that all assumptions of multiple linear regression were almost met with the help of using standardized values of all predictors and target variables in all models. The exception for slight violation was observed as variances in error terms in P-P plots and predicted values against observed values scatter plot for homoscedasticity that could be questionable if other test results were unsatisfactory.

We conducted multiple linear regression analyses for different situations/scenarios and created predictive models to determine indicators aligned with these scenarios. After having tested the dataset against the assumptions of multiple linear regression, we investigate the dataset using some considerations for each scenario/model through a best-subset approach that analyzes data in terms of a predictors subset that could create applicable models that illuminates the next step for finding the best one. We could decide the better one by inspecting each model in terms of cost-benefit ratios such as R2, Std. Error or Mallows' Cp (min is preferred), AIC (Akaike Information Criterion: min is preferred), and BIC (Bayesian Info Criterion: min is preferred).[11]

After determining the incorporated variables in a model, we applied a hierarchical regression model with forward selection and stepwise methods. There is no constant in our models because they were all insignificant statistically. Therefore, they were discarded. Using the information from best-subset, literature, and best practices; we tried a variety of models by evaluating the usage of intercept, handling missing values, handling outliers, applying the best technique for regression analysis (e.g., enter, stepwise, forward/backward selection), determining thresholds for alpha-to-enter (0.05 through 0.15) and remove (0.05 through 0.15) for

accepting or rejecting the next variable to the incorporated model. We utilized the SPSS 24 and Minitab 19-trial version for analyzing the data.

RESULTS

The descriptive statistics of all variables were provided in Table 1. We have considered possible different models, explaining the TotalCase-COV19 and TotalDeath-COV19, including both common and different estimators that could explain the variations in countries. Therefore, we have listed all possible significant models via the best subset method widely used in regression analyses for TotalCase-COV19 and TotalDeath-COV19 are demonstrated in Supplementary Table A and Table B.

Based on best-subset analysis of TotalCase-COV19 and TotalDeath-COV19 modeling, the indicators "Sunshine", "Obesity", "Hypertension", "Urbanization", "Schooling", "Alcohol", "Democracy", "Transparency", and "HDI" are commonly found significant for the most of models. Additionally, for modeling the "TotalDeath-COV19", "Age", "GINI" and "Undernourishment" were found significant; and for modeling "TotalCase-COV19", "Cholesterol" was found significant. However, it was remarkable that the best-subset model shows the results for only 139 countries with complete data for all variables at issue. That is why the recommended model could not be significant as in best-subset because of missing values and their handling strategies. In all these regression analyses, we found that regression model fit and all parameter fits are significant ($p<0.05$).

Urbanization, schooling, HDI, transparency, democracy, and the GINI index are all interrelated factors, despite the lack of their multicollinearity. Thus, countries with desirable socioeconomic conditions are likely to have high levels of urbanization, education, human development, transparency, and democracy, as well as low-income inequality.

Table 2 shows the results of the regression model indicating the determinants of the TotalCase-COV19 worldwide. Here is the regression model indicating that transparency, sunshine, and obesity, and hypertension in countries could help to explain the variability of the TotalCase-COV19 across countries. In this regression analysis R2 value shows that the model with these four indicators explains 44.9% of all variability. As a result, the generated model is acceptable and comes out with expected roles of variables as aligned with the literature.

[Table 2 is near here]

Table 3 shows the results of the regression model indicating the determinants of the TotalDeath-COV19 worldwide. The regression model revealed that urbanization and alcohol in addition to obesity and hypertension might help to explain the variation in the TotalDeath-COV19 across countries. In this regression analysis, the R2

value shows that the model with these four predictors explains 40.9% of all variability. Durbin-Watson value shows that there is no evidence for autocorrelation at 0,05 significance level.

[Table 3 is near here]

Finally, we analyzed the TBDeath by utilizing the same 17 macro-scale indicators for validating-controlling purposes of our model proposals for the TotalCase-COV19 and TotalDeath-COV19. As an infectious disease quite dissimilar from COVID-19; the TBDeath model was expected to be different from our proposed model. As parallel in our expectations, it was quite different in terms of their involved variables except for obesity where "Air Pollution", "Undernourishment", and "GINI" are other significant factors rather than our proposed models (Table 4). Exceptionally, VIF values of Age and Schooling variables are large, relatively and still tolerable since their VIF values are not exceeding a threshold of 5.[12]

[Table 4 is near here]

DISCUSSION

Principal findings

The present study, which aims to explain the heterogeneity in the TotalCase-COV19 and TotalDeath-COV19 worldwide, by using an analysis of the health-related, socio-economic, climatic, and political macro indicators of 170 countries showed two regression models. Well-established comorbidities, obesity and hypertension, were significant in both these models. Additionally, indicators of sunshine and transparency for the TotalCase-COV19 model, as well as indicators of alcohol and urbanization for the TotalDeath-COV19, were important.

Interpretation and comparison with previous studies

Previous studies proposed that the comorbidities (obesity, hypertension, pulmonary disease, etc.) contributed to the risk for COVID-19 and progression to severe disease by increasing the expression of angiotensin-converting enzyme 2 (ACE2) and/or transmembrane protease serine 2 (TMPRSS2) on host lung cells and heightening the permissiveness of viral infection2.[13-15] In this study, the TotalCase-COV19 and TotalDeath-COV19 were greater in countries with a higher prevalence of obesity and hypertension. Global analyzes of the variability of COVID-19 mortality among countries, as in our study, revealed that TotalDeath-COV19 were positively correlated with deaths due to comorbidities, including cardiovascular, chronic respiratory, and kidney diseases, obesity, and cancer.[5 7 16]

The association between sunshine and the COVID-19 is arguable in the literature. Studies revealing sunshine hours positively linked to COVID-19 growth claimed that on sunny days outdoor activities increased

along with a decline in adherence to lockdown rules, resulting in increased virus exposure and transmission.[17-19] Other studies, on the other hand, suggested that sunny weather could help prevent the spread of COVID-19 by minimizing air pollution and increasing vitamin D production.[20-22] Those studies revealed that UV light from the sun can inactivate SARS-CoV-2, reduce outdoor transmission or increase immune resistance through vitamin D production. These results are supported by our research, which found a high TotalCase-COV19 in countries with low sunshine.

Public corruption and GINI reflecting corruption in the flow of products, resources, and services inside a nation were found to be associated with the TotalCase-COV19 and TotalDeath-COV19.[23 24] The present study further supports these studies by indicating that there are more TotalCase-COV19 in countries with higher transparency (i.e., lower corruption) levels. Thus, not only socio-economic indicators but also political indicators could play a critical role in the COVID-19 outbreak. Lack of access to reliable sources of information, as well as misinformation and inadequate communication, might lead to people disregarding government health alerts. Policies that promote the effective distribution of government budgets in public goods and services like healthcare and education, as well as policies that encourage transparency and information flow, could serve information accuracy about COVID-19's spread.[23 25]

There is no country-level analysis demonstrating an association between alcohol consumption and the TotalCase-COV19 and TotalDeath-COV19 in the literature. Recent studies have emphasized that increasing alcohol consumption is linked to excessive production of pro-inflammatory cytokines by hepatic cells, resulting in higher levels of inflammatory markers (Interleukin-8 and tumor necrosis factor [TNF-a]) which has also been observed in COVID-19 patients.[26-28]. As a result, it was hypothesized that acute and chronic alcohol consumption might suppress the immune system, leading to reduced resistance to acute respiratory disease and SARS-CoV-2 infection, as well as facilitating the progression of COVID-19.[29 30] The present study supports these findings, demonstrating that TotalDeath-COV19 are higher in countries with high alcohol consumption.

Urbanization has the potential to exacerbate disease and mortality by creating a slew of challenges such as public transportation, health, and economic disparities, substandard living facilities, insufficient freshwater supplies, and ineffective sanitation and ventilation systems.[4 6] Existing literature has shown that dense urban populations could cause less social distance, more trade, and more human mobility, leading to multiple infection routes and more rapid spread of COVID-19.[31 32] However, no significant relation was observed between the urbanization and the TotalCase-COV19 whereas some of the best subset models involve it. One

explanation for this may be the difficulty of identifying infected people in countries with low socioeconomic status due to inadequate test conditions for COVID-19. Therefore, the effect of urbanization on the TotalCase-COV19 can be hard to observe. Besides that, developed countries with a larger urban population and advanced healthcare systems have longer life expectancies and a higher percentage of the population over 65, which could lead to higher mortality rates in high-income countries8. Moreover, as tougher restriction policies for elderly people have been implemented, urbanization might be a significant factor in TotalDeath-COV19 rather than TotalCaseCOV-19. Our study thus showed that countries with higher urban populations have higher TotalDeath-COV19 in line with other global COVID-19 studies.[6 33]

Best subset regression analyses show that heterogeneity in the TotalCase-COV19 and the TotalDeathCOV-19 across countries may be attributed to variations in macro indicators including obesity and hypertension, urbanization, schooling, transparency, democracy, and HDI, as well as alcohol consumption and sunshine. Additionally, while the elderly population, the GINI index, and undernourishment were related to the TotalDeathCOV-19, hypercholesterolemia was related to the TotalCase-COV19. While the elderly population was a major parameter in nearly all TotalDeathCOV-19 models in our analysis, however, it was left out of the proposed models because it suppresses other estimators.

Other worldwide and national studies suggested that socio-economic, political, climatic, environmental, and ecological factors are correlated with the TotalCase-COV19 and the TotalDeathCOV-19 at different contexts and levels.[4-8 34-37] These studies revealed that countries with a high elderly population and level of democracy, as well as low levels of schooling and human development, had the highest mortality rate. Socioeconomic inequalities like low socioeconomic status, inadequate schooling, limited access to healthcare, and income inequality induce difficulty in accessing and affording healthy food, use of tobacco and alcohol, low physical activity, and lower use of preventive medicine. This situation might increase the prevalence of comorbidities by altering the gut microbiome, increasing local inflammation, and compromising immunity. Hence, the high-risk population becomes more susceptible to COVID-19.[2]

To control the COVID-19 models, we perform a regression analysis to estimate the TBDeath consisting of the same macro-scale estimators. In this regression model, tuberculosis is associated with GINI, air pollution, undernourishment, age, democracy, schooling, and obesity following the literature.[38-40] There were remarkable differences between our Covid models and Tuberculosis as we expected since they are different infections and they should have their own (disease-specific) characteristics.

Limitations

There are certain limitations of the present study. Firstly, not all health-related, socio-economic, climatic, and political indicators were up to date like COVID-19 deaths and cases in 2020. At that point, it is reasonable to expect that macro-scale indicators would not change significantly over the few years. Secondly, just 139 countries out of 170 were involved in the best subset regression analyses since only these countries had complete data. Thirdly, definitions of COVID-19 cases and illness for people based on symptoms and diagnosis varied by country over a certain time lead some country data on the number of cases to be biased, and so it increases the residual in the model. Additionally, some interruptions on data flow about COVID-19 deaths and cases from countries could lead the model to be biased. To some extent, these error impacts (i.e., bias) were minimized by standardizing the data's relative position to other countries' data by considering overall mean and standard deviation as well. In deciding the cause of death, a similar condition occurred in that some deaths could be recorded such as due to respiratory problems, other infections, or multiple organ failure. Finally, the reliability of the PCR test, that is used for diagnosing COVID-19 cases as a most preferred procedure across the world, could also be regarded as an important limitation of the study since false positive and false negative results of PCR test might cause deviations both in the COVID-19 cases and models.

Conclusion and implications of findings

In conclusion, this original study reveals that health-related, as well as social, economic, climatic, and political macro indicators have been noteworthy in the COVID-19 outbreak. The findings of this study do not claim that these macro indicators cause disease or death directly because it does not conduct any clinical or laboratory research. Essentially, the present study, demonstrating that these macro indicators explain the variation in the number of COVID-19 cases and deaths across 170 countries, may serve as a basis for future clinical trials. Additionally, if it is assumed that COVID-19 would have a lasting effect in the world for an extended period, this study might assist policymakers in developing short- and long-term health and social strategies to enhance these factors associated with COVID-19. Following the complete vaccination in nations, further research can be done to show how the vaccine affects these models.

Footnotes

Author Contribution: SE and SA had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: SE, SA, and FI. Acquisition, analysis, or interpretation of data: SE and FI. Drafting of the manuscript: SE and FI. Critical revision of the

manuscript for important intellectual content: SE and FI. Statistical analysis: SE. Administrative, technical, or material support: FI, AB, VG, EE, SM, and AA. Supervision: SA.

Competing of interest statement: The authors declare that they have no conflict of interest.

Data sharing: The data that support the findings of this study are openly available in Open Science Framework at https://osf.io/fx3pj/?view_only=5d226a0d346a488c9492512688a85334.

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

Ethical approval: This study is registered at ClinicalTrials.gov, NCT04757870 and ethical committee approval was received from İzmir Katip Çelebi University/Non-interventional Clinical Researches Ethics Board with the 2021-GOKAE-0287 protocol number.

Public and patient involvement statement: The authors of the study plan to share the results with communities through presentations and discussions with public health leaders, some of whom have been aware of the research but were not directly involved in it.

Transparency statement: The lead authors affirm that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Table 1. The list and the definitions of indicators and the outcome variable used in the analysis

Factors	Variables	Description	Year	N	Mean±Standard Deviation	Min	Max
Health-Related	Obesity[41]	Obesity (defined as of Body Mass Index [BMI]>= 30, crude estimate %) prevalence in ≥ 18 years old people	2016	170	10.59±5.84	2.10	37.30
	Hypertension[42]	High Blood Pressure (Systolic Blood Pressure [SBP] ≥ 140) or (Diastolic Blood Pressure [DBP] ≥ 90) Prevalence in 18 ≥ Years Old People	2015	170	23.45±5.31	12.60	41.00
	Alcohol[43]	Annual Alcohol Consumption Per Capita (Liter) in ≥ 15 Years Old People	2018	170	4.96±3.87	0.00	20.05
	Diabetes[44]	Diabetes, raised fasting blood glucose (Millimoles Per Liter [mmol/L] ≥ 7.0) Prevalence in ≥ 18 Years Old People	2014	170	7.70±3.99	1.00	22.10
	Cholesterol[45]	Hypercholesterolemia, raised total Cholesterol (Millimoles Per Liter [mmol/L] ≥ 6.2) prevalence in ≥25Years Old People	2008	170	10.59±5.84	2.50	29.10
	AirPollution[46]	Mortality attributable to joint effects of household and ambient air pollution in adults such as acute respiratory infections, cerebrovascular diseases, ischemic heart diseases, chronic obstructive pulmonary disease, and lung cancer in adults (Ambient and household air pollution attributable death rate (per 100 000 population)	2016	169	10.59±5.84	9.00	184.00
Climatic	Sunshine[47]	Total Sunshine Hours (Duration of sunlight in a year for countries)	2019	170	2519.18±576.3	1177.00	3737.00
Socio Economic	Vegetable[48]	Annual Vegetable Consumption for countries Kg/Per Capita	2017	154	91.93±69.73	6.28	377.17
	HDI[49]	Human Development Index. A measure prepared in line with the life span, literacy rate, education and life level for countries around the world (Rate between 0.800 - 1000 is very high, rate between 0.350 - 0.549 low)	2020	169	727.75±151.3	394.00	957.00
	Urbanization[50]	% of population living in urban areas for Countries around the world	2019	169	60.31±22.11	13.25	100.00
	Age[51]	Elderly people ratio whose age 65 & over for countries around the world	2019	168	9.22±6.5	1.16	28.00
	Schooling[52]	Average number of completed years of education of a countries' population aged 25 and higher around the world	2020	169	8.44±3.12	1.40	13.40
	Undernourishment[53]	The proportion of people who are malnourished as a percentage of the population that reflects the share of the population with insufficient caloric intake around the world	2019	170	0.11±0.12	0.01	0.60

	Happiness[54]	Happiness Index that shows life satisfaction in different nations, nationally representative samples of respondents are asked to imagine a ladder, (Rate between 10 representing the best possible life and a 0 representing the worst possible life.)	2019	139	5.55±1.09	2.52	7.84
	GINI[55]	Gini Inequality Index measures whether the distribution of national income in a country is equal or not. The coefficient takes values between 0 and 1, and higher values correspond to greater inequality.	2019	170	38.86±8.04	25.00	63.00
Political	Democracy[56]	Index that measures the state of democracy in 167 countries. It takes values between 0 and 10. 10 represents the perfect democracy and 0 represents the dictator regime.	2020	155	5.52±2.18	1.43	9.81
	Transparency[57]	Transparency measured by the Corruption Perception Index for each country, scored between 100 represents very clean and 0 represents highly corrupt.	2020	166	43.92±18.74	12.00	88.00
Target Variables	TotalDeath-COV19[9, 58]	Cumulative Total Deaths per Million People by 31th December 2020	2020	170	294.91±375.79	0.00	1673.00
	TotalCase-COV19[9, 59]	Cumulative Total Cases per Million People by 31th December 2020	2020	170	16874.29±20449.21	4.00	11568.00
Control Variable	TBDeath[60]	Deaths Related with Tuberculosis per 100.000 population yearly	2020	170	101.79±135.30	0.40	611.00

Table 2. Parameter testing results* for Linear Regression for Predicting "Total Cases Per Million People"

	Standardized Coefficients	95.0% Confidence Interval for β		Collinearity Statistics	p-value
	Beta	Lower Bound	Upper Bound	VIF	
Transparency	0.147	0.012	0.264	1.363	0.032
Obesity	0.460	0.311	0.547	1.213	0.001
Sunshine	-0.157	-0.266	-0.026	1.266	0.018
Hypertension	0.214	0.085	0.312	1.134	0.001

*R^2=0.449, Se=0.70, Durbin-Watson=2.01

*TotalCase-COV19= 0.214*Hypertension+ 0.460*Obesity-0.157*Sunshine+0.147*Transparency

Table 3. Parameter testing results* for Linear Regression for Predicting "Total Deaths Per Million People"

	Standardized Coefficients	95.0% Confidence Interval for β		Collinearity Statistics	p-value
	Beta	Lower Bound	Upper Bound	VIF	
Alcohol	0.173	0.030	0.315	1.461	0.018
Urban	0.204	0.040	0.368	1.920	0.015
Obesity	0.279	0.116	0.442	1.915	0.001
Hypertension	0.285	0.146	0.423	1.380	0.000

*R^2=0.409, Se=0.78, Durbin-Watson=2.29

*TotalDeath-COV19= 0.204*Urban+0.285*Hypertension+0.173*Alcohol+0.279*Obesity

Table 4. Parameter testing results* for Linear Regression for Predicting "Deaths Due to Tuberculosis"

	Standardized Coefficients	95.0% Confidence Interval for β		Collinearity Statistics	p-value
	Beta	Lower Bound	Upper Bound	VIF	
Obesity	-0.246	-0.427	-0.085	2.129	0.004
GINI	0.23	0.096	0.389	1.528	0.001
AirPollution	0.284	0.148	0.44	1.566	0.001
Undernourishment	0.245	0.084	0.402	2.014	0.003
Age	-0.453	-0.690	-0.224	4.193	0.001
Democracy	0.280	0.110	0.477	2.399	0.002
Schooling	0.253	0.048	0.459	3.303	0.016

*R^2=0.535, Se=0.72, Durbin-Watson=1.79

*TBDeath=-0.246*Obesity+0.230*GINI+0.284*AirPollution+0.245*Undernourishment-0.453*Age+0.280*Democracy+0.253*Schooling

REFERENCES

1 Steenhuysen J, Kelland K. 'When will it end?': How a changing virus is reshaping scientists' views on COVID-19. Reuters, March 3, 2021. https://www.reuters.com/article/us-health-coronavirus-variants-insight-idUSKBN2AV1T1.

2 Carethers JM. Insights into disparities observed with COVID-19. J Intern Med 2021;289(4):463-73. doi: 10.1111/joim.13199 [published Online First: 2020/11/10]

3 Wollenstein-Betech S, Cassandras CG, Paschalidis IC. Personalized predictive models for symptomatic COVID-19 patients using basic preconditions: Hospitalizations, mortality, and the need for an ICU or ventilator. Int J Med Inform 2020;142:104258. doi: 10.1016/j.ijmedinf.2020.104258 [published Online First: 2020/09/15]

4 Ficetola GF, Rubolini D. Containment measures limit environmental effects on COVID-19 early outbreak dynamics. Sci Total Environ 2021;761:144432. doi: 10.1016/j.scitotenv.2020.144432 [published Online First: 2020/12/29]

5 Arsalan M, Mubin O, Alnajjar F, et al. COVID-19 Global Risk: Expectation vs. Reality. Int J Environ Res Public Health 2020;17(15) doi: 10.3390/ijerph17155592 [published Online First: 2020/08/07]

6 Su D, Chen Y, He K, et al. Influence of socio-ecological factors on COVID-19 risk: a cross-sectional study based on 178 countries/regions worldwide. medRxiv 2020 doi: 10.1101/2020.04.23.20077545 [published Online First: 2020/06/09]

7 Sorci G, Faivre B, Morand S. Explaining among-country variation in COVID-19 case fatality rate. Sci Rep 2020;10(1):18909. doi: 10.1038/s41598-020-75848-2 [published Online First: 2020/11/05]

8 Rahman M, Islam M, Shimanto MH, et al. A global analysis on the effect of temperature, socio-economic and environmental factors on the spread and mortality rate of the COVID-19 pandemic. Environ Dev Sustain 2020:1-15. doi: 10.1007/s10668-020-01028-x [published Online First: 2020/10/13]

9 Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020;20(5):533-34. doi: 10.1016/s1473-3099(20)30120-1 [published Online First: 2020/02/23]

10 Newbold P, Carlson WL, Thorne BM. Statistics for Business and Economics: Global Edition. 8th ed. Pearson Education 2012.

11 Graybill FA, Iyer HK. Regression Analysis: Concepts and Applications. 1st ed. Duxbury Pr 1994.

12. Akinwande MO, Dikko HG, Samson A. Variance inflation factor: as a condition for the inclusion of suppressor variable (s) in regression analysis. Open J Stat 2015;5:754-67. doi: 10.4236/ojs.2015.57075

13 Pinto BGG, Oliveira AER, Singh Y, et al. ACE2 Expression Is Increased in the Lungs of Patients With Comorbidities Associated With Severe COVID-19. J Infect Dis 2020;222(4):556-63. doi: 10.1093/infdis/jiaa332 [published Online First: 2020/06/12]

14 Radzikowska U, Ding M, Tan G, et al. Distribution of ACE2, CD147, CD26, and other SARS-CoV-2 associated molecules in tissues and immune cells in health and in asthma, COPD, obesity, hypertension, and COVID-19 risk factors. Allergy 2020;75(11):2829-45. doi: 10.1111/all.14429 [published Online First: 2020/06/05]

15 Wang B, Li R, Lu Z, et al. Does comorbidity increase the risk of patients with COVID-19: evidence from meta-analysis. Aging (Albany NY) 2020;12(7):6049-57. doi: 10.18632/aging.103000 [published Online First: 2020/04/09]

16 Wang J, Sato T, Sakuraba A. Coronavirus Disease 2019 (COVID-19) Meets Obesity: Strong Association between the Global Overweight Population and COVID-19 Mortality. J Nutr 2021;151(1):9-10. doi: 10.1093/jn/nxaa375 [published Online First: 2020/11/15]

17 Azuma K, Kagi N, Kim H, et al. Impact of climate and ambient air pollution on the epidemic growth during COVID-19 outbreak in Japan. Environ Res 2020;190:110042. doi: 10.1016/j.envres.2020.110042 [published Online First: 2020/08/18]

18 Menebo MM. Temperature and precipitation associate with Covid-19 new daily cases: A correlation study between weather and Covid-19 pandemic in Oslo, Norway. Sci Total Environ 2020;737:139659. doi: 10.1016/j.scitotenv.2020.139659 [published Online First: 2020/06/04]
19 Paez A, Lopez FA, Menezes T, et al. A Spatio-Temporal Analysis of the Environmental Correlates of COVID-19 Incidence in Spain. Geogr Anal 2021;53(3):397-421 doi: 10.1111/gean.12241 [published Online First: 2020/08/25]
20 Biasin M, Bianco A, Pareschi G, et al. UV-C irradiation is highly effective in inactivating SARS-CoV-2 replication. Sci Rep 2021;11(1):6260. doi: 10.1038/s41598-021-85425-w [published Online First: 2021/03/20]
21 Coccia M. Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. Sci Total Environ 2020;729:138474. doi: 10.1016/j.scitotenv.2020.138474 [published Online First: 2020/06/06]
22 Merow C, Urban MC. Seasonality and uncertainty in global COVID-19 growth rates. Proc Natl Acad Sci 2020;117(44):27456-64. doi: 10.1073/pnas.2008590117 [published Online First: 2020/10/15]
23 Chaudhry R, Dranitsaris G, Mubashir T, et al. A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes. EClinicalMedicine 2020;25:100464. doi: 10.1016/j.eclinm.2020.100464 [published Online First: 2020/08/25]
24 Farzanegan MR. The Effect of Public Corruption on COVID-19 Fatality Rate: A Cross-Country Examination. CESifo Working Paper No. 8938 2021 doi: 10.2139/ssrn.3805464 [published Online First: 16/03/2021]
25 Cartaxo ANS, Barbosa FIC, de Souza Bermejo PH, et al. The exposure risk to COVID-19 in most affected countries: A vulnerability assessment model. PLoS One 2021;16(3):e0248075. doi: 10.1371/journal.pone.0248075 [published Online First: 2021/03/05]
26 Bilal B, Saleem F, Fatima SS. Alcohol consumption and obesity: The hidden scare with COVID-19 severity. Med Hypotheses 2020;144:110272. doi: 10.1016/j.mehy.2020.110272 [published Online First: 2020/12/02]
27 Pilotto A, Odolini S, Masciocchi S, et al. Steroid-Responsive Encephalitis in Coronavirus Disease 2019. Ann Neurol 2020;88(2):423-27. doi: 10.1002/ana.25783 [published Online First: 2020/05/18]
28 Casillo GM, Mansour AA, Raucci F, et al. Could IL-17 represent a new therapeutic target for the treatment and/or management of COVID-19-related respiratory syndrome? Pharmacol Res 2020;156:104791. doi: 10.1016/j.phrs.2020.104791 [published Online First: 2020/04/18]
29 Althobaiti YS, Alzahrani MA, Alsharif NA, et al. The Possible Relationship between the Abuse of Tobacco, Opioid, or Alcohol with COVID-19. Healthcare (Basel) 2020;9(1):2 doi: 10.3390/healthcare9010002 [published Online First: 2020/12/31]
30. Lassen MCH, Skaarup KG, Sengeløv M, et al. Alcohol Consumption and the Risk of Acute Respiratory Distress Syndrome in COVID-19. Ann Am Thorac Soc 2021;18(6):1074-76. doi: 10.1513/AnnalsATS.202008-988RL [published Online First: 2020/12/15]
31 Ahmadi M, Sharifi A, Dorosti S, et al. Investigation of effective climatology parameters on COVID-19 outbreak in Iran. Sci Total Environ 2020;729:138705. doi: 10.1016/j.scitotenv.2020.138705 [published Online First: 2020/05/04]
32 Kodera S, Rashed EA, Hirata A. Correlation between COVID-19 Morbidity and Mortality Rates in Japan and Local Population Density, Temperature, and Absolute Humidity. Int J Environ Res Public Health 2020;17(15):5477. doi: 10.3390/ijerph17155477 [published Online First: 2020/08/06]
33 Squalli J. Evaluating the determinants of COVID-19 mortality: A cross-country study. medRxiv 2020:2020.05.12.20099093. doi: 10.1101/2020.05.12.20099093

34 Coelho FC, Lana RM, Cruz OG, et al. Assessing the spread of COVID-19 in Brazil: Mobility, morbidity and social vulnerability. PLoS One 2020;15(9):e0238214. doi: 10.1371/journal.pone.0238214 [published Online First: 2020/09/19]

35 Hawkins RB, Charles EJ, Mehaffey JH. Socio-economic status and COVID-19-related cases and fatalities. Public Health 2020;189:129-34. doi: 10.1016/j.puhe.2020.09.016 [published Online First: 2020/11/24]

36 Wollenstein-Betech S, Silva AAB, Fleck JL, et al. Physiological and socioeconomic characteristics predict COVID-19 mortality and resource utilization in Brazil. PLoS One 2020;15(10):e0240346. doi: 10.1371/journal.pone.0240346 [published Online First: 2020/10/15]

37 Martins LD, da Silva I, Batista WV, et al. How socio-economic and atmospheric variables impact COVID-19 and influenza outbreaks in tropical and subtropical regions of Brazil. Environ Res 2020;191:110184. doi: 10.1016/j.envres.2020.110184 [published Online First: 2020/09/19]

38 Dos Santos DT, Seles Alves L, Moraes Arcoverde MA, et al. Social risk and its association with tuberculosis mortality in a context of high inequality in South Brazil: A geo-epidemiology analysis. J Infect Public Health 2020;13(8):1148-55. doi: 10.1016/j.jiph.2020.03.010 [published Online First: 2020/04/17]

39 Lönnroth K, Jaramillo E, Williams BG, et al. Drivers of tuberculosis epidemics: the role of risk factors and social determinants. Soc Sci Med 2009;68(12):2240-6. doi: 10.1016/j.socscimed.2009.03.041 [published Online First: 2009/04/28]

40 Murray M, Oxlade O, Lin HH. Modeling social, environmental and biological determinants of tuberculosis. Int J Tuberc Lung Dis 2011;15 Suppl 2:64-70. doi: 10.5588/ijtld.10.0535 [published Online First: 2011/07/16]

[dataset] [41] NCD Risk Factor Collaboration (NCD-RisC). Data from: Prevalence of obesity among adults, BMI >= 30 (crude estimate) (%). Global Health Observatory Data Repository, 2017. https://www.who.int/data/gho/data/indicators/indicator-details/GHO/prevalence-of-obesity-among-adults-bmi-=-30-(crude-estimate)-(-)

[dataset] [42] NCD Risk Factor Collaboration (NCD-RisC). Data from: Raised blood pressure (SBP ≥ 140 OR DBP ≥ 90), age-standardized (%) Estimates by country. Global Health Observatory Data Repository, 2017. https://apps.who.int/gho/data/node.main.A875STANDARD?lang=en.

[dataset] [43] World Health Organization. Data from: Alcohol, recorded per capita (15+ years) consumption (in litres of pure alcohol). Global Health Observatory Data Repository, 2020. https://www.who.int/data/gho/data/indicators/indicator-details/GHO/alcohol-recorded-per-capita-(15-)-consumption-(in-litres-of-pure-alcohol)

[dataset] [44] World Health Organization. Data from: Raised fasting blood glucose (≥ 7.0 mmol/L or on medication) (crude estimate). Global Health Observatory Data Repository, 2021. https://apps.who.int/gho/data/node.main.A869?lang=en

[dataset] [45] World Health Organization. Data from: Prevalence of raised total cholesterol (≥ 240 mg/dl). Global Health Observatory Data Repository, 2021. https://www.who.int/data/gho/indicator-metadata-registry/imr-details/2383

[dataset] [46] World Health Organization. Data from: Joint effects of air pollution. Global Health Observatory Data Repository, 2018. https://apps.who.int/gho/data/node.main.ENVHEALTHJOINTAAPHAP?lang=en%0A

[dataset] [47] Wikipedia. Data from: 2019/W44: World Cities Ranked by Annual Sunshine Hours. Data.world Repository, 2019. https://data.world/makeovermonday/2019w44

[dataset] [48] United Nations Food and Agricultural Organization (FAO). Data from: Vegetable supply per person kilograms per year. Our World in Data Repository, 2020. https://ourworldindata.org/grapher/vegetable-consumption-per-capita?tab=table.

[dataset] [49] United Nations Development Programme. Data from: Human Development Report 2020. Human Development Reports Data Repository, 2021. http://hdr.undp.org/en/content/download-data

[dataset] [50] The World Bank. Data from: Urban population (% of total population). DataBank World Development Indicators Data Repository, 2021. https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS

[dataset] [51] The World Bank. Data from: Population ages 65 and above (% of total population). DataBank World Development Indicators Data Repository, 2021. https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS

[dataset] [52] United Nations Development Programme. Data from: Mean years of schooling (years). Human Development Reports Data Repository, 2020. http://hdr.undp.org/en/indicators/103006#b

[dataset] [53] United Nations Food and Agricultural Organization (FAO). Data from: Prevalence of undernourishment (percent). FAOSTAT Data Repository, 2021. https://www.fao.org/faostat/en/#data/SDGB

[dataset] [54] Helliwell JF, Layard. R, Sachs. J, et al. Data from: World Happiness Report 2021. New York: Sustainable Development Solutions Network 2021. https://worldhappiness.report/ed/2021/#appendices-and-data

[dataset] [55] Central Intelligence Agency. Data from: Gini Index coefficient – distribution of family income. The World Factbook Data Repository, 2021. https://www.cia.gov/the-world-factbook/field/gini-index-coefficient-distribution-of-family-income/

[dataset] [56] Economist Intelligence Unit. Data from: Democracy Index 2020: In sickness and in health? The Economist Intelligence Unit Limited Data Repository, 2021. https://www.eiu.com/n/campaigns/democracy-index-2020/#mktoForm_anchor

[dataset] [57] Transparency International. Data from: Corruption Perceptions Index, 2020. Transparency International Data Repository, 2021. https://www.transparency.org/en/cpi/2020/table/nzl

[dataset] [58] Worldometer. Data from: Cumulative Total Deaths per Million People by 31th December 2020. Worldometer's COVID-19 Data Repository, February 26, 2021. https://www.worldometers.info/coronavirus/

[dataset] [59] Worldometer. Data from: Cumulative Total Cases per Million People by 31th December 2020. Worldometer's COVID-19 Data Repository, February 26, 2021. https://www.worldometers.info/coronavirus/

[dataset] [60] World Health Organization. Data from: Incidence of tuberculosis (per 100 000 population per year). Global Health Observatory Data Repository, 2020. https://www.who.int/data/gho/data/indicators/indicator-details/GHO/incidence-of-tuberculosis-(per-100-000-population-per-year)

Supplementary Table A. Best Subset Models for Total Cases Per Million

Number of Variables	R^2	R^2 (adjusted)	R^2 (pred)	Mallows Cp	S	Age	Happiness	Gini	Transparency	Democracy	Urbanization	Schooling	AirPollution	HDI	Vegetable	SunShine	Undernourishment	Obesity	Cholesterol	Hypertension	Alcohol	Diabet
1	38.5	38.1	36.1	16.0	0.78	X																
1	38.5	38	36.7	16.1	0.78									X								
2	46.5	45.6	43.5	-0.4	0.73	X												X				
2	44.5	43.6	41.6	4.1	0.74									X					X			
3	48.1	46.8	44.3	-2.1	0.72	X									X			X				
3	47.9	46.7	44.4	-1.7	0.72									X				X	X			
4	49.5	47.9	45.2	-3.6	0.71									X	X			X	X			
4	49.2	47.6	44.2	-2.7	0.71	X									X			X	X			
5	50.0	47.9	45.1	-2.5	0.71					X				X	X			X	X			
5	49.9	47.9	45.0	-2.4	0.71									X	X			X	X		X	
6	50.4	48.0	44.5	-1.6	0.71					X				X	X			X	X	X		
6	50.4	48.0	44.5	-1.5	0.71									X	X			X	X	X	X	
7	50.7	47.9	44.3	-0.4	0.71					X				X	X			X	X	X	X	
7	50.6	47.8	44.0	-0.1	0.71					X	X			X	X			X	X	X		
8	50.9	47.7	43.8	1.2	0.71					X	X			X	X			X	X	X	X	
8	50.9	47.6	43.8	1.4	0.71					X			X	X	X			X	X	X	X	
9	51.1	47.4	43.3	2.8	0.72					X	X		X	X	X			X	X	X	X	
9	51.0	47.3	42.6	3.0	0.72				X	X	X			X	X			X	X	X	X	
10	51.2	47.1	42.2	4.6	0.72				X	X	X		X	X	X			X	X	X	X	
10	51.1	47.1	42.7	4.7	0.72			X		X	X		X	X	X			X	X	X	X	
11	51.2	46.7	41.5	6.5	0.72		X		X	X	X		X	X	X			X	X	X	X	
11	51.2	46.7	41.3	6.5	0.72				X	X	X		X	X	X			X	X	X	X	X
12	51.3	46.3	40.7	8.4	0.72		X		X	X	X		X	X	X			X	X	X	X	X
12	51.3	46.3	40.2	8.4	0.72	X	X		X	X	X		X	X	X			X	X	X	X	
13	51.3	45.9	40.0	10.3	0.73		X	X	X	X	X		X	X	X			X	X	X	X	X
13	51.3	45.9	39.6	10.3	0.73	X	X	X	X	X	X		X	X	X			X	X	X	X	
14	51.4	45.5	38.6	12.2	0.73	X	X	X	X	X	X		X	X	X			X	X	X	X	X
14	51.4	45.5	38.6	12.2	0.73		X	X	X	X	X		X	X	X	X		X	X	X	X	X
15	51.4	45.1	37.7	14.1	0.73	X		X	X	X	X	X	X	X	X		X	X	X	X	X	X
15	51.4	45.1	37.4	14.1	0.73	X	X	X	X	X	X	X	X	X	X			X	X	X	X	X
16	51.4	44.6	37.0	16.0	0.73	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X
16	51.4	44.6	35.7	16.0	0.73	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X
17	51.4	44.1	35.3	18.0	0.74	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

139 cases used; 31 cases contain missing values.

Supplementary Table B. Best Subset Models for Total Deaths Per Million

Number of Variables	R^2	R^2 (adj)	R^2 (pred)	Mallows Cp	S	Age	Happiness	Gini	Transparency	Democracy	Urbanization	Schooling	AirPollution	HDI	Vegetable	SunShine	Undernourishment	Obesity	Cholesterol	Hypertension	Alcohol	Diabet
1	35.6	35.1	33.2	28.1	0.86	X																
1	30.5	30.0	28.6	40.3	0.89									X								
2	42.3	41.4	39.3	13.8	0.81	X												X				
2	41.4	40.5	38.9	16.0	0.82											X		X				
3	45.6	44.3	41.9	7.9	0.79											X		X		X		
3	45.4	44.1	41.8	8.4	0.79	X										X		X				
4	48.4	46.7	43.8	3.3	0.78	X			X							X		X				
4	46.9	45.2	42.6	6.8	0.79			X								X		X		X		
5	49.7	47.6	44.5	2.2	0.77	X		X	X							X		X				
5	49.1	47.1	43.9	3.4	0.77	X			X	X						X		X				
6	50.5	48.1	45.0	2.1	0.77	X		X	X							X	X	X				
6	50.3	47.9	44.1	2.6	0.77	X		X	X							X		X		X		
7	51.2	48.4	44.7	2.5	0.76	X		X	X							X	X	X		X		
7	51.1	48.3	44.8	2.7	0.76	X		X	X		X					X	X	X				
8	51.6	48.4	45.2	3.5	0.76	X		X	X			X				X	X	X			X	
8	51.5	48.3	44.4	3.7	0.76	X		X	X	X						X	X	X		X		
9	52.0	48.4	44.8	4.6	0.76	X		X	X		X	X				X	X	X			X	
9	51.9	48.3	44.5	4.8	0.76	X		X	X			X				X	X	X		X	X	
10	52.3	48.3	44.4	5.9	0.76	X		X	X	X	X	X				X	X	X			X	
10	52.2	48.2	44.1	6.1	0.76	X		X	X	X		X				X	X	X		X	X	
11	52.5	48.1	44.1	7.4	0.77	X		X	X	X	X	X		X		X	X	X			X	
11	52.4	48.0	43.8	7.5	0.77	X		X	X	X		X		X		X	X	X		X	X	
12	52.7	47.9	43.4	8.8	0.77	X		X	X	X	X	X		X		X	X	X		X	X	
12	52.6	47.8	42.8	9.0	0.77	X		X	X	X		X		X	X	X	X	X		X	X	
13	52.9	47.6	42.4	10.5	0.77	X		X	X	X	X	X		X	X	X	X	X		X	X	
13	52.8	47.6	42.5	10.5	0.77	X		X	X	X	X	X		X		X	X	X		X	X	X
14	53.0	47.3	41.4	12.2	0.77	X		X	X	X	X	X		X	X	X	X	X		X	X	X
14	52.9	47.2	41.1	12.4	0.77	X		X	X	X	X	X		X	X	X	X	X	X	X	X	
15	53.0	46.9	40.1	14.1	0.77	X		X	X	X	X	X		X	X	X	X	X	X	X	X	X
15	53.0	46.8	40.8	14.2	0.77	X	X	X	X	X	X	X		X	X	X	X	X		X	X	X
16	53.0	46.5	39.5	16.0	0.78	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X
16	53.0	46.4	39.4	16.1	0.78	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
17	53.1	46.0	38.8	18.0	0.78	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

139 cases used; 31 cases contain missing values.

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5
Bias	9	Describe any efforts to address potential sources of bias	NA
Study size	10	Explain how the study size was arrived at	5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5-7
		(*b*) Describe any methods used to examine subgroups and interactions	5-7
		(*c*) Explain how missing data were addressed	5-7
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	NA
		(*e*) Describe any sensitivity analyses	NA
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7-8
		(b) Give reasons for non-participation at each stage	7-8
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7-8
		(b) Indicate number of participants with missing data for each variable of interest	7-8
Outcome data	15*	Report numbers of outcome events or summary measures	7-8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	7-8

		(*b*) Report category boundaries when continuous variables were categorized	7-8
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	7-8
Discussion			
Key results	18	Summarise key results with reference to study objectives	8-9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	8-11
Generalisability	21	Discuss the generalisability (external validity) of the study results	8-12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	NA

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.