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

Sabri Erdem,[1] Fulya Ipek ,[2] Aybars Bars,[3] Volkan Genç,[3] Esra Erpek,[4] Shabnam Mohammadi,[3] Anıl Altınata,[3] Servet Akar[4]

To cite: Erdem S, Ipek F, Bars A, *et al*. Investigating the effect of macro-scale estimators on worldwide COVID-19 occurrence and mortality through regression analysis using online country-based data sources. *BMJ Open* 2022;**12**:e055562. doi:10.1136/bmjopen-2021-055562

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-055562).

Received 15 July 2021
Accepted 24 January 2022

Check for updates

© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

[1]Department of Business Administration, Dokuz Eylül University, Izmir, Turkey
[2]Faculty of Physical Therapy and Rehabilitation, Hacettepe University, Ankara, Turkey
[3]Social Sciences Institute, Dokuz Eylül University, Izmir, Turkey
[4]Department of Internal Medicine, Division of Rheumatology Atatürk Education and Research Hospital, Izmir Katip Celebi University, Izmir, Turkey

Correspondence to
Fulya Ipek;
fulya.ipek@hacettepe.edu.tr

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 organisations.

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 hypercholesterolaemia).

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 analysed by 17 macro-scale estimators around the health-related, socioeconomic, 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 urbanisation (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 Inequality Index.

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

Trial registration number ClinicalTrials.gov Registry (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 1 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

COVID-19 caused by SARS-CoV-2 has spread over countries worldwide. On 11 March 2020, the 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 would also 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 (Total-Case-COV19) and mortality (total number of deaths per million) of COVID-19 (Total-Death-COV19) vary significantly among 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

BMJ Open: first published as 10.1136/bmjopen-2021-055562 on 14 February 2022. Downloaded from http://bmjopen.bmj.com/ on May 1, 2022 by 193.140.151.30. Protected by copyright.

differentiating countries in terms of TotalCase-COV19 and TotalDeath-COV19.

The term 'macro-level data' (eg, Human Development Index (HDI) 2020 by United Nations (UN)) could be defined as periodical (eg, weekly, monthly, quarterly or annual) and long-term data that were gathered, summarised, and published by the government, private, public, national and international organisations about a specific subject (eg, health, energy, society, economy, politics) 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 organisations 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, socioeconomic, 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 because of covering data for a longer period. Therefore, the current study aims to identify health-related and socioeconomic 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.[9] While the process of the pandemic was still ongoing, mass immunisation has started in several countries as of December 2019. To eliminate the effect of immunisation, in this study, TotalCase-COV19 and TotalDeath-COV19 data were censored by 31 December 2020, when vaccination has not yet become widespread. Our dataset is available in Open Science Framework. Health-related indicators from the WHO and socioeconomic, climatic, and political indicators from UN World Inequality Database and World Bank explaining variability in TotalCase-COV19 and TotalDeath-COV19 among countries were summarised in table 1. Our dataset involves 17 macro-scale variables of 170 countries.

Statistical analyses

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

- Drawing 'actual target variable versus predicted target value' scatter plot diagram for testing linearity.
- Calculating standardised residual values for each case for detecting outlier as outside +3.
- Drawing a partial regression plot diagram for each predictor versus 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 versus regression standardised residual' plot diagram for detecting heteroscedasticity.
- Drawing 'observed cumulative probability versus expected cumulative probability (P–P)' plot diagram of standardised 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 standardised form $[(x-\text{mean}(x))/SD(x)]$, so that we could get a more acceptable VIF ($VIF=1/(1-R^2)<2.0$) 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.[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 standardised 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/

BMJ Open: first published as 10.1136/bmjopen-2021-055562 on 14 February 2022. Downloaded from http://bmjopen.bmj.com/ on May 1, 2022 by 193.140.151.30. Protected by copyright.

BMJ Open: first published as 10.1136/bmjopen-2021-055562 on 14 February 2022. Downloaded from http://bmjopen.bmj.com/ on May 1, 2022 by 193.140.151.30. Protected by copyright.

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

Factors	Variables	Description	Year	N	Mean±SD	Min	Max
Health related	Obesity[41]	Obesity (defined as body mass index ≥30, crude estimate %) prevalence in ≥18-year-old people	2016	170	10.59±5.84	2.10	37.30
	Hypertension[42]	High blood pressure (systolic blood pressure ≥140) or (diastolic blood pressure ≥90) prevalence in ≥18-year-old people	2015	170	23.45±5.31	12.60	41.00
	Alcohol[43]	Annual alcohol consumption per capita (litre) in ≥15-year-old people	2018	170	4.96±3.87	0.00	20.05
	Diabetes[44]	Diabetes, raised fasting blood glucose (≥7.0 mmol/L) prevalence in ≥18-year-old people	2014	170	7.70±3.99	1.00	22.10
	Cholesterol[45]	Hypercholesterolaemia, raised total cholesterol (≥6.2 mmol/L) prevalence in ≥25-year-old people	2008	170	10.59±5.84	2.50	29.10
	Air pollution[46]	Mortality attributable to joint effects of household and ambient air pollution in adults such as acute respiratory infections, cerebrovascular diseases, ischaemic 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

Continued

Table 1 Continued

Factors	Variables	Description	Year	N	Mean±SD	Min	Max
Socioeconomic	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 and 1000 is very high, rate between 0.350 and 0.549 low)	2020	169	727.75±151.3	394.00	957.00
	Urbanisation[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 and 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 country's 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, score 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 31 December 2020	2020	170	294.91±375.79	0.00	1673.00
	TotalCase-COV19[9 59]	Cumulative total cases per million people by 31 December 2020	2020	170	16874.29±20 449.21	4.00	11 568.00
Control variable	TBDeath[60]	Deaths related to tuberculosis per 100.000 population yearly	2020	170	101.79±135.30	0.40	611.00

BMJ Open: first published as 10.1136/bmjopen-2021-055562 on 14 February 2022. Downloaded from http://bmjopen.bmj.com/ on May 1, 2022 by 193.140.151.30. Protected by copyright.

Table 2 Parameter testing results for linear regression for predicting 'total cases per million people'

	Standardised coefficients	95% CI for β		Collinearity statistics	
	Beta	Lower bound	Upper bound	VIF	P value
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

TotalCase-COV19=0.214×hypertension+0.460×obesity−0.157×sunshine+0.147×transparency.
R^2=0.449, SE=0.70, Durbin-Watson=2.01.
VIF, variance inflation factor.

model through a best subset approach that analyses data in terms of a predictor's subset that could create applicable models that illuminate 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, SE or Mallows' Cp (min is preferred), Akaike Information Criterion (min is preferred) and Bayesian Information 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 (eg, enter, stepwise, forward/backward selection), determining thresholds for alpha to enter (0.05 through 0.15) and remove (0.05 through 0.15), and for accepting or rejecting the next variable to the incorporated model. We used the SPSS V.24 and Minitab V.19 trial version for analysing 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 as demonstrated in online supplemental tables A and B.

Based on best subset analysis of TotalCase-COV19 and TotalDeath-COV19 modelling, the indicators 'sunshine', 'obesity', 'hypertension', 'urbanisation', 'schooling', 'alcohol', 'democracy', 'transparency' and 'HDI' are commonly found significant for the most of models. Additionally, for modelling the 'TotalDeath-COV19', 'age', 'GINI' (Gini Inequality Index) and 'undernourishment' were found significant; and for modelling '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 ($p<0.05$).

Urbanisation, schooling, HDI, transparency, democracy and GINI are all inter-related factors, despite the lack of their multicollinearity. Thus, countries with desirable socioeconomic conditions are likely to have high levels of urbanisation, 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, obesity and hypertension in

Table 3 Parameter testing results for linear regression for predicting 'total deaths per million people'

	Standardised coefficients	95% CI for β		Collinearity statistics	
	Beta	Lower bound	Upper bound	VIF	P value
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

TotalDeath-COV19=0.204×urban+0.285×hypertension+0.173×alcohol+0.279×obesity.
R^2=0.409, SE=0.78, Durbin-Watson=2.29.
VIF, variance inflation factor.

BMJ Open: first published as 10.1136/bmjopen-2021-055562 on 14 February 2022. Downloaded from http://bmjopen.bmj.com/ on May 1, 2022 by 193.140.151.30. Protected by copyright.

Table 4 Parameter testing results for linear regression for predicting 'deaths due to tuberculosis'

	Standardised coefficients	95% CI for β		Collinearity statistics	
	Beta	Lower bound	Upper bound	VIF	P value
Obesity	−0.246	−0.427	−0.085	2.129	0.004
GINI	0.23	0.096	0.389	1.528	0.001
Air pollution	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

TBDeath=−0.246×obesity+0.230×GINI+0.284×air pollution+0.245×undernourishment−0.453×age+0.280×democracy+0.253×schooling.
R^2=0.535, SE=0.72, Durbin-Watson=1.79.
GINI, Gini Inequality Index; VIF, variance inflation factor.

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 3 shows the results of the regression model indicating the determinants of the TotalDeath-COV19 worldwide. The regression model revealed that urbanisation 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.

Finally, we analysed the TBDeath by using 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]

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, socioeconomic, 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 sunshine and transparency for the TotalCase-COV19 model, as well as indicators alcohol and urbanisation 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 of COVID-19 and progression to severe disease by increasing the expression of ACE2 and/or transmembrane protease serine 2 on host lung cells and heightening the permissiveness of viral infection.[13–15] In this study, the TotalCase-COV19 and TotalDeath-COV19 were greater in countries with a higher prevalence of obesity and hypertension. Global analyses of the variability of COVID-19 mortality among countries, as in our study, revealed that TotalDeath-COV19 was 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 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 minimising air pollution and increasing vitamin D production.[20–22] Those studies revealed that ultraviolet 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 (ie, lower corruption) levels. Thus, not only socioeconomic indicators but also political indicators could play a critical role in the COVID-19 outbreak. Lack of access

BMJ Open: first published as 10.1136/bmjopen-2021-055562 on 14 February 2022. Downloaded from http://bmjopen.bmj.com/ on May 1, 2022 by 193.140.151.30. Protected by copyright.

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 emphasised that increasing alcohol consumption is linked to excessive production of proinflammatory cytokines by hepatic cells, resulting in higher levels of inflammatory markers (interleukin-8 and tumour necrosis factor-α), which has also been observed in patients with COVID-19.[26–28] As a result, it was hypothesised 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 is higher in countries with high alcohol consumption.

Urbanisation 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 urbanisation and 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 urbanisation 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 years, which could lead to higher mortality rates in high-income countries.[8] Moreover, as tougher restriction policies for elderly people have been implemented, urbanisation 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, urbanisation, schooling, transparency, democracy and HDI, as well as alcohol consumption and sunshine. Additionally, while the elderly population, GINI and undernourishment were related to the TotalDeathCOV-19, hypercholesterolaemia was related to the TotalCase-COV19. While the elderly population was a major parameter in nearly all TotalDeath-COV19 models in our analysis, it was left out of the proposed models because it suppresses other estimators.

Other worldwide and national studies suggested that socioeconomic, 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-19 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. First, not all health-related, socioeconomic, 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. Second, just 139 countries out of 170 were involved in the best subset regression analyses since only these countries had complete data. Third, definitions of COVID-19 cases and illness of people based on symptoms and diagnosis varied by country over a certain time that led 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 (ie, bias) were minimised by standardising the data's relative position to other countries' data by considering overall mean and SD 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 the most preferred procedure across the world could also be regarded as an important limitation of the

BMJ Open: first published as 10.1136/bmjopen-2021-055562 on 14 February 2022. Downloaded from http://bmjopen.bmj.com/ on May 1, 2022 by 193.140.151.30. Protected by copyright.

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-term 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.

Contributors SE is responsible for the overall content as guarantor. SE and SA had full access to all of the data in the study and take responsibility for the integrity 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.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Ethical committee approval was received from İzmir Katip Çelebi University/Non-interventional Clinical Researches Ethics Board (protocol number: 2021-GOKAE-0287).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. The data that support the findings of this study are openly available in Open Science Framework at https://osf.io/fx3pj/?view_only=5d226a0d346a488c9492512688a85334.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD

Fulya Ipek http://orcid.org/0000-0001-7606-042X

REFERENCES

1 Steenhuysen J, Kelland K. 'When will it end?': How a changing virus is reshaping scientists' views on COVID-19, 2021. Available: 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;289:463–73.
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.
4 Ficetola GF, Rubolini D. Containment measures limit environmental effects on COVID-19 early outbreak dynamics. *Sci Total Environ* 2021;761:144432.
5 Arsalan M, Mubin O, Alnajjar F, *et al*. COVID-19 global risk: expectation vs. reality. *Int J Environ Res Public Health* 2020;17:5592.
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:10.2139/ssrn.3582852. [Epub ahead of print: 29 Apr 2020].
7 Sorci G, Faivre B, Morand S. Explaining among-country variation in COVID-19 case fatality rate. *Sci Rep* 2020;10:18909.
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.
9 Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020;20:533–4.
10 Newbold P, Carlson WL, Thorne BM. *Statistics for business and economics: global edition*. 8th edn. 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;05:754–67.
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;222:556–63.
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;75:2829–45.
15 Wang B, Li R, Lu Z, *et al*. Does comorbidity increase the risk of patients with COVID-19: evidence from meta-analysis. *Aging* 2020;12:6049–57.
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;151:9–10.
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.
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.
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;53:397–421.
20 Biasin M, Bianco A, Pareschi G, *et al*. Uv-C irradiation is highly effective in inactivating SARS-CoV-2 replication. *Sci Rep* 2021;11:6260.
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.
22 Merow C, Urban MC. Seasonality and uncertainty in global COVID-19 growth rates. *Proc Natl Acad Sci U S A* 2020;117:27456–64.
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.
24 Farzanegan MR. The effect of public corruption on COVID-19 fatality rate: a Cross-Country examination. *SSRN Electronic Journal* 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:e0248075.
26 Bilal B, Saleem F, Fatima SS. Alcohol consumption and obesity: the hidden scare with COVID-19 severity. *Med Hypotheses* 2020;144:110272.
27 Pilotto A, Odolini S, Masciocchi S, *et al*. Steroid-Responsive encephalitis in coronavirus disease 2019. *Ann Neurol* 2020;88:423–7.

BMJ Open: first published as 10.1136/bmjopen-2021-055562 on 14 February 2022. Downloaded from http://bmjopen.bmj.com/ on May 1, 2022 by 193.140.151.30. Protected by copyright.

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.
29 Althobaiti YS, Alzahrani MA, Alsharif NA, *et al*. The possible relationship between the abuse of tobacco, opioid, or alcohol with COVID-19. *Health Care* 2020;9:2.
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;18:1074–6.
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.
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;17:5477.
33 Squalli J. Evaluating the determinants of COVID-19 mortality: a cross-country study. *medRxiv*2020.
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:e0238214.
35 Hawkins RB, Charles EJ, Mehaffey JH. Socio-Economic status and COVID-19-related cases and fatalities. *Public Health* 2020;189:129–34.
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:e0240346.
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.
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;13:1148–55.
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;68:2240–6.
40 Murray M, Oxlade O, Lin H-H. Modeling social, environmental and biological determinants of tuberculosis. *Int J Tuberc Lung Dis* 2011;15 Suppl 2:64–70.
41 NCD Risk Factor Collaboration (NCD-RisC). Data from: Prevalence of obesity among adults, BMI >= 30 (crude estimate) (%). Global Health Observatory Data Repository, 2017. Available: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/prevalence-of-obesity-among-adults-bmi-=-30-(crude-estimate)-(-)
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. Available: https://apps.who.int/gho/data/node.main.A875STANDARD?lang=en
43 World Health Organization. Data from: alcohol, recorded per capita (15+ years) consumption (in litres of pure alcohol). Global health Observatory data Repository, 2020. Available: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/alcohol-recorded-per-capita-(15-)-consumption-(in-litres-of-pure-alcohol)
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. Available: https://apps.who.int/gho/data/node.main.A869?lang=en
45 World Health Organization. Data from: prevalence of raised total cholesterol (≥ 240 mg/dl). Global health Observatory data Repository, 2021. Available: https://www.who.int/data/gho/indicator-metadata-registry/imr-details/2383
46 World Health Organization. Data from: joint effects of air pollution. global health Observatory data Repository, 2018. Available: https://apps.who.int/gho/data/node.main.ENVHEALTHJOINTAAPHAP?lang=en%0A
47 Wikipedia. Data from: 2019/W44: world cities Ranked by annual sunshine hours. Data.world Repository, 2019. Available: https://data.world/makeovermonday/2019w44
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
49 United Nations Development Programme. Data from: human development report 2020. human development reports data Repository, 2021. Available: http://hdr.undp.org/en/content/download-data
50 The World Bank. Data from: urban population (% of total population). DataBank world development indicators data Repository, 2021. Available: https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS
51 The World Bank. Data from: population ages 65 and above (% of total population). DataBank world development indicators data Repository, 2021. Available: https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS
52 United Nations Development Programme. Data from: mean years of schooling (years). human development reports data Repository, 2020. Available: http://hdr.undp.org/en/indicators/103006#b
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
54 Helliwell JF, R L, J S. *Data from: world Happiness report 2021*. New York: Sustainable Development Solutions Network, 2021. https://worldhappiness.report/ed/2021/#appendices-and-data
55 Central Intelligence Agency. Data from: Gini index coefficient – distribution of family income. the world Factbook data Repository, 2021. Available: https://www.cia.gov/the-world-factbook/field/gini-index-coefficient-distribution-of-family-income/
56 Economist Intelligence Unit. Data from: democracy index 2020: in sickness and in health? the Economist intelligence unit limited data Repository, 2021. Available: https://www.eiu.com/n/campaigns/democracy-index-2020/#mktoForm_anchor
57 Transparency International. Data from: corruption perceptions index, 2020. transparency international data Repository, 2021. Available: https://www.transparency.org/en/cpi/2020/table/nzl
58 Worldometer. Data from: Cumulative Total Deaths per Million People by 31th December 2020. Worldometer's COVID-19 Data Repository, 2021. Available: https://www.worldometers.info/coronavirus/
59 Worldometer. Data from: Cumulative Total Cases per Million People by 31th December 2020. Worldometer's COVID-19 Data Repository, 2021. Available: https://www.worldometers.info/coronavirus/
60 World Health Organization. Data from: incidence of tuberculosis (per 100 000 population per year). global health Observatory data Repository, 2020. Available: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/incidence-of-tuberculosis-(per-100-000-population-per-year)

BMJ Open: first published as 10.1136/bmjopen-2021-055562 on 14 February 2022. Downloaded from http://bmjopen.bmj.com/ on May 1, 2022 by 193.140.151.30. Protected by copyright.