

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/361196900>

COVID-19 vaccine dosages and government factors role on the global variation in COVID-19 mortality: A statistical and regression analysis

Article in International Journal of ADVANCED AND APPLIED SCIENCES · June 2022

DOI: 10.21833/ijaas.2022.05.003

CITATION

1

READS

114

6 authors, including:



Asif Hassan

King Abdulaziz University

20 PUBLICATIONS 129 CITATIONS

SEE PROFILE



Tabrej Khan

King Abdulaziz University

26 PUBLICATIONS 135 CITATIONS

SEE PROFILE



Nashwan Alromema

King Abdulaziz University

25 PUBLICATIONS 108 CITATIONS

SEE PROFILE



Lal Bihari Barik

King Abdulaziz University

43 PUBLICATIONS 162 CITATIONS

SEE PROFILE

COVID-19 vaccine dosages and government factors role on the global variation in COVID-19 mortality: A statistical and regression analysis



Asif Hassan Syed^{1,*}, Tabrej Khan², Nashwan A. Alromema¹, Lalbihari Barik², Ahmad Abdul Qadir AlRababah¹, Murad M. Aljiffry³

¹Department of Computer Science, Faculty of Computing and Information Technology Rabigh (FCITR), King Abdulaziz University, Jeddah, Saudi Arabia

²Department of Information Sciences, Faculty of Computing and Information Technology Rabigh (FCITR), King Abdulaziz University, Jeddah, Saudi Arabia

³Department of Surgery, Faculty of Medicine King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:

Received 22 October 2021

Received in revised form

1 February 2022

Accepted 1 March 2022

Keywords:

COVID-19

COVID-19 mortality variation

Cross-sectional data

Partial-correlation

Controlling factor

COVID-19 vaccine dosages

Stepwise regression model

ABSTRACT

The objective of our study was to explore the influence of the current vaccination program and other relevant government factors to explain the variation in COVID-19 mortality in the world. The study involves a cross-sectional survey of COVID-19 related and government factors from 161 countries. We retrieved and processed publically available coronavirus pandemic data (July 17, 2021) from several online databases, excluding countries' data violating correlation and regression analysis assumptions. In addition, partial correlations studies and multivariate analysis were performed to explore the influence current vaccination program and other relevant government factors on the relationship between the explanatory variable and the total deaths due to COVID-19. The partial-correlation studies revealed that controlling for a complete dosage of COVID-19 vaccine per 100 people in the population had a significant ($P < 0.001$) impact on the strength of the relationship between some explanatory variables and the response variable (total COVID-19 mortality). Furthermore, the Stepwise Linear Regression (SLR) model shows that the covariates, namely total_cases, hospital patients per million, hospital beds per thousand, male smokers, and people fully vaccinated per hundred, added significantly ($P < 0.001$) to the prediction of the response variable. Our SLR model validation study revealed that the observed total COVID-19 mortality was highly correlated with the predicted total COVID-19 mortality in various countries ($r = 0.977$, $P < 0.001$). Our Stepwise Linear Regression model performs significantly better with an R-squared value of 0.958 and adjusted R-squared value of 0.956 than other related regression models built to predict COVID-19 mortality. Based on our current findings, we conclude that governments with better hospital infrastructure and people with complete dosages of the COVID-19 vaccine will have minimal COVID-19 fatalities.

© 2022 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The highly infectious COVID-19 has infested over 190,565,973 people, and more than 4,095,485 people have died across the globe as of July 17, 2021 (Lipsitch et al., 2020). However, the total COVID-19 based death differs significantly at the country level. For instance, in July 2021, the total deaths attributed

to COVID-19 in Laos, Fiji, and Burundi were 3, 4, and 8, respectively. On the other hand, on July 17, 2021, the total COVID-19 based mortality in countries, namely Mexico, India, Brazil, and the United States of America, were 230,959, 386,708, 500,800, and 601,741, respectively.

Such a significant country-wise discrepancy in the total deaths attributed to COVID-19 implies that factors apart from patient characteristics, namely government policies, play a substantial role in fatalities due to COVID-19. In addition, several COVID-19 patient-based cohort studies have discussed that deaths due to COVID-19 can be determined by age, obesity, and other chronic ailments, such as diabetes, hypertension and heart

* Corresponding Author.

Email Address: shassan1@kau.edu.sa (A. H. Syed)

<https://doi.org/10.21833/ijaas.2022.05.003>

Corresponding author's ORCID profile:

<https://orcid.org/0000-0002-7288-3098>

2313-626X/© 2022 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

illnesses, hospital care, clinical symptoms, prior immunity, and mutations in virus (Chen et al., 2020; Huang et al., 2020; Rubino et al., 2020; Wang et al., 2020; Zhou et al., 2020). Moreover, as mentioned above, the clinical factors help the clinicians classify people at high risk of COVID-19 infection.

Nevertheless, the factors mentioned earlier for explaining the mortality of individuals from COVID-19 are insufficient to support effective policymaking by governments where the COVID-19 related mortality is very high (Liang et al., 2020). Several studies have addressed this gap in COVID-19 research. Few academics have discussed the effect of government policies, namely lockdown or quarantine, to control the spread of COVID-19 (Hou et al., 2020; Iacobucci, 2020).

Additionally, an increase in mass COVID-19 testing and vaccination has been promoted by various countries to decrease the spread of COVID-19 (Dagan et al., 2021; Peto, 2020). However, few researchers have also favored utilization of the hospital resources through the COVID-19 pandemic as it assures the sufficiency of means available to provide treatment to many patients suffering from COVID-19 (Moghadas et al., 2020). Scholars have recently analyzed the correlation between healthcare resources and mortality readiness because of COVID-19 (Ji et al., 2020). Yet, the available evidence in the literature to curtail the spread of COVID-19 has not been applied to accurately explain the significant country-wise deviation in the COVID-19 attributed deaths. Moreover, countries differ extensively in capabilities to detect, prevent, and respond to pandemics (Kandel et al., 2020).

Therefore, Liang et al. 2020, aimed to explore factors associated with COVID-19 associated mortalities at the country level. However, several limitations were related to assessing the total COVID-19 mortality at the country level. The selection of a limited number of COVID-19 related and country-based factors that possibly determine the COVID-19 mortality rate in a country was one of the significant limitations of the study conducted by Liang et al. (2020). Therefore, we aim to explore and explain the influence of several new governments and COVID-19 related factors on COVID-19 based mortalities in the world. In this context, in the current study, cross-sectional data comprising various COVID-19 related and country-based attributes of 181 countries were retrieved and analyzed to screen the most informative factors explaining the variation in the COVID-19 mortalities in the world. Furthermore, the partial correlation studies keeping the partial and complete dosages of COVID-19 vaccines as controlling attributes were performed to understand the possible impact of the partial and complete dosages of the COVID-19 vaccine on the total deaths attributed to COVID-19 in different countries. Specifically, the study examined the association of covariates such as critical cases, new COVID-19 cases per million, aged 65 older, aged 70 older, and total confirmed COVID-19 cases with

the response variable (total deaths attributed to COVID-19), while controlling for partial and complete dosage of COVID-19 vaccine. In addition, we have also examined and discussed the role of attributes, namely total confirmed cases of COVID-19, hospital patients per million, hospital beds per thousand, male smokers, and people fully vaccinated per hundred, explaining the discrepancy in total COVID-19 based mortalities across the globe. Further, the factors screened in the present study might help countries severely affected by COVID-19 formulate policies to attenuate the higher fatality of COVID-19.

2. Materials and methods

The flow diagram for implementing the Machine learning-based approaches to explain the attributes related to discrepancies in total COVID-19 based fatality across countries is shown in Fig. 1.

2.1. Data source and variable description

The COVID-19 cross-sectional dataset is a collection of attributes, namely Vaccinations, Tests and positivity, Hospital and ICU, Confirmed cases, Confirmed deaths, Policy responses, and other variables of interest were being retrieved from 'https://ourworldindata.org/coronavirus', an open-access database for Coronavirus Pandemic (COVID-19). Furthermore, the data regarding the variables, namely "the number of variant sequences" and "total sequences since the first variant sequence," were retrieved from https://cov-lineages.org/ (O'Toole et al., 2021).

2.2. Data preprocessing

2.2.1. Multiple imputations

Missing data is a usual occurrence in cross-sectional datasets. Multiple Imputations (Chang et al., 2020; Buuren, 2018) due to their easiness of usage are possibly the most popular approaches for addressing missing data from sample data. Multiple Imputation (MI) techniques for imputing data can be used where the data are missing at random, missing completely at random, or still when the data are not missing at random. In the present study, understudy cross-sectional data of 181 countries comprising COVID-19 related and country-specific attributes were processed for multiple missing values using the Multiple Imputation by Chained Equations (MICE) package in Azure Machine studio, assuming that the data are missing at random (MAR). MICE has been widely accepted for data imputation and has displayed better performance in practice (Buuren and Groothuis-Oudshoorn, 2011). The imputed cross-sectional data was further tested to satisfy the five regression analysis assumptions (Casson and Farmer, 2014).

2.3. Descriptive statistics for sample data

Descriptive and inferential statistics are employed in scientific data analysis and are crucial in statistics. In the present study, we have discussed descriptive statistics measures to describe the data and popular methods to test the normality of the data (Kim, 2013; Mishra et al., 2019). Three primary

types of descriptive statistics, namely 1) measures of frequency (e.g., Number of occurrences, percent), 2) measure of central tendency (e.g., mean), and 3) measures of variation (variance, standard error, and Standard Deviation) were assessed to provide an understanding of the simple statistical measures of the cross-sectional data.

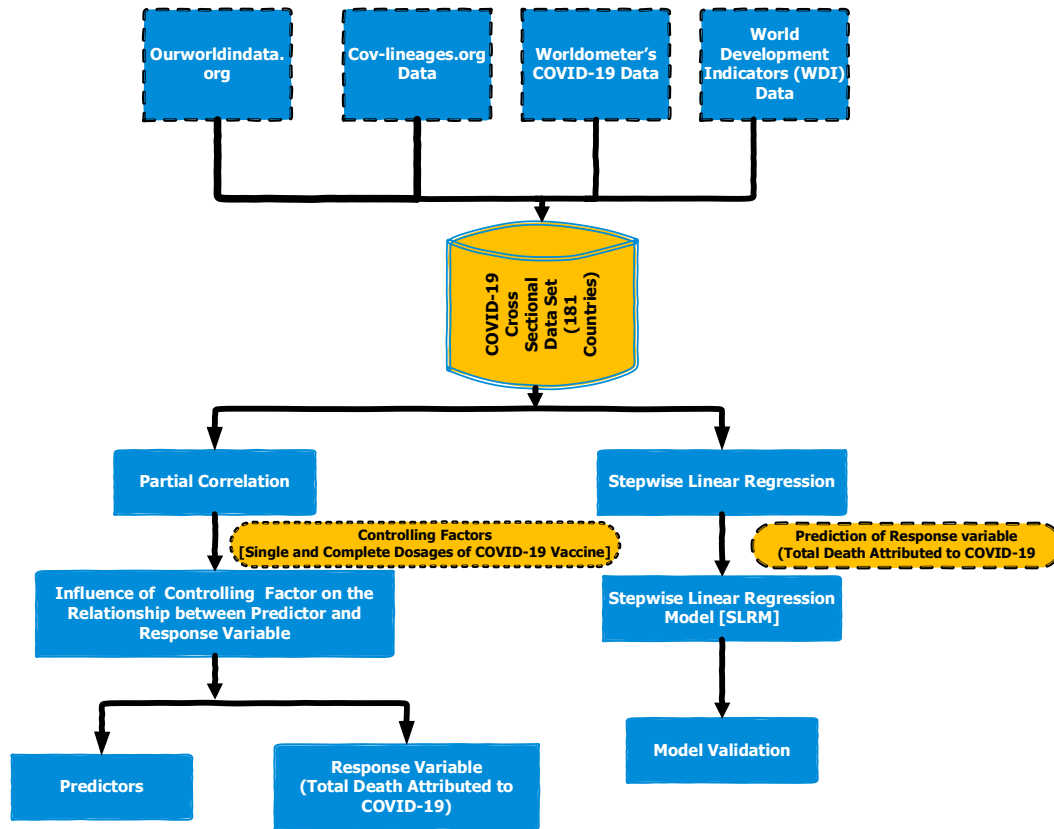


Fig. 1: A flow diagram of the statistical approaches used to evaluate various factors explaining discrepancies in the total deaths attributed to COVID-19 across countries

2.4. Checking for regression assumptions

2.4.1. Testing the normality of data

An evaluation of the normality of data is essential because normally distributed data is a basic assumption for parametric testing. Therefore, the most popular normality testing methods of continuous data are Kolmogorov–Smirnov test (Stephens, 1974), Shapiro–Wilk test (Shapiro and Wilk, 1965), kurtosis and skewness (Kim, 2017), Z-score (Ghasemi and Zahediasl, 2012), and mean with SD (Davis, 2008). The factors under study were analyzed to detect the distribution pattern factors. Normality tests (Shapiro–Wilk test and Kolmogorov–Smirnov test) were performed using the statistical “SPSS” software package (analyze→descriptive statistics→explore).

2.4.2. Probability-based inverse distribution function transformations

Inverse distribution Function Normal (IDF. Normal) transforms a continuous variable's sample

distribution to appear more normally distributed (Beasley et al., 2009). Upon checking for the normality of the sample distribution, the IDF-based transformation was performed using SPSS on the cross-sectional data of 181 countries to control the skewness and satisfy the normality assumption.

2.4.3. Multivariate outlier detection, multicollinearity, homoscedasticity, and normality of the error distribution

In cross-sectional data with multiple factors, the chance of the unusual observation increases, and only a few outliers are enough to alter the mean performance, thereby distorting data results. Therefore, in multivariate statistics, the Mahalanobis distance is one of the most famous methods for detecting outliers in multivariate data (Maesschalck et al., 2000; Grentzelos et al., 2021). Finally, all the outliers detected were removed, and a pruned cross-sectional dataset was created for further analysis. Multicollinearity happens when there is a high correlation between the independent attributes in the dataset (Vatcheva et al., 2016). Multicollinear

features need to be removed as multicollinearity destabilizes the statistical significance of a predictor variable in a model. Multicollinearity between the variables in the cross-sectional data was assessed using the following methods (James et al., 2013) (1) correlation between two independent variables with a cutoff of 0.7. A higher correlation ($r \geq \pm 0.7$), i.e., closer to a positive or negative one between the predictor variables, indicates multicollinearity, (2) Variance Inflation Factor (VIF) with a cutoff value less than five, and (3) Tolerance (TOI) of each variable with a cut off value less than 0.2. Breusch-Pagan Test (Breusch and Pagan, 1979) was performed to test the assumption that the residuals are distributed independently with the predictor variables (homoscedasticity). The Breusch-Pagan Test uses the following hypotheses:

Null Hypothesis (H_0)=Residuals are distributed with equal error variance (Homoscedasticity is present)

Alternative Hypothesis (H_A)= the residual is not distributed with equal variance (Heteroscedasticity is present)

The normality of the error distribution (unstandardized residuals) was tested by performing Kolmogorov-Smirnov and Shapiro-Wilk tests (Ghasemi and Zahediasl, 2012). The Kolmogorov-Smirnov test and Shapiro-Wilk tests use the following hypotheses:

Null Hypothesis (H_0)= the unstandardized residuals are normally distributed

Alternative Hypothesis (H_A)= the unstandardized residuals are not normally distributed

If the p-value of the Kolmogorov-Smirnov and Shapiro-Wilk normality tests are more than the significance level of 0.05, accept the null hypothesis and conclude that unstandardized residuals are normally distributed.

2.5. Partial correlation: The relationship between total deaths and the independent variable

Partial correlation was used to estimate the strength of the relationship between the predictors and the response variable (total deaths attributed to COVID-19), keeping dosages of the COVID-19 vaccines as controlling factors. The COVID-19 vaccines approved in the United States of America and European countries have shown effectiveness against hospitalization and death in trials and real-time across the globe (Breusch and Pagan, 1979; Chodcik et al., 2021). Thus, there is consensus that the COVID-19 vaccination may end the COVID-19 pandemic by 2022. However, in the current setting, the influence of immunization on the relationship between total deaths attributed to COVID-19 (response variable) and government factors (predictor variable) is still not apparent. Therefore,

in this study, we intend to explore the influence of the vaccination dosages (single and complete dosages of COVID-19 vaccine) on the relationship between the total death attributed to COVID-19 and the COVID-19 related government factors. The Partial Correlation between the predictors and the response variables while controlling for dosages of the COVID-19 vaccine was performed using the statistical "SPSS" software package (Correlate>partial correlation).

2.6. The stepwise linear regression model

Stepwise Linear regression analyses (Hocking, 1976) were performed to study the relationship between the total deaths attributed to COVID-19 and seventeen normally distributed and non-collinear predictors as tabulated in Section 3.3.

2.7. Validation study of the stepwise regression model

The validity of our final stepwise regression model was considered by evaluating the observed total COVID-19 deaths against the predicted total deaths attributed to COVID-19 for each country. First, we drew a graph with observed and predicted total deaths due to COVID-19 on the axis of the two-dimensional graph. We anticipated seeing the sample points distributed around the 45-degree cross line on the two-dimensional chart provided the model fits well.

3. Results

3.1. Multiple imputations using mice

The missing data in our cross-sectional dataset were artificially generated, and the imputed cross-sectional dataset was tested for the five assumptions of the regression analysis.

3.2. Normality tests and descriptive statistics for sample data

The cross-sectional data of 181 countries comprising COVID-19 related and country-based factors exhibited marked skewness. Therefore, using the inverse distribution function normal transformation, the skewed cross-sectional data of 181 countries was transformed to correct the skewness and fulfill the assumptions of normalized data. In addition, the normality of the transformed cross-sectional data was assessed using the different normality testing methods. Table 1 summarizes the sample characteristics of the predictors (Standard error, mean, 95% Confidence Interval for Mean, Lower and upper bound, and standard deviation value). In addition, Table 2 summarizes the transformed cross-sectional data's normality test and kurtosis, skewness, and normality test values. The results observed in Tables 1 and 2, respectively,

showed that the transformed data of various COVID-19 related and country-based variables were

normally distributed.

Table 1: Descriptive statistics of COVID-19 and country-related factors post-probability-based inverse distribution function normal transformation

S#	Attributes (transformed using IDF)	Mean	Std. Error	95% Confidence Interval for Mean		Std. Deviation
				Lower Bound	Upper Bound	
1	New_cases_per_million	30609.76	2566.57	25445.70	35674.39	34434.16
2	Human_development_index	0.7262	.01116	.7042	.7482	.15018
3	Life_expectancy	72.8317	.56163	71.7234	73.940	7.53508
4	Hospital_beds_per_thousand	2.9207	.17285	2.5796	3.2618	2.31904
5	Handwashing_facilities	74.2352	4.3971	65.5582	82.9122	58.99423
6	male_smokers	31.5393	.90032	29.7627	33.3159	12.0790
7	Female_smokers	9.6926	.69970	8.3119	11.0733	9.38744
8	Diabetes_prevalence	7.6438	.27403	7.1031	8.1846	3.67656
9	Gdp_per_capita	18909.84	1413.375	16120.82	21698.86	18962.427
10	Aged_70_older	5.6091	.30910	4.9992	6.2191	4.14704
11	Aged_65_older	8.7948	.45075	7.9054	9.6843	6.04743
12	Population_density	312.3852	113.39690	88.6185	536.1519	1521.37906
13	Population	42766008.39	11135465.14	20792334.51	64739682.27	149397942.16
14	Stringency_index	84.4384	.96951	82.5254	86.3515	13.04336
15	People_fully_vaccinated_per_hundred	4.9959	.52332	3.9632	6.0285	7.02102
16	Vaccination_Single_per_Hundred	9.6815	.91116	7.8835	11.4795	12.22453
17	Total_vaccinations_per_hundred	13.8042	1.45356	10.9358	16.6725	19.50157
18	positive_rate	.2685	.00867	.2514	.2856	.11628
19	Total_Test	14585020.40	2845999.24	8968994.63	20201046.16	38183086.59
20	Hospital_patients_per_million	350.1778	20.60461	309.5186	390.8370	276.43988
21	Reproduction_rate	1.9960	.05445	1.8886	2.1035	.73051
22	Total_cases_per_million	30663.6929	2568.26478	25595.7221	35731.6637	34456.88
23	total_deaths	16696.3128	4157.58	8492.12	24900.50	55779.898
24	total_cases	781237.3039	210609.28	365640.85	1196833.75	2825620.10
25	Mortality_rate	.0203	.00136	.0176	.0230	.01820
26	Number of variant sequences	2708.0580	1124.61759	488.8440	4927.2720	15088.32827
27	Total sequences since first variant sequence	5772.7552	2066.48642	1694.9463	9850.5640	27724.82463
28	LPI	2.8314	.03864	2.7551	2.9076	.51845
29	Communicable diseases 2016	22.4214	1.47904	19.5028	25.3400	19.84338
31	Gov-Effectiveness-2019	-.0553	.07308	-.1995	.0889	.98323
32	Critical_Cases	770.6191	99.71267	573.855	967.382	1337.78

Table 2: Summarizes the transformed cross-sectional data's normality test and kurtosis, skewness, and normality test values

S#	Attributes (transformed using IDF)	Skewness			Kurtosis			Kolmogorov-Smirnov		Shapiro-Wilk	
		Skewness	Std. Error	Z-value	Kurtosis	Std. Error	Z-value	df	Sig.	df	Sig.
1	New_cases_per_million	.002	.181	0.011	-.242	.360	-0.672	181	.200*	181	1.00
2	Human_development_index	.070	.181	0.3867	-.152	.359	-0.423	181	.200*	181	.925
3	Life_expectancy	.006	.181	0.0331	-.242	.360	-0.672	181	.200*	181	1.00
4	Hospital_beds_per_thousand	.003	.181	0.0165	-.250	.360	-0.694	181	.200*	181	0.998
5	Handwashing_facilities	.006	.181	0.0331	-.242	.360	-0.672	181	.200*	181	1.00
6	male_smokers	.006	.181	0.0331	-.242	.360	-0.672	181	.200*	181	1.00
7	Female_smokers	.003	.181	0.0165	-.248	.360	-0.688	181	.200*	181	1.00
8	Diabetes_prevalence	.029	.181	0.1602	-.280	.360	-0.777	181	.200*	181	.847
9	Gdp_per_capita	.004	.181	0.0220	-.242	.360	-0.672	181	.200*	181	1.00
10	Aged_70_older	.006	.181	0.0331	-.242	.360	-0.672	181	.200*	181	1.00
11	Aged_65_older	.006	.181	0.0331	-.242	.360	-0.672	181	.200*	181	1.00
12	Population_density	.002	.181	0.0110	-.242	.360	-0.672	181	.200*	181	1.00
13	Population	.007	.181	0.0386	-.242	.360	-0.672	181	.200*	181	1.00
14	Stringency_index	-.076	.181	-0.419	-.435	.359	-1.211	181	.200*	181	0.43
15	People_fully_vaccinated_per_hundred	.002	.181	0.0110	-.246	.360	-0.683	181	.200*	181	1.00
16	Vaccination_Single_per_Hundred	.008	.181	0.0441	-.259	.360	-0.719	181	.200*	181	0.999
17	Total_vaccinations_per_hundred	.008	.181	0.0441	-.259	.360	-0.719	181	.200*	181	0.999
18	positive_rate	.001	.181	0.0055	-.244	.360	-0.677	181	.200*	181	0.989
19	Total_Test	0.005	.181	0.0276	-.242	.360	-.6722	181	.200*	181	1.00
20	Hospital_patients_per_million	0.001	.181	0.0055	-.242	.360	-0.672	181	.200*	181	1.00
21	Reproduction_rate	0.00	.181	0.0110	-.241	.360	-0.669	181	.200*	181	1.00
22	Total_cases_per_million	0.002	.181	0.0111	-.241	.360	-0.669	181	.200*	181	1.00
23	total_deaths	.006	.181	0.0331	-.257	.360	-0.713	181	.200*	181	0.988
24	total_cases	0.001	.181	0.0055	-.242	.360	-0.672	181	.200*	181	1.00
25	Mortality_rate	0.001	.181	0.0055	-.242	.360	-0.672	181	.200*	181	1.00
26	Number of variant sequences	.053	.181	0.2928	-.358	.360	-0.994	181	.200*	181	0.464*
27	Total sequences since first variant sequence	.004	.181	0.0220	-.250	.360	-0.694	181	.200*	181	1.00
28	LPI	-.001	.181	-0.0055	-.243	.360	-0.67	181	.200*	181	1.00
29	Communicable diseases 2016	-.001	.181	-0.0055	-.248	.360	-0.688	181	.200*	181	1.00
31	Gov-Effectiveness-2019	.131	.181	0.72375	-.335	.359	-0.933	181	.200*	181	.068*
32	Critical_Cases	.035	.181	0.1933	-.324	.359	-0.902	181	.200*	181	.748

3.3. Regression assumptions analysis of the cross-sectional data

The multivariate outliers detected using the Mahalanobis distances were removed, and the pruned dataset was processed further for regression assumption analysis. The pruned cross-sectional dataset with thirty-one variables was further checked for multicollinearity, homoscedasticity, and normality of the error distribution of the residuals. As summarized in Table 3, we can observe that the

values of the VIF and Tolerance for the seventeen explanatory variables are below five ($1 < \text{VIF} < 5$), and more than 0.2, respectively; this shows that the features space as shown in Table 3 are low to moderately correlated. Upon performing the Breusch-Pagan Test, a p-value of 0.165954 corresponds to a Chi-Square (χ^2) 21.336 with 16 degrees of freedom was obtained on the pruned cross-sectional data. Since the p-value is larger than the significance level ($\alpha=0.05$), we accept the null hypothesis, which states that the residuals are

distributed with equal variance (homoscedasticity is present in the cross-sectional data with 16 predictor

variables).

Table 3: Attributes with low to moderate correlation after multicollinearity analysis

S#	Features	Coefficients ^a	
		Collinearity Statistics	
		Tolerance	VIF
1	Hospital Bed Per million (IDF)	.291	3.432
2	Male Smoker (IDF)	.604	1.656
3	Female Smoker (IDF)	.424	2.357
4	Diabetes Prevalence (IDF)	.615	1.627
5	GDP (IDF)	.257	3.893
6	Population_density (IDF)	.838	1.193
7	Stringency_index (IDF)	.782	1.279
8	Positive_rate (IDF)	.514	1.944
9	Total_Test (IDF)	.695	1.439
10	Hospital Patient Per Million (IDF)	.203	4.921
11	Reproduction Rate (IDF)	.438	2.282
12	Total Cases (IDF)	.303	3.305
13	LPI (IDF)	.411	2.433
14	Communicable Disease (IDF)	.352	2.840
15	Critical Cases (IDF)	.634	1.576
16	Fully Vaccinated Per 100 (IDF)	.363	2.756

^aDependent Variable: Total Death (IDF)

Graphically the homoscedasticity of the cross-sectional data is also represented in Fig. 2. As depicted in Fig. 2, the variance of the residuals is constant across all the standardized predicted values. Thus, the constant residual variance signifies the homoscedastic nature of the cross-sectional data. Finally, the normality of the error distribution (unstandardized residuals) was tested by performing Kolmogorov-Smirnov and Shapiro-Wilk

tests. The p-value of the unstandardized residuals was 0.200 and 0.892, respectively. Thus, the p-values of the unstandardized residual calculated for the Kolmogorov-Smirnov and Shapiro-Wilk tests were greater than the significance level of 0.05. Therefore, we accept the null hypothesis, which states “the unstandardized residuals are normally distributed,” and reject the alternate hypothesis.

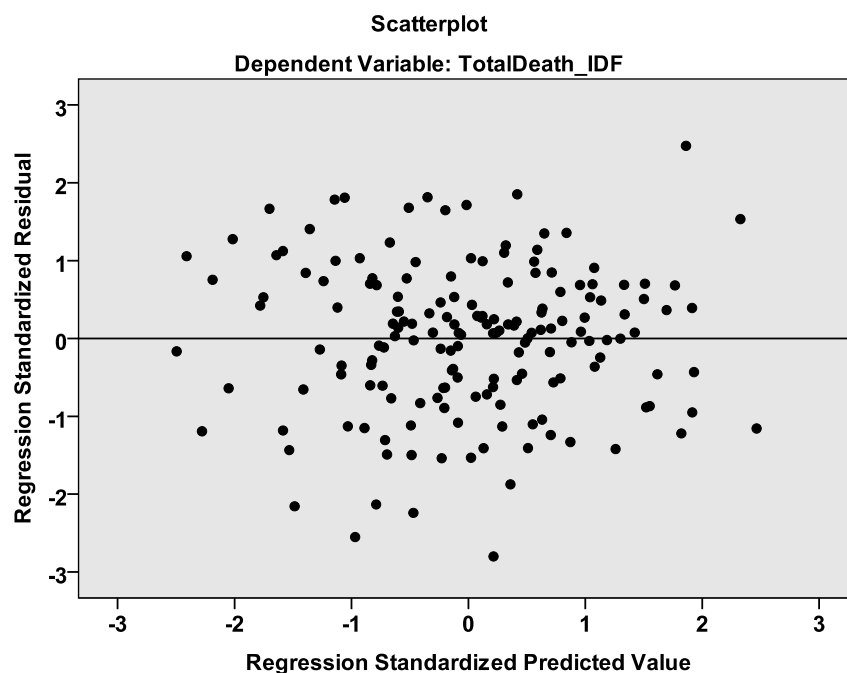


Fig. 2: Linearity analysis of the cross-sectional data using the residuals against the predicted values

3.4. Partial correlation: The relationship between total deaths and the explanatory variables

The relationship between the total death attributed to COVID-19 and the seventeen independent variables in the pruned cross-sectional data was explored using partial and zero-order

correlation while controlling for partial and complete dosages of the COVID-19 vaccine. Partial correlation of independent variable(s), namely critical cases, new COVID-19 cases per million, aged_65_older, aged_70_older, and total confirmed COVID-19 cases, using a partial and complete dosage of COVID-19 vaccine dosage, respectively as the controlling factor, presented a significant

relationship with the total deaths attributed to COVID-19 (dependent variable) as shown in Figs. 3a-

3j and Figs. 4a-4j, respectively.

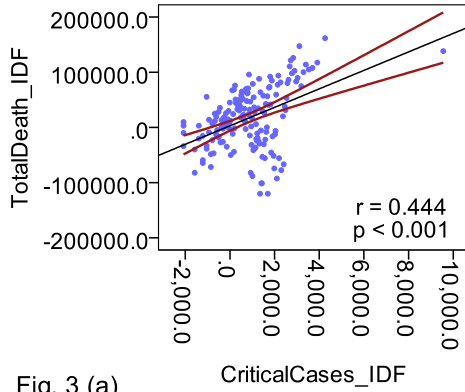


Fig. 3 (a)

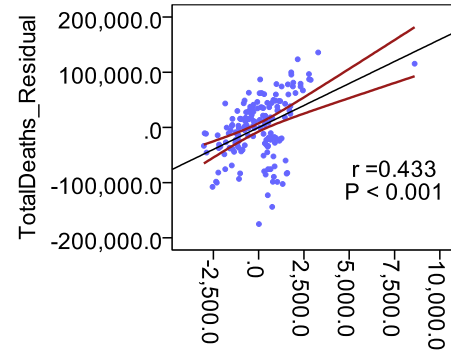


Fig. 3 (b)

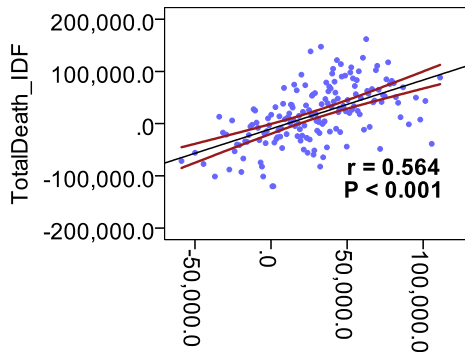


Fig. 3 (c)

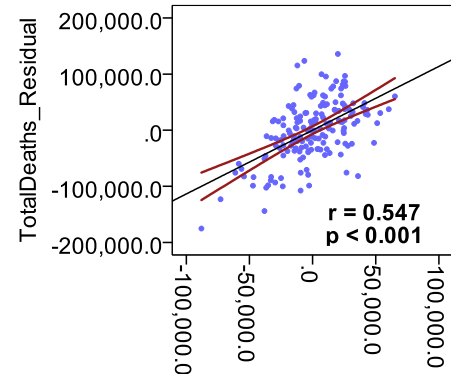


Fig. 3 (d)

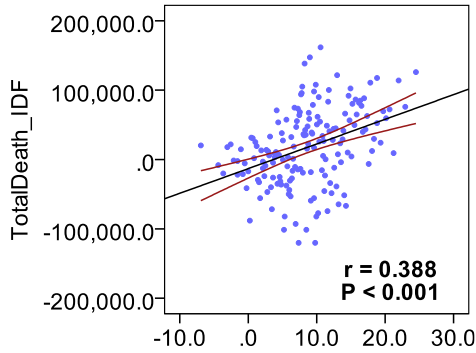


Fig. 3 (e)

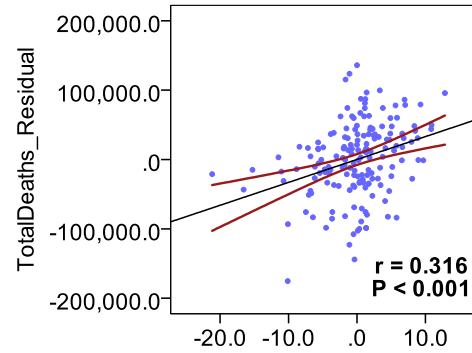


Fig. 3 (f)

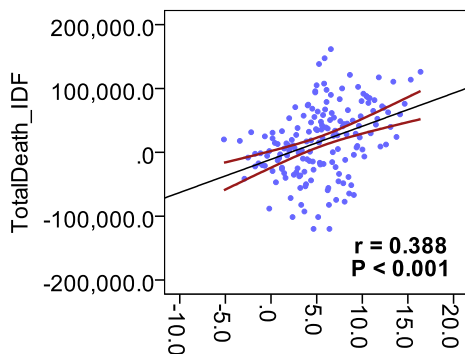


Fig. 3 (g)

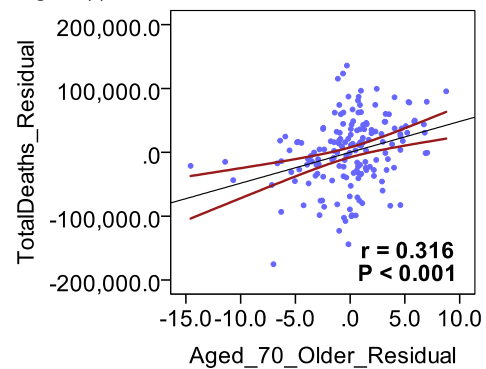


Fig. 3 (h)

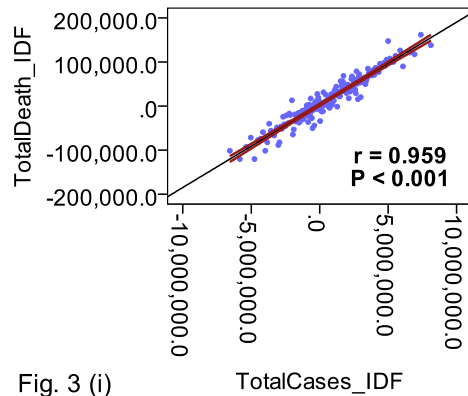


Fig. 3 (i)

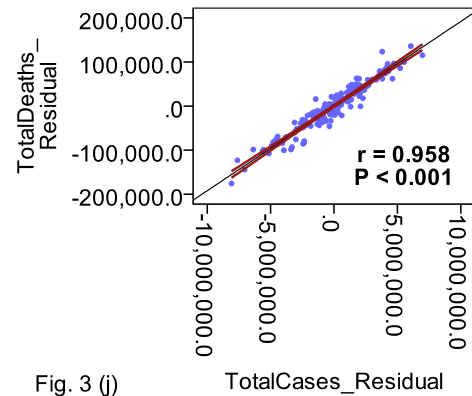


Fig. 3 (j)

Fig. 3: Figs. 3 (a-j): Where Fig. 3a represents the Zero-order and Fig. 3b displays a Partial Correlation between total deaths attributed to COVID-19 and critical cases while controlling for a partial dosage of COVID-19 vaccine; While Fig. 3c shows the Zero-order and Fig. 3d represents a Partial Correlation between total deaths attributed to COVID-19 and new COVID-19 cases per million while controlling for a partial dosage of COVID-19 vaccine; Fig. 3e depicts the Zero-order and Fig. 3f displays a Partial Correlation between total death attributed to COVID-19 and aged_65_older while controlling for a partial dosage of COVID-19 vaccine; Fig. 3g represents the Zero-order and Fig. 3h illustrates a Partial Correlation between total death attributed to COVID-19 and aged_70_older while controlling for a partial dosage of COVID-19 vaccine and total confirmed COVID-19 cases; Fig. 3i displays the Zero-order and Fig. 3j represents a Partial Correlation between total deaths attributed to COVID-19 and total cases while controlling for a partial dosage of COVID-19 vaccine. Linear Lines in Figs. 3(a-j) are linear predictions of the dependent and the specific independent variable. The area between the two red fitted lines represents the 95% confidence intervals of the fitted values across the cross-sectional data (r: correlation coefficient)

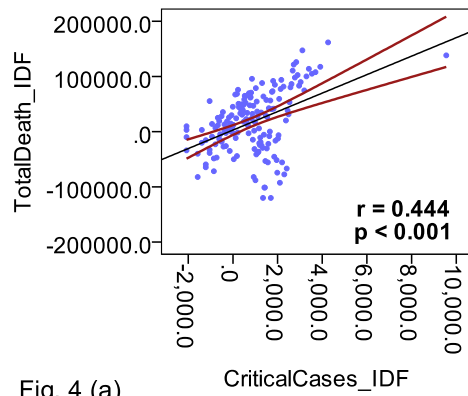


Fig. 4 (a)

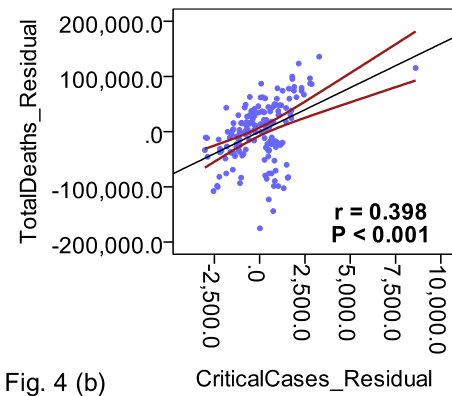


Fig. 4 (b)

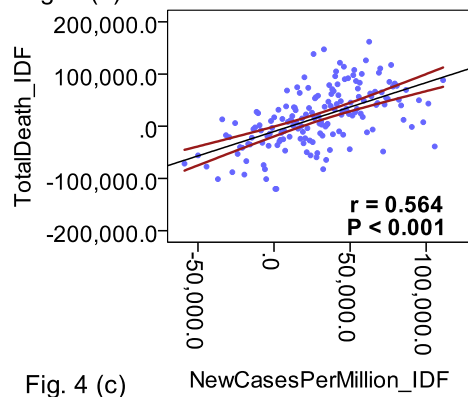


Fig. 4 (c)

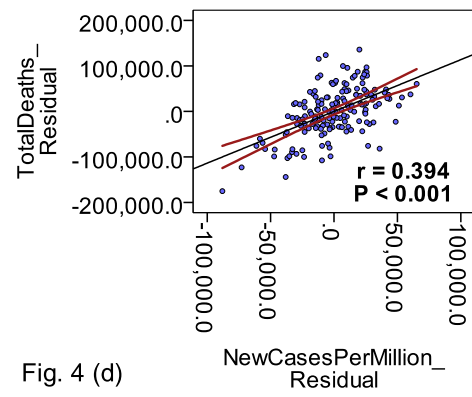


Fig. 4 (d)

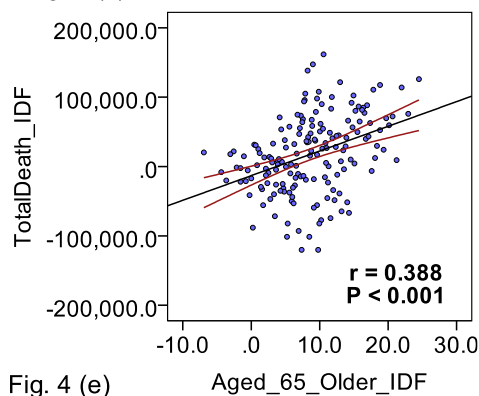


Fig. 4 (e)

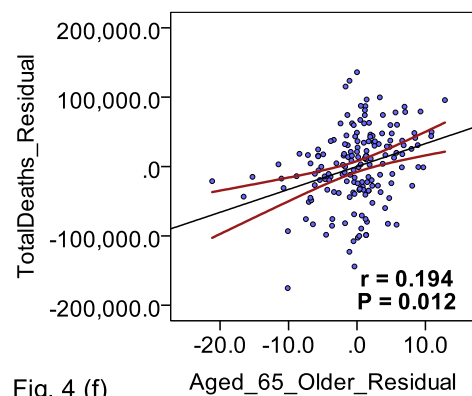


Fig. 4 (f)

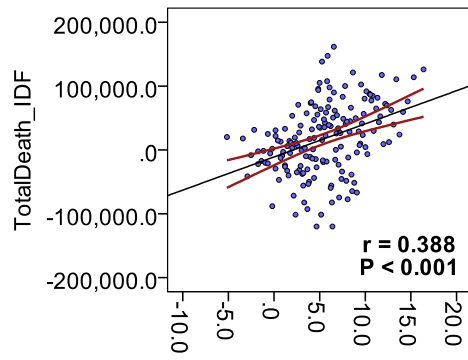


Fig. 4 (g)

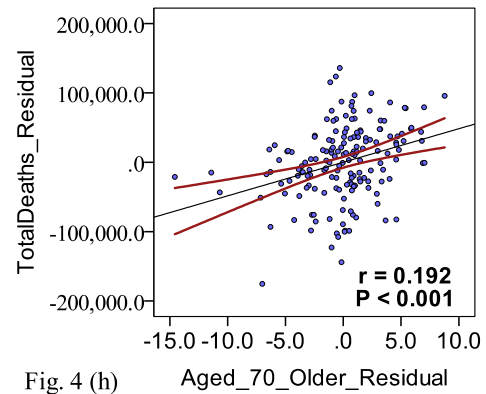


Fig. 4 (h)

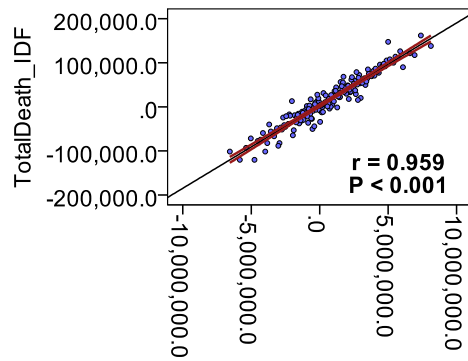


Fig. 4 (i)

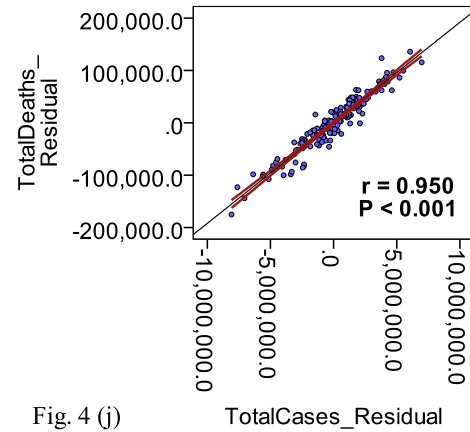


Fig. 4 (j)

Fig. 4: Fig. 4a represents the Zero-order, and Fig. 4b shows the Partial Correlation between total death attributed to COVID-19 and critical cases while controlling for complete dosage of COVID-19 vaccine; Fig. 4c represents the Zero-order, and Fig. 4d depicts the Partial Correlation between total death attributed to COVID-19 and total new confirmed COVID-19 cases per million while controlling for complete dosage of COVID-19 vaccine; Fig. 4e represents the Zero-order, and Fig. 4f represents the Partial Correlation between total death attributed to COVID-19 and aged_65_older while controlling for complete dosage of COVID-19 vaccine. Fig. 4g illustrates the Zero-order, and Fig. 4h depicts the Partial Correlation between total death attributed to COVID-19 and aged_70_older while controlling for the complete dosage of the COVID-19 vaccine. Finally, Fig. 4i represents the Zero-order, and Fig. 4j illustrates the Partial Correlation between total death attributed to COVID-19 and total cases while controlling for the complete dosage of the COVID-19 vaccine. Linear Lines in Figs. 4 (a-j) are linear predictions of the dependent and the specific independent variable. The area between the two red fitted lines represents the 95% confidence intervals of the fitted values across the cross-sectional data (r: correlation coefficient)

3.5. Stepwise linear regression analysis

Stepwise linear regression was performed using a set of attributes tabulated in Table 4 to build a regression model with variables that significantly predict the response variable (total death attributed to COVID-19). Table 4 summarizes the findings of the

stepwise linear regression model. The variables total_cases, hospital patients per million, hospital beds per thousand, male smokers, and people fully vaccinated per hundred in combination significantly predicted the total deaths attributed to COVID-19, ($F(6, 161)=617.691$, $p<0.001$), $R\text{-Square}=0.9584$.

Table 4: Final predictive model summary

Stepwise regression model Predictors	R Square	Adjusted R Square	df	F	Sig.
Final Predictive Model	0.958367	0.956816	161	617.6914	2.26E-108

Table 5 summarizes the final stepwise regression model predictors for predicting the total death attributed to COVID-19. In addition, the absolute value of the standardized regression coefficients can be compared, thereby providing a rough indication of the importance of the variables. For example, among the COVID-19 related factors, the most significant absolute standardized value is 0.9503 for the total confirmed cases of COVID-19 (95% CI 0.0178 to 0.0193, $P<0.001$), suggesting that the total confirmed cases of COVID-19 are the most important of the five model predictors in predicting the total deaths attributed to COVID-19. Next is hospital

patients per million, with a positive and statistically significant standardized coefficient value of 0.307 (95% CI 51.165510 to 72.318343, $P<0.001$). Then again, among the government-related factors, the Hospital bed per thousand is the most critical factor with a negative standardized coefficient value of -0.194 (95% CI -6010.567860 to -3431.576739, $P<0.001$), suggesting that a decrease of one standard deviation in hospital beds per thousand will result in an expected increase of total deaths attributed to COVID-19 by 0.194 standard deviations. Next most contributing government-related factors, with a negative standardized coefficient value of -0.114297

for fully vaccinated people per hundred (95% CI -1381.021441 to -579.197698, $P < 0.001$), suggesting that a decrease of one standard deviation in hospital beds per thousand will result in an expected rise of total deaths attributed to COVID-19 by 0.114297 standard deviations. Finally, the smallest absolute

value is $-0.112518 = 0.113$ for male smokers (95% CI -647.070996 to -315.907098, $P < 0.001$), suggesting that the male smokers are the least among the five predictors in predicting the total deaths attributed to COVID-19 for an individual country.

Table 5: Stepwise linear regression model for predicting the total deaths attributed to COVID-19

Final Stepwise regression model Predictors	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	10288.95	3553.24	2.895651	4.58755	0.000009	6405.1931	16087.743688
TotalCases (IDF)	0.0185	0.00038	0.9503	48.916	1.52E-98	0.0178	0.019337
Hospital Patient Per Million (IDF)	61.7419	5.355668	0.30679	11.528334	8.43E-23	51.165510	72.318343
Hopital bed per thousand (IDF)	-4721.072	652.9725	-0.19437	-7.230123	1.84E-11	-6010.5678	-3431.576739
Male Smoker (IDF)	481.48904	83.847109	-0.112518	-5.742465	4.53E-8	-647.07099	-315.907098
Fully Vaccinated Per hundred (IDF)	-980.1095	203.013079	-0.114297	-4.827815	0.000003	-1381.02144	-579.197698

3.6. Validation of the stepwise regression model

The validation of the current stepwise regression model is pictorially represented in Fig. 5 by plotting the standardized predicted values of the total death on the x-axis and the observed total deaths due to

COVID-19 on the y-axis. The standardized predicted value of the total deaths was positively and significantly correlated with the observed total deaths due to COVID-19 ($r = 0.977$, $P < 0.001$).

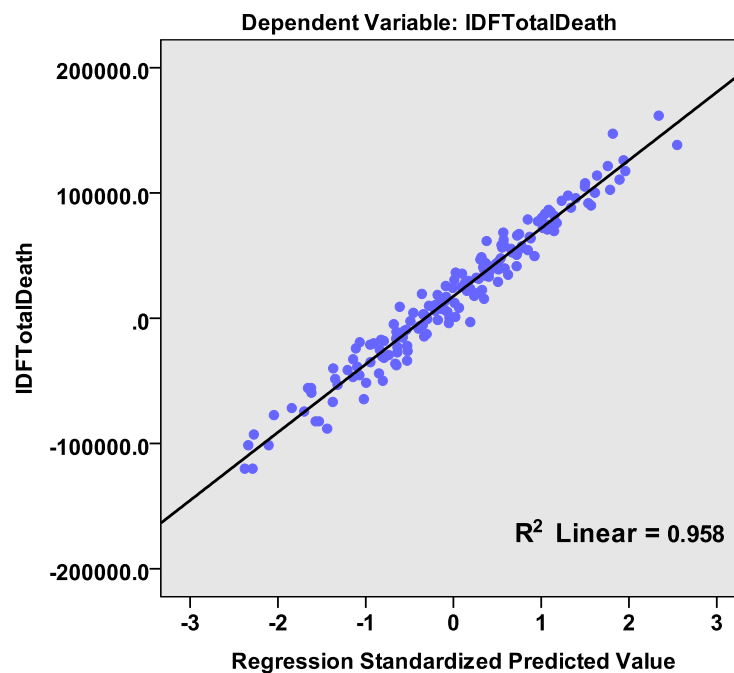


Fig. 5: Correlation between predicted and observed total deaths attributed to COVID-19

3.7. Comparison with similar studies

The performance of our stepwise linear regression model in predicting the total mortality of countries was compared to model performance from a related study conducted by Liang et al. (2020), shown in Table 5. We can observe from Table 6 that the R-squared value and the adjusted R-squared value of our stepwise MLR model are 0.958 and 0.9568, respectively. Our model's R-squared value

and the adjusted R-squared value are significantly better than the proposed regression model proposed by Liang et al. (2020) (R-squared value was 0.58; adjusted R-squared value was 0.54). Moreover, the correlation coefficient value ($r = 0.977$, $P < 0.001$) between the predicted value and the observed total death of our model was comparatively better than the model proposed by Liang et al. (2020) ($r = 0.77$; $P < 0.001$).

Table 6: Comparative analysis of R-squared, adjusted R-square value, and the correlation coefficient between the predicted and the observed total deaths between regression models

Model	R-squared value	Adjusted R-squared value	The correlation coefficient of the correlation between predicted and observed total deaths at a significance level of $p < 0.001$
Liang et al., 2020	0.58	0.54	0.77
Our Proposed Stepwise MLR model	0.958	0.9568	0.977

4. Discussion

Cross-sectional data from 181 countries with various COVID-19 and Government-related factors were preprocessed for the missing values in the cross-sectional data of 18 countries. First, the missing values were artificially created using the MICE package in Azure Machine Learning Studio Classic. Then, the imputed cross-sectional data were further processed to satisfy the correlation and regression analysis assumptions, namely normality, multivariate outliers, multicollinearity, linearity, homoscedasticity, and normality of the error distribution of the cross-sectional data. Finally, the pruned preprocessed cross-sectional dataset comprising sixteen attributes from 161 countries was further used for correlation and regression studies.

Notably, the zero-order and partial correlation analyses as illustrated in Figs. 3a-3j and Figs. 4a-4j showed that controlling for a complete dosage of COVID-19 vaccine over a partial dosage of COVID-19 vaccine per 100 people in the population had a moderate but significant influence on the strength of correlation between the total death attributed to COVID-19 (response variable) and the independent variables, namely critical cases, new COVID-19 cases per million, aged_65_older, and aged_70_older.

However, controlling for each partial and complete dosage of the COVID-19 vaccine had a statistically significant but negligible effect on the relationship between the total deaths due to COVID-19 (dependent variable) and the total confirmed COVID-19 cases (independent variable). The negligible impact of the partial and complete dosage of the COVID-19 vaccine can be attributed to the emergence of novel variants of COVID-19 (Grubaugh et al., 2021; Winblad et al., 2004) that have a selective advantage of enhancing transmission dynamics (Davies et al., 2021a; Dorp et al., 2021) and the capability to reduce prompt neutralization by the host (Planas et al., 2021; Singh et al., 2021). Thus, for example, the present restrictions in India and Europe (Priesemann et al., 2021; Singh et al., 2021) are in because of a more communicable (Davies et al., 2021a; Dorp et al., 2021) and principally more pathogenic (Challen et al., 2021; Davies et al., 2021b) B.1.1.7 variant that originated in the United Kingdom and is rapidly gaining dominance in a country such as India (Singh et al., 2021).

Therefore, based on our current observation of the overall results of the partial correlation, we recommend that governments of different countries should emphasize more on immunizing their residents with all doses prescribed by the vaccination protocol rather than a partial vaccine

dose per 100 people in the total population to counter the surge in the total death attributed to COVID-19.

Additionally, the current country-level study comprehensively examines the association of various COVID-19 related and country-specific factors with COVID-19 mortality. The stepwise multiple regression model analysis reveals that the estimated regression line (predictors) of the model explains 95.8 % (R-square=0.958) variation in the response variable (total COVID-19 mortality). A statistically significant and higher F-statistics of the final stepwise regression model shows that the independent variables help explain total deaths attributed to COVID-19 (response variable) with a confidence greater than 99.999. The detailed analysis of the attributes contributing significantly to the stepwise multiple regression model has shown that total deaths attributed to COVID-19 are positively associated with the total confirmed cases of COVID-19 and the number of COVID-19 patients in the hospital on a given day per 1,000,000 people. Therefore more confirmed cases of COVID-19 are associated with an increase in COVID-19 mortality (Sarkodie and Owusu, 2020).

Moreover, more COVID-19 patients admitted to a hospital on a given day are associated with increased total deaths attributed to COVID-19. Thus, an increase in total deaths of patients admitted to hospital with COVID-19 infection may be attributed to a rapid emergence of new cases, leading to rapidly increasing demand for patients' facilities and multiple negative factors and symptoms (end-stage renal diseases, age, diabetes, heart diseases, etc.) associated with hospitalized COVID-19 patients (Alwafi et al., 2021; Rossman et al., 2021). Alternatively, our stepwise regression model also showed a negative association between total deaths attributed to COVID-19 (response variable) and the independent variables (hospital beds per thousand, male smokers, and people fully vaccinated).

The people fully vaccinated have a significant negative unstandardized coefficient, which means that in countries where people received all doses of the COVID-19 vaccine prescribed by the vaccination protocol per 100 people in the total population, the total death due to COVID-19 significantly decreases (Haas et al., 2021), offering hopefulness that COVID-19 vaccination will ultimately stem the COVID-19 pandemic as vaccination programs rise across the globe. In addition, a negative coefficient value for a male smoker signifies a negative association between smoking and COVID-19 infection, indicating a reduced mortality risk in smokers due to COVID-19. Several recent studies have shown that smoking reduces the risk of people getting infected with

COVID-19 disease (Ward et al., 2021). However, the current findings that smoking is inversely associated with COVID-19 infection challenge the fact that smokers are more susceptible to respiratory illnesses, including COVID-19 infection (Hopkinson et al., 2021; Westen-Lagerweij et al., 2021). Therefore, further investigations are needed to clarify the role of cigarette smoking on COVID-19 disease and the reported low prevalence of smokers amongst the current patients diagnosed with COVID-19 infection.

Moreover, the multiple regression analyses indicated a negative association between COVID-19 mortality and hospital beds per thousand people. Thus, the negative association between COVID-19 mortality and hospital beds per thousand people suggests increasing hospital beds per thousand people might serve as a practical methodology to decrease deaths attributed to COVID-19 mortality by governments that were less effective in controlling disease outbreaks when hospital beds were not sufficient in providing proper healthcare to many patients with COVID-19 (Liang et al., 2020; Sen-Crowe et al., 2021). Additionally, from the absolute t-value of predictors, as shown in Table 2, we can presume that countries should focus more on hospital beds per thousand, people fully vaccinated per hundred, and male smokers, respectively counter the increase in total mortality attributed to COVID-19.

Hence we can conclude from the comprehensive analysis of the stepwise multiple regression that the predictors of total cases, hospital patients per million, hospital beds per thousand, male smokers, and people fully vaccinated per hundred were added statistically to predict the total deaths attributed to COVID-19. Furthermore, the present final stepwise regression model was validated by plotting the standardized predicted values of the total death due to COVID-19 on the x-axis and the observed total deaths due to COVID-19 on the y-axis, as shown in Fig. 5. A very strong positive and significant correlation ($r=0.977$, $P<0.001$) was observed between the standardized predicted values and the observed values of the response variable (total deaths attributed to COVID-19). A significant correlation observed between the predicted, and the experimental values of the response variable (total deaths attributed to COVID-19) can be attributed to the 45° angle line between the observed and predicted total deaths due to COVID-19. Moreover, the comparative model validation study revealed that the correlation between the observed total COVID-19 mortality and the predicted total COVID-19 mortality in various countries of the Stepwise Linear Regression model ($r=0.977$, $P<0.001$) was significantly better than Liang et al., 2020 regression model ($r=0.77$; $P<0.001$). Our Stepwise Linear Regression model performs significantly better in predicting the total COVID-19 mortality of countries with an R-squared value of 0.958 and an adjusted R-squared value of 0.956 than the regression model

proposed by Liang et al. (2020) (R-squared value was 0.58; adjusted R-squared value was 0.54).

However, there are certain limitations to the present study. First, the presence of multivariate outliers limits our analyses to cross-sectional data from 161 countries instead of 183 countries. Second, is the lack of completeness of countries' data (missing values). The missing values in the cross-sectional data were artificially created using the MICE module, assuming that the data are missing at random. However, real-time data instead of artificially synthesized data would have provided a more real personification of the factor(s) influencing the total death attributed to COVID-19 (response variable). Third, a sudden surge in COVID-19 related deaths and inaccurate reporting of fatalities attributed to COVID-19 from different countries may significantly influence the predictive nature of our model. However, the prognostic factors and their trends for predicting the total deaths attributed to COVID-19 might not change. Finally, the immunity acquired after the global spread of COVID-19 might influence the prediction accuracy of our stepwise regression model.

This study also has its strengths. This study involves analyzing many COVID-19 and government factors positively and negatively impacting the total deaths attributed to COVID-19 in a country. The results of our research possibly will contribute toward effective policymaking at the country level to counter the sudden surge in the number of deaths attributed to COVID-19 and variants of COVID-19. Based on our current findings, we conclude that governments with better hospital infrastructure and people with complete dosages of the COVID-19 vaccine will have minimal COVID-19 fatalities.

5. Conclusion

In conclusion, we can conclude that a higher number of fatalities attributed to COVID-19 and its variants are positively associated with the total confirmed cases of COVID-19 and the number of COVID-19 patients in the hospital on a given day per 1,000,000 people. While the total COVID-19 mortality is negatively associated with hospital beds per thousand, male smokers, and people fully vaccinated per 100 people from the total population. So, based on the current regression and partial-correlation studies, we presume that the countries which focus on improving the infrastructure of the hospitals by providing more hospital beds, providing better governance policies to counter the spread of COVID-19 and its variants, thereby a load of patients in hospitals are reduced. Moreover, providing a complete vaccination dosage of the COVID-19 vaccine to the entire population will undoubtedly lower the COVID-19 related mortalities.

Acknowledgment

This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University,

Jeddah, under Grant G:-186-132-1442. The authors, therefore, gratefully acknowledge DSR's technical and financial support.

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

- Alwafi H, Naser AY, Qanash S, Brinji AS, Ghazawi MA, Alotaibi B, and Shabrawishi M (2021). Predictors of length of hospital stay, mortality, and outcomes among hospitalised COVID-19 patients in Saudi Arabia: A cross-sectional study. *Journal of Multidisciplinary Healthcare*, 14: 839-852. <https://doi.org/10.2147/JMDH.S304788> PMID:33883900 PMCID:PMC8055273
- Beasley TM, Erickson S, and Allison DB (2009). Rank-based inverse normal transformations are increasingly used, but are they merited? *Behavior Genetics*, 39(5): 580-595. <https://doi.org/10.1007/s10519-009-9281-0> PMID:19526352 PMCID:PMC2921808
- Breusch TS and Pagan AR (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, 47: 1287-1294. <https://doi.org/10.2307/1911963>
- Buuren VS (2018). *Flexible imputation of missing data*. CRC Press, Boca Raton, USA.
- Buuren VS and Groothuis-Oudshoorn K (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3): 1-67. <https://doi.org/10.18637/jss.v045.i03>
- Casson RJ and Farmer LD (2014). Understanding and checking the assumptions of linear regression: A primer for medical researchers. *Clinical and Experimental Ophthalmology*, 42(6): 590-596. <https://doi.org/10.1111/ceo.12358> PMID:24801277
- Challen R, Brooks-Pollock E, Read JM, Dyson L, Tsaneva-Atanasova K, and Danon L (2021). Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: Matched cohort study. *British Medical Association: Peer-Reviewed Journal*, 372: n579. <https://doi.org/10.1136/bmj.n579> PMID:33687922 PMCID:PMC7941603
- Chang C, Deng Y, Jiang X, and Long Q (2020). Multiple imputation for analysis of incomplete data in distributed health data networks. *Nature Communications*, 11(1): 1-11. <https://doi.org/10.1038/s41467-020-19270-2> PMID:33122624 PMCID:PMC7596726
- Chen T, Wu DI, Chen H, Yan W, Yang D, Chen G, and Ning Q (2020). Clinical characteristics of 113 deceased patients with coronavirus disease 2019: Retrospective study. *British Medical Association: Peer-Reviewed Journal*, 2020: 368. <https://doi.org/10.1136/bmj.m1091> PMID:32217556 PMCID:PMC7190011
- Chodick G, Tene L, Patalon T, Gazit S, Tov AB, Cohen D, and Muhsen K (2021). Assessment of effectiveness of 1 dose of BNT162b2 vaccine for SARS-CoV-2 infection 13 to 24 days after immunization. *JAMA Network Open*, 4(6): e2115985. <https://doi.org/10.1001/jamanetworkopen.2021.15985> PMID:34097044 PMCID:PMC8185600
- Dagan N, Barda N, Kepten E, Miron O, Perchik S, Katz MA, and Balicer RD (2021). BNT162b2 mRNA COVID-19 vaccine in a nationwide mass vaccination setting. *New England Journal of Medicine*, 384(15): 1412-1423. <https://doi.org/10.1056/NEJMoa2101765> PMID:33626250 PMCID:PMC7944975
- Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, and Edmunds WJ (2021b). Estimated transmissibility and impact of SARS-CoV-2 lineage B. 1.1. 7 in England. *Science*, 372(6538). <https://doi.org/10.1126/science.abg3055> PMID:33658326 PMCID:PMC8128288
- Davies NG, Jarvis CI, Edmunds WJ, Jewell NP, Diaz-Ordaz K, and Keogh RH (2021a). Increased mortality in community-tested cases of SARS-CoV-2 lineage B. 1.1. 7. *Nature*, 593(7858): 270-274. <https://doi.org/10.1038/s41586-021-03426-1> PMID:33723411
- Davis JW (2008). *Medical statistics: A textbook for the health sciences*. The American Statistician, 62(4): 362. <https://doi.org/10.1198/tas.2008.s274>
- Dorp, C. H. V., Dorp CH, Goldberg EE, Hengartner N, Ke R, and Romero-Severson EO (2021). Estimating the strength of selection for new SARS-CoV-2 variants. *Nature Communications*, 12(1): 1-13. <https://doi.org/10.1038/s41467-021-27369-3> PMID:34907182 PMCID:PMC8671537
- Ghasemi A and Zahediasl S (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2): 486-489. <https://doi.org/10.5812/ijem.3505> PMID:23843808 PMCID:PMC3693611
- Grentzelos C, Caroni C, and Barranco-Chamorro I (2021). A comparative study of methods to handle outliers in multivariate data analysis. *Computational and Mathematical Methods*, 3(3): e1129. <https://doi.org/10.1002/cmm4.1129>
- Grubaugh ND, Hodcroft EB, Fauver JR, Phelan AL, and Cevik M (2021). Public health actions to control new SARS-CoV-2 variants. *Cell*, 184(5): 1127-1132. <https://doi.org/10.1016/j.cell.2021.01.044> PMID:33581746 PMCID:PMC7846239
- Haas EJ, Angulo FJ, McLaughlin JM, Anis E, Singer SR, Khan F, and Alroy-Preis S (2021). Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: An observational study using national surveillance data. *The Lancet*, 397(10287): 1819-1829. [https://doi.org/10.1016/S0140-6736\(21\)00947-8](https://doi.org/10.1016/S0140-6736(21)00947-8)
- Hocking RR (1976). A Biometrics invited paper: The analysis and selection of variables in linear regression. *Biometrics*, 32: 1-49. <https://doi.org/10.2307/2529336>
- Hopkinson NS, Rossi N, El-Sayed_Moustafa J, Laverty AA, Quint JK, Freidin M, and Falchi M (2021). Current smoking and COVID-19 risk: Results from a population symptom app in over 2.4 million people. *Thorax*, 76(7): 714-722. <https://doi.org/10.1136/thoraxjnl-2020-216422> PMID:33402392 PMCID:PMC7789201
- Hou C, Chen J, Zhou Y, Hua L, Yuan J, He S, and Jia E (2020). The effectiveness of quarantine of Wuhan city against the Corona Virus Disease 2019 (COVID-19): A well-mixed SEIR model analysis. *Journal of Medical Virology*, 92(7): 841-848. <https://doi.org/10.1002/jmv.25827> PMID:32243599
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, and Cao B (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, 395(10223): 497-506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
- Iacobucci G (2020). COVID-19: UK lockdown is "crucial" to saving lives, say doctors and scientists. *British Medical Association: Peer-Reviewed Journal*, 368: m1091. <https://doi.org/10.1136/bmj.m1204> PMID:32209548

- James G, Witten D, Hastie T, and Tibshirani R (2013). An introduction to statistical learning. Volume 112, Springer, New York, USA. <https://doi.org/10.1007/978-1-4614-7138-7>
- Ji Y, Ma Z, Peppelenbosch MP, and Pan Q (2020). Potential association between COVID-19 mortality and health-care resource availability. The Lancet Global Health, 8(4): e480. [https://doi.org/10.1016/S2214-109X\(20\)30068-1](https://doi.org/10.1016/S2214-109X(20)30068-1)
- Kandel N, Chungong S, Omaar A, and Xing J (2020). Health security capacities in the context of COVID-19 outbreak: An analysis of International Health Regulations annual report data from 182 countries. The Lancet, 395(10229): 1047-1053. [https://doi.org/10.1016/S0140-6736\(20\)30553-5](https://doi.org/10.1016/S0140-6736(20)30553-5)
- Kim HY (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. Restorative Dentistry and Endodontics, 38(1): 52-54. <https://doi.org/10.5395/rde.2013.38.1.52> PMID:23495371 PMCID:PMC3591587
- Kim HY (2017). Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. Restorative Dentistry and Endodontics, 42(2): 152-155. <https://doi.org/10.5395/rde.2017.42.2.152> PMID:28503482 PMCID:PMC5426219
- Liang LL, Tseng CH, Ho HJ, and Wu CY (2020). COVID-19 mortality is negatively associated with test number and government effectiveness. Scientific Reports, 10(1): 1-7. <https://doi.org/10.1038/s41598-020-68862-x> PMID:32709854 PMCID:PMC7381657
- Lipsitch M, Swerdlow DL, and Finelli L (2020). Defining the epidemiology of COVID-19-Studies needed. New England Journal of Medicine, 382(13): 1194-1196. <https://doi.org/10.1056/NEJMp2002125> PMID:32074416
- Maesschalck DR, Jouan-Rimbaud D, and Massart DL (2000). The mahalanobis distance. Chemometrics and Intelligent Laboratory Systems, 50(1): 1-18. [https://doi.org/10.1016/S0169-7439\(99\)00047-7](https://doi.org/10.1016/S0169-7439(99)00047-7)
- Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, and Keshri A (2019). Descriptive statistics and normality tests for statistical data. Annals of Cardiac Anaesthesia, 22(1): 67-72. https://doi.org/10.4103/aca.ACA_157_18 PMID:30648682 PMCID:PMC6350423
- Moghadas SM, Shoukat A, Fitzpatrick MC, Wells CR, Sah P, Pandey A, and Galvani AP (2020). Projecting hospital utilization during the COVID-19 outbreaks in the United States. Proceedings of the National Academy of Sciences, 117(16): 9122-9126. <https://doi.org/10.1073/pnas.2004064117> PMID:32245814 PMCID:PMC7183199
- O'Toole Á, Hill V, Pybus OG, Watts A, Bogoch II, Khan K, and COVID T (2021). Tracking the international spread of SARS-CoV-2 lineages B. 1.1. 7 and B. 1.351/501Y-V2. Wellcome Open Research, 6: 121.
- Peto J (2020). COVID-19 mass testing facilities could end the epidemic rapidly. British Medical Association: Peer-Reviewed Journal, 368: m1163. <https://doi.org/10.1136/bmj.m1163> PMID:32201376
- Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, and Schwartz O (2021). Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. Nature, 596(7871): 276-280. <https://doi.org/10.1038/s41586-021-03777-9> PMID:34237773
- Priesemann V, Balling R, Brinkmann MM, Ciesek S, Cypionka T, Eckerle I, and Szczurek E (2021). An action plan for pan-European defence against new SARS-CoV-2 variants. The Lancet, 397(10273): 469-470. [https://doi.org/10.1016/S0140-6736\(21\)00150-1](https://doi.org/10.1016/S0140-6736(21)00150-1)
- Rossmann H, Meir T, Somer J, Shilo S, Gutman R, Arie AB, and Gorfine M (2021). Hospital load and increased COVID-19 related mortality in Israel. Nature Communications, 12: 1904. <https://doi.org/10.1038/s41467-021-22214-z> PMID:33771988 PMCID:PMC7997985
- Rubino S, Kelvin N, Bermejo-Martin JF, and Kelvin D (2020). As COVID-19 cases, deaths and fatality rates surge in Italy, underlying causes require investigation. The Journal of Infection in Developing Countries, 14(03): 265-267. <https://doi.org/10.3855/jidc.12734> PMID:32235086
- Sarkodie SA and Owusu PA (2020). Investigating the cases of novel coronavirus disease (COVID-19) in China using dynamic statistical techniques. Heliyon, 6(4): e03747. <https://doi.org/10.1016/j.heliyon.2020.e03747> PMID:32289090 PMCID:PMC7128585
- Sen-Crowe B, Sutherland M, McKenney M, and Elkbuli A (2021). A closer look into global hospital beds capacity and resource shortages during the COVID-19 pandemic. Journal of Surgical Research, 260: 56-63. <https://doi.org/10.1016/j.jss.2020.11.062> PMID:33321393 PMCID:PMC7685049
- Shapiro SS and Wilk MB (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3/4): 591-611. <https://doi.org/10.1093/biomet/52.3-4.591>
- Singh J, Rahman SA, Ehtesham NZ, Hira S, and Hasnain SE (2021). SARS-CoV-2 variants of concern are emerging in India. Nature Medicine, 27: 131-133. <https://doi.org/10.1038/s41591-021-01397-4> PMID:34045737
- Stephens MA (1974). EDF statistics for goodness of fit and some comparisons. Journal of the American Statistical Association, 69(347): 730-737. <https://doi.org/10.1080/01621459.1974.10480196>
- Vatcheva KP, Lee M, McCormick JB, and Rahbar MH (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. Epidemiology (Sunnyvale, Calif.), 6(2): 227. <https://doi.org/10.4172/2161-1165.1000227> PMID:27274911 PMCID:PMC4888898
- Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, and Peng Z (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. Journal of the American Medical Association: Peer-Reviewed Journal, 323(11): 1061-1069. <https://doi.org/10.1001/jama.2020.1585> PMID:32031570 PMCID:PMC7042881
- Ward H, Atchison C, Whitaker M, Ainslie KE, Elliott J, Okell L, and Elliott P (2021). SARS-CoV-2 antibody prevalence in England following the first peak of the pandemic. Nature Communications, 12: 905. <https://doi.org/10.1038/s41467-021-21237-w> PMID:33568663 PMCID:PMC7876103
- Westen-Lagerweij VNA, Meijer E, Meeuwse EG, Chavannes NH, Willemssen MC, and Croes EA (2021). Are smokers protected against SARS-CoV-2 infection (COVID-19)? The origins of the myth. NPJ Primary Care Respiratory Medicine, 31: 10. <https://doi.org/10.1038/s41533-021-00223-1> PMID:33637750 PMCID:PMC7910565
- Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund LO, and Petersen RC (2004). Mild cognitive impairment-Beyond controversies, towards a consensus: Report of the international working group on mild cognitive impairment. Journal of Internal Medicine, 256(3): 240-246. <https://doi.org/10.1111/j.1365-2796.2004.01380.x> PMID:15324367
- Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, and Cao B (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. The Lancet, 395(10229): 1054-1062. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)