Jagannath University Journal of Science
Volume 12, Number II, Dec. 2025, pp. 99−118
https://jnu.ac.bd/journal/portal/archives/science.jsp
ISSN 3005-4486 (Online), ISSN 2224-1698 (Print)

Predicting COVID-19 Daily Mortality Using Multivariate Linear Regression: Comparative Evidence from Bangladesh and the UK

Research Article

Tajnin Jahan*

Department of Management Studies, Jagannath University, Dhaka-1100, Bangladesh

DOI: https://doi.org/10.3329/jnujsci.v12i2.89624

Received: 26.11.2025, Accepted: 01.03.2026, Published: July 2026

ABSTRACT

Covid-19, a global epidemic, has created a health crisis all over the world.This study examines the correlation between daily new COVID-19 cases, vaccination rates, and daily new fatalities in Bangladesh and the United Kingdom, utilizing secondary data from the Our World in Data COVID-19 dataset spanning from 8 March 2020 to 19 December 2023. Following data cleaning, transformation, and normalization, correlation and multivariate regression analyses were performed utilizing RStudio. The findings indicated a strong positive correlation between daily new cases and daily new deaths in Bangladesh, while the United Kingdom exhibited a moderate positive correlation between both variables. In both nations, there was a negative correlation between vaccination rates and new fatalities, demonstrating the beneficial impact of immunization on mortality reduction. Vaccination has shown a negative correlation with new cases in Bangladesh, although it demonstrated a positive correlation in the UK samples. Regression models for both nations demonstrated a strong fit, accounting for roughly 84% of the variability in daily new fatalities. The study concludes that vaccination markedly decreased COVID-19 fatalities in both nations. The results offer significant insights for policymakers to formulate data-driven public health strategies and enhance pandemic preparedness.

Keywords: *COVID-19, Vaccination, Mortality Prediction, Correlation Analysis, Bangladesh and United Kingdom*

1. Introduction

The SARS-CoV-2 virus caused coronavirus disease, and it was spread from Wuhan, China, for the first time in December 2019. It was declared as a public health emergency on 11 January 2020 (WHO, 2020). The first case of a COVID-19-positive patient in Bangladesh was declared by the Institute of Epidemiology, Disease Control and Research

***Correspondence:** Tajnin Jahan
E-mail: tajnin@mgt.jnu.ac.bd

(IEDCR) on March 8, 2020, and the first death was announced on March 18, 2020. However, on March 11, 2020, the World Health Organization (WHO) announced Covid-19 as an epidemic (Khan et al., 2021). The United Kingdom (UK) has also faced huge health hazards due to covid-19. The UK detected its first COVID case and death on 11th January 2020 and 1st February 2020, respectively (Merchant et al., 2021). Since covid-19 was a pandemic all over the world, a huge number of studies have been conducted in every country. After the coronavirus outbreak, vaccination was articulately mentioned by the global community as a countermeasure effort. As the first country in the world, the United Kingdom launched the vaccination program on 8th December 2020 (Pritchard et al., 2021), and in Bangladesh, vaccination was inaugurated on 27th January 2021 (Abedin et al., 2021). There was a positive impact of vaccination on decreasing COVID-19 infection and death rates (Moghadas et al., 2021). It was also found that an increase of vaccination allocation helped to reduce deaths in many countries of the world (Bertsimas et al., 2020; Liang et al., 2021; Gupta et al., 2021). In Bangladesh, vaccinated people were more protected and less infected than those who were not vaccinated (Rahman et al., 2022). These findings, however, were from primary data; it would be more robust and precise if analysis were conducted on accumulated data from the ‘Our World in Data COVID-19 dataset’ (Ritchie et al., 2023). Conversely, there is no remarkable analysis between Bangladesh and the UK regarding COVID-19 deaths, cases, and vaccination. Since Bangladesh and the United Kingdom are very different in various economic and social indicators, findings from a comparative study between these two countries will give substantial direction to the policymakers on how to deal with infectious diseases in the future. Consequently, this study will predict daily COVID deaths by using a regression model where the independent variables are people vaccinated and new COVID-19 cases. Apart from this, the relationship among people_vaccinated, new_cases, and new_deaths has been assessed to understand the relationship between these variables in both countries.

1.1 Aim of the Study

The primary aim of this study is to examine the association among daily new COVID-19 cases, the number of people vaccinated, and daily new deaths in the United Kingdom and Bangladesh. In addition, the study seeks to develop a regression-based prediction model for daily new COVID-19 deaths in both countries.

1.2 Research Questions

To achieve the above aim, the study seeks to answer the following research questions:

1. What is the relationship between daily new COVID-19 cases and daily new deaths in Bangladesh and the United Kingdom?
2. How does the number of people vaccinated influence daily new deaths in both countries?
3. What is the association between daily new COVID-19 cases and the number of people vaccinated in Bangladesh and the United Kingdom?
4. Is there a significant overall association among daily new cases, daily new deaths, and the number of people vaccinated in both countries?
5. To what extent can daily new deaths be predicted using daily new cases and the number of people vaccinated through a regression model?
6. How do these associations and predictive outcomes differ between Bangladesh and the United Kingdom?

2. Literature Review

In order to formulate public health planning and policy response, allocate healthcare resources, and improve preparedness for future pandemics, it is very crucial to understand the relationship among infection rate, vaccination coverage, and mortality rate of covid-19. Therefore, all over the world, extensive research has been conducted to examine the determinants of the covid-19 pandemic and the prediction of mortality.

2.1 COVID-19 Cases and Mortality Dynamics

A substantial number of recent studies have deployed statistical modeling techniques to examine the relationship between infection rates and mortality trends (Chatterjee et al., 2020; Onovo et al., 2021). Epidemiological theory indicates that increases in confirmed cases precede rises in daily deaths, which has reflected the biological progression from infection to severe disease. Time-series and regression-based analyses consistently show a strong association between new infections and subsequent mortality uptake, particularly during periods of variant emergence and healthcare system strain (Davies et al., 2021; Rahman & Sathi, 2021).

2.2 Vaccination Coverage and Mortality Reduction

Vaccination has been identified as one of the most significant protective factors in reducing COVID-19 mortality worldwide. Studies across Europe and North America report substantial declines in deaths following large-scale vaccination rollout, especially among older and high-risk populations (Haas et al., 2021; Andrews et al., 2022). Similarly, cross-national evidence suggests that higher vaccination coverage is associated with lower fatality burdens, even in the presence of continued infection waves (Our World in Data, 2023).

2.3 Predictive Modeling of COVID-19 Mortality

A growing body of literature has employed statistical modeling approaches, including multivariate regression frameworks, to examine the association between COVID-19 cases, vaccination progress, and mortality trends. For example, researchers have demonstrated that surges in new infections are strongly linked with subsequent increases in daily deaths, particularly during periods of variant emergence and limited healthcare resources (Davies et al., 2021; Rahman & Sathi, 2022). Multivariate regression frameworks are widely employed in this context because they enable simultaneous evaluation of multiple predictors while controlling for confounding influences and facilitating cross-country comparison (Montgomery et al., 2021). Although advanced time-series models such as ARIMA, distributed lag models, and machine learning algorithms have been widely used for COVID-19 forecasting (Kumar et al., 2020), traditional regression techniques remain prevalent in public health research because of their ability to evaluate predictive performance across countries and to assess associations between key epidemiological indicators (Olowe et al., 2024).

2.4 Research Gap

Although numerous studies have analyzed COVID-19 mortality determinants within individual countries, fewer studies provide a comparative predictive analysis between developing and developed economies using a consistent modeling framework (Chang et al., 2022). Bangladesh and the United Kingdom present contrasting healthcare systems, demographic structures, and vaccination rollout strategies. A comparative assessment using a unified analytical approach can therefore offer valuable insights into mortality dynamics across different public health contexts. This study contributes to the literature by developing a multivariate regression-based predictive model of daily COVID-19 mortality and applying it comparatively to Bangladesh and the UK using a chronological validation framework.

3. Methodology

3.1 Study area

Bangladesh, a densely populated country, has also experienced a huge number of COVID-19 infections, and 29530 people died of COVID-19 in Bangladesh until 20th October 2025 (Directorate General of Health Services, 2025). On the other hand, the United Kingdom, a developed country with low density, also experienced 232112 cases of deaths until 15th October 2025 (UK Health Security Agency, 2025). As a result, the study areas in this research have been selected from two countries: Bangladesh and the United Kingdom, in light of covid-19 new cases, new deaths, and people vaccinated.

3.2 Data Source and data preparation for analysis

This research employed secondary data from the "Our World in Data COVID-19 dataset" to examine COVID-19 trends in Bangladesh and the United Kingdom (Ritchie et al., 2020). Data for each country were refined to encompass three variables—new_cases, new_deaths, and people_vaccinated—spanning the period from 8 March 2020 to 19 December 2023. Upon filtering for Bangladesh, the dataset originally had 1,447 rows and 67 columns, which were further refined to 490 complete observations via the R function na.omit() and further sanitized for precision, yielding 299 consistent records. Likewise, the UK data were retrieved and refined using identical procedures to guarantee comparability. Both datasets underwent normality testing, and neither exhibited a normal distribution. Multivariate linear regression necessitates normally distributed data (Osborne & Waters, 2019; Uyanık & Güler, 2013); hence, data transformation was implemented according to skewness values (Hammouri et al., 2020; Osborne, 2019). In Bangladesh, new cases and new deaths exhibited significant skewness and were consequently log-transformed (McGuinness et al., 1997); however, people who were vaccinated remained untransformed, as other approaches did not enhance results. In the UK, new_cases and new_deaths underwent log transformation, but people_vaccinated need a square transformation to properly mitigate skewness (Noel et al., 2021). Ultimately, two datasets—Covid_Data_ Analysis_ BD and Covid_Data_Analysis_UK—were established, and a uniform sample of 250 daily observations from each was chosen for comparative analysis

3.3 Variable description

The outcome variable in this study is:

- Daily New Deaths (new_deaths): the number of newly reported COVID-19 deaths per day.

The explanatory variables include:

- Daily New Cases (new_cases): newly confirmed COVID-19 infections reported each day.
- Vaccination Coverage (people_vaccinated): daily vaccination progress, representing the level of immunization within the population.

These predictors were selected based on prior evidence that infection rates are strongly associated with mortality outcomes, while vaccination is a key factor in reducing fatality risk (Haas et al., 2021; Andrews et al., 2022).

3.4 Model Justification and Analytical Approach

To examine the determinants of daily COVID-19 mortality, this study employs multivariate linear regression. This approach is appropriate when the objective is to quantify the relationship between one dependent variable and multiple independent variables while controlling for their joint effects (Montgomery et al., 2021). The regression framework provides an interpretable method for assessing the relative contribution of infection intensity and vaccination progress in shaping mortality outcomes in two different national settings.

The regression model is specified as follows:

$$\text{new\_deaths} = \beta_0 + \beta_1 \text{new\_cases} + \beta_2 \text{people\_vaccinated} + \varepsilon_t$$

where:

- new_deaths represents daily new deaths,
- new_cases denotes daily new confirmed cases,
- people_vaccinated represents vaccination coverage,
- β_0 is the intercept,

- β_1 and β_2 are regression coefficients,
- ε_t is the error term.

R studio has been used for data preparation and analysis. Firstly, correlation analysis was conducted among three variables. Then, multivariate regression has been done for both countries: UK and Bangladesh. Finally, prediction of new_deaths has been made depending on regression model.

Table 1: R library used in this study

R library	Purposes of Use
Tidyverse	Data preparation, wrangling, tidying, visualizing and so forth
Rgl	3D visualization
GGally	Correlation Plot
readxl	To load excel file

4. Results

4.1 Analysis of Bangladesh Covid-19 Data:

4.1.1 Descriptive statistics:

Table 2: Comparison between original data and transformed data of Bangladesh covid-19 data

Name	**new_cases**	**new_deaths**	**people_vaccinated**	**log_transformed_new_cases**	**log_transformed_new_deaths**	**log_transformed_people_vaccinated**
Minimum	6	6	26	1.79	0.00	3.25
Median	1105	15	23164422	7	2.71	16.95
Mean	2714.9	45.2	53655125	6.84	2.64	16.90
Maximum	16230	264	151502489	9.69	5.58	18.84

Here, a comparative scenario of descriptive statistics between original COVID-19 data and transformed COVID-19 data is displayed. In Bangladesh, daily covid_cases, daily covid_deaths, and people vaccinated daily were not normally distributed. According to the current research aim, multivariate regression will be conducted. As a result, our selected variables have to be symmetrical.

Table 3: Interpretation of Skewness

Range	**Interpretation**
-0.5 to 0.5	Almost symmetrical
-1 to -0.5 (negatively skewed) 0.5 to 1 (positively skewed)	Moderately skewed
Lower than -1 (negatively skewed) Greater than 1 (positively skewed)	Highly skewed

(Menon, 2023)

Table 4: Comparison of skewness of original data vs. transformed data

new_cases	new_deaths	People_vaccinated	Log_transformed_new_cases	Log_transformed_new_cases	Log_transformed_people_vaccinated
1.93	1.92	0.69	-0.36	-0.04	-2.35

In the case of Bangladesh's COVID-19 data, new_deaths and new_cases were transformed by the log transformation method. After that, these two variables transformed into symmetrical ones. Nevertheless, if we transformed people_vaccinated by the log transform method, the transformed data was more asymmetric. If people_vaccinated could be transformed in the following way, possible skewness in every transformation process is like this:

Table 5: Skewness value of "people_vaccinated" in different transformation methods

reciprocal_transform_people_vaccinated	square_transform_people_vaccinated	cube_transform_people_vaccinated
17.15	1.01	1.24

Table 6: Comparison of P value by Shapiro wilks normality test

Original data			Transformed data		
New_cases	**New_deaths**	**People_vaccinated**	**Log_transformed_new_cases**	**Log_transformed_new_deaths**	**Log_transformed_people_vaccinated**
< 2.2e-16	< 2.2e-16	< 2.2e-16	8.035e-05	5.437e-09	< 2.2e-16

In all aspects, it is observed that the p-value has been increased after the transformation. Usually, P values are very tiny in every variable's original data. If we chose,

H1: The distribution is not normally distributed.

In that case, if the p-value is greater than 0.05, then we can accept null hypotheses and reject alternative hypotheses. So, in the original data, the p-value of each distribution is very close to zero. However, after transformation, p-values were increased, though these were not big enough to accept null hypotheses. Nonetheless, the distribution seemed more symmetrical than its previous value.

4.1.2 Correlation Test

Correlation between new_cases and new_deaths

H_0= There is no relationship between daily new_cases and new_deaths of COVID-19 in Bangladesh.

H_1= There is a significant correlation between daily new_cases and new_deaths of COVID-19 in Bangladesh.

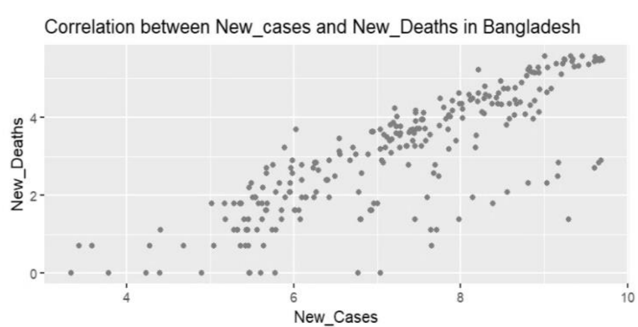

Figure 1: Scatter plot comparing New_cases and new_deaths of Covid_19 in Bangladesh.

Table 7: Correlation Coefficient

Correlation coefficient value	**Relationship**
-0.3 to 0.3	Weak
-0.5 to -0.3 / 0.3 to 0.5	Moderate
-0.9 to -0.5 / 0.5 to 0.9	Strong
-1.0 to -0.9 / 0.9 to 1.0	Very strong

(Asuero et al., 2006)

Table 8: Correlation analysis result between new_cases and new_deaths in Bangladesh.

t value	Df	p-value	95 percent confidence interval		correlation coefficient
			Lower limit	upper limit	
21.87	248	< 2.2e-16	0.76	0.85	0.81

In Bangladesh, there was a significant relationship between new_cases and new_deaths. Here, the p-value is very close to zero, which is smaller than the threshold value of p=0.05. So, we can reject null hypotheses and accept alternative hypotheses. Moreover, the correlation coefficient is 0.81, which represents that for one unit of covid-19 new_cases increase, new_deaths tends to increase 0.81 directions.

Correlation between people_vaccinated and new_deaths:

H_0= There is no relationship between people_vaccinated and new_deaths of covid-19 in Bangladesh.

H_1= There is a significant correlation between people_vaccinated and new_deaths of covid-19 in Bangladesh.

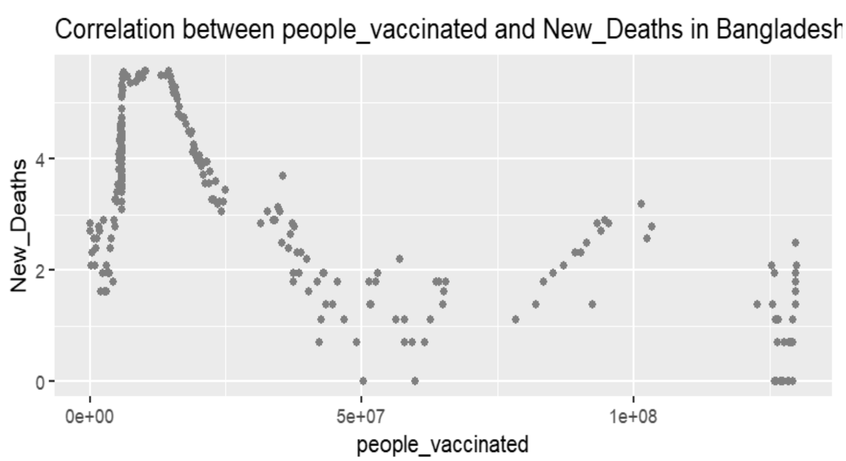

Figure 2: Scatter plot comparing people_vaccinated and new_deaths of Covid_19 in Bangladesh

Table 9: Correlation analysis result between people_vaccinated and new_deaths in Bangladesh.

t value	**Df**	**p-value**	**95 percent confidence interval**		**correlation coefficient**
			Lower limit	upper limit	
-15.372	248	< 2.2e-16	-0.76	-0.63	-0.70

It is observed that there is a strong negative correlation between people_vaccinated and new_deaths in Bangladesh since the p-value is < 2.2e-16, which is much less than 0.05. So, we can reject the null hypothesis and accept the alternative hypothesis.

4.1.3 Multi-variate linear regression

The prime goal of this research is to predict the covid-19 daily deaths in Bangladesh depending on two independent variables: new_cases and people_vaccinated. In order to perform a multivariate regression model for prediction, the total rows of the data file have been split into two parts. One is for analysis and another one for testing whether the prediction has been made perfect or not. So, for conducting a prediction model using multivariate regression, the correlation between two independent variables has to be checked first to assess collinearity. If we get that the two independent variables are perfectly correlated, then the second independent variable in the regression model does not contribute on a large scale.

Table 10: Correlation between new_cases and people_vaccinated of Bangladesh

t value	Df	p-value	95 percent confidence interval		correlation coefficient	VIF
			Lower limit	upper limit		
-6.8997	248	$4.323e^{-11}$	-0.50	-0.29	-0.40	1.19

From the previous table, it is clear that the correlation coefficient between new_cases and people_vaccinated is -0.40. As a result, there was a negative correlation between these two independent variables, but they were not perfect predictors of each other. Moreover, VIF < 5; there is no serious multicollinearity problem between new_cases and people_vaccinated. The following scatter plot will clarify this relationship.

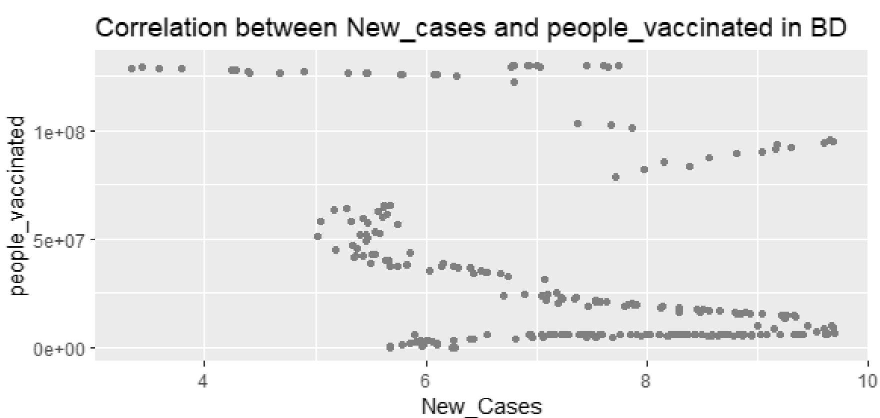

Figure 3: Scatter plot comparing people_vaccinated and new_cases of Covid_19 in Bangladesh.

There are lots of scatter in the plot though there is some correlation. So, if these two independent variables are added to our regression model, there is no violation of the assumption of multicollinearity.

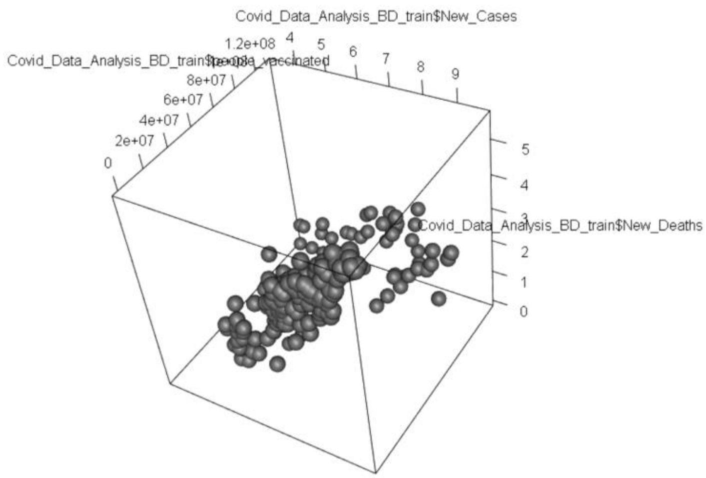

Figure 4: 3D plot to show the linear relationship among variables

In the previous 3D plot, it seems that there is an approximate linear relationship among two independent variables (people_vaccinated, new_cases) and covid_19 new deaths.

Regression model

Table 11: Residuals of mod_BD

Min	1Q	Median	3Q	Max
-2.1650	-0.3790	0.1041	0.4309	1.6174

Residuals

Here, first quarter(1Q) and third quarter(3Q) residuals are quite similar and residuals are normally distributed as median is near to zero. As a result, it can be said that model fits the data very well.

Table 12: Coefficients of mod_BD

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.163e+00	2.425e-01	-4.795	2.82e-06
New_Cases	6.647e-01	3.059e-02	21.732	< 2e-16
people_vaccinated	-1.585e-08	1.039e-09	-15.257	< 2e-16

Coefficients

H_0: b1=b2=0

$H_{1:}$ is that at least one b≠0

y = a + b1x1 + b2x2

y= -1.163e+6.647e-01*(New_cases) + -1.585e-08*people_vaccinated

Slope (b1) = 6.647e-01

Slope2 (b2)=-1.585e-08

Intercept =-1.163e

Table 13: Summary of mod_BD.

Residual standard error: 0.63 on 247 degrees of freedom
Multiple R-squared: 0.82, Adjusted R-squared: 0.82
F-statistic: 578[...8 on 2 and 247 DF, p-value: < 2.2e-16

In this regression model, the prediction of new deaths would be incorrect by 0.633. In our model, the mean value of new_deaths is 3.07. So, the prediction error can be nearly 21%, which is not huge because the prediction will be matched about 80 percent. The R-squared value of this regression model is 0.8242, which means 82.42% of the variability of Covid-19 daily new deaths in Bangladesh is explained by these two independent variables. Therefore, this model is a good fit. Moreover, the adjusted R-square is the same as the R-square value, so the second independent variable did not decrease the goodness of fit in this regression model. Since the P-value of this model is less than 0.05, it's statistically significant. We can accept the alternative hypothesis. Consequently, this model is reliable for further prediction.

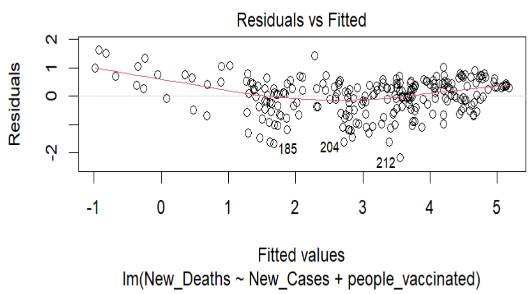

Figure 5: Plot of residuals vs. fitted (mod_BD)

4.1.4 Prediction of new_deaths with the regression model (Bangladesh)

Table 14: 95% confidence intervals around these predictions (mod_BD)

No.	**Fit**	**95% confidence Interval**	
		Lower limit	**Upper limit**
1	1.29	1.09	1.49
2	0.81	0.61	1.00
3	0.97	0.77	1.16
4	1.05	0.86	1.25
5	0.67	0.48	0.86
6	0.67	0.47	0.86
7	0.72	0.52	0.91
8	0.36	0.16	0.56
9	0.36	0.18	0.57
10	0.38	0.18	0.57

Table 15: True values, prediction and Residuals of (mod_BD)

No.	New_Cases	people_vaccinated	New_Deaths	predicted	Residuals
1	1.79	129987682	1.79	1.29	-0.50
2	1.39	130000824	1.39	0.81	-0.58
3	1.61	130010268	1.61	0.97	-0.64
4	1.39	130017083	1.39	1.05	-0.34
5	1.10	130049697	1.10	0.67	-0.43
6	0	130065446	0	0.67	0.67
7	1.10	130099847	1.10	0.72	-0.38
8	0.693	130137673	0.693	0.36	-0.33
9	0	130179063	0	0.36	0.36
10	0	130878488	0	0.38	0.38

Although the model does not predict the exact correct values of new_deaths, the prediction is within the range of the 95% confidence interval. Thus, the strength of the model is good, and it could make a nearly perfect prediction of new deaths from covid-19 in Bangladesh.

Table 16: sum of squared errors (SSE) of BD_model

sum of squared errors (SSE)	46.54

Finally, SSE of this model 46.54 which represent good fit of the model.

4.2 Analysis of United Kingdom Covid-19 Data

4.2.1 Descriptive statistics

Table 17: Comparison between original data and transformed data of United Kingdom covid-19 data

Name	new_cases	new_deaths	people_vaccinated	log_transformed_new_cases	log_transformed_new_deaths	squareroot_transformed_people_vaccinated
Minimum	1433	4	2286572	7.27	1.386	5.228e+12
Median	25170	118	50455318	10.133	4.771	2.546e+15
Mean	33436	174	44615354	9.845	4.603	2.153e+15
Maximum	275647	1487	53813491	12.527	7.305	2.896e+15

This table represents comparative descriptive statistics of original data and transformed COVID-19 data in the UK to assess the normality of data. Here, the data were not normally distributed in the original format. Thus, the data was transformed to fulfill the assumption of multivariate linear regression.

Table 18: Comparison of skewness of original data vs. transformed data

new_cases	new_deaths	People_vaccinated	Log_transformed_new_cases	Log_transformed_new_cases	Log_transformed_people_vaccinated
2.73	3.47	-1.74	-0.24	-0.06	-3.09

Table 19: Skewness value of "people_vaccinated" in different transformation methods

reciprocal_transform_people_vaccinated	square_transform_people_vaccinated
6.05	-1.20

Here, it is visible that the original data was skewed in all three variables. So, the log transformation method was applied to transform the original data into symmetrical data. However, while the log transformation produced good results for new_cases and new_deaths, it converted the people_vaccinated variable into a more asymmetrical format. As a result, the square root transformation method was applied to the people_vaccinated variable because it produced data with less skewness compared to other transformation methods.

Table 20: Comparison of P value by Shapiro wilks normality test

Original_data			Transformed_data		
New_cases	**New_deaths**	**People_vaccinated**	**Log_transformed_new_cases**	**Log_transformed_new_deaths**	**Square_root_transformed_people_vaccinated**
< 2.2e-16	< 2.2e-16	< 2.2e-16	8.601e-10	3.531e-13	< 2.2e-16

Here, the p-value of new_cases and new_deaths has been increased, but the people_vaccinated p-value is still the same. Finally, transformed data for three variables will be used to conduct regression analysis in this study to confirm the assumption of multivariate regression.

4.2.2 Correlation Test

Correlation between new_cases and new_deaths:

H_0= There is no relationship between covid-19 daily new_cases and new_deaths in the UK.

H_1= There is significant correlation between covid-19 daily new_cases and new_deaths in the UK.

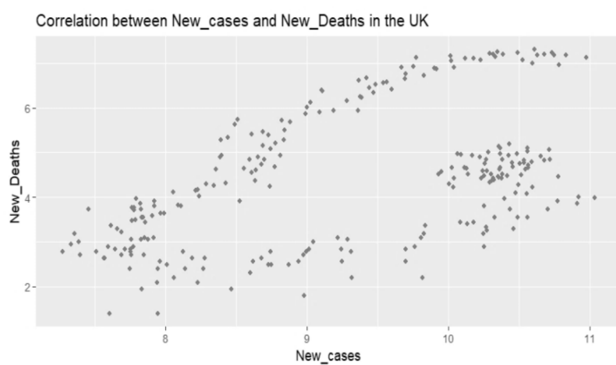

Figure 6: Scatter plot comparing new_cases and new_deaths of Covid_19 in The United Kingdom.

Table 21: Correlation analysis result between new_cases and new_deaths in The United Kingdom.

t value	Df	p-value	95 percent confidence interval		correlation coefficient
			Lower limit	**upper limit**	
9.19	248	< 2.2e-16	0.41	0.59	0.50

From the previous table, it is seen that the p-value is lower than 0.05. So, we can reject the null hypothesis, and it is proven that there is a positive correlation between new_cases and deaths in the UK.

Correlation between people_vaccinated and new_deaths

H_0= There is no relationship between people_vaccinated and new_deaths of covid-19 in the UK.

H_1= There is a significant correlation between people_vaccinated and new_deaths of covid-19 in The UK.

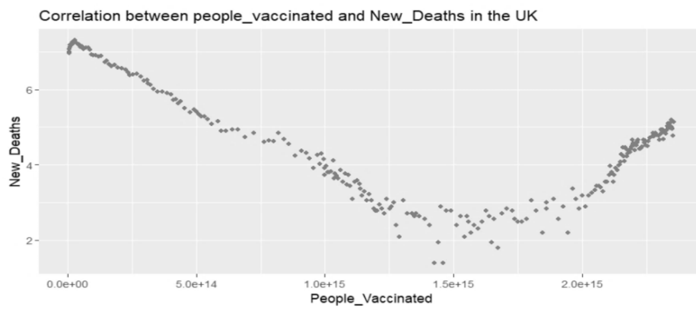

Figure 7: Scatter plot comparing people_vaccinated and new_deaths of Covid_19 in the UK.

Table 22: Correlation analysis result between People_vaccinated and new_deaths in The United Kingdom.

t value	**Df**	**p-value**	**95 percent confidence interval**		**correlation coefficient**
			Lower limit	upper limit	
-11.191	248	< 2.2e-16	-0.66	-0.49	-0.58

There was a strong negative correlation between covid-19 new_ deaths and people_vaccinated. Moreover, the P value is quite low compared to 0.05. So, the alternative hypothesis is accepted.

4.2.3 Multi-variate linear regression

In the case of the UK, this study will predict the daily new_ deaths. Here, dependent variable is new_deaths whereas independent variables are new_cases and people_vaccinated. So, for conducting regression analysis, the collinearity of two independent variables must be checked before the analysis.

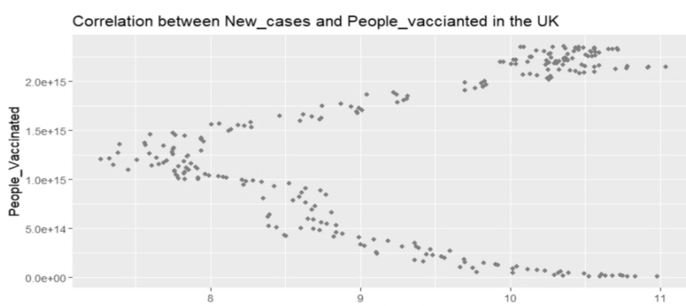

Figure 8: Scatter plot comparing people_vaccinated and new_cases of Covid_19 in the UK.

Table 23: Correlation between new_cases and people_vaccinated (Covid-19) of the United Kingdom

t value	**Df**	**p-value**	**95 percent confidence interval**		**correlation coefficient**	**VIF**
			Lower limit	upper limit		
4.78	248	2.96e-06	0.17	0.40	0.29	1.09

So, there is a correlation between these two independent variables, but they are not perfect predictors of each other because the correlation coefficient value of these two variables is 0.29, which is a weak positive correlation. Furthermore, VIF < 5; there is no multicollinearity problem between the variables. Consequently, the regression model can be conducted with these two independent variables.

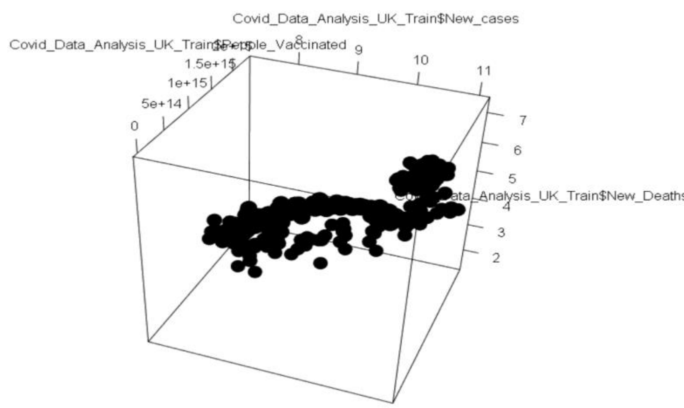

Figure 9: 3D plot to show the linear relationship among variables

So, it is visible that an approximate linear relationship is also available among three variables.

Regression model:

Table 24: Residuals of mod_Uk

Min	**1Q**	**Median**	**3Q**	**Max**
-1.81	-0.41	0.15	0.41	1.31

Residuals

In this model, median of residuals is near to zero. Also, the values of 1st quarter and 3rd quarter are exactly same. Therefore, it can be concluded that this model fits the data.

Table 25: Coefficients of mod_UK

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	-2.954e+00	3.374e-01	-8.755	3.28e-16
New_Cases	1.003e+00	3.753e-02	26.726	< 2e-16
people_vaccinated	-1.481e-15	5.132e-17	-28.850	< 2e-16

Coefficients

H_0: b1=b2=0

$H_{1:}$ is that at least one b≠0

y = a + b1x1 + b2x2

y=-2.954e+00 +1.003e+00*(New_cases) + -1.481e-15*people_vaccinated

Table 26: Summary of mod_UK.

Residual standard error: 0.6138 on 247 degrees of freedom	
Multiple R-squared: 0.83,	Adjusted R-squared: 0.83
F-statistic: 599.9 on 2 and 247 DF, p-value: < 2.2e-16	

In the case of the UK covid-19 data model of our study, adjusted R-squared has not been changed in a negative direction, which means that the addition of the people_vaccinated variable in the model has not reduced the goodness of fit of this model. Furthermore, 83 percent of the variability of covid-19 new_deaths in the UK can be explained by these two variables, which is excellent. Since the P value is much lower than 0.05, the null hypothesis can be rejected, and it means that this model is highly significant. The mean value of death in this model is 4.41, and the standard error is 0.61. As a result, prediction error is only approximately 13%, which confirms that this model is a good fit model.

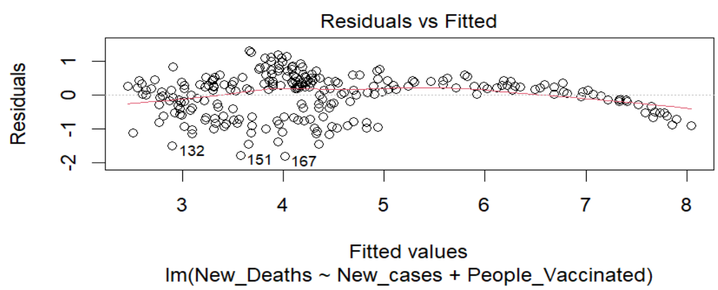

Figure 10: Plot of residuals vs. fitted (mod_UK)

4.2.4. Prediction of new_deaths with the regression model (United Kingdom)

Table 27: 95% confidence intervals around these predictions (mod_UK)

No.	Fit	95% confidence Interval	
		Lower limit	**Upper limit**
1	3.91	3.77	4.04
2	3.88	3.75	4.01
3	3.87	3.74	4.00
4	3.77	3.64	3.90
5	3.92	3.79	4.05
6	4.19	4.05	4.33
7	4.14	3.99	4.28
8	4.08	3.94	4.21
9	4.05	3.91	4.19
10	3.94	3.80	4.07

Table 28: True values, prediction and Residuals of (mod_UK)

No	New_Cases	people_vaccinated	New_Deaths	predicted	Residuals
1	10.3	2.36e15	5.12	3.91	-1.21
2	10.3	2.36e15	5.09	3.88	-1.20
3	10.3	2.36e15	5.21	3.87	-1.34
4	10.2	2.36e15	4.92	3.77	-1.15
5	10.3	2.36e15	4.92	3.92	-0.998
6	10.6	2.37e15	5.07	4.19	-0.879
7	10.6	2.37e15	4.93	4.14	-0.796
8	10.5	2.37e15	5.06	4.08	-0.984
9	10.5	2.37e15	5.00	4.05	-0.954
10	10.4	2.38e15	4.93	3.94	-0.997

Although the model does not predict the exact correct values of new_deaths, the prediction is within the range of the 95% confidence interval. Thus, the strength of the model is excellent, and it could make a nearly perfect prediction of new deaths from covid-19 in the UK.

Table 29: Sum of squared errors (SSE) of mod_UK

sum of squared errors (SSE)	1046.654

Finally, the SSE of this model is 46.54, which represents a good fit of the model.

5. Discussion

5.1 Comparison between Bangladesh and the UK:

Table 30: Correlation of variables of this study in the UK and Bangladesh.

Bangladesh			United Kingdom		
New_cases$New_Deaths	**People_vaccinated$New_Deaths**	**New_cases$People_vaccinated**	**New_cases$New_Deaths**	**People_vaccinated$New_Deaths**	**New_cases$People_vaccinated**
0.81	-0.70	-0.40	0.50	-0.58	0.29

In Bangladesh, there was a strong positive correlation between COVID-19 daily new_cases and new_deaths whereas the UK experienced a moderate positive correlation between these two variables. Nonetheless, in both countries, people_vaccinated and daily new_deaths were negatively correlated with high value. Yet, in Bangladesh, one unit of people_vaccinated increased, and new_deaths tended to decrease by -0.70 directions, which is higher than that of the UK correlation coefficient by -0.12. Furthermore, the correlation between new_cases and people_vaccinated in Bangladesh was negatively correlated with a moderate coefficient. Conversely, in the UK, people_vaccinated and new_cases are still positively correlated in our sample data. However, previous research has shown that Covid vaccination helped to decrease the new_cases and new_deaths in the real world (Li et al., 2021). In Tennessee, United States, vaccination data from December 17, 2020, to March 3, 2021, has shown that vaccination could reduce daily cases, daily deaths, and hospitalization rates among older people. On the other hand, vaccination did not significantly change the death rate of other age groups there (Roghani, 2021). In the context of Bangladesh and the UK, both regression models were a good fit, and the prediction of covid deaths by these two variables is nearly appropriate. In both countries, nearly 84 percent of the variability of COVID-19 new deaths could be explained by these two independent variables.

5.2 Limitation of the study

Although daily COVID-19 data are time-series in nature and may exhibit temporal dependence, the present study adopts a regression-based predictive framework for comparative analysis. The model does not explicitly incorporate lag structures or autoregressive components. Future research may extend this framework using distributed lag or time-series models to capture dynamic temporal effects more comprehensively.

5.3 Future Research Direction:

Table 31: Comparative study of Bangladesh and the UK in different indicators

Indicator	**United Kingdom**	**Bangladesh**
Total covid_cases	24812582	2046165
Total Covid_deaths	232112	29477
Population	67508936	171186368
Population Density	286.50	1265.034
Life_expentency	81.32	72.59
Median_Age	40.8	27.5
GDP-per _capita	$39753.24	$523.98
Cardiovase-death-rate	122.137	298
Diabetes_prevalence	8.38	8.38
Hospital_bed_Per_thousand	2.58	0.8
Human_development_index	0.932	0.632

Source: Our World in Data COVID-19 dataset

However, despite having a better position in all economic and social aspects, COVID-19 deaths and cases in the United Kingdom were much higher than those in Bangladesh. Overall, more exploratory research could be conducted in the future to examine the factors and reasons associated with more covid-19 deaths in the UK compared to Bangladesh

6. Conclusion

In summary, there was a significant association between COVID-19 new deaths, new_cases, and people vaccinated both in the UK and Bangladesh. Covid-19 vaccine has reduced the number of deaths in both the United Kingdom and Bangladesh. Nevertheless, covid cases in the United Kingdom have not been negatively correlated with people_vaccinated which was seen in our sample data whereas in Bangladesh vaccination is negatively correlated with new_cases. On the other hand, the main limitation of this study was lots of missing data available in the data set. Due to missing data, so many rows had to be deleted which might impact the regression model. Another remarkable limitation was the unavailability of people-vaccinated values in many cells after vaccination was started. If data from the beginning to till date was available without missing data, the model could have produced better results. Moreover, in the multiple linear regression model, more variables could be added as independent variables, but other possible factors like diabetes prevalence, Cardio vase-death-rate, Hospital_bed_Per_thousand and so forth had the same value throughout the Covid period which was not intuitive. As a result, these variables could not be considered.

References

Abedin, M., Islam, M. A., Rahman, F. N., Reza, H. M., Hossain, M. Z., Hossain, M. A., & Hossain, A. 2021. Willingness to vaccinate against COVID-19 among Bangladeshi adults: Understanding the strategies to optimize vaccination coverage. *PLOS ONE, 16*(4), e0250495. https://doi.org/10.1371/journal.pone.0250495

Andrews, N., Stowe, J., Kirsebom, F., Toffa, S., Rickeard, T., Gallagher, E., & Lopez Bernal, J. 2022. COVID-19 vaccine effectiveness against the Omicron (B.1.1.529) variant. *New England Journal of Medicine, 386*(16), 1532–1546. https://doi.org/10.1056/NEJMoa2119451

Anwar, S., Nasrullah, M., & Hosen, M. J. 2021. COVID-19 and Bangladesh: Challenges and how to address them. *Frontiers in Public Health, 8*, 154. https://doi.org/10.3389/fpubh.2020.00154

Asuero, A. G., Sayago, A., & González, A. 2006. The correlation coefficient: An overview. *Critical Reviews in Analytical Chemistry, 36*(1), 41–59. https://doi.org/10.1080/10408340500420303

Bertsimas, D., Ivanhoe, J., Jacquillat, A., Li, M., Previero, A., Lami, O. S., & Bouardi, H. T. 2020. Optimizing vaccine allocation to combat the COVID-19 pandemic. *medRxiv*. https://doi.org/10.1101/2020.11.17.20233213

Chang, D., Chang, X., He, Y., & Tan, K. J. K. 2022. The determinants of COVID-19 morbidity and mortality across countries. *Scientific Reports, 12*(1), 5888. https:// doi.org/10.1038/s41598-022-09783-9

Chatterjee, A., Gerdes, M. W., & Martinez, S. G. 2020. Statistical explorations and univariate time-series analysis on COVID-19 datasets to understand the trend of disease spreading and death. *Sensors, 20*(11), 3089. https://doi.org/10.3390/s20113089

Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., & Edmunds, W. J. 2021. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science, 372*(6538), eabg3055. https://doi.org/10.1126/science.abg3055

Directorate General of Health Services. 2025. *COVID-19 dashboard: Bangladesh COVID-19 update*. Government of the People's Republic of Bangladesh. Retrieved October 20, 2025, from https://dashboard.dghs.gov.bd/pages/covid19.php

Gupta, S., Cantor, J., Simon, K. I., Bento, A. I., Wing, C., & Whaley, C. M. 2021. Vaccinations against COVID-19 may have averted up to 140,000 deaths in the United States. *Health Affairs, 40*(9), 1465–1472. https://doi.org/10.1377/hlthaff.2021.00758

Hammouri, H. M., Sabo, R. T., Alsaadawi, R., & Kheirallah, K. A. 2020. Handling skewed data: A comparison of two popular methods. *Applied Sciences, 10*(18), 6247. https://doi.org/10.3390/app10186247

Khan, A. U., Akter, R., Khan, F. U., Khanom, S., Khan, A. U., & Afsana, A. S. 2021. Second wave and pandemic situation of COVID-19 (2020–2021) in Bangladesh. *Qubahan Academic Journal, 1*(4), 25–31.

Li, Z., Liu, X., Liu, M., Wu, Z., Liu, Y., Li, W., & Luo, Y. 2021. The effect of the COVID-19 vaccine on daily cases and deaths based on global vaccine data. *Vaccines, 9*(11), 1328. https://doi.org/10.3390/vaccines9111328

Liang, L.-L., Kuo, H.-S., Ho, H. J., & Wu, C.-Y. 2021. COVID-19 vaccinations are associated with reduced fatality rates: Evidence from cross-country quasi-experiments. *Journal of Global Health, 11*, 05015. https://doi.org/10.7189/jogh.11.05015

McGuinness, D., Bennett, S., & Riley, E. 1997. Statistical analysis of highly skewed immune response data. *Journal of Immunological Methods, 201*(1), 99–114. https://doi.org/10.1016/S0022-1759(96)00150-1

Menon, K. 2023. The complete guide to skewness and kurtosis. *Simplilearn*. Retrieved December 31, 2023, from https://www.simplilearn.com/tutorials/statistics-tutorial/skewness-and-kurtosis

Merchant, H. A., Kow, C. S., & Hasan, S. S. 2021. COVID-19 first anniversary review of cases, hospitalization, and mortality in the UK. *Expert Review of Respiratory Medicine*, *15*(8), 973–978. https://doi.org/10.1080/17476348.2021.1965678

Moghadas, S. M., Vilches, T. N., Zhang, K., Wells, C. R., Shoukat, A., Singer, B. H., & Fitzpatrick, M. C. 2021. The impact of vaccination on coronavirus disease 2019 (COVID-19) outbreaks in the United States. *Clinical Infectious Diseases*, *73*(12), 2257–2264. https://doi.org/10.1093/cid/ciab266

Montgomery, D. C., Peck, E. A., & Vining, G. G. 2021. *Introduction to linear regression analysis* (6th ed.). Wiley.

Noel, D. D., Justin, K. G. A., Alphonse, A. K., Désiré, L. H., Dramane, D., Nafan, D., & Malerba, G. 2021. Normality assessment of several quantitative data transformation procedures. *Biostatistics and Biometrics Open Access Journal*, *10*(3), 51–65.

Olowe, K. J., Edoh, N. L., Zouo, S. J. C., & Olamijuwon, J. 2024. Comprehensive review of logistic regression techniques in predicting health outcomes and trends. *World Journal of Advanced Pharmaceutical and Life Sciences*, *7*(2), 16–26. https://doi.org/10.53346/wjapls.2024.7.2.0039

Onovo, A. A., Kalaiwo, A., Obanubi, C., Odezugo, G., Estill, J., & Keiser, O. 2021. Estimates of the COVID-19 infection fatality rate for 48 African countries: A model-based analysis. *BioMed*, *1*(1), 63–79.

Osborne, J. 2019. Notes on the use of data transformations. *Practical Assessment, Research, and Evaluation, 8*(1), 1–5.

Osborne, J. W., & Waters, E. 2019. Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research, and Evaluation, 8*(1), 2.

Pritchard, E., Matthews, P. C., Stoesser, N., Eyre, D. W., Gethings, O., Vihta, K.-D., & Bell, I. 2021. Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom. *Nature Medicine*, *27*(8), 1370–1378. https://doi.org/10.1038/s41591-021-01438-2

Rahman, A., & Sathi, N. J. 2021. Risk factors of the severity of COVID-19: A meta-analysis. *International Journal of Clinical Practice, 75*(7), e13916. https://doi.org/ 10.1111/ijcp.13916

Rahman, M. S., Harun, M. G. D., Sumon, S. A., Mohona, T. M., Abdullah, S. A. H. M., Khan, M. N. H., & Anwar, M. M. U. 2022. Hospitalization and mortality by vaccination status among COVID-19 patients aged ≥25 years in Bangladesh. *Vaccines*, *10*(12), 1987. https://doi.org/10.3390/vaccines10121987

Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., & Roser, M. 2023. Coronavirus pandemic (COVID-19). *Our World in Data*. https://ourworldindata.org/coronavirus

Roghani, A. 2021. The influence of COVID-19 vaccination on daily cases, hospitalization, and death rate in Tennessee, United States: Case study. *JMIRx Med*, *2*(3), e29324. https://doi.org/10.2196/29324

UK Health Security Agency. 2025. *National flu and COVID-19 surveillance report: 11 September 2025 (week 37)*. GOV.UK. https://www.gov.uk/government/publications/covid-19-weekly-surveillance-reports

Uyanık, G. K., & Güler, N. 2013. A study on multiple linear regression analysis. *Procedia – Social and Behavioral Sciences*, *106*, 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027

World Health Organization. 2020. *Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19)*. https://www.who.int/publications/i/item/who-china-joint-mission-on-covid-19