Computational Journal of Mathematical and Statistical Sciences
2(2), 223–239
DOI:10.21608/CJMSS.2023.229207.1014
https://cjmss.journals.ekb.eg/

A Statistical Analysis of Excess Mortality Mean at Covid-19

Md Nurul Raihen[1*], **Sultana Akter**[2], **Fariha Tabassum**[3] , **Farjana Jahan**[2], **Shakera Begum**[2]

[1] Department of Mathematics and Computer Science, Fontbonne University, Saint Louis, 63105, MO, USA; nraihen@fontbonne.edu

[2] Department of Statistics, Western Michigan University, Kalamazoo 49006, MI, USA; sbg2612@wmich.edu

[3] Department of Sociology, Western Michigan University, Kalamazoo, 49006, MI, USA; fbv2349@wmich.edu

* **Correspondence:** nraihen@fontbonne.edu

Abstract: When it comes to making assessments about public health, the mortality rate is a very important factor. The COVID-19 pandemic has exacerbated well-known biases that affect the measurement of mortality, which varies with time and place. The COVID-19 pandemic took the world off surveillance, and since the outbreak, it has caused damage that many would have thought unthinkable in the present era. By estimating excess mortality for 2020 and 2021, we provide a thorough and consistent evaluation of the COVID-19 pandemic's effects. Excess mortality is a term used in epidemiology and public health to describe the number of fatalities from all causes during a crisis that exceeds what would be expected under 'normal' circumstances. Excess mortality has been used for thousands of years to estimate health emergencies and pandemics like the 1918 "Spanish Flu"6. Excess mortality occurs when actual deaths exceed previous data or recognized patterns. It could demonstrate how a pandemic affected mortality rate. The estimates of excess mortality presented in this research are generated using the procedure, data, and methods described in detail in the methods section and briefly summarized in this study. We explored different regression models in order to find the most effective factor for our estimates. We predict the pandemic period all-cause deaths in locations lacking complete reported data using the Binary logistic regression, and Probit regression analysis count framework. Standardized residual plots, AIC, and Variance Inflation Factor (VIF) after checking all of those, we found some significant predictors from our choosing model , and the coefficient of all predictors gave the information that some factors have positive effect, and some has a negative effect at excess mortality at COVID-19 (2020-2021).

Keywords: COVID-19, Excess Mortality, Pandemic, Probit Regression, Logistic regression.
Mathematics Subject Classification: 62J12; 62G08.
Received: 14 August 2023; **Revised:** 15 September 2023; **Accepted:** 20 September 2023; **Online:** 24 September 2023.

1. Introduction

Coronaviruses are a type of virus. A coronavirus identified in 2019, SARS-CoV-2, has caused a pandemic of respiratory illness, called COVID-19. The World Health Organization has been tracking the impact of COVID-19 as the pandemic has evolved over time. The pandemic's conditions, which have overburdened some health systems and caused some patients to avoid care, have caused further deaths. Due to COVID-19 data difficulties, excess mortality is a more objective and comparable indicator. The WHO defines excess mortality as "the mortality above what would be expected based on the non-crisis mortality rate in the population of interest" (See [1]).

The excess mortality associated with the COVID-19 pandemic is utilized to quantify the pandemic's direct and indirect effects. Excess mortality is the difference between the predicted total number of deaths for a particular location and time period and the number that would have been expected in the absence of a crisis (e.g., the COVID-19 pandemic). In accordance with the difficulties associated with using reported data on COVID-19 cases and fatalities, excess mortality is viewed as a more objective and comparable measure that accounts for both the direct and indirect effects of the pandemic (See [2]).

The World Health Organization is being updated with overall case and mortality rates. These figures obviously don't give a clear picture of the COVID-19-related health burden, or the number of lives lost as a result of the epidemic. Some of the fatalities that can be attributed to COVID-19 have not been confirmed as being caused by the virus because premortem tests were not carried out. Raihen et al. [3] there has been a wide range of approaches to death certification across countries, especially in the face of co-morbidities and the emergence of COVID-19.

The primary outcome to evaluate excess death was excess mortality during the pandemic, defined as the difference between the number of reported all-cause deaths and the expected number of deaths during the pandemic. The formula for calculating excess mortality was as follows: excess mortality = estimated deaths − expected deaths population.

What is the meaning of positive excess mortality and negative excess mortality? Positive excess mortality occurs when the number of expected deaths from prehistoric pandemic data is less than the number of deaths observed during the Covid-19 period. This positive excess mortality can be caused by several factors, including the seriousness of the disease, the burden on the healthcare system, changes in behavior or social connections, and other indirect effects. If deaths that would have occurred in the absence of the pandemic were prevented as a result of the actions taken to combat the pandemic, negative excess deaths might have occurred. The number of mortality from causes other than COVID-19 has decreased as a result of some public health interventions (such as lockdown, social withdrawal, mask use, and working from home) [2].

1.1. Research Question

- The major goal of this study is to determine which factors (region, year, sex, age_group, and pop) are significant with different aspects of excess mortality presence of Covid-19.
- In order to put this idea various quantitative research strategies were used and figure out which regression model was best fit for the excess mortality mean considering the people of COVID-19 disease

2. Literature Review

According to the World Health Organization (WHO), there were an estimated 14.9 million excess deaths caused by COVID-19 worldwide between 2020 and 2021; See Jha [4]. The data provided by WHO for deaths worldwide are lower than the estimates provided by the Institute for Health Metrics and Evaluation (IHME)2, which reported 18.2 million deaths (17.1 million–19.6 million), and the estimates provided by The Economist, which provided 17.7 million deaths (13.9 million–21.1 million) for the same time period. On the other hand, Rocco et al [5] estimates from the government based on data from the Coronavirus App on the number of deaths caused by COVID-19 over the world in 2020–2021 imply that the number is less than 6 million.

When making public health decisions, mortality rates are crucial (See [6]). On the other hand by Karlinsky and Ramírez [7, 8], countries, health systems, and individual physicians all classify deaths differently. Achilleos et al. [9] inadequate tests and overloaded health systems, caused by a sudden increase in COVID-19 symptom patients in most countries, may have led to an underestimation of COVID-19-related fatalities in the early stages of the pandemic.

Furthermore, deaths that are indirect during the pandemic, such as those caused by resource restrictions in health care systems, unnatural causes, or severe occurrences, are likely to be misclassified as direct mortality of COVID-19 due to misdiagnosis and pandemic "bias" (See [6, 8, 10]). Even Prior to the massive reorganization of death causes caused by COVID-19, death certificates already had a bad reputation for being inaccurate. Gobiņa, Kiang and Koffman [11, 12, 13] COVID-19 and other diagnoses on the death certificate may be more difficult to assign if comorbidities are present.

Death rates, death counts, and life expectancy are only few of the metrics that can be used to provide an overview of mortality in a given location. Death rates and death counts are the most widely used indices for determining excess mortality [14, 15, 16]. Researchers Aburto, Basellini, Németh and Stokes [17, 18, 19, 20] have used a wide variety of death rates, including crude death rates (CDRs) and age-specific death rates; See Németh et al. [19] and age-standardized death dates (SDRs), Islam and Krieger [21, 22] variation in the predicted mortality level used to assess excess mortality may emerge from the fact that these indices capture a wide range of mortality levels and trends.

In this paper, we introduce the World death Dataset, an effort to collect and maintain global death statistics that are updated on a regular basis; See Raihen and Akter [23]. The dataset is publicly available and is almost daily updated at *https://www.who.int/data/sets/global-excess-deaths-associated-with-covid-19-modelled-estimates*. Since the work of our research, the dataset has been added to Our World in Data's [24] and The Economist's and Financial Times' excess mortality trackers. We conclude that data from all different locations is reliable enough to permit computation of excess mortality (see Discussion), while not all countries give data of the same quality or detail.

In addition to deaths caused by COVID-19 infection, Mungmunpuntipantip, Rozenfeld and Salottolo [25, 26, 27] it is possible that social distancing mandates and other pandemic restrictions reduced the number of deaths caused by certain diseases and injuries, such as those caused by traffic accidents, while increasing the number of deaths caused by others, such as those caused by chronic and acute conditions affected by delayed care-seeking in overstretched health-care systems, by Zubiri and Folino [28, 29] in comparison to expected or baseline conditions. It is difficult to determine how much of the excess mortality is attributable to COVID-19 infection and how much is attributable to other societal, economic, or behavioral changes associated with the pandemic. This is especially the case because

there is a lack of detailed data on the specific causes of death in many countries. Understanding the total mortality impact that the pandemic has is a key first step, even though it will be extremely necessary to identify the factors that contributed to the excess mortality that was observed.

3. Methodology

3.1. Data Source

As the COVID-19 epidemic has grown over time, the World Health Organization (WHO) has monitored its effects. WHO (World Health Organization) has been informed how many people have been infected and how many have died. Besides the deaths that can be directly linked to the pandemic, there are also deaths that can be linked to conditions that have been around since the pandemic started and have caused some health systems to be overloaded or some patients to avoid getting care. Since using reported COVID-19 data can be hard, excess mortality is thought to be a more objective and similar measure. The World Health Organization (WHO) says that "excess mortality" is a death rate that is higher than what would be predicted based on the normal death rate in the population of interest. According to data collection history, the official WHO estimate published the excess mortality data expected as yearly based frequencies on May 20, 2021.

3.2. Data Structures

The data on excess mortality is a binary data here (positive and negative excess mortality), and its spatial coverage is Global. In this article, we will detail the various models we performed on this dataset, as well as the methods we used to draw conclusions about the excess mortality of COVID-19 and to predict the efficient predictors. In order to properly conduct any data analysis or data operation, we must first possess extensive background knowledge in the field in question. Therefore, we will discuss data set characteristics and how they relate to one another. Five of the attributes in this data collection are classified, and three are numerical attributes. In the latest estimated data at WHO, there are 224 observations collected about the excess deaths associated with the COVID-19 pandemic from all causes by age, sex, and year.

3.3. Data Manipulation

There is a total of 224 data points in the WHO collection (word.csv data); perhaps here I got 184 observations, which contain only positive excess mortality, and 40 observations which have negative excess mortality. For my analysis perspective, I considered my explanatory variables to be location, year, age group, sex, and pop, and the response variable to be excess.mean.binary (positive excess mortality=1, and negative excess mortality=0).

3.4. Model Procedure

We use the programming language R to fit logistic regression, odds ratio of each coefficient and Probit models to the Mortality data because the data is of the binary type. Excel, Microsoft Word, and R-code have all been put to use at various points throughout our study to perform manipulations on raw data. We provide an outline of the theory as well as its implementation in R (R Development Core

Table 1. A list of the features, along with a description of each one, that are contained in the dataset

Feature name	**Type**	**Description and values**	**missing(%)**
location	Nominal	Region (Global or WHO region: AFR: African Region; AMR: Region of the Americas; EMR: Eastern Mediterranean Region; EUR: European Region; WPR: Western Pacific Region; SEAR: South-East Asia Region)	0%
year	Nominal	year of death (2020 and 2021)	0%
sex	Nominal	gender (male and female)	0%
age	Nominal	age- group from 0 to 85+ (0-24,25-34,35-44,45-54,55-64,65-74,75-84,85+)	0%
Pop	Numeric	Sex-and age-specific population number	0%
type	Nominal	estimated type for select year (reported or predicted)	0%
expected. mean	Numeric	expected deaths from all-causes by age, sex and year (mean)	0%
acm. mean	Numeric	estimated deaths from all-causes by age, sex, and year (mean)	0%
Excess. mean.	Numeric	excess deaths associated with COVID-19 pandemic from all-causes by age sex and year (positive and negative excess mortality mean)	0%

Team 2008) for certain fundamental count data regression models such as logit and probit model (see Table 1 for an overview of the model's components).

The function in R that we are using is called glm(). The word GLM stands for Generalized Linear Model. Logistic Regression, Probit Regression fall under GLM. GLM has three components:

1. Random Component: The response variable. For Logistic Regression and Probit Regression, we have binary variables and we assume that they follow Bernoulli Distribution.
2. Systematic Component: The covariates or explanatory variables.
3. Link function: The function that links between the systematic component and the random component. In Logistic regression, it is log odds, in Probit regression it is probit of the probability of success.
   They are implemented in R by the glm() function (See Chambers et al. [30] in the stats package and the glm() function in the nlme package (See Venables et al. [31]), and logit regression is an extended of GLM. The classical Poisson model and the logit and probit models are both described in a generalized linear model (GLM) framework. Although all GLMs utilize the same log-linear mean function ($\log \mu = x > \beta$), each GLM makes a unique set of assumptions regarding the likelihood of the remaining variables; See Raihen and Akter [32].

4. Analysis

4.1. Binary logistic Model

To investigate the excess mortality at COVID-19 in 2020-2021

- In the logistics model with more than one independent variable, the model can be written as:
- Where Z is a linear function of the explanatory variables.
- If $X_1, \ldots, X_k$ represent various determining characteristics of loyalty, then Z equation would be as follows.

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k,$$

where, X_i = explanatory variables for all $i = 1, 2, \ldots, k$ and β_i = parameters of the model for all $i = 1, 2, \ldots, k$.

The Functional Form of the Logistic Regression model:

$$logit\ [p(Y \leq j)] = \frac{\log(p(Y \leq j))}{(1 - p(Y \leq j))} = \alpha + \beta x$$

where $j = 1, 2, \ldots, j - 1$, and X_i's are categorical variables or continuous but Y are categorical with order.

For our finding: Z = excess mortality at COVID-19, β_0 = intercept, and $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are our coefficient of location, year, age_group, sex, pop respectively.

Table 2. level of response variable for binary logistic model

Dependent Variable	**Measurements**
Excess Mortality Mean	1 = Positive excess mortality mean 0 = Negative excess mortality mean

Binary Probit Regression Model: In Probit Regression, we model the probit of probability of having a disease $P(D = 1)$ in terms of the covariates. The covariate terms can contain main risk factors, confounding terms, as well as interaction terms. Mathematically probit is the inverse of the cumulative distribution function (CDF) of a standard normal distribution. If X is a random variable that follows a normal distribution with mean 0 and variance 1, then CDF of X to be less than u is defined as: $\Pi(u) = P(X \leq u)$. The probit is inverse of the function $\Pi()$, i.e., $\Pi^{-1}()$.

In this section, we discuss the data processing, data summary, and methodology used to get estimates of increased mortality based only on the presence of Covid-19. In this investigation, we use both direct and indirect COVID-19 death. In addition to Stein and Raihen [33] estimating the number of direct and indirect deaths caused by COVID-19 required the Department of Economic and Social Affairs of the United Nations (UN DESA) and the WHO Regulations for Scientific and Advisory Groups to work together to produce harmonized excess mortality procedures.

4.2. The Logistic Regression Model

Since our data is binary type, we fitted our data with the generalized linear model (glm) of the Binomial regression family.

Model 1: Since our data is binary type, we fitted our data with the generalized linear model (glm) of the Binomial regression family.

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -7.301e+03  1.615e+03  -4.520 6.19e-06 ***
locationAMR     1.486e+00  1.232e+00   1.206 0.227715
locationEMR     1.496e+00  1.233e+00   1.214 0.224871
locationEUR     2.382e+00  1.332e+00   1.788 0.073762 .
locationGlobal  3.008e-01  1.761e+00   0.171 0.864381
locationSEAR    2.103e+00  1.288e+00   1.634 0.102325
locationWPR    -4.270e+00  1.240e+00  -3.443 0.000575 ***
year            3.616e+00  7.997e-01   4.522 6.13e-06 ***
sexMale         1.852e+00  6.944e-01   2.667 0.007653 **
age_group0-24  -8.397e+00  2.211e+00  -3.798 0.000146 ***
age_group25-34 -7.319e+00  1.933e+00  -3.785 0.000153 ***
age_group35-44 -2.744e+00  1.767e+00  -1.553 0.120436
age_group45-54 -1.981e-01  1.969e+00  -0.101 0.919856
age_group55-64 -1.463e-01  1.960e+00  -0.075 0.940526
age_group65-74 -9.837e-02  1.954e+00  -0.050 0.959840
age_group75-84 -2.570e+00  1.739e+00  -1.478 0.139321
pop             1.444e-09  1.912e-09   0.756 0.449920
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 1. Coefficient Table for the Logistic regression model

From the coefficient table of the logisitc regression model, it shows that some predictors are significant at a p-value¡0.05. The asterisks and dot denote the significance (different from zero) of each regression parameter as follows: No asterisk and no dot: NOT significant, no asterisk and a single dot: significant at the 10% level and one or more asterisks: significant at the 5% by Mansfield et al. [34].

Assumption Checking:

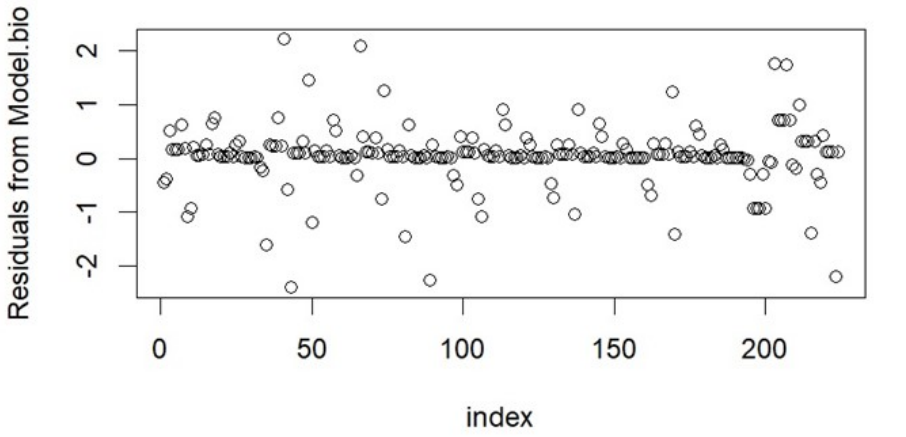

Figure 2. Residula plot of poisson regression

From the above figure, we can see from the residuals plot of fitting model1 (logistic Regression) that the residuals have spread out in a random way with no properly clear pattern. This shows that the model fitting for model1 has no systematic bias, which is a good sign for a model. But we must consider more things for assumption checking.

Log odds of logistic regression analysis:

term <chr>	estimate <dbl>	std.error <dbl>	statistic <dbl>	p.value <dbl>
(Intercept)	0.000000e+00	1.615361e+03	-4.51993860	6.185757e-06
locationAMR	4.418105e+00	1.231662e+00	1.20626558	2.277151e-01
locationEMR	4.464381e+00	1.232725e+00	1.21367713	2.248710e-01
locationEUR	1.082902e+01	1.332277e+00	1.78808845	7.376174e-02
locationGlobal	1.350900e+00	1.760953e+00	0.17079994	8.643811e-01
locationSEAR	8.194390e+00	1.287551e+00	1.63368331	1.023254e-01
locationWPR	1.398247e-02	1.240139e+00	-3.44312152	5.750409e-04
year	3.720060e+01	7.997287e-01	4.52193981	6.127548e-06
sexMale	6.373268e+00	6.944482e-01	2.66702737	7.652544e-03
age_group0-24	2.255001e-04	2.211112e+00	-3.79772261	1.460316e-04
age_group25-34	6.630868e-04	1.933338e+00	-3.78547618	1.534145e-04
age_group35-44	6.429612e-02	1.767127e+00	-1.55294757	1.204357e-01
age_group45-54	8.202810e-01	1.968980e+00	-0.10061467	9.198564e-01
age_group55-64	8.639299e-01	1.960395e+00	-0.07460926	9.405256e-01
age_group65-74	9.063096e-01	1.953646e+00	-0.05035423	9.598401e-01
age_group75-84	7.650586e-02	1.738719e+00	-1.47832316	1.393213e-01
pop	1.000000e+00	1.911659e-09	0.75554793	4.499203e-01

Figure 3. log odds of logistic regression model

4.3. The probit Regression Model

They mainly differ in the link function.

In Logit:

$$Pr(Y = 1 \mid X) = (1 + e^{X'\beta})^{-1}$$

In Probit:

$$Pr(Y = 1 \mid X) = \Phi(X'\beta)$$

And logistics have slightly flatter tails. i.e., the probit curve approaches the axes more quickly than the logit curve.

A popular generalization of logistic regression is probit regression. And logistics have slightly flatter tails. i.e., the probit curve approaches the axes more quickly than the logit curve. Occasionally, data exhibit more variation that exceeds the mean. Consequently by Charro and Raihen [35], the Probit model is nested within the Logistic binomial model.

Model 2: Comparison of the Poisson Regression Model and the Negative Binomial Regression Model

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -1.587e+03  1.780e+02  -8.915  < 2e-16 ***
locationAMR     9.427e-01  1.597e-01   5.902 3.59e-09 ***
locationEMR     3.312e-02  1.577e-01   0.210 0.833637
locationEUR     6.571e-01  1.566e-01   4.198 2.70e-05 ***
locationGlobal  2.502e+00  2.047e-01  12.223  < 2e-16 ***
locationSEAR    1.357e+00  1.573e-01   8.627  < 2e-16 ***
locationWPR    -6.501e-01  1.881e-01  -3.457 0.000546 ***
year            7.908e-01  8.809e-02   8.977  < 2e-16 ***
age_group0-24  -3.497e+00  2.799e-01 -12.495  < 2e-16 ***
age_group25-34 -2.452e+00  2.020e-01 -12.141  < 2e-16 ***
age_group35-44 -1.556e+00  1.689e-01  -9.214  < 2e-16 ***
age_group45-54 -5.925e-01  1.634e-01  -3.627 0.000287 ***
age_group55-64 -9.698e-02  1.607e-01  -0.604 0.546120
age_group65-74  1.577e-01  1.584e-01   0.995 0.319546
age_group75-84 -9.865e-02  1.607e-01  -0.614 0.539202
sexMale         3.104e-01  8.594e-02   3.612 0.000304 ***
pop            -7.307e-10  3.719e-10  -1.965 0.049457 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4. Coefficient table of Probit regression model

From this result we got some significant predictors for positive excess mortality at COVID-19. The asterisks and dot denote the significance (different from zero) of each regression parameter as follows: No asterisk and no dot: NOT significant, No asterisk and a single dot: significant at 10% level and one or more asterisks: significant at 5% level by Mansfield et al. [34].

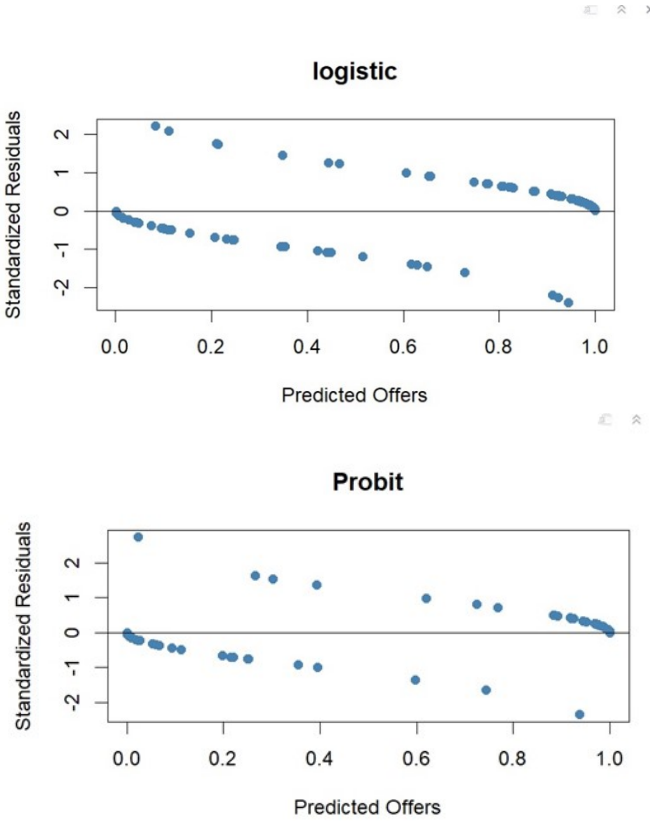

Figure 5. Standardized residual plot for comparing of Logit and Probit regression model.

Table 3. Akaike Information Criteria (AIC), and Residual deviance

Model	df	AIC	Residual Deviance
Logit Model	207	105.59	71.594
Probit Model	207	105	71.003

AIC and residual deviance for both models are approximately equal in Table 3. Yamaoka et al. [36] model with smaller AIC, and smaller residual deviance is a better model.

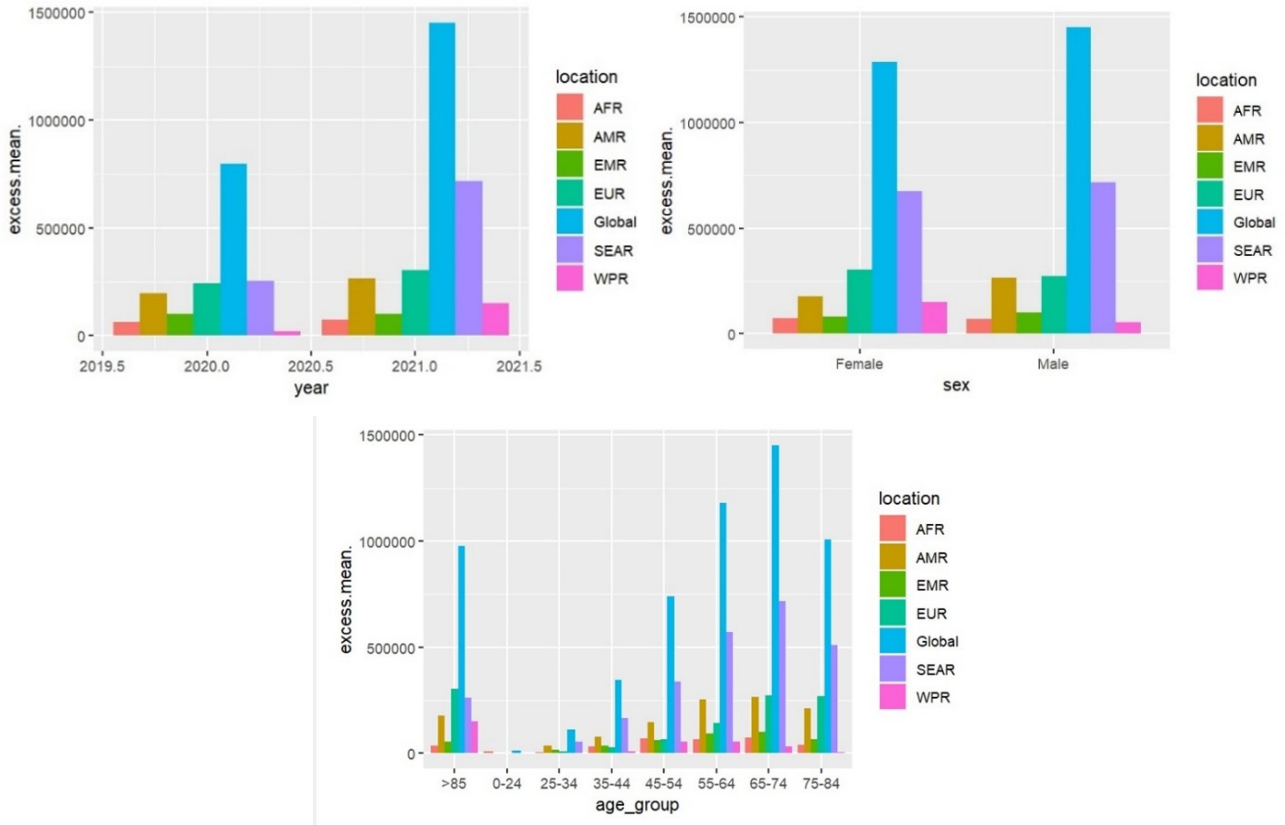

Figure 6. Histogram plot for each level of predictors.

Model 3:

```
Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)         -2.051e+05  5.747e+07  -0.004    0.997
locationAMR         -7.374e+04  1.838e+07  -0.004    0.997
locationEMR         -2.324e+05  5.128e+07  -0.005    0.996
locationEUR         -3.164e+05  5.896e+07  -0.005    0.996
locationGlobal       1.707e+05  4.953e+07   0.003    0.997
locationSEAR         2.784e+04  4.332e+07   0.001    0.999
locationWPR          1.817e+05  3.425e+07   0.005    0.996
year                 1.015e+02  2.845e+04   0.004    0.997
sexMale             -6.660e+04  1.315e+07  -0.005    0.996
age_group0-24        6.665e+05  1.030e+08   0.006    0.995
age_group25-34       3.520e+05  6.754e+07   0.005    0.996
age_group35-44       1.644e+05  7.188e+07   0.002    0.998
age_group45-54       1.917e+05  7.386e+07   0.003    0.998
age_group55-64       1.480e+05  7.090e+07   0.002    0.998
age_group65-74       1.049e+05  6.870e+07   0.002    0.999
age_group75-84       1.786e+05  5.356e+07   0.003    0.997
pop                 -1.462e-03  2.422e-01  -0.006    0.995
locationAMR:year     3.651e+01  9.101e+03   0.004    0.997
locationEMR:year     1.151e+02  2.538e+04   0.005    0.996
locationEUR:year     1.567e+02  2.919e+04   0.005    0.996
locationGlobal:year -8.452e+01  2.452e+04  -0.003    0.997
locationSEAR:year   -1.378e+01  2.145e+04  -0.001    0.999
locationWPR:year    -8.996e+01  1.696e+04  -0.005    0.996
year:sexMale         3.297e+01  6.511e+03   0.005    0.996
year:age_group0-24  -3.300e+02  5.097e+04  -0.006    0.995
year:age_group25-34 -1.743e+02  3.343e+04  -0.005    0.996
year:age_group35-44 -8.137e+01  3.558e+04  -0.002    0.998
year:age_group45-54 -9.490e+01  3.656e+04  -0.003    0.998
year:age_group55-64 -7.328e+01  3.510e+04  -0.002    0.998
year:age_group65-74 -5.192e+01  3.401e+04  -0.002    0.999
year:age_group75-84 -8.844e+01  2.652e+04  -0.003    0.997
year:pop             7.238e-07  1.199e-04   0.006    0.995
```

Figure 7. An overview of the statistically significant model (of logistic regression with interaction term)

It represents our statistically significant factors considering interaction term to estimate the excess mortality at COVID-19 at all causes of effect. The figures mentioned above illustrate that the levels of

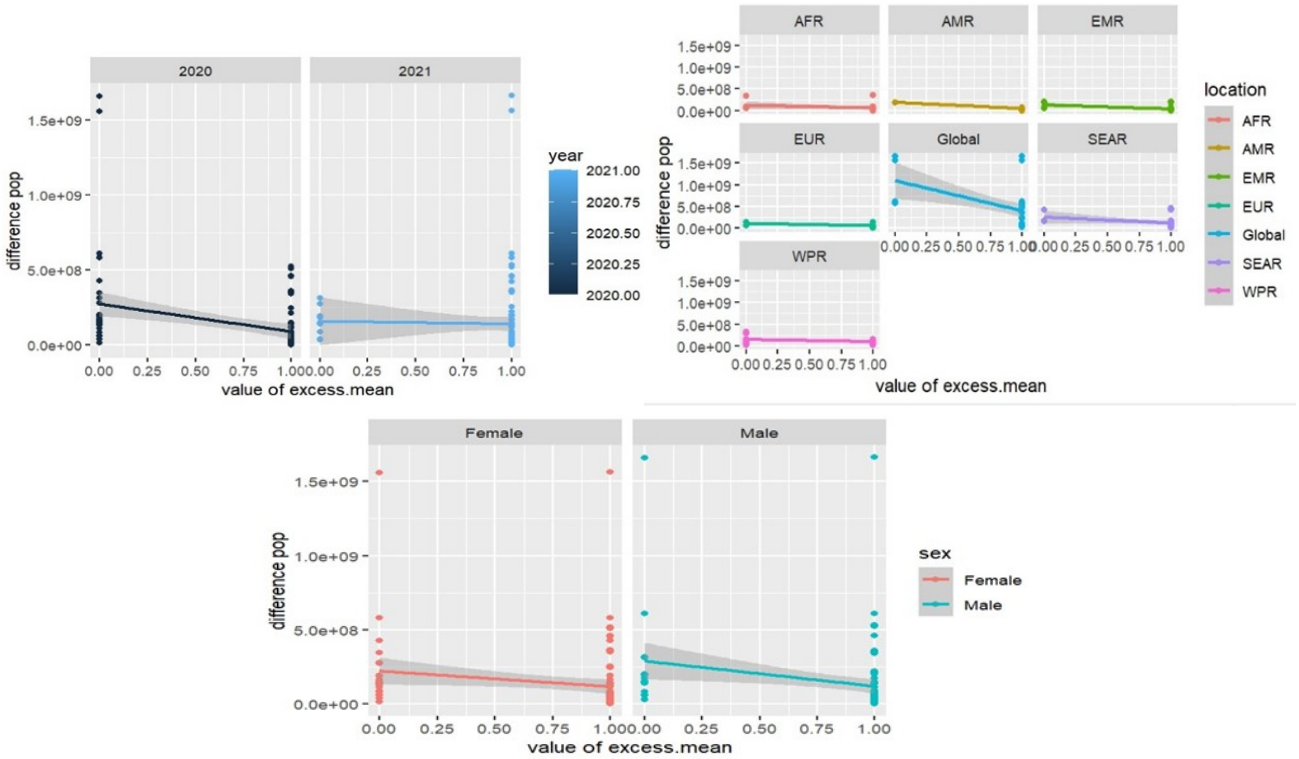

Figure 8. Plots of Comparing According to levels of explanatory variables.

each factor correspond to which belongs the largest and least positive excess mortality.

```
   locationAMR    locationEMR    locationEUR locationGlobal
      1.672710       1.679367       1.756140       3.877599
  locationSEAR    locationWPR           year        sexMale
      1.671254       3.353920       1.566893       1.288854
 age_group0-24 age_group25-34 age_group35-44 age_group45-54
     10.961058       8.410773       3.288121       1.939230
age_group55-64 age_group65-74 age_group75-84            pop
      1.922148       1.912329       3.217050       4.741024
```

Figure 9. Variance Inflation factor (VIF) (logit model)

This is the result of the Variance Inflation factor (VIF) which is used to test multicollinearity in Raihen et al. [37].

Table 4. An overall Comparison among of all models

Model	**df**	**AIC**	**Residual Deviance**
Logit Model	207	105.59	71.594
Probit Model	207	105	71.003
Logit Model (Interaction term)	37	101.02	37.017

As Table 4 shows that AIC and residual deviance fore both models Logit and Probit models are approximately equal though AIC for Logit model with interaction term is much smaller.

5. Results and Discussion

Our research was to forecast the outcome using only the four input factors and the single output variable. In order to determine something, we carried out a series of models. Comparing observed number of deaths with the mean number of deaths from our data there were 48.28% female population and 51.71% male population , there were 13.13% African, 16.21 AMR,14.21% EMR, 14.76% EUR, 17.76% SEAR, 8.69% and Global 15.21% , there were 48.02% population from 2020 year, and 51.97% population from 2021, there were 49.28% female and 50.71% male, there were 14.67% population age > 85, 6.52% population age between 0-24, 7.60% population age between 25-34, 13.58% population age between 35-44, 14.67% population age between 45-54, 14.67% population age between 55-64, 14.67% population age between 65-74, 13.58% population age between 75-84, and 14.67% population age > 85 to analysis of excess mortality at COVID-19.

Table 1 shows an overview of the structure of data for the excess mortality mean. We found that there are no missing observations. **Table** 2 summarizes the level of our response variable (excess.mortality.mean). **Figure** 1 The coefficient result under logistic regression model of location provides that significant log odds of excess mortality mean significantly increases under location factors AMR, EMR,EUR, Global, and SEAR compared to the factor of AFR when all other variables are held constant(p-value < 0.05, so they are statistically insignificant factors), and the log odds of mean of excess mortality significantly decreases under location factors WPR compared to the factor of AFR

when all other variables are held constant, since (p-value < 0.05) for factor WPR of location has a significance level of predictor for excess mortality mean.

For the year, the log odds of the mean of the number of excess mortality is higher 3.6 times under the year 2021 compared to the year 2020 when all other variables are held constant, and since p-value < 0 so year has a significant factor for the excess mortality of COVID-19. The coefficient of sex shows that the log odds of mean of male of excess mortality mean increased by 1.852 times compared to the female when all other is held constant, and through a p-value < 0.05 sex has a significant effect on the excess death at COVID-19.

The coefficient values of age-group shows that for the age groups of 0-24,25-34,35-44,45-54, 55-64, 65-74,75-84 the log odds of mean decreases compared to age group more than 85 age, and the age group 0-24, and 25-34 are significant factors for the excess mortality mean of the presence of COVID pandemic (hence, p-value < 0.05). Population has a positive effect and is a significant predictor for our analysis of the positive excess mortality mean. (Thus, the coefficient is 1.444×10^{-9} and the p-value < 0.05).

In **Model 2 Figure** 3, we applied Logistic Probit regression model for our excess mortality data at COVID-19. **Figure** 4 is the summary of Logistic Probit regression model result. Result from "R" display in **Figure** 4 we find that the regression coefficients for each of the variables along with the standard errors, z scores, and P-values.The coefficients from **Figure** 4 indicates the probabilities of success of excess mortality means at COVID-19 for location, agegroup, year, gender and population.

From the **Table** 3, thus the AIC and residual deviance of the Logistic regression and Probit model both are approximately same value so that I considered interaction term for logistic model to better fit of our model for our data of COVID-19. **Figure** 6, and **figure** 8 show that which level of each predictor have higher excess mortality mean at COVID-19. **Figure** 7 is the result of all the coefficients of Logistic regression model (**Model 3)** with the interaction term, and it shows that all predictors are insignificant hence p-value for all coefficients are greater than 0.05,and we can not choose this **Model 3** is a better model than the without interaction model of **Model 1**.

Figure 9 is the result of the Variance Inflation factor (VIF) to check the multicollinearity among the dependent variables, hence it shows that almost all predictors have VIF < 10, which means that multicollinearity does not exist at the logistic regression model in this analysis, so we do not need standardized at Logistic regression model for our analysis. A study can be conducted using **Table** 4 to compare various models and determine which one is the most suitable in terms of Excess Mortality Mean during the Covid-19 pandemic. The **Model 3**, which includes an interaction term, exhibits a lower AIC value compared to the other two models. However, upon examining **Figure** 7, it is evident that **Model 3** does not possess statistically significant coefficients for the mean excess mortality associated with COVID-19. Among all the models considered, it is evident that **Model 1** is the most suitable for conducting an analysis of the mean excess mortality during the Covid-19 pandemic in the years 2020-2021.

6. Conclusion

By Raihen and Akter [38], the present study undertook the computation of excess death rates for the period encompassing the global pandemic in the years 2020 and 2021, focusing on both the Global and World Health Organization (WHO) Member State Regions. Instead of presenting novel estimations,

our objective was to rigorously evaluate more efficient methodologies and ascertain the key factors that impact COVID-19's Excess Mortality, comparisons, and limitations, along with the implications of the dataset; See Raihen et al. [39]. One of the strategies was used to display the results of this analysis in Shi et al. [40]. Overall we demonstrate that Logistic Regression Model is a better model than others for achieving the study's purpose. locationWPR, year, age-group (0-24, 25-34,), and sex of excess mortality are the most effective predictors. It also revealed that the Global of location predictors, year 2021, age-group 65-74 has more excess mortality mean, and male has more rate of success of excess mortality than female. Our Logistic Regression Model represents the chance of Excess mortality occurring during the Covid-19 pandemic, and it shows us which predictors are more efficient during that pandemic era in 2020-2021. Logistic Regression without interaction could produce a superior outcome for our study; nonetheless, it does not reveal any efficient predictors for our study's target (Excess Mortality Mean at Covid-19, 2020-2021).

7. Limitation

Unfortunately, few countries have excess mortality data accessible, and this situation will persist due to a lack of prior data. When tracking a pandemic around the world, this is a major disadvantage. Excess mortality calculations require precise, high-frequency mortality data from prior years. Usually, these statistics are only available to wealthier countries that can afford high-quality data reporting systems. To make more precise and useful conclusions about the positive excess mortality mean, more thorough data is required. Future study could be conducted to calculate the extra mortality from all causes attributable to the COVID-19 pandemic more precisely.

8. Future Work

Excessive mortality serves as a significant measure that encapsulates the impact of COVID-19, providing valuable insights for public health policy and the identification of future planning requirements. The research on COVID-19-related excess deaths is expected to be important in the future. Research on excess mortality sheds light on the COVID-19 pandemic's true effects. It considers not just the virus's direct fatalities but also its indirect impacts on mortality, including things like delayed medical attention, mental health problems, and other things.

Researchers can evaluate the efficacy of public health measures like lock downs, mask mandates, and vaccination campaigns during a pandemic by accessing data on excess mortality. Responses to future crises of a similar nature can be improved using this data. The long-term health implications of COVID-19, including any problems and mortality risks related to post-acute sequelae of SARS-CoV-2 infection (PASC or "long COVID"), can also be clarified through research on excess mortality.

Studying excess mortality helps researchers identify susceptible communities who may have been disproportionately affected by the pandemic. This can influence future targeted initiatives and support for these populations. Comparative studies of excess mortality across countries can improve global best practices by revealing the efficacy of pandemic response methods and healthcare systems. Governments, health organizations, and researchers must gather and disseminate precise and thorough data on mortality that is connected to COVID-19, both directly and indirectly, for this study to proceed effectively. This continuous endeavor will enhance comprehension of the pandemic's consequences

and direct forthcoming public health determinations.

Conflict of interest: The authors state that they have no financial or other conflicts of interest to disclose with connection to this research.

References

1. World Health Organization, https://www.who.int/data/sets/global-excess-deaths.
2. World Health Organization, https://www.who.int/data/technical-advisory-group.
3. Raihen, M. N., Akter, S., & Sardar, M. N. (2023). *Food Satisfaction among Students: A Study of Present Public University Students in Bangladesh*. Journal of Mathematics and Statistics Studies, 4(1), 01-18.
4. Jha, P., Brown, P. E., & Ansumana, R. (2022). *Counting the global COVID-19 dead*. The Lancet, 399(10339), 1937-1938.
5. Rocco, P., Rich, J. A., Klasa, K., Dubin, K. A., & Béland, D. (2021). *Who counts where? COVID-19 surveillance in federal countries*. Journal of Health Politics, Policy and Law, 46(6), 959-987.
6. Collaborators, C. E. M. (2022). *COVID-19 Excess Mortality Collaborators: Estimating excess mortality due to the COVID-19 pandemic: A systematic analysis of COVID-19-related mortality*, 2020-21. Lancet, 399, 1513-1536.
7. Karlinsky, A., & Kobak, D. (2021). *Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset*. Elife, 10, e69336.
8. Ramírez-Soto, M. C., & Ortega-Cáceres, G. (2022). *Analysis of excess all-cause mortality and COVID-19 mortality in Peru: observational study*. Tropical Medicine and Infectious Disease, 7(3), 44.
9. Achilleos, S., Quattrocchi, A., Gabel, J., Heraclides, A., Kolokotroni, O., Constantinou, C., ... & Demetriou, C. A. (2022). *Excess all-cause mortality and COVID-19-related mortality: a temporal analysis in 22 countries, from January until August 2020*. International journal of epidemiology, 51(1), 35-53.
10. Ioannidis, J. P. (2021). *Over-and under-estimation of COVID-19 deaths. European journal of epidemiology*, 36, 581-588.
11. Gobiņa, I., Avotiņš, A., Kojalo, U., Strēle, I., Pildava, S., Villeruša, A., & Briģis, Ģ. (2022). *Excess mortality associated with the COVID-19 pandemic in Latvia: a population-level analysis of all-cause and noncommunicable disease deaths in 2020*. BMC public health, 22(1), 1-12.
12. Kiang, M. V., Irizarry, R. A., Buckee, C. O., & Balsari, S. (2020). *Every body counts: measuring mortality from the COVID-19 pandemic. Annals of internal medicine*, 173(12), 1004-1007.
13. Koffman, J., Gross, J., Etkind, S. N., & Selman, L. (2020). *Uncertainty and COVID-19: how are we to respond*. Journal of the Royal Society of Medicine, 113(6), 211-216.
14. Bilinski, A., & Emanuel, E. J. (2020). *COVID-19 and excess all-cause mortality in the US and 18 comparison countries*. Jama, 324(20), 2100-2102.

15. Faust, J. S., Krumholz, H. M., Du, C., Mayes, K. D., Lin, Z., Gilman, C., & Walensky, R. P. (2021). *All-cause excess mortality and COVID-19–related mortality among US adults aged 25-44 years, March-July 2020*. Jama, 325(8), 785-787.

16. Schöley, J. (2021). *Robustness and bias of European excess death estimates in 2020 under varying model specifications*. MedRxiv, 2021-06.

17. Aburto, J. M., Kashyap, R., Schöley, J., Angus, C., Ermisch, J., Mills, M. C., & Dowd, J. B. (2021). *Estimating the burden of the COVID-19 pandemic on mortality, life expectancy and lifespan inequality in England and Wales: a population-level analysi*s. J Epidemiol Community Health, 75(8), 735-740.

18. Basellini, U., Alburez-Gutierrez, D., Del Fava, E., Perrotta, D., Bonetti, M., Camarda, C. G., & Zagheni, E. (2021). *Linking excess mortality to mobility data during the first wave of COVID-19 in England and Wales*. SSM-Population Health, 14, 100799.

19. Németh, L., Jdanov, D. A., & Shkolnikov, V. M. (2021). *An open-sourced, web-based application to analyze weekly excess mortality based on the Short-term Mortality Fluctuations data series*. PLoS One, 16(2), e0246663.

20. Stokes, A. C., Lundberg, D. J., Elo, I. T., Hempstead, K., Bor, J., & Preston, S. H. (2021). It is made available under a CC-BY-NC-ND 4.0 International license . (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. medRxiv preprint doi: https://doi.org/10.1101/2021.07.20.21260869; this version posted July 23, 2021. The copyright holder for this preprint 28 *COVID-19 and excess mortality in the United States: A county-level analysis*. PLOS Medicine, 18(5), e1003571.

21. Islam, N., Shkolnikov, V. M., Acosta, R. J., Klimkin, I., Kawachi, I., Irizarry, R. A., ... & Lacey, B. (2021). *Excess deaths associated with covid-19 pandemic in 2020: age and sex disaggregated time series analysis in 29 high income countries*. bmj, 373.

22. Krieger, N., Chen, J. T., & Waterman, P. D. (2020). *Excess mortality in men and women in Massachusetts during the COVID-19 pandemic*. The Lancet, 395(10240), 1829.

23. Raihen, M. N., & Akter, S. (2023). *Forecasting Breast Cancer: A Study of Classifying Patients' Post-Surgical Survival Rates with Breast Cancer*. Journal of Mathematics and Statistics Studies, 4(2), 70-78.

24. Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., ... & Roser, M. (2020). *Coronavirus pandemic (COVID-19)*. Our world in data.

25. . Mungmunpuntipantip, R., & Wiwanitkit, V. (2021). *The COVID-19 pandemic and traffic accidents. South African Medical Journal*, 111(3), 192-192.

26. Rozenfeld, M., Peleg, K., Givon, A., Bala, M., Shaked, G., Bahouth, H., & Bodas, M. (2021). *COVID-19 changed the injury patterns of hospitalized patients*. Prehospital and disaster medicine, 36(3), 251-259.

27. Salottolo, K., Caiafa, R., Mueller, J., Tanner, A., Carrick, M. M., Lieser, M., ... & Bar-Or, D. (2021). *Multicenter study of US trauma centers examining the effect of the COVID-19 pandemic on injury causes, diagnoses and procedures*. Trauma Surgery & Acute Care Open, 6(1), e000655.

28. Zubiri, L., Rosovsky, R. P., Mooradian, M. J., Piper-Vallillo, A. J., Gainor, J. F., Sullivan, R. J., ... & Reynolds, K. L. (2021). *Temporal trends in inpatient oncology census before and during the COVID-19 pandemic and rates of nosocomial COVID-19 among patients with cancer at a large academic center*. The Oncologist, 26(8), e1427-e1433.

29. Folino, A. F., Zorzi, A., Cernetti, C., Marchese, D., Pasquetto, G., Roncon, L., ... & Iliceto, S. (2020). *Impact of COVID-19 epidemic on coronary care unit accesses for acute coronary syndrome in Veneto region, Italy*. American Heart Journal, 226, 26.

30. Chambers, J., Hastie, T., & Pregibon, D. (1992). *Statistical models in S. In Compstat: proceedings in computational statistics*, 9th symposium held at Dubrovnik, Yugoslavia, 1990 (pp. 317-321). Heidelberg: Physica-Verlag HD.

31. Venables, W. N., Ripley, B. D., Venables, W. N., & Ripley, B. D. (2002). *Random and mixed effects*. Modern applied statistics with S, 271-300.

32. Raihen, M. N., Akter, S., & Sardar, M. N. (2023). *Women's Career Challenges and Opportunities (A Study of Career and Job Satisfaction among Bangladeshi Women)*. Academic Journal of Research and Scientific Publishing, 5(51), 05-22.

33. Stein Jr, I., & Raihen, M. N. (2023). *Convergence Rates for Hestenes' Gram–Schmidt Conjugate Direction Method without Derivatives in Numerical Optimization*. AppliedMath, 3(2), 268-285.

34. Mansfield, E. R., & Helms, B. P. (1982). *Detecting multicollinearity*. The American Statistician, 36(3a), 158-160.

35. Charro, F., Ali, A. H., Raihen, N., Torres, M., & Wang, P. (2023). *A bifurcation phenomenon in a singularly perturbed two-phase free boundary problem of phase transition. Nonlinear Analysis*: Real World Applications, 73, 103911.

36. Yamaoka, K., Nakagawa, T., & Uno, T. (1978). *Application of Akaike's information criterion (AIC) in the evaluation of linear pharmacokinetic equations*. Journal of pharmacokinetics and biopharmaceutics, 6(2), 165-175.

37. . Raihen, N. (2017). *Convergence Rates for Hestenes' Gram-Schmidt Conjugate Direction Method without Derivatives in Numerical Optimization* (Master Thesis, University of Toledo).

38. Raihen, M. N., Akter, S., Tabassum, F., Jahan, F., & Sardar, M. N. (2023). *A Statistical Analysis of Positive Excess Mortality at Covid-19 in 2020-2021*. Journal of Mathematics and Statistics Studies, 4(3), 07-17.

39. Raihen, M. N. I. (2022). *A Bifurcation Phenomenon of Regularized Free Boundary Problems of Two-Phase Elliptic-Parabolic Type (Doctoral dissertation*, Wayne State University).

40. Shi, Y., Wang, G., Cai, X. P., Deng, J. W., Zheng, L., Zhu, H. H., ... & Chen, Z. (2020). *An overview of COVID-19. Journal of Zhejiang University*. Science. B, 21(5), 343.