Post-Pandemic Analysis of COVID-19 Vaccines and Deaths in the Philippines

Elijah Miguel Anacta, Julia Marie Federizo, Ralph Vincent Ortiz
Department of Physics, College of Science, De la Salle University
Corresponding Author, elijah_anacta@dlsu.edu.ph, julia_federizo@dlsu.edu.ph, ralph_ortiz@dlsu.edu.ph

Abstract: The COVID-19 pandemic has impacted developing countries like the Philippines in which livelihood and lifestyle has changed drastically. The government has been in constant discussion with experts and sector leaders in the management of the pandemic crisis. With the emergence of medical intervention in mitigating the effects of the pandemic, analysis and studies have been developed to support claims on the effectiveness of vaccines and policies. After months of pandemic, records and data concerning the Philippine response to COVID-19 pandemic have been reviewed to this study in which descriptive statistics, time series graphs, correlation heatmap, simple and multiple linear regression has been produced using Python's Jupyter and SAS software Results are interpreted in order to analyze the situation and recommend alternative response.

Key Words: COVID-19; Time Series; Linear Regression; Correlation Heatmap

1. INTRODUCTION

The COVID-19 pandemic marked the year 2020 as a significant point in history, bringing the world to a temporary standstill and impacting the economy and overall civility. In December of 2019, the novel CoronaVirus SARS-CoV-2 was initially identified in Wuhan, China, which then eventually spread to numerous countries worldwide at exponential speed. The World Health Organization declared it as a global pandemic on March 11, 2020, and the world has never been the same since.

The rapid spread of the disease can be attributed to the characteristic of the virus; there was none like it before. First, the SARS-CoV-2 virus primarily spreads through respiratory droplets. It can easily enter one's respiratory system by being in close proximity with infected individuals. Moreover, airborne transmission is also of high potential due to virus-laden aerosols. Infected individuals release droplets of various sizes: the bigger ones fall on the ground while the smaller droplets, which are the aerosols, can remain suspended in the air for longer periods of time. Individuals in the same vicinity can inhale these virus-laden aerosols and can easily get infected.

Furthermore, the virus has asymptomatic and pre-symptomatic transmission features. Also called the "silent-spread", infected individuals can transmit the virus even before they develop the symptoms – or even if they do not (asymptomatic). Depending on the material, the virus can also persist on various surfaces for differing amounts of time. When people touch contaminated surfaces and then their face, nose, or mouth, this can act as a source of transmission. There have also been emerging variants of the virus, like the Alpha, Beta and Delta variants, making the virus more transmissible in comparison to the original strain.

All these factors can be the primary reason as to why the virus sprawled across the growth exponentially, making the number of deaths and infected cases seem like just numbers alone rather than actual living bodies. [insert economics info]. The state of the pandemic was determined through these data, by the consistent effort of the administration to record, collect, analyze these numbers. Data such as infection rates, mortality rates, vaccination rates, and numerous other epidemiological and clinical variables have all been produced to have a deeper understanding of the pandemic, hence effectively mitigating the phenomena. Public health policies and initiatives are derived from such information, making the statistics behind all these play a central role in this widespread war.

Studying the relationship between COVID-19 vaccination and death rates in the context of the Philippines, or any other country, is of significant importance for several reasons:

1. Public Health Impact:

2. Policy Development

3. Vaccine Confidence

4. Allocation of Resources

5. Variants and Booster Shot

6. Data-Driven Decision-Making

7. Global Health

8. Long-Term Health Impact

It's essential to conduct rigorous research, analyze the data, and communicate findings transparently to ensure that the information can be effectively used to combat the pandemic and save lives in the Philippines and beyond. This research should be a collaborative effort between healthcare professionals, science

1.1 Objectives

1. To extract and filter data from two CSV files to focus on the COVID-19 situation in the Philippines, allowing for analysis.
2. To visually depict trends and variations in data from use of time-series, heatmap, and multiple-linear regression scatter plot offering valuable insights into the dynamics of COVID-19.
3. To assess the impact of vaccination and government management by analyzing deaths and cases from the 'COVID_VACCINATION.csv' dataset, with a focus on understanding trends in daily cases of vaccination.
4. To evaluate the effectiveness of vaccines in reducing COVID-19 fatalities

2. METHODOLOGY

2.1 Data Gathering

The dataset titled “Covid-19 deaths and vaccinations Dataset” was obtained from Kaggle. It consisted of two (2) CSV files, namely “COVID Deaths” and “COVID Vaccinations”. The “COVID Deaths” dataset comprises 26 variables wherein four (4) are qualitative and the remaining 22 are quantitative variables; meanwhile, the “COVID Vaccination” is comprised of 21 variables, wherein only one (1) is qualitative and the remaining 20 are quantitative variables.

For this study, the researchers utilized only some of the quantitative data specified from the Philippines, which are: number of new deaths, people vaccinated, people fully vaccinated, total boosters administered, new vaccinations , total boosters, new vaccinations, new vaccinations smoothed, new deaths smoothed and lastly, stringency index.

2.2 Python Syntax

SYNTAX: Reading the 2 csv files that will be the datasets serving as the foundation for the researcher's data analysis and interpretation. Both CSV files have been filtered to include only rows related to the Philippines, ensuring that the analysis focused on this specific region.

In the 'COVID_DEATHS.csv' file, the columns that are extracted for analysis are: location, date, population, total_cases, new_cases, total_deaths, new_deaths, and reproduction_rate. These columns are needed in understanding the progression of COVID-19 and its impact on the Philippines.

Similarly, in the 'COVID_VACCINATIONS.csv' file, they have narrowed down the data to specific columns, including location, date, total_tests, new_tests,, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, new_vaccinations, and stringency_index. These selected columns had provided the researchers with valuable insights into the vaccination and testing efforts, as well as policy aspects related to the pandemic in the Philippines.

2.3 Data Analysis

2.3.1 Time Series Modeling

In the analysis of the 'COVID_DEATHS.csv' dataset, they ought to see the progression of COVID-19 fatality and the rapid spread of the virus over time within the Philippines. To achieve this, the researchers have initially preprocessed the data by filtering it for the Philippines, selecting key columns like 'date,' 'total_cases,' 'new_deaths,' and 'reproduction_rate,'. This allows them to compute daily statistics, specifically the number of deaths and new cases per day. By graphing these values in a time-series, it allowed them to visually depict how these metrics have evolved over time, providing a clear understanding of the fatality and spread trends. To further enhance the analysis, they have also derived descriptive statistics, offering a more comprehensive insight into the dataset, in the identification of key patterns and variations in COVID-19's impact on the Philippines.

In the provision of an analysis on the severity of the cases, the researchers had done an analysis by computing the ratio of the time series of deaths to cases. This ratio serves as a vital metric for gauging the severity of COVID-19 cases at specific points in time. From the result, higher ratio, especially close to 3, signifies a more severe impact, while a lower ratio indicates less severity. This approach offers a nuanced perspective on the gravity of the situation, contributing to a more nuanced and insightful interpretation of the dataset

2.3.2 Correlation Heatmap

To assess the impact of vaccination on the COVID-19 situation as well as the effect of the strictness of policy, from the 'COVID_VACCINATION.csv' dataset, the researchers have extracted the 'people_vaccinated', 'people_fully_vaccinated', 'total_booster', 'new_vaccinations', and 'stringency_index' data, referencing to the 'new_deaths' data, indicating the correlation of these parameters to the number of death, all within the same time frame. This analysis aims to discern if there's a notable decline/correlation in the number of deaths per day following the implementation of mandatory vaccination and its strict implementation. It is also to gauge the effectiveness of vaccines in reducing fatalities.

Summarizing the analysis, Visualization by heatmap correlation will be used to see the patterns in the data. These visual representations will help them gain a clearer understanding of how these key factors correlate to each case and their impact. This comprehensive approach will provide valuable insights for public health assessments and policymaking.

2.3.3 Linear Regression Analysis

The researchers conducted both simple and multiple linear regression analysis. For the simple linear regression fitting, variables 'new cases' and 'stringency index' were compared.

3. RESULTS AND DISCUSSION

This chapter showcases the processed data and interpretation results regarding COVID-19 Situation in the Philippines. Analysis was done through Python's Jupyter, thus allowing the proponents to create descriptive statistics of the data set, Time Series Graph of Daily COVID-19 Mortality, Daily Case Count and Death to Case Ratio over Time, Correlation Heatmap and Simple Linear Regression. Furthermore, the study expands to the use of SAS software for simple and multiple linear regression relationships.

	new_deaths	new_cases	reproduction_rate
count	1194.000000	1194.000000	1024.000000
mean	55.628141	3420.165829	1.045332
std	70.637861	5457.111210	0.415240
min	0.000000	0.000000	0.350000
25%	5.000000	233.000000	0.870000
50%	26.000000	1460.500000	0.980000
75%	83.000000	4001.000000	1.140000
max	484.000000	38867.000000	4.220000

Figure 1: Descriptive Statistics of the Data Set

The table above shows that New deaths exhibit a considerable range, from a minimum of zero to a maximum of 484, with an average of 55.63 deaths and a notable standard deviation of 70.64. Moreover, New cases also demonstrate variability, ranging from zero to a maximum of 38867, and exhibit a mean of 3420.17 cases and a standard deviation of 5457.11. Lastly, reproduction rate, based on 1024 observations, hovers around 1.045 on average, with a standard deviation of 0.415.

3.1 Time Series Modelling

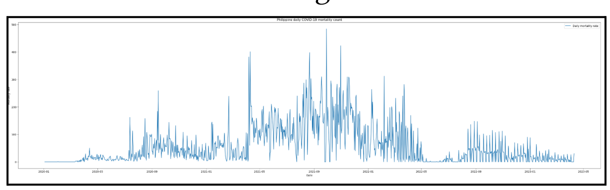

Figure 2: Time Series Graph - Daily COVID-19 Mortality

The time-series graph depicting daily death counts showed a dynamic pattern characterized by a bell curve. The peak occurred on September 29, 2021, registering 484 death cases. Throughout the observed time frame, three prominent peaks emerge, offering distinct insights. The initial peak in September 2020 sees daily deaths stabilize around 270. A rapid increase in April 2021 surpasses 400 daily deaths, marking the second notable peak. The third and highest peak occurs on September 29, 2021. Post this apex, the trend experiences a gradual decline, reaching near-zero levels by May 2022. However, a resurgence in deaths is noted in September 2022, followed by a subsequent decline.These fluctuations emphasize the temporal dynamics of mortality rates, warranting further exploration into potential contributing factors and implications for public health strategies.

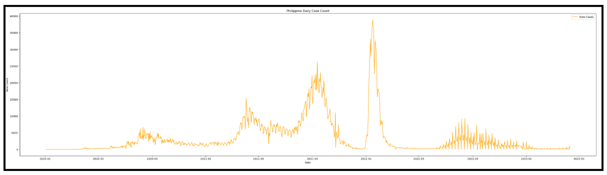

Figure 3: Time Series Graph - Daily Case Count

From the graph above, the initial crest,around August 2020, witnessed a peak nearing 7000 cases. As the year turned, there was a temporary only to be succeeded by a second surge in April 2021, where cases spiked to approximately 15000. Following a brief respite, the narrative continued with a third peak in September 2021, reaching a peak of 25000 cases. Approaching the subsequent year, the trend gradually subsided, but a noteworthy resurgence ensued, culminating in the highest peak to date—around 39000 cases in late January and early February 2022. This upward trajectory persisted until an abrupt downturn in March 2022, marking the end of the fourth peak. A subsequent resurgence in new cases was noted from July 2022 to January 2023, characterized by comparatively lower peaks stabilizing around 10000
cases.

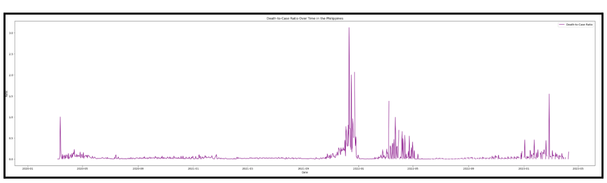

Figure 4: Death to Case Ratio Over Time

The ratio of new deaths to new cases provides insights into case severity over time. March 2020 saw a ratio of 1, indicating a slight favoring towards deaths. From May 2020 to November 2021, the ratio remained close to zero, reflecting a balanced proportion of deaths to cases. An alarming spike occurred from December 2021 to January 2022, with a peak ratio of 3, signifying significant severity. After a brief return to near-zero levels, another spike emerged from March 2022 to May 2022, peaking at 1.5. A relatively stable period followed from the end of May 2022 to early January 2023, with the ratio close to zero. However, a second-to-highest peak in the ratio was observed from January to May 2023, reaching 1.75.

3.2 Correlation Heatmap

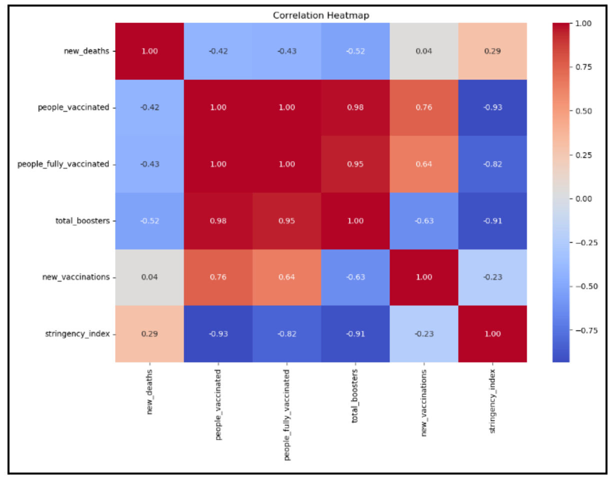

Figure 5: Correlation Heatmap of Vaccination Metrics and Stringency Index

The heatmap visually represents the correlation coefficients among various COVID-19-related variables, offering a detailed breakdown of the key findings. In the analysis of New Death, it is observed that there is a positive correlation with new vaccinations (0.04) and stringency index (0.29), while negative correlations are noted with people vaccinated (-0.42), people fully vaccinated (-0.43), and total boosters (-0.52). Similarly, examining People Vaccinated reveals positive correlations with people fully vaccinated (1) and new vaccinations (0.76), along with negative correlations with new death (-0.42) and stringency index (-0.93). Notably, there is a strong positive correlation with total boosters (0.98). People Fully Vaccinated also displays positive correlations with people vaccinated (1) and new vaccinations (0.64), coupled with negative correlations with new death (-0.43) and stringency index (-0.82). A moderate positive correlation with total boosters (0.95) is observed. Total Boosters exhibit positive correlations with people vaccinated (0.98), people fully vaccinated (0.95), and new vaccinations (-0.63), along with negative correlations with new death (-0.52) and stringency index (-0.91). Lastly, New Vaccinations indicate positive correlations with people vaccinated (0.76) and people fully vaccinated (0.64), while displaying negative correlations with new death (0.04), total boosters (-0.63), and stringency index (-0.23).

The subsequent discussion interprets the findings, emphasizing the negative correlation between new deaths and vaccination-related variables, implying that higher vaccination rates correspond to lower new death numbers. The negative correlations between the stringency index and vaccination-related metrics suggest that higher vaccination rates are associated with less stringent measures. Despite a relatively weak positive correlation (0.04) between new vaccinations and new deaths, the heatmap effectively illustrates the intricate interplay between vaccination efforts, COVID-19-related metrics, and the stringency of implemented measures.

3.3 Simple Linear Regression

The linear regression fitting was illustrated as a scatter plot to take into account the missing data in certain parts of the graph. Several observations can be derived from the figure; first, a good number of the actual data is saturated at lower y values, saying that there are fewer new cases within the stringency indices of 60 - 80. Stringency index is the measurement of how strict a country is in terms of policy implementation to mitigate COVID-19, and an index of around 60-80 indicates a relatively strict policy operation.

It can also be observed that in certain stringency indices within the range of 60 - 80, there's an aggressive increase in the number of new COVID-19 cases. This can be attributed to the fact that policies and mitigation measures were enforced only once COVID-19 cases started to surge. This causes the delay between the time of the policy implementation and the reduction in the number of new cases as these courses of action take days to take effect—even longer for developing countries like the Philippines. The data points falling at the stringency index of 100 can be disregarded; the Philippines attained the index during the first lockdown and the information regarding the pandemic was freshly disseminated to the public and only a few COVID cases were discovered yet.

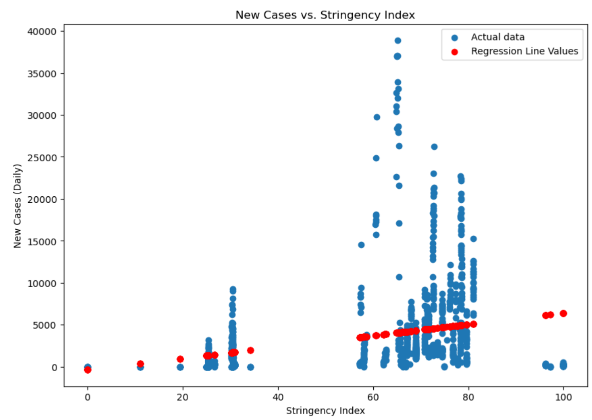

Figure 6: Simple Linear Regression of new daily cases vs. stringency index

A statistical software called SAS was also utilized in the study to further analyze the linear relationship of the variables. Based on the graph of the residual and predicted value, the data points of the model are not fully normally distributed and follow some kind of pattern. Meanwhile, the majority of the residuals fall within the +-2 range, as the data outside of this range illustrate the amount of outliers of the model. It can also be seen that not all of the data are considered useful based on the Rstudent vs. leverage plot, and the actual data did not follow the predicted value based on the 'positive' vs. 'predicted value' graph.

When further assessed, it can be derived that the simple linear regression model is not an accurate model to use when predicting the number of new cases based on the stringency index; in real life, actual pandemic data fluctuate a lot — making it all the more unpredictable by simple linear regression models.

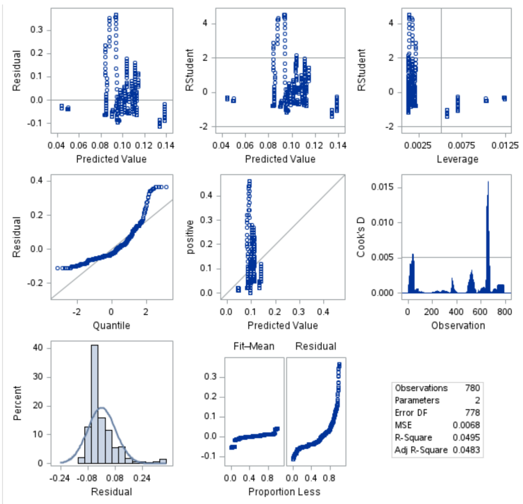

Figure 7: SAS Analysis for SLRM

3.4 Multiple Linear Regression

Figure 8 shows the results of the multiple regression analysis. The researchers assessed how the number of people vaccinated and total boosters administered could affect the number of new deaths per day. Numerous deaths were observed in the left-most part of the x-axis – when few people were still fully vaccinated. Although the actual data points are scattered, the trendline indicates that continuous administration of COVID-19 vaccination to the public had aided in halting the continuous increase in the number of daily deaths tallied; moreover, it continuously declined as the number of total vaccinated people increased. This can also be attributed to additional COVID-19 mitigation accelerators such as the administration of vaccine boosters shortly prior to the continuous deterioration of the dependent variable. It can also be observed that the actual data points started to cluster at the lowest part of the plane around the section where the total number of vaccinated people is at the maximum.

On the other hand, the relationship between new deaths and total boosters alludes to a more precise finding. The consistent administration of the COVID-19 vaccine booster led to a continuous decline in the number of new deaths over time.

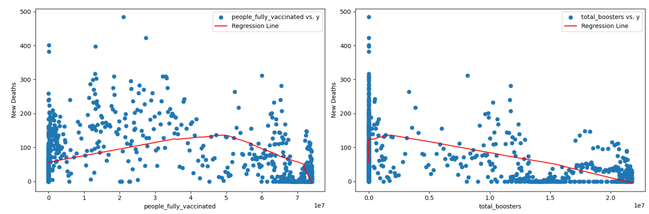

Figure 8: Multiple Linear Regression Analysis for new deaths vs. people fully vaccinated vs. total boosters

Figure 9 shows the multiple linear regression results for the new deaths smoothed vs. new vaccinations smoothed vs. total boosters. The smoothed datasets represent the 7-day smoothed average of the variable. It can be observed that the actual data points are more clustered in the early parts of the graph and behave in an uptrend manner to a specific section of the graph. This translates to the administration's behavior that the number of vaccinations administered per week increases when the number of weekly deaths is deemed to increase as well.

Meanwhile, the figure on the right illustrates the relationship between new deaths smoothed vs. total boosters administered. The graph displays fluctuations in specific sections, yet the general trend shows a decrease in the number of new deaths due to the continuous administration of vaccine boosters. Additionally, the total number of weekly death cases is observed to have nearly plateaued at zero when the maximum total boosters were administered.

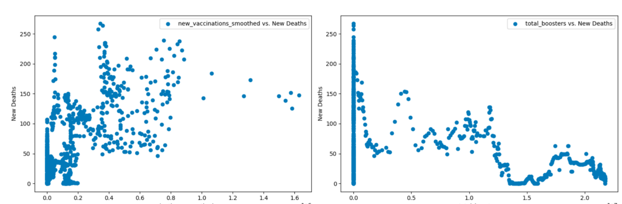

Figure 9: SAS Analysis for MLRM

Figure 10 shows the SAS results for the multiple linear regression analysis for the variables in Figure 9. It can be observed that the residuals of the model are not randomly scattered or not fully normally distributed; a specific section is also deemed to be following a specific pattern. The majority of the data points fall within the desired scope and only a few are considered outliers; however, the predicted values for the observed data were significantly not followed. By further assessing the SAS output, it can be derived that the model is not the most ideal and accurate model to use in predicting the behaviors of the mentioned variables.

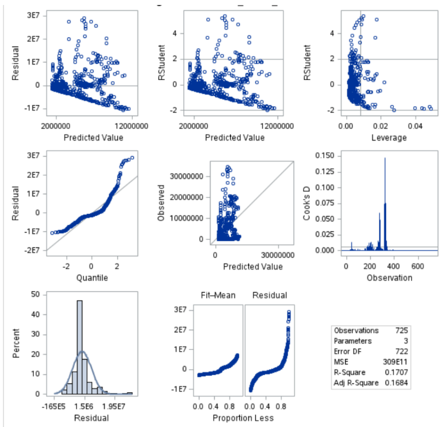

Figure 10: SAS Analysis for MLRM

4. CONCLUSIONS

Through this study, the effectiveness of vaccination was evident in the reduction of cases based on the time series graphs. The time period of increased spread of the virus was presented in which specific seasons are attributed to high numbers of cases following the events. Furthermore, data collection and recording manifests effects on the trend of graphs, in which inconsistency with the remittance of data was observed. Effectiveness of strictness of government policies in pandemic management was also observed with the processed data in which particular seasons show high levels of strictness with high number of cases. The phenomenon was attributed to the lack of anticipation of the government with the surges of cases in which policies are enforced during high levels of cases.

For the recommendations, it is best to review and enhance the healthcare capacity of the nation, in which the review must be in-depth. It should be more than the infrastructure that extends to the skills and population of healthcare workers in order to suffice the possible volume in times of health emergencies. The country must also foster international cooperation that would lead to partnership in the development of studies and expansion of the latest technologies in relation to health. Apart from this, the Philippines should also integrate local production of medical equipment and supplies which will allow self-sustaining capacity and economic growth. Lastly, efforts for research and development in the country should be boosted with a greater emphasis on the benefits and possibilities of advancements.

5. ACKNOWLEDGMENTS

The successful implementation of this programming and analysis and the accomplishment of this paper would not have been possible without the help and guidance of various people and the institution.

We recognize and sincerely thank our professor, Mr. Jude Maria Antenorcruz, for his supervision throughout the course and for teaching us the concepts behind all of this.

Above all, we give credit to our institution, DLSU-Manila, for giving us this opportunity to learn, explore, and enjoy Physics not only through lectures and assessments, but also through programming analysis like this.

6. REFERENCES

Baclig, C.E. (2021, March). TIMELINE: One year of Covid-19 in the Philippines https://newsinfo.inquirer.net/1406004/timeline-one-year-of-covid-19-in-the-philippines#ixzz8LCsroVy6

Covid-19 deaths and vaccinations Dataset. (2023, July 19). Kaggle. https://www.kaggle.com/datasets/tohidkhanbagani/covid-19-deaths-and-vaccinations-dataset/