Impact of COVID-19 Vaccination on New Infection Cases and Mortalities in Different Countries and Regions

Prepared by:
Hermann Junhao Fan, Xiangzhi Chen, Qingyang Wang, Yiqin Zhou

Submission Date: December 12, 2024

Department: Data Science and Analytics

Institution: Georgetown University

This report is prepared as part of academic requirements for the study of data-driven insights into COVID-19 vaccination impacts.

Abstract

This project aims to investigate the impact of vaccines on the spread of COVID-19 during the pandemic. The primary research question addresses whether vaccination has played a significant role in controlling new COVID-19 cases and deaths in countries across different continents. To explore this, exploratory data analysis (EDA) was conducted, followed by hypothesis testing using bootstrapping, permutation tests, ANOVA, and t-tests. The analysis reveals a complex relationship between vaccination and new confirmed cases, where the number of new cases increases as the number of daily administered vaccinations rises in regions such as the U.S., Europe, and Asia. This challenges the common belief that vaccination directly reduces case numbers. In Africa, vaccination seems to have reduced the number of new cases, but this trend may be caused by the problem with data completeness. In addition, when examining daily new deaths, vaccination appears to have a diminishing effect on deaths across all continents except Asia, which suggests that other important factors, such as government policies, may also influence the effect of vaccination. Overall, this study highlights the nuanced role of vaccination in the pandemic and suggests that its effectiveness may vary depending on complex interactions such as human behavior, new variants, the readiness of healthcare systems, and demographics.

1 Introduction

Throughout human history, many dangerous pandemics have shaped societies, economies, and cultures, often with devastating effects. The Black Death for instance, which swept across Europe in the 14th century, killed an estimated 50 million people, resulting in drastic social and economic upheaval. Similarly, the Spanish flu of 1918 took millions of lives worldwide, and more recently, the H1N1 influenza pandemic profoundly impacted global health, especially among young people. Each of these pandemics highlighted the vulnerabilities of human societies to infectious diseases and the importance of effective public health measures. Moving forward to the 21st century, there came the COVID-19. From the end of 2019 to 2024, the COVID-19 pandemic had a profound impact on our world, disrupting people's daily lives, overwhelming healthcare systems in nearly all countries, and more importantly, causing millions of deaths. Based on cumulative data reported to WHO, until November 24, 2024, COVID-19 was responsible for the deaths of more than seven million. In response to this deadly virus, governments from many nations, health organizations, and scientists scrambled to develop effective interventions to control the spread of the virus, with vaccination emerging as one of the most crucial solutions. As vaccines were developed and deployed around the world, a major question arose: did vaccination truly make a significant difference in mitigating the spread of COVID-19, or was its impact overstated?

As years went by, the effectiveness of vaccines had become a topic of intense debate. While vaccination initially appeared to reduce severe illness and deaths, people in many regions were still suffering from the pandemic, which raised doubt in the minds of ordinary people. Moreover, this discrepancy has prompted public health experts to reconsider the assumptions about vaccine efficacy and how different regions experienced the effects of vaccination during the war against COVID-19. Looking through a biological view, before the end of 2022, around the initial stage of COVID-19, administered vaccines seemed to be somewhat protective against SARS-CoV-2, which is the virus that caused COVID-19. Based on Figure 1. created by Zheng et al. (2022), all of the examined vaccines had accuracy above 87 percent, which was relatively high enough to prove their effectiveness. But, for symptomatic COVID-19 disease, the efficacy of vaccines decreased by 24.9 percent among people of all ages and 32.0 percent for elders (Feikin et al., 2022). Additionally, with the emergence of more dangerous variants of SARS-CoV-2 like omicron, vaccine effectiveness on average was 70 percent against infections, and the effect will continue to be dampened with time (Wu et al. 2023). Based on all these past studies, one can tell that the true effectiveness of COVID-19 vaccination is vague. Hence, exploring the true relationship between vaccination and COVID-19 outcomes is crucial for shaping future public health strategies, especially in the face of potential future pandemics.

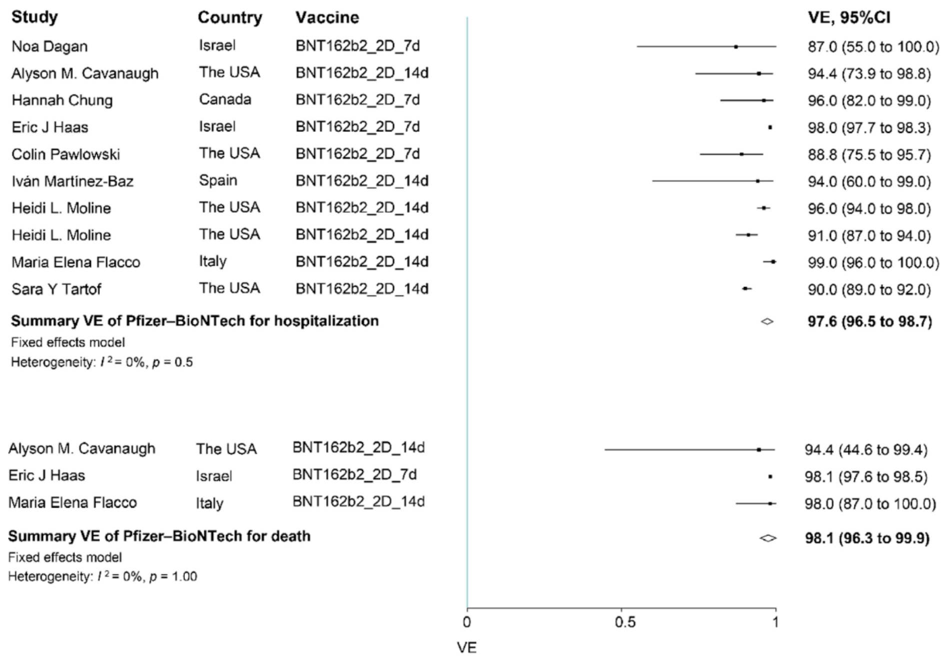

Figure 1: (Zheng et al.)

Therefore, this study aims to explore the effectiveness of vaccination by analyzing the relationship between the number of daily administered vaccinations and COVID-19 outcomes like cases and deaths in countries across different continents. By focusing on the number of new confirmed cases and daily deaths, we seek to uncover patterns that can inform future strategies for pandemic response. The findings will also contribute to ongoing debates about vaccine efficacy and public health policy during the pandemic, highlighting regions where vaccines were successful and where additional interventions may be necessary to reduce the burden of infectious diseases like COVID-19 in the future.

2 Analysis

2.1 Additional Background and Dataset Description

In this research, two major datasets were utilized for data analysis: Our World in Data (OWID) COVID-19 Dataset and the European Centre of Disease Prevention and Control (EU CDC) Dataset. Both datasets were selected for their reliability, comprehensive features, and depth of information they provide, which together offer a global and regional (by continent) perspective on the pandemic's trajectory and impact. For thousands of years, tenacious humans have been fighting against diseases. From smallpox virus, malaria, Black Death, tuberculosis and other diseases, people have used their superior wisdom to defeat these frightening viruses. The weapon to defeat them is the development of vaccines and the advancement of medical science. However, in the wake of the coronavirus outbreak, concerns about vaccines have also emerged. According to the report "60 percent of Americans say they probably won't get an updated COVID-19 vaccine' from Pew Research Center in 2024, 60 percent of U.S Adults claim that they "probably not get a vaccine", and only 15 percent of them claims with confident that they have received the vaccine. The reason for this situation comes from the general public's distrust of the new m-RNA vaccine and doubts about its effect. Therefore, in order to verify the reliability of the vaccine, we designed and analyzed the relevant data mentioned above. The OWID dataset is one of the most detailed and regularly updated datasets people could find on Google. This dataset contains features including and not limited to demographic data (population density, age more than 60 years old, age more than 75 years old, etc.), Regional economic data (GDP per Capita), Infrastructure data (hospital beds per thousand), and most importantly, the government stringency data. The word "government stringency" refers to the Stringency Index, a composite metric aiming to represent the strictness of government regulations and initiatives in response to the COVID-19 pandemic. This index provides a consistent method for comparing the severity of government-imposed restrictions across countries and throughout time. Such dictation like "government stringency" sounds intimidating in some cases; however, there are lots of incentives that the population can study from it.

In November 2002, a Severe Acute Respiratory Syndrome (SARS) was found in Shunde City, Guangzhou, China. 8000 people were being infected and 774 patients died in that pandemic. Three months after the outbreak was discovered, the infection rate in Beijing soared. Beijing was also listed as an epidemic area by the Chinese government and WHO. With the assistance of Chinese epidemic prevention personnel and international volunteers, people found that the habit of isolation and wearing masks could greatly reduce the number of overall infections. On April 23, 2003, the Chinese government mobilized military personnels to build the People's Liberation Army Xiaotangshan Hospital to provide centralized treatment for infected patients. Less than a month after this measure, the epidemic was greatly alleviated, and on May 9, the Chinese Premier Wen Jiabao announced that the number of infections had been controlled and greatly reduced. By April 2004, the death toll was almost zero and the epidemic was basically over. From the response to the SARS crisis, we can see that the government played a tremendous role in controlling the epidemic. The SARS epidemic also provided a "template" for China's later response to the COVID-19 epidemic. As a result, China was able to promptly control the spread of the domestic epidemic in early 2020. In this study, government stringency was therefore also included in the research focus. In public health crises, government functions often play a very important role, including in vaccines. The government not only needs to develop or organize the development of vaccines, but more importantly, it can adjust and control the vaccination rate of the public, for example, give official suggestions to the public and convince them to take the vaccination. Therefore, the government is also an indispensable part in the research on vaccines. For the Government Stringency, and you will get a more intuitive understanding in the EDA part later.

In addition to the OWID dataset, we also used data from the EU CDC. This database can provide more regional references in later data visualizations, such as the epidemic situation and vaccination status of EU countries. Also, it provides specific information on cases and fatalities by age and gender, which adds a valuable dimension to demographic study. The collection also contains metadata on testing strategies, healthcare capacity, and vaccination efforts across the EU, allowing for a more detailed examination of regional disparities. By analyzing the two datasets, we acquire a comprehensive perspective of the pandemic that includes both macro trends and micro details necessary for effective analysis.

2.2 Data Processing and Cleaning

In the OWID dataset, we removed missing values (NA). We concentrated on critical variables such as immunization rates and test results. These factors are critical for interpreting the data. Rather than filling in missing values, we chose to exclude rows with NAs. This approach ensured the dataset's accuracy and consistency. We began by finding columns containing key metrics. These comprised daily case counts, deaths, testing rates, and immunization statistics. We used Python's Pandas package to filter out rows with missing values. Before we did this, we determined how much data was missing and searched for patterns in the missing data, such as if specific nations or historical periods had more gaps. Removing NAs will reduce the total number of rows in data. However, since sufficient data was provided in this study, removing NA will be the most efficient way to make the dataset more reliable for analysis. By working only with complete data, we avoided making assumptions about the missing values. This was important for comparing countries and regions accurately. The European Union CDC dataset is another reliable dataset. The EU CDC database provides us with a more detailed analysis of each country in the EU, such as the most accurate mortality rate, total number of deaths, number of infections and infection rate in different countries in the EU. On the official EU CDC website, Covid-19 datasets have lots of sub-branches with different focuses. For example, "Data on COVID-19 Vaccination on EU/EEA", "Data on the daily number of new reported COVID-19 cases and deaths by EU/EEA country", and so on. For a better and more convenient data analysis, all missing values were removed, and different datasets have been merged into one dataset, which includes all features that we need for Data Analysis.

3 Exploratory Data Analysis (EDA)

From a broad perspective, the global attitude and responses towards the pandemic was quite united. Early on in the pandemic, the urgency to develop and distribute vaccines was universally recognized, and many governments prioritized manufacturing vaccine doses. People around the world were also open to COVID-19 vaccination. Looking at Table 1., in Europe and North America, where public health infrastructure and economic resources were relatively robust, more than four thirds of the total population had taken the vaccination at least once. And the number of vaccinated people in Asia was also of great amount, though there were variations. Countries like Japan had large-scale vaccination drives, often supported by government mandates. However, the dynamics were not uniform across the entire continent. While East Asia quickly scaled up vaccination efforts, countries such as India faced organizing and infrastructural challenges due to their large populations and resource constraints. But India still managed to achieve a result of 72.68 percent. In the case of Asia, China resulted in a relatively poor performance in vaccination, which might have to do with incomplete data recording caused by national security and poverty-stricken areas.

Continents	People Vaccinated (At least once)	Total Population (by 2021)	Percent (%)
Africa	550,433,516	1,393,676,444	39.50
Asia	3,065,284,498	4,694,576,167	65.29
Europe	572,613,838	745,173,774	76.84
North America	458,857,285	595,783,465	77.02

Table 1: Vaccination Coverage by Continent

Country	People Vaccinated	Total Population (by 2021)	Percent (%)	Population Percent (%)
China	705,833,217	1,426,437,267	49.48	30.38
India	1,027,912,606	1,414,203,896	72.68	30.12
Japan	104,710,374	125,679,338	83.32	2.68

Table 2: Vaccination Coverage in Selected Countries

People's willingness to inject COVID-19 vaccination can also be shown in Figure 2., in which the dashed lines represent the approximated release date of vaccination for each of the four continents. Therefore from the plot, it is clear that in Asia, Europe, and North America, people were very active about getting vaccinated since the number of new administered doses reached the peak at the beginning stage of the pandemic. This was particularly true in the early stages of vaccine rollout when public health campaigns were heavily promoted, and many people were eager to protect themselves and their family members from the rapidly spreading coronavirus. Again, for Africa, the climax happened later in the timeline since Africa was located at the end of the waiting line due to a short supply of vaccinations globally. Overall, even though there were doubts about the efficacy of vaccines, the number of vaccinated people around the world is still prominent.

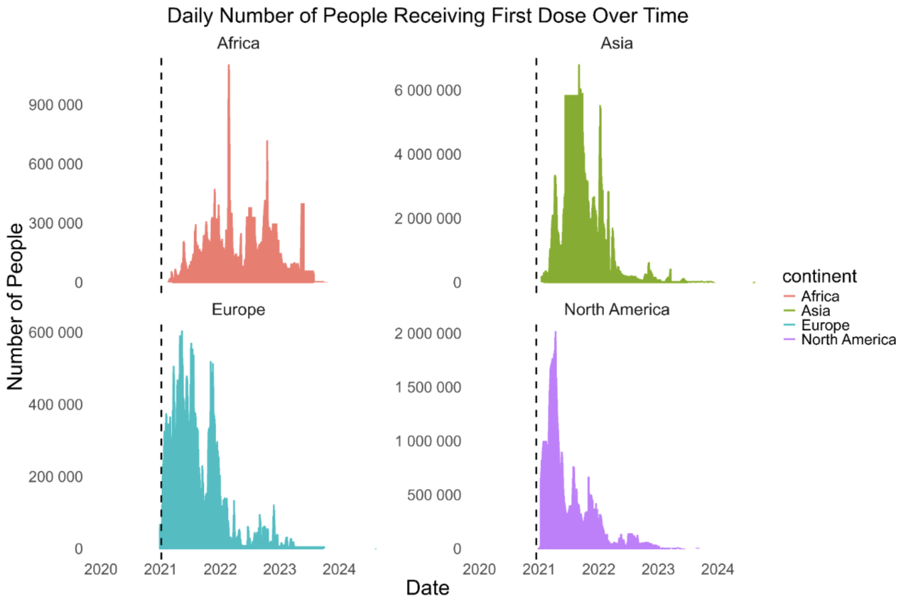

Figure 2.

Figure 2: Daily Number of People Receiving doses over time

Continuing, the next chart, Figure 3., shows deeper details about the daily confirmed deaths of COVID-19. As previously mentioned, the dash line represents the vaccination date of that continent. Throughout the world, almost all countries conducted the vaccination process in Late December 2020 or Early January 2021. From the chart we can see that within a short period of time after vaccination, the number of deaths had dropped significantly. However, as time went by, the effect of the vaccine gradually weakened. From the perspective of the history of world health, governments of various countries require or advocate people to get the latest influenza virus vaccine in time every year. The main reason is the drug resistance of the virus. If people start to use a large amount of antibiotics or high-intensity virus vaccines in the early stage of the disease, the effect will be significant in a short period. However, after the virus adapts to this treatment method, the effect of the vaccine will not be so strong. This is also one of the reasons why the number of deaths in various continents in 2022 and 2023 has soared in the chart.

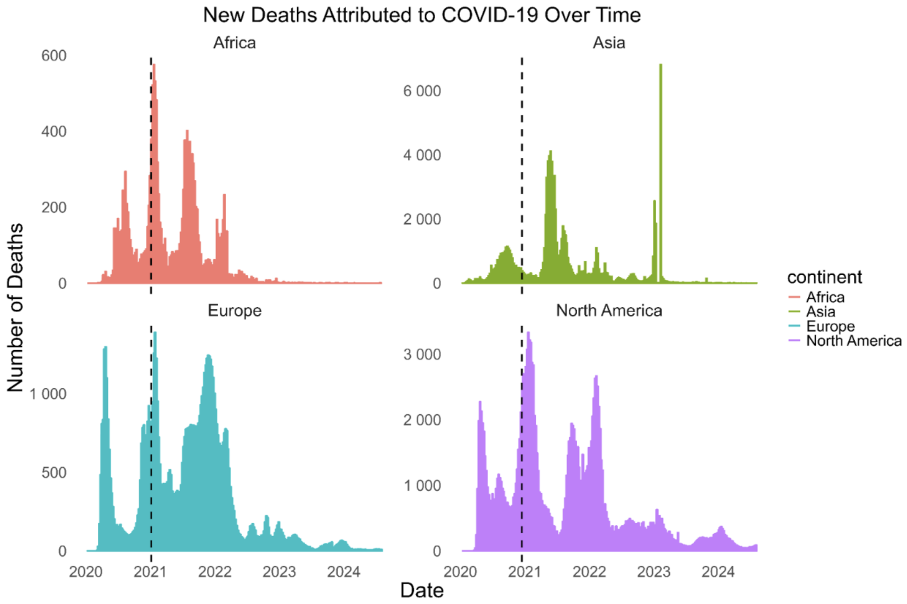

Figure 3.

Figure 3: New Deaths Attributed to COVID-19 Over Time

Without doubt, there are other potential reasons for this situation. According to statistics from the World Health Organization, there are more than 1,000 new coronavirus variants in the world, among which the most noteworthy variants are Alpha, Beta, Gamma, Delta, and Omicron. Peaks across continents correspond to the emergence of variants like Delta (mid-2021) and Omicron (late 2021 to early 2022). The rapid spread of these variants led to record-high daily deaths, particularly in Europe, Africa, and North America.

In general, with numerous people being vaccinated, logically, if the vaccinations were so effective, coronavirus must have been eliminated. But the truth is, it took four years for human society to overcome the pandemic. Digging deeper into the number of daily cases and deaths during the pandemic, we can actually gain some perspective about the effectiveness of vaccines. Figure 4. is a time series plot that marks the number of new confirmed cases of COVID-19. For all the four continents, the peaks of daily positive cases are all located after the release date of vaccination, which at some point marked the relative powerlessness of vaccinations while facing various variants.

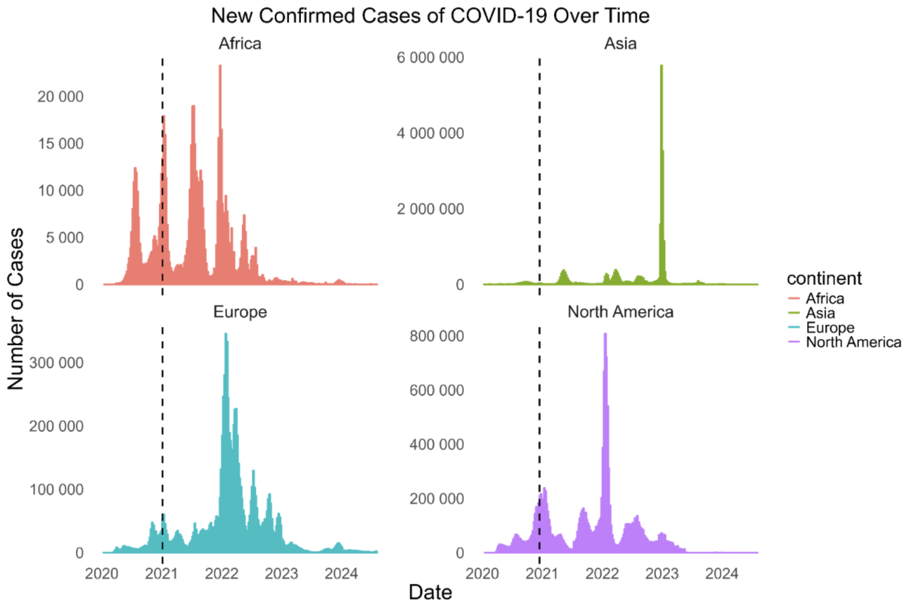

Figure 4.

Figure 4: New Confirm Cases of COVID-19 Over Time

Based on all the visualizations, vaccinations didn't seem to be so effective. Since it is impossible to draw conclusions by simply looking at the graphs, in the following sections, hypothesis testing will be employed to closely examine the potential relationship between vaccination and the pandemic.

4 Statistical Methods and Results

4.1 Overview

In the previous EDA section, we mentioned that as time goes by, the effectiveness of the vaccine will decrease, and the infection rate of people will become higher after a certain period of time. In order to investigate the effect of vaccination, for each continent(For North America, the United States is extracted based on its more well-rounded dataset. Additionally, for Africa, since many countries have numerous missing values, we combined data from 10 countries, including Algeria, Botswana, Egypt, Gabon, Liberia, Mauritius, Morocco, Nigeria, South Africa, and Seychelles, to form a dataset that has quite complete data), the dataset is divided into parts: data before the release date of vaccines and the ones after that, which allows for further comparison of continental datasets. Then, to examine whether the effect held during the pandemic, multiple hypothesis testing will be employed based on the number of confirmed cases and the number of deaths.

To investigate the impact of vaccination rollout on new COVID-19 deaths, the following hypotheses are formulated: the null hypothesis H0: Vaccination made no difference to the death rate of COVID-19, and the alternative hypothesis H1: Vaccination had a significant effect on the death rate. According to these hypotheses, we performed bootstrap analysis and permutation testing on the smoothed daily new deaths data from the OWID dataset. The specific method of the bootstrap method involves repeatedly sampling with replacement from the two datasets before and after vaccination to generate 10,000 simulated samples. In each bootstrap iteration, we calculated the means of the resampled "before" and "after" datasets and found their difference. Through this process, we obtained the distribution of the mean difference, and thus estimated the population mean difference and its uncertainty range. Mathematically, the mean difference is calculated in this form:

$$\text{Mean difference} = \text{Mean value of new infections before vaccination} - \text{Mean value of new infections after vaccination}$$

Based on this empirical distribution, we calculated the mean difference and its 95 percent confidence interval, where the confidence interval was obtained by taking the 2.5th and 97.5th percentiles of the distribution.

In order to solidify our thoughts, the permutation test is also an effective way to conduct the analysis. To deploy this, the permutation process repeatedly randomly rearranges the "before" and "after" labels of the data while keeping the actual number of deaths constant. After each permutation, a new difference between the group means is calculated. This process is repeated 10,000 times, generating a distribution of mean differences that represent the differences we might observe if the null hypothesis is true. On the basis of these modelings, the following Table 3. is generated.

Nation/Continent	**Bootstrapping (2.5%, 97.5%)**	**Permutation Test (P-value)**
U.S.	104.9207, 269.6053	p-value $= 3 \times 10^{-4}$
Europe	4.382306, 7.673467	p-value $= 3 \times 10^{-4}$
Asia	-75.07168, 59.64493	p-value $= 0.6449$
Africa	9.185213, 12.737823	$p - value < 0.05$

Table 3: Vaccine v.s. New Deaths

Apart from daily deaths during the pandemic, the relationship between COVID-19 vaccination and new cases is also explored. Following the same logic, a table for bootstrapping and permutation testing results can be formulated. In addition to the two statistical methods, t-test is also implemented as further validation of the potential relationship. The combined results are stored in Table 4.

Nation/Continent	**Bootstrapping (2.5% 97.5%)**	**Permutation Test (P-value)**	**T-Test (P-value)**
U.S.	-65289.63, -46414.92	$p - value < 0.05$	2.2×10^{-16}
Europe	-2445.525, -2193.695	$p - value < 0.05$	0.037
Asia	-44942.71, -25777.70	$p - value < 0.05$	2.2×10^{-16}
Africa	134.7785, 250.9538	$p - value < 0.05$	2.2×10^{-16}

Table 4: Vaccine v.s. New Cases

In the following sections, closer analysis on countries in each continent will be conducted.

4.2 Statistical analysis of the United States:

As a representative sample of North America, the result for the United States in Table 3. matched our former expectations. The number of daily administered vaccines is correlated with the daily number of deaths. With both sides of the confidence interval surpassing zero which indicates a positive mean difference between data before the publishing date of vaccines and the ones after that date, the bootstrapping statistics demonstrate the positive effect of vaccination on reducing the death rates. Moreover, Table 4. shows that the number of new positive cases is also correlated with vaccination since zero is located outside of the bootstrapping CI and the two p-values are also smaller than 0.05. However, the number of new cases actually rises with the increasing number of administered vaccines according to the CI by the bootstrap method. State responses to lockdowns can cause the rising number of positive cases after the publishing date of vaccines, since not all states had strict restrictions and the duration of the lockdown was also shorter in the United States. In addition, since vaccination couldn't guarantee one hundred percent immunity, the number of cases rose when lockdown was over. This explains the climax of the daily cases in North America in 2022 in Figure 4. in the EDA section.

4.3 Statistical analysis of Asia:

In the EDA section, we observed that the number of new confirmed cases and new deaths in Asia did not decline as expected.. In contrast, Asia experienced a peak in new cases and deaths in 2023.

For further exploration, we used bootstrap, permutation tests, one-way ANOVA, and T-test as hypothesis testing methods. In terms of data selection, in addition to the overall Asia data, we also selected the data of China, Japan, and India as representative samples of Asia. As the two most populous countries in Asia, China and India have an important influence on Asia. At the same time, Japan and China, as the two largest Asian economies, also play a crucial role in Asia. Moreover, by analyzing the composition of the surge in new cases and new deaths in 2023, we find that Japan and China account for a significant portion of this increase. This finding provides additional justification for choosing these two countries as the focus of the study.

Therefore, by selecting these three countries as a subset, we can have a good knowledge of the whole situation in Asia.

Bootstrap Analysis of the Impact of Vaccination on Covid-19 New deaths:

From Table 3, we can see that the 95 percent CI for the mean difference in the number of new deaths before and after vaccination ranges from -75.07168 to 59.64493. 0 in this CI with a p-value greater than 0.05 indicate that there is no statistically significant association between vaccination and new deaths at the overall level in Asia.

Country	Bootstrapping (2.5% 97.5%)	Permutation Test (P-value)
China	-107.10460, -47.64657	p-value = 0.0115
Japan	-45.02909, -29.84858	p-value = 0.2139
India	80.39175, 193.19802	p-value = 0.0013

Table 5: Vaccine v.s. New Deaths (China, Japan, and India)

From Table 5, we can get bootstrap results for China, Japan, and India. For China, the 95 percent CI for the mean difference in the number of new deaths before and after vaccination ranged from -107.10460 to -47.64657, with p-value (0.00115) much smaller than 0.05. This shows a significant negative association between vaccination and the number of new deaths in China, suggesting that vaccination is associated with a reduction in the number of new deaths.

For Japan, the 95 percent CI for the mean difference in the number of new deaths before and after vaccination ranged from -45.02909 to -29.84858, with p-value (0.2139) greater than 0.05. This means that while vaccination may be inversely associated with new deaths in Japan, the relationship is not statistically significant.

For India, the 95 percent CI for the mean difference in the number of new deaths before and after vaccination ranged from 80.39175 to 193.19802, with p-value (0.0013) of less than 0.05. Interestingly, this result, in contrast to China, shows that in India, vaccination presents a positive association with the number of new deaths, meaning that vaccination was associated with an increase in the number of new deaths.

Bootstrap Analysis of the Impact of Vaccination on Covid-19 New cases:

From Table 4, the 95 percent CI for the difference in the mean number of new cases before and after vaccination ranges from -44942.71 to -25777.70. Plus, the P-values of both statistical tests (permutation and T-test) were less than 0.05, suggesting that the reduction in new cases was statistically significant and unlikely to have occurred by chance. Therefore, these statistical results provide strong evidence that vaccination campaigns are having a significant effect in reducing the number of new cases in Asia.

Country	Bootstrapping (2.5% 97.5%)	Permutation Test (P-value)	T-Test (P-value)
China	-103145.22, -49271.73	$p - value < 0.05$	2.2×10^{-16}
Japan	-29168.75, -24111.31	$p - value > 0.05$	2.2×10^{-16}
India	-1622.296, 8005.391	$p - value > 0.05$	2.2×10^{-16}

Table 6: Vaccine v.s. New Cases

For Table 6, we can get bootstrap results for China, Japan, and India. For China, the 95 percent CI of the mean difference between the new cases before and after vaccination was -103145.22 to -49271.73. The p-values of the permutation test and the t-test were less than 0.05, indicating that the decrease of new cases was statistically significant. In other words, the vaccination had a significant effect on the reduction of new cases.

For Japan, the 95 percent CI of the mean difference in the number of new cases before and after vaccination ranged from -29168.75 to -24111.31. The p-value of the permutation test was greater than 0.05, but the p-value of the t-test was less than 0.05, which meant that vaccination did not have a statistically significant effect on the reduction in the number of new cases.

For India, the 95 percent CI of the mean difference of new cases before and after vaccination ranged from -1622.296 to 8005.391. The p-value of both tests were greater than 0.05, showing that vaccination did not have a crucial effect on decreasing new cases.

One-way ANOVA Results for Covid-19 Death Rates by Stringency Phase:

Region	Df	F-value	P-value	Sig.
Asia	2	128.8	2.2e-16	***
China	2	9.197	0.000107	***
Japan	1	23.2	1.59e-06	***

Table 7: ANOVA Results for Covid-19 Death Rates by Stringency Phase

For Asia, when comparing the three different stages of stringency, the F-value was 128.8 and the p-value was much less than 0.05, indicating a highly significant association between the stringency of the immunization policy and the mortality rate throughout Asia.

For China, the 3 stages were also compared, even the F-value was 9.1, which was relatively small, it still statistically significant (p-value=0.000107). This may indicate that factors other than policy stringency are influencing changes in mortality rates.

For Japan, only two stages were compared, but showed a strong correlation (F-value=23.2) with high statistical significance (p=1.59e-06). This suggests that the effect of policy stringency remains significant even when comparing a much smaller number of stages.

4.4 Statistical analysis of Africa:

We observed from the EDA part that Africa has the lowest vaccination rate, thus we suspect that for Africa, the impact of the vaccines may have a slighter impact on the pandemic than how much vaccination impacts COVID-19 in other regions.

We find out the date of when the first dose of vaccination was taken, which we used 2021-04-01, and split the data into two parts, labeled as 'before' and 'after' based on the date, to do the comparison.

Bootstrap Analysis of the Impact of Vaccination on Covid-19 New cases:

First, we focused on the new cases. We applied a bootstrap method to our Africa data, by simulating 10,000 times, each time we randomly sampled our splitted dataset and got the means of each dataset, then calculated the difference between means of before-data and after-data. We will obtain the distribution of the 10000 mean differences.

Diving into it, we calculate its 95 percent confidence interval, which is in the range of about 134.78 to 250.95, and the mean of the mean differences is approximately 192.23. We can observe that the entire 95 percent level CI of the mean difference is above 0 and the mean of it is much larger than 0. The number of new cases after vaccine release is significantly lower than before the release.

Bootstrap Analysis of the Impact of Vaccination on Covid-19 New deaths:

Then, we focus on new deaths. We use the similar bootstrap method as we used for new cases.

We got the result that the 95 percent confidence interval of the mean difference for new deaths before and after vaccination release is in the range of about 9.2 to 12.7, and the mean of the mean difference is about 11. The entire 95 percent level CI is above 0 and the mean of it is also larger than 0. Therefore, we can conclude that on the 95 percent significance level, we can say that the death situation had improved after the vaccination release, even though the difference is not significant.

Significance Testing with Permutation and T-tests:

To further validate our findings, we conducted permutation tests and t-tests for new cases and new deaths.

- **Permutation Tests**

For both new cases and new deaths, we obtained small p-values of 0, which are far less than 0.05, allowing us to reject the null hypothesis at the 0.05 level of significance. This means that vaccinations do make a difference to the death rate and new cases of COVID-19.

- **T-tests**

  The t-tests between new cases and new people vaccinated, as well as between new deaths and new people vaccinated, yielded p-values smaller than 2.2e-16, less than 0.05, confirming the findings from the permutation tests.

At the 0.05 significance level, we reject the null hypothesis, further supporting the conclusion that there is a significant impact of vaccinations on reducing new cases and deaths.

4.5 Statistical analysis of European Union Countries:

When analyzing European affairs, many scholars often think about which European country's data can best represent Europe. Or which countries' data should be selected to analyze Europe? In fact, the EU region itself can be analyzed as a "country" as a whole. The handling of most public emergencies within the EU is almost uniform, and EU citizens can also live and work freely within the Schengen area, so I choose to use the overall data of EU area countries for analysis. Since we focus on analyzing EU countries, the United Kingdom, which previously left the EU, and Türkiye, which has not yet joined the EU, are not included in the analysis of the data. At the same time, Russia and certain Eastern European countries were excluded from the data analysis. Since most of Europe's population is concentrated in the Schengen Area, we use the EU as the country for our analysis. (Although Switzerland is a neutral non-EU country, it is still a Schengen area country, so Swiss data are included here.)

Bootstrap Analysis of the Impact of Vaccination Rollout on COVID-19 New infections:

The analysis results show that the average number of new daily cases before vaccination was lower than after vaccination, with an average difference of -2320.35 and a 95 percent confidence interval of [-2445.53, -2193.70]. This negative value indicates that the number of new daily cases in Europe has increased significantly after vaccination began. Since this confidence interval's range doesn't include zero, it suggests a meaningful reduction in deaths after vaccinations began.

This result is consistent with what we analyzed from the chart. But what is the reason for this? In addition to vaccine resistance and body immunity, we have other opinions.

First, in the early stages of vaccination, the vaccine coverage rate was low, and only high-risk groups were given priority, which was not enough to form herd immunity. At the same time, the emergence of virus variants (such as Alpha and Delta) greatly increased the transmissibility and aggravated the increase in the number of cases. Second, the promotion of vaccines may have triggered the relaxation of public health measures (reduced government stringency), such as the lifting of blockades or the reduction of epidemic prevention efforts in some areas, which may also have accelerated the spread of the virus. In addition, the improvement in testing capabilities has enabled more mild and asymptomatic infections to be diagnosed, further pushing up the number of new cases reported. Finally, the protective effect of the vaccine takes some time to manifest, and in the early stages of vaccination, new cases may still increase due to transmission dynamics. This result does not mean that the vaccine is ineffective, but reflects the complexity of the early stages of the epidemic. Future studies should further analyze it in conjunction with the mortality rate.

Bootstrap Analysis of the Impact of Vaccination Rollout on COVID-19 Mortality:

This process gave us two main results. First, it calculated the average difference in COVID-19 deaths between the two periods, based on the bootstrap samples. Second, it provided a 95 percent confidence interval for this difference. To get this interval, we took the 2.5th and 97.5th percentiles of the bootstrap data. This shows the range where we can be 95 percent sure the true average difference falls. The analysis found that the average difference in new daily deaths before and after vaccinations started was about 6.03. The 95 percent confidence interval for this difference was between 4.38 and 7.67. Since this range doesn't include zero, it suggests a meaningful reduction in deaths after vaccinations began. This drop likely highlights how effective vaccines are at preventing severe illness and saving lives.

Verification of the results:

In order to solidify our thoughts, the permutation test is also an effective way to conduct the analysis. To do this, the permutation process repeatedly randomly rearranges the "before" and "after" labels of the data while keeping the actual number of deaths constant. After each permutation, a new difference between the group means is calculated. This process is repeated 10,000 times, generating a distribution of mean differences that represent the differences we might observe if the null hypothesis is true.

The p-value is then obtained by calculating the proportion of these permuted differences that are as extreme or more extreme in absolute value than the observed difference. A p-value of 0 (or substantially less than the common significance level of 0.05) indicates that the observed difference is extremely unlikely to occur if the null hypothesis is true, meaning that the difference between the two groups is statistically significant.

Since the P-value of both cases is almost 0, which is smaller than the significant level of 0.05, indicates that the observed difference is highly unlikely to have occurred under the null hypothesis, suggesting that there is a statistically significant difference between the groups. In addition to the permutation test, T-test also can provide additional evidence for the significance of difference. Since both T-tests of new infection cases and mortality are smaller than 5 percent (significance level), we can, again, conclude that there is a difference made by the implementation of vaccines. [The results are shown in Table 4.]

5 Conclusion

Based on our test results, we can conclude that for all continents, there is a relationship between vaccination and new confirmed cases of Covid-19.

From the fact that in the U.S., EU, and Asia, the number of new cases rose with the increasing number of administered vaccines, we conclude that vaccination did not weaken the growing cases of COVID-19 like many thought.

We also hypothesize that the observed relationship between vaccination and the increase in new COVID-19 cases might be influenced by several confounding factors. One is that after vaccination, people believed the personal risk of infection was reduced, which led to a decrease in self-awareness of pandemic preparedness, and the number of cases increased after the vaccine was released. Another factor is that new variants might also impact the effectiveness of vaccinations. Some variants such as Delta and Omicron, which were more transmissible and emerged after vaccine release, might offset the benefits of vaccination, especially if vaccine-induced immunity was less effective against these strains.

In Africa, the number of new cases dropped with the increase in vaccination numbers, which is a totally different trend from other regions. We suspect that the data for Africa may be incomplete. Although we have used the data of 10 African countries that have relatively complete data, there might still be some missing data issues.

In terms of new daily deaths, vaccination is having a diminishing impact on every continent except Asia. However, data analysis shows a soar in new deaths in Asia in 2023, a trend driven mainly by China, followed by Japan.

In the case of China, it is well known that China has always maintained a cautious approach to COVID-19 outbreak control, and then when it opened in early January 2023, it instead led to a surge in new cases and new deaths due to the shortage of healthcare resources and inadequate preparation, as well as China's own large population base. We can conclude that even though vaccination usually has a dampening effect on new deaths, sudden policy changes can still lead to serious consequences.

On the other hand, Japan, as the second largest contributor, has long faced the problem of an aging population. During the pandemic, even with a sound healthcare system, many elderly people died before receiving treatment due to the shortage of emergency medical transportation services.

In summary, while vaccination generally plays an important role in reducing the number of new cases and deaths from COVID-19, the interplay of factors such as human behavior, the emergence of new variants, the readiness of healthcare systems, and demographics can significantly influence observed trends, underscoring the importance of developing a comprehensive and flexible outbreak response strategy.

6 Appendix

During the COVID-19 epidemic, the government's role on the vaccine issue is crucial because it not only involves the formulation of public health policies, but also reflects a country's ability and efficiency to respond to the crisis. First, vaccine development, production and distribution are highly dependent on government support. An effective vaccination system requires strong financial investment, strict supervision and efficient resource allocation, all of which are inseparable from the leading role of the government. For example, by concentrating its efforts on promoting vaccine research and development, China has not only rapidly launched a variety of vaccines, but also achieved large-scale vaccination in a short period of time. This efficiency fully demonstrates its institutional advantages. However, mandatory policies alone are not enough to solve all problems. For the masses, the significance of vaccination must be scientifically communicated to build trust. The government needs to dispel public doubts about the safety and effectiveness of vaccines through transparent information disclosure and scientific publicity, thereby increasing vaccination rates.

At the same time, a key issue in epidemic prevention and control is how to strike a balance between strict public health policies and civil liberties. China's closed management measures have achieved remarkable results in controlling the spread of the virus, but they have also had a certain impact on people's daily life and mental health. My understanding is that the government can avoid the "comprehensive blockade" approach through more accurate risk assessment, and gradually shift to "precision prevention and control", focusing limited resources on high-risk areas and reducing intervention in low-risk areas. In addition, the application from China of digital management methods (such as health codes and vaccine tracking systems) provides the possibility to achieve more scientific and efficient epidemic monitoring. However, the promotion of any digitalization measures should take into account issues of privacy protection and social acceptance, which are also important factors that the government should consider during policy implementation.

Economic and social problems during the epidemic also require comprehensive support from the government. Strict epidemic prevention policies may lead to a stagnation of economic activities, especially for small and medium-sized enterprises and low-income groups. Until now, China's economy is still affected by strict control of the epidemic. Therefore, the government needs to alleviate people's economic pressure through economic subsidies, employment support, etc., and ensure that basic living standards are not affected by epidemic policies. This comprehensive approach can not only improve the effectiveness of epidemic prevention policies, but also help enhance social trust and lay a solid foundation for long-term anti-epidemic efforts. In summary, I believe that the role of the government is to find a dynamic balance between the science of vaccination, the flexibility of policies, and the well-being of the people, in order to achieve the dual goals of public health and socioeconomics.

Project Repository

The source code are available on GitHub. You can access it using the following link: GitHub Repository.

References

1. Zheng, C., Shao, W., Chen, X., et al. (2022). *Real-world effectiveness of COVID-19 vaccines: A literature review and meta-analysis.* International Journal of Infectious Diseases.

2. Feikin, D., et al. (2022). *Duration of effectiveness of vaccines against SARS-CoV-2 infection and COVID-19 disease.* The Lancet.

3. Wu, N., et al. (2023). *Long-term effectiveness of COVID-19 vaccines.* The Lancet.

4. WHO, COVID-19 Vaccines. Retrieved from `https://www.afro.who.int/health-topics/coronavirus-covid-19/vaccines`.

5. Pew Research Center (2024). *60% of Americans say they probably won't get an updated COVID-19 vaccine.*

6. Wikipedia contributors. (n.d.). *severe acute respiratory syndrome epidemic*. Wikipedia. Retrieved December 10, 2024, from https://zh.wikipedia.org/zh-hans/%E5%9A%B4%E9%87%8D%E6%80%A5%E6%80%A7%E5%91%BC%E5%90%B8%E7%B3%BB%E7%B5%B1%E7%B6%9C%E5%90%88%E7%97%87%E7%96%AB%E6%83%85#%E5%8F%82%E8%80%83%E6%96%87%E7%8C%AE.