Analyzing Global COVID-19 Trends: The Impact of Vaccination and Regional Disparities

Ashley Stevens, Xuanxuan Chen, Ziyi Tao

Introduction

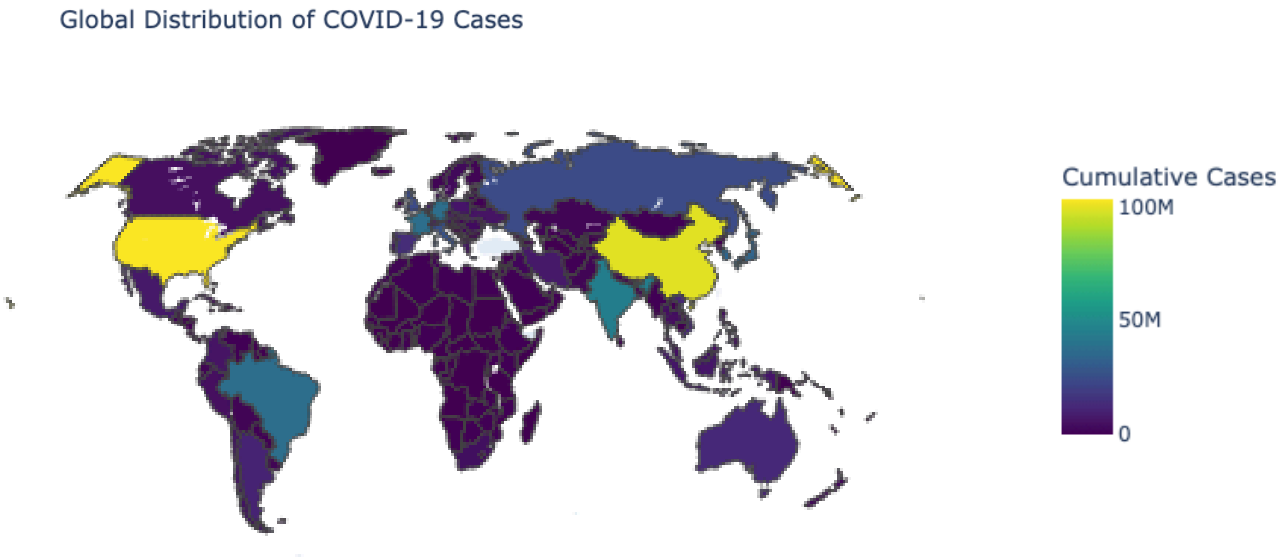

Figure 1: This map provides an overview of COVID-19 cases for each country around the world. Darker blue colors reflect higher numbers of cases, whereas lighter yellow colors indicate lower numbers of cases. Our codes show more details about this interactive map about regional disparities in the spread of COVID-19.

When COVID-19 hit the world in late 2019, no one really knew just how differently it would affect different parts of the globe. While some countries quickly rolled out mass vaccination programs, others struggled to get even basic supplies. In this report, we analyze these differences by looking at COVID-19 data from the six World Health Organization (WHO) regions, focusing on three main questions: How well did vaccines work in preventing deaths? Which regions handled the pandemic better than others? And why did some countries have a much harder time than others?

Using real data from the WHO, we tracked important numbers like how many people got vaccinated, how many people got sick, and sadly, how many people died. By creating graphs and charts to visualize this information, such as the global distribution map of COVID-19 cases presented above, we discovered some distinct patterns. The map in Figure 1 highlights the striking disparities in cumulative cases across countries with certain regions like North America and parts of Europe, exhibiting the highest case counts in stark contrast to lower counts in many parts of Africa. For more detail, an interactive version of this map (Figure 1) is included in the

Jupyter notebook, allowing further exploration of case counts by country and region. For example, places that vaccinated more people generally had fewer deaths. But we also found something concerning - poorer regions often had fewer vaccines available and worse health outcomes, showing how unfair our global health system can be.

These findings are not just interesting statistics - they tell us something important about how the world handles big health crises. Some regions had strong healthcare systems and plenty of resources to fight COVID-19, while others were left scrambling for basic medical supplies. Understanding these differences is not just about learning from the past; it is about figuring out how we can be better prepared for the next global health emergency and make sure everyone - no matter where they live - has a fair chance at staying healthy.

Motivation

This analysis is motivated by the major differences in how the COVID-19 pandemic affected different parts of the world and the need to understand how public health measures, like vaccination campaigns, played a role in these outcomes. The pandemic revealed big gaps in healthcare systems, testing abilities, and access to vaccines across countries. By looking at important numbers like vaccination rates, case fatality rates, and total deaths, this project aims to find patterns that can help create better global health strategies in the future, especially to make sure everyone has fair access to life-saving resources.

Another goal of this analysis is to explore how vaccination efforts are connected to pandemic outcomes. Vaccines are one of the most important tools for reducing severe illness and deaths, but how successful they were depended on things like healthcare systems, government policies, and the makeup of populations. This project highlights why timely and focused efforts are essential to managing global health crises and saving lives.

Lastly, this study looks at how COVID-19 cases and deaths changed over time in different parts of the world. By examining the links between vaccination rates, case fatality rates, and total deaths, the project gives insights into what worked and what did not in pandemic responses. The goal is to help countries prepare for future health emergencies and make healthcare systems stronger and more fair for everyone.

Data Cleaning and Preparation

The data cleaning process for the WHO COVID-19 and vaccination datasets involved several crucial steps to ensure data quality and consistency. For the COVID-19 dataset, the cleaning began with converting date fields to proper datetime format to enable temporal analysis. Negative values in numeric columns which likely represented reporting errors, were removed by setting a lower bound of zero. The process also addressed missing values by dropping entirely empty rows and filling remaining numeric missing values with zeros, as these likely represented no reported cases or vaccinations. For the vaccination dataset, similar datetime conversions were

performed on all date-related columns. Country names were standardized by converting them to uppercase and removing extra whitespace to ensure consistent matching between datasets. Percentage values, which were originally stored as strings with '%' symbols, were converted to proper floating-point numbers for mathematical operations. Missing values in categorical columns were filled with 'Unknown' while numeric missing values were set to zero. To track data quality, a scoring system was implemented that calculates the percentage of non-missing values for each record. Finally, duplicate entries were removed from both datasets to prevent double-counting of records. This comprehensive cleaning process resulted in two structured datasets ready for analysis, with standardized formats and consistent handling of missing or problematic values.

However, the limitations of WHO COVID-19 and vaccination datasets reflect broader challenges in global health data collection and reporting. These datasets inherently suffer from reporting inconsistencies across different countries and regions, as nations vary significantly in their testing capacity, reporting standards, and health infrastructure. Data quality often correlates with a country's economic resources, potentially leading to systematic underreporting in lower-income regions. Additionally, the dynamic nature of the pandemic means that reporting criteria and methods evolved over time, making temporal comparisons challenging. Political factors may have influenced reporting practices, with some countries potentially underreporting cases or deaths for various reasons. The vaccination data faces similar challenges, with inconsistent reporting intervals and varying definitions of vaccination status across jurisdictions. There are also significant time lags in data reporting, and some countries may have gaps in their reporting during periods of crisis or system overload. The data might not capture informal healthcare settings or rural areas effectively, and differences in testing strategies, case definitions, and death attribution methods between countries make direct comparisons problematic. Furthermore, these datasets typically don't account for unreported cases, asymptomatic infections, or deaths that occurred outside of formal healthcare settings, suggesting that the true impact of the pandemic may be significantly different from what the data indicates.

Analysis

Our analysis focuses on the global COVID-19 pandemic, examining it through three key perspectives: regional vaccination differences, the connection between vaccination efforts and health outcomes, and how the pandemic evolved over time in different regions and countries. Using data from the WHO COVID-19 dataset and a global vaccination dataset, we explored important metrics like vaccination rates (doses per 100 people), cumulative cases, and case fatality rates (CFR).

From this analysis, three main findings emerged:

1. Regional Disparities: There is a clear gap between regions with strong healthcare systems and those with fewer resources. Regions like the Western Pacific and Europe consistently led in vaccination rates while Africa faced challenges in both vaccine access and accurate case reporting.
2. Vaccination and Health Outcomes: We found an inverse relationship between vaccination rates and case fatality rates—regions with higher vaccination rates tended to have lower death rates. However, this connection is influenced by other factors, such as the quality of healthcare and the accuracy of reporting.
3. Different Pandemic Timelines: The way the pandemic progressed varied greatly across regions. Some areas experienced clear waves of infections, while others had slower, steadier increases. These differences often depended on policies, healthcare systems, and public health strategies.

Through visualizations and in-depth analysis, we explored how these patterns played out in different WHO regions and countries. This helped identify both successes and challenges in managing the pandemic, offering a clearer understanding of what factors contributed to different outcomes worldwide.

Regional Analysis of Vaccination and Cases

Having identified these three major patterns in our initial analysis, we can now examine how they manifested specifically across different World Health Organization regions. The following regional analysis provides a detailed examination of vaccination efforts and case trajectories, offering concrete examples of the disparities and relationships identified above.

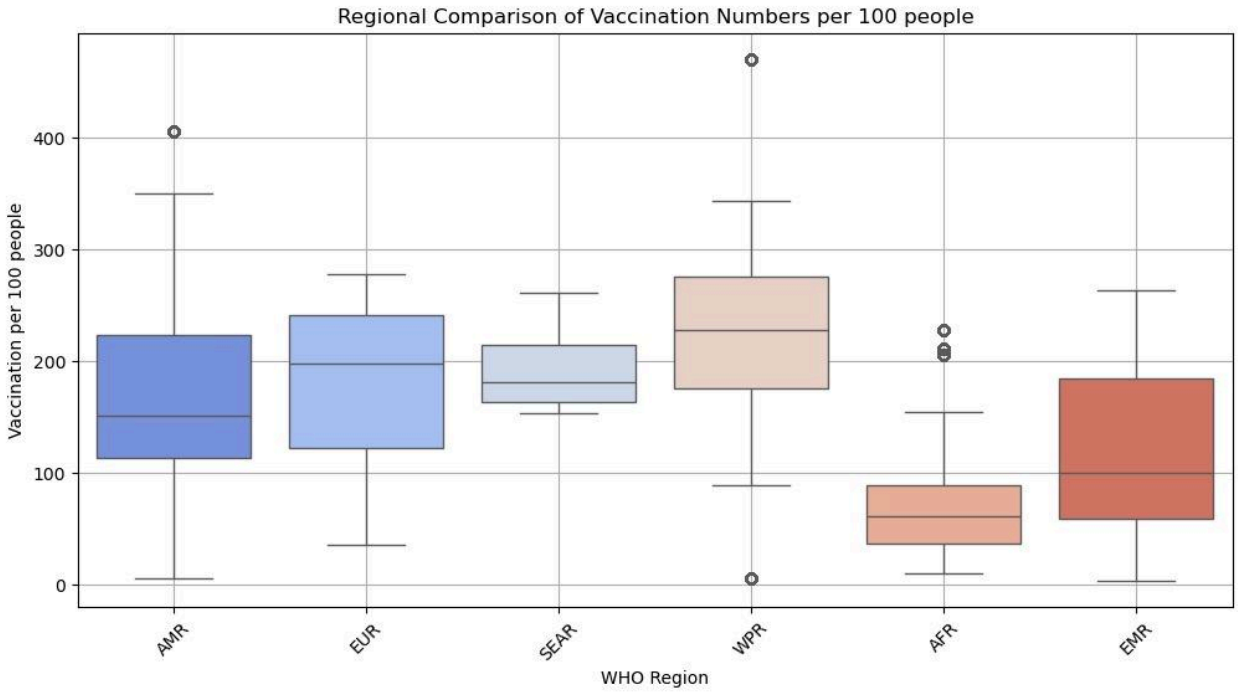

Figure 2: The boxplot shows COVID-19 vaccination rates across six World Health Organization (WHO) regions: Americas (AMR), Europe (EUR), South-East Asia (SEAR), and Western Pacific (WPR), Africa (AFR), Eastern Mediterranean (EMR).

The boxplot in Figure 2 shows COVID-19 vaccination rates across six World Health Organization (WHO) regions: Africa (AFR), Americas (AMR), Eastern Mediterranean (EMR), Europe (EUR), South-East Asia (SEAR), and Western Pacific (WPR). The x-axis represents the WHO regions, and the y-axis shows vaccination rates per 100 people. It highlights the differences in vaccination efforts across the world, showing both the average rates (medians) and how much they vary within each region.

Furthermore, The Western Pacific Region (WPR) and European Region (EUR) in Figure 2 stand out with the highest median vaccination rates, meaning most countries in these regions were able to vaccinate large portions of their populations. WPR also has some outliers—countries with even higher vaccination rates—showing that targeted efforts can lead to exceptional results. The Americas (AMR) and South-East Asia (SEAR) are in the middle range for vaccination rates but their results are more inconsistent. For example, AMR has a large spread, suggesting big differences in vaccination success between countries in that region. The African Region (AFR) has the lowest vaccination rates overall, showing the major challenges African countries face in getting vaccines and distributing them. The Eastern Mediterranean Region (EMR) also has lower median rates but with more variation meaning some countries did better than others. Even in AFR and EMR, some countries achieved higher vaccination rates showing that well-planned efforts can still make a big difference even in regions with fewer resources.

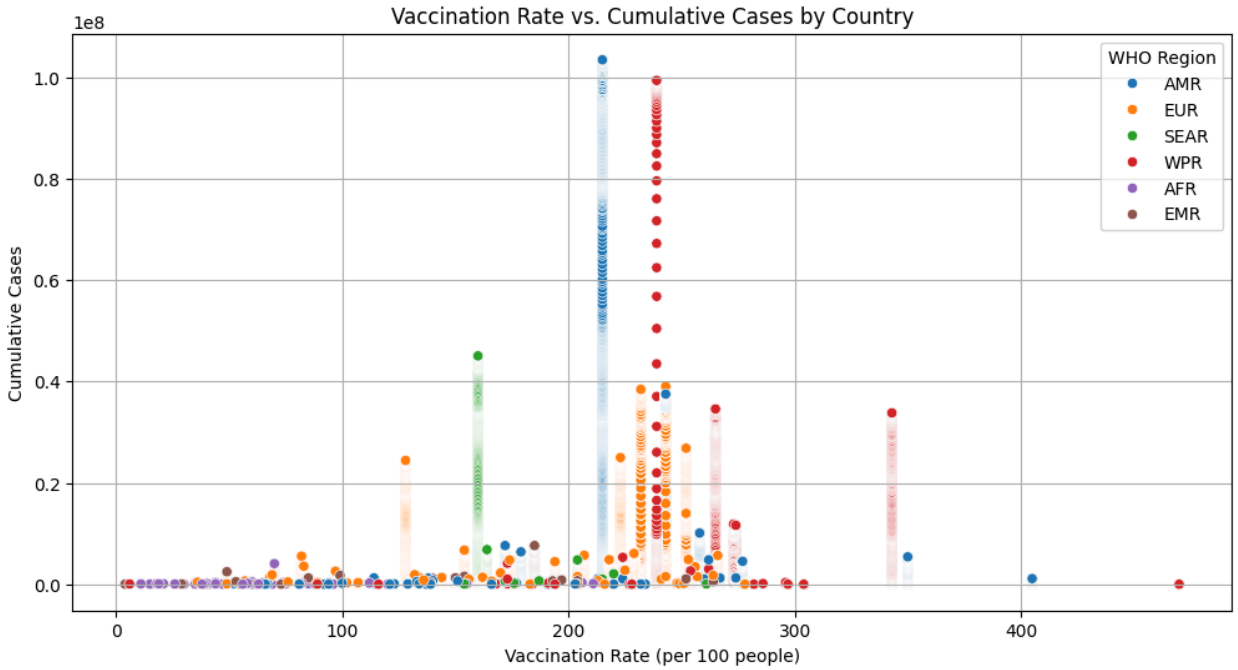

Figure 3: The scatter plot illustrates the relationship between vaccination rates and cumulative COVID-19 cases by country. Each point corresponds to a country, with colors indicating their respective WHO regions: Americas (AMR), Europe (EUR), South-East Asia (SEAR), Western Pacific (WPR), Africa (AFR), Eastern Mediterranean (EMR).

The scatter plot in Figure 3 shows the relationship between vaccination rates and total COVID-19 cases. The x-axis represents vaccination rates per 100 people, and the y-axis shows the total number of COVID-19 cases in each country. Each dot represents a country, and the different colors represent WHO regions, Americas (AMR), Europe (EUR), South-East Asia (SEAR), Western Pacific (WPR), Africa (AFR), Eastern Mediterranean (EMR).

This reveals an interesting pattern: countries with higher case numbers, like those in the Americas (AMR) and Western Pacific Region (WPR), tend to have higher vaccination rates. This suggests that countries with more severe outbreaks responded with strong vaccination efforts, including widespread booster campaigns, as seen in vaccination rates exceeding 100 per 100 people. Two clear clusters appear at 200-250 and 300-350 vaccinations per 100 people showing the intensity of global vaccination efforts in heavily affected regions.

On the other hand, African countries show both lower case numbers and lower vaccination rates, which reflects challenges like limited vaccine access, testing, and healthcare infrastructure. Other regions, like Europe and the Eastern Mediterranean, show a mix of vaccination rates and case counts, highlighting more variability. This visualization points to the major role of healthcare access and public health responses in shaping both vaccination trends and the impact of the pandemic across the world.

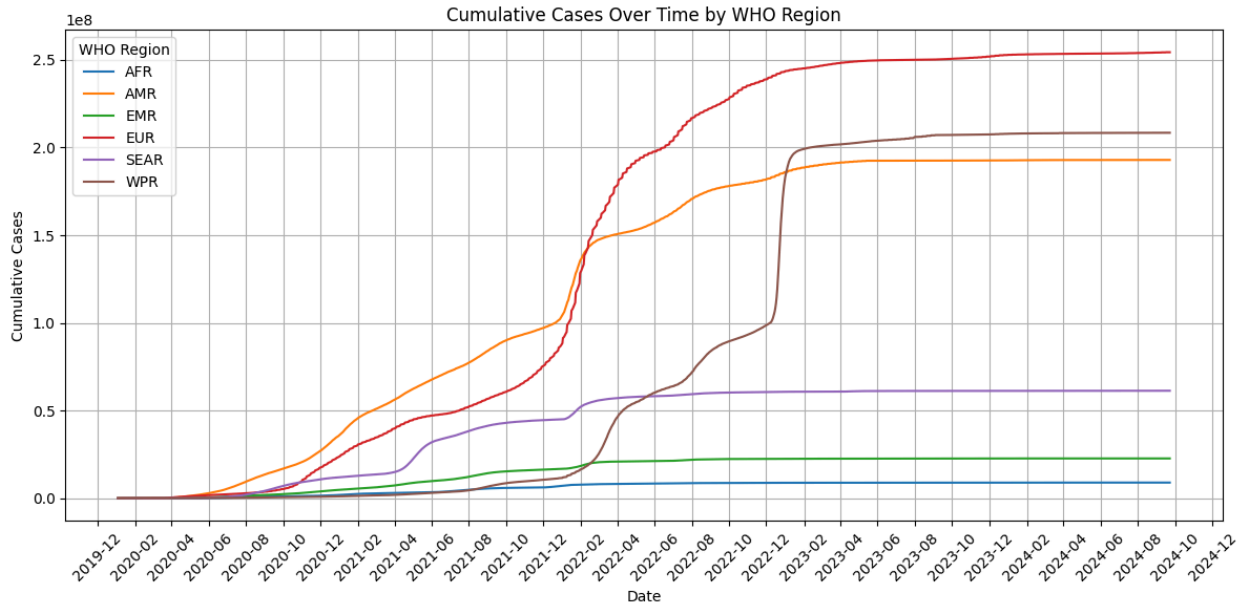

Figure 4: The line graph shows the trajectory of cumulative COVID-19 cases across different WHO regions from late 2019 through December 2024. Each colored line corresponds to one of the six World Health Organization regions: Africa (AFR), Americas (AMR), Eastern Mediterranean (EMR), Europe (EUR), South-East Asia (SEAR), and Western Pacific (WPR).

The line graph in Figure 4 shows the trajectory of cumulative COVID-19 cases across different WHO regions from late 2019 through projected estimates into 2024. The x-axis represents the timeline, starting in December 2019 and extending to December 2024, while the y-axis indicates cumulative COVID-19 case counts, measured in millions. Each colored line corresponds to one of the six WHO regions: Africa (AFR), Americas (AMR), Eastern Mediterranean (EMR), Europe (EUR), South-East Asia (SEAR), and Western Pacific (WPR). The visualization presents a compelling story of how the pandemic evolved differently across global regions.

The Americas (AMR) and European Region (EUR) in Figure 3 also stand out with the highest cumulative case counts. Their curves show similar patterns, with steep increases during major waves of infection, particularly in late 2021 and early 2022. After this period, both regions' lines flatten, reflecting a slowdown in new cases. The European Region maintains a slightly higher total, highlighting the extended impact of the pandemic in this region. The Western Pacific Region (WPR) exhibits a distinctive pattern with relatively low case counts until early 2022 when the curve rises sharply. This dramatic vertical increase aligns with significant policy changes or delayed outbreak waves in key countries within the region. Post-surge, the curve flattens suggesting effective measures to curb further spread. The South-East Asia Region (SEAR), Eastern Mediterranean Region (EMR), and African Region (AFR) display much lower cumulative case counts. Africa's line remains notably flat throughout the entire period, likely

reflecting underreporting due to limited testing capacities rather than actual infection rates. SEAR shows moderate growth with visible waves of infection while EMR exhibits a steady but slower increase in case counts. Both regions' curves taper off towards the end, indicating stabilization.

Figure 4 emphasizes the global nature of COVID-19 while highlighting stark regional differences in its impact and reporting. Disparities in testing capacities, public health infrastructure, and pandemic responses across regions are evident and underscore the unequal burden of the pandemic worldwide.

Temporal and Outcome Trends in COVID-19

While the regional analysis reveals significant geographical disparities in pandemic response and outcomes, these patterns did not remain static. The following temporal analysis explores how these regional differences evolved over time, revealing dynamic relationships between vaccination efforts and health outcomes. By examining cumulative case trends, vaccination rates in relation to case fatality rates, and the total death tolls in the most impacted nations, we can better understand how initial regional disparities influenced long-term pandemic trajectories.

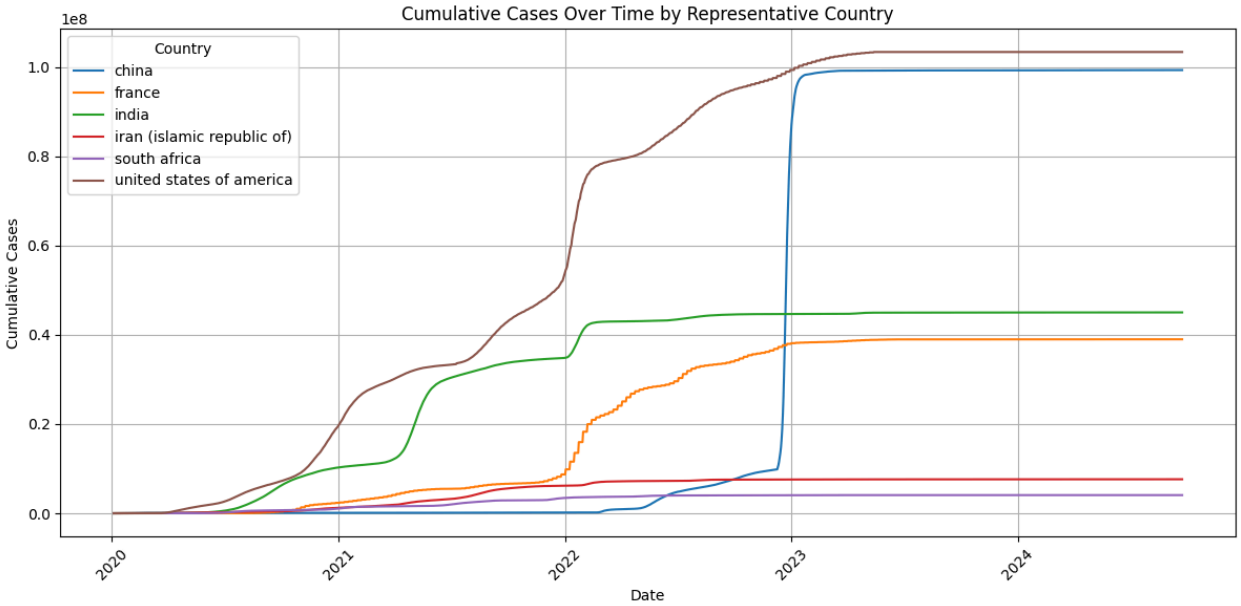

Figure 5: The line graph shows the progression of cumulative COVID-19 cases across six representative countries—China, France, India, Iran, South Africa, and the United States—from 2020 through 2024.

This line graph in Figure 5 illustrates the progression of cumulative COVID-19 cases across six representative countries—China, France, India, Iran, South Africa, and the United

States—from 2020 through projections into 2024. The x-axis represents the timeline, starting in 2020 and extending to 2024, while the y-axis shows cumulative COVID-19 case counts, measured in millions. Each line corresponds to a specific country, highlighting their distinct trajectories during the pandemic.

The United States stands out with the highest reported cumulative cases, its line showing a steady increase beginning in early 2020. The curve features periods of sharp acceleration, particularly in late 2021 and early 2022, before gradually flattening as new case growth slowed. In contrast, China presents an unusual pattern with consistently low reported case counts until early 2023, when its line spikes dramatically, reflecting a potential policy shift from its "zero-COVID" strategy. After this vertical surge, the curve plateaus, suggesting either containment or changes in reporting practices. India and France demonstrate similar overall trends but on different scales. India's line features distinct waves, with periods of rapid growth followed by slower increases, ultimately reaching a middle-range case count. France's trajectory is steadier with a consistent upward slope and less pronounced surges reflecting a gradual accumulation of cases over time. Iran and South Africa maintain relatively lower cumulative case counts throughout the pandemic and their lines showing earlier plateaus and modest increases compared to the other countries. These flatter curves may reflect differences in testing capacity, healthcare infrastructure, and reporting systems, as well as potentially lower infection rates.

Figure 5 underscores how national strategies, healthcare systems, and public health policies influenced the pandemic's impact and its documentation across different countries. It highlights the variability in pandemic experiences, from aggressive outbreaks to controlled spread, shaped by diverse global contexts.

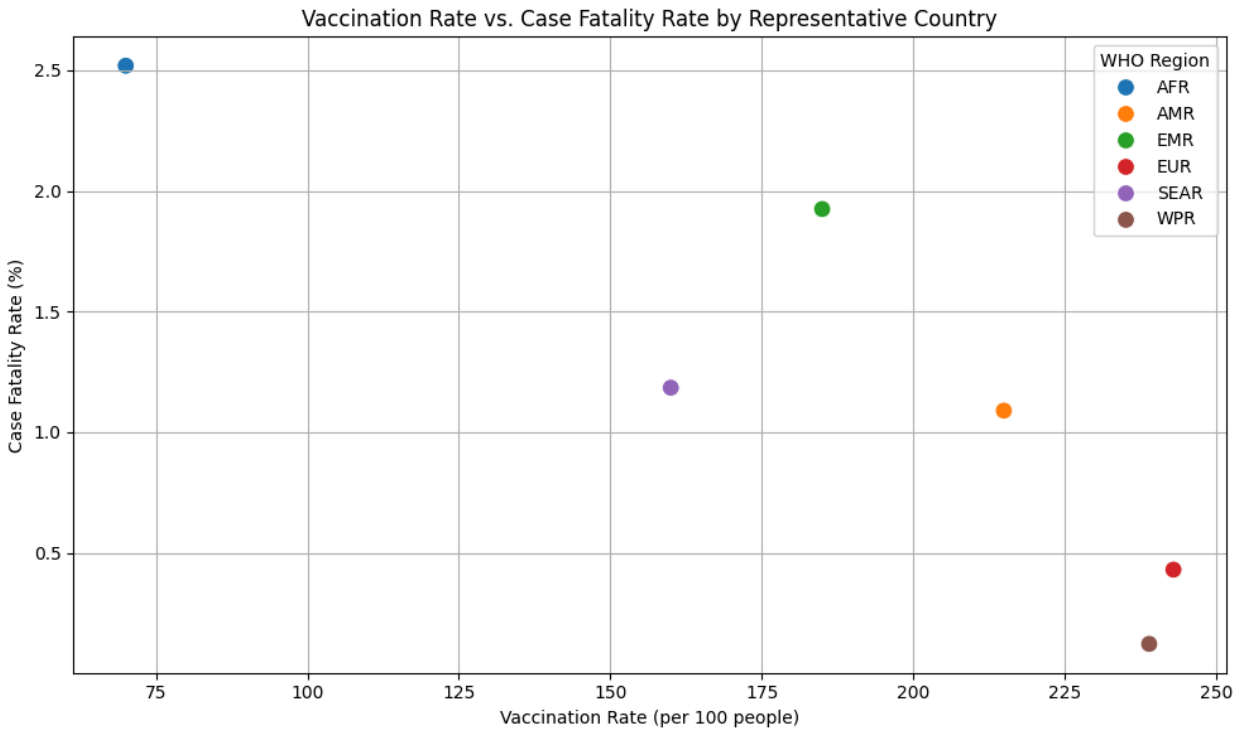

Figure 6: The scatterplot visualizes the relationship between vaccination rates and case fatality rates (CFR) across countries grouped by WHO regions. Each dot represents a country, with different colors indicating WHO regions: Africa (AFR), Americas (AMR), Eastern Mediterranean (EMR), Europe (EUR), South-East Asia (SEAR), and Western Pacific (WPR).

This scatterplot in Figure 6 visualizes the relationship between vaccination rates and case fatality rates across different WHO regions. The x-axis represents vaccination rates, measured as doses per 100 people, while the y-axis shows case fatality rates (CFR) as a percentage. Each dot on the graph represents a country, with colors distinguishing the six WHO regions: Africa (AFR), Americas (AMR), Eastern Mediterranean (EMR), Europe (EUR), South-East Asia (SEAR), and Western Pacific (WPR).

A striking inverse relationship emerges between vaccination rates and case fatality rates as you can see in Figure 6. The African region (AFR), despite having the lowest vaccination rate at around 75 per 100 people, shows the highest case fatality rate at approximately 2.5%. In contrast, the European (EUR) and Western Pacific (WPR) regions, with the highest vaccination rates approaching 250 per 100 people, demonstrate the lowest fatality rates at under 0.5%. The middle range presents an interesting pattern, with the Eastern Mediterranean region (EMR) showing a relatively high fatality rate of nearly 2% despite moderate vaccination levels around 180 per 100 people. The Americas (AMR) and South-East Asia (SEAR) regions fall between these extremes, with fatality rates around 1-1.2% and varying vaccination rates.

Figure 6 powerfully illustrates how higher vaccination rates generally correlate with lower case fatality rates, though other factors such as healthcare infrastructure, population

demographics, and reporting accuracy likely influence these outcomes. The stark contrast between regions with high and low vaccination rates underscores the critical role of vaccine access and distribution in reducing COVID-19 mortality.

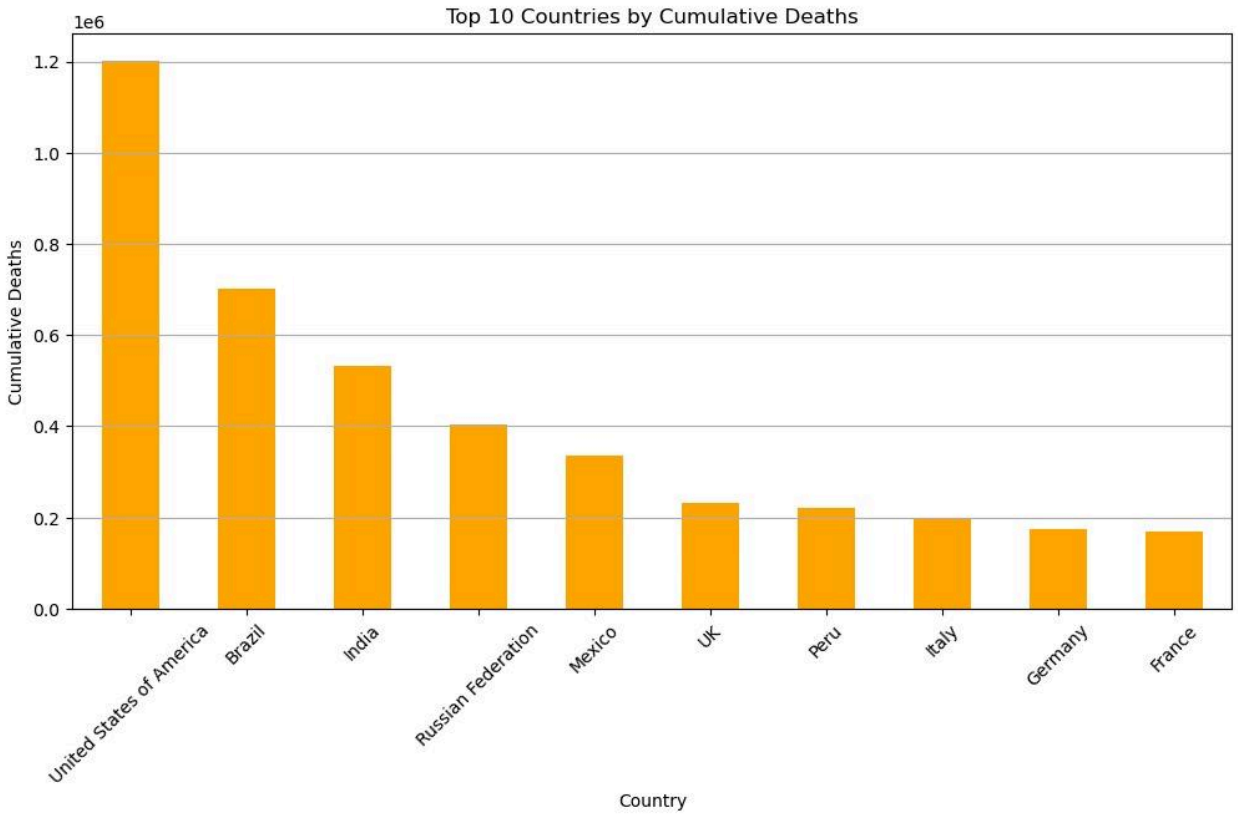

Figure 7: The bar chart displays the cumulative deaths caused by COVID-19 across the top 10 most impacted countries.

This bar chart in Figure 7 illustrates the cumulative deaths caused by COVID-19 across the top 10 most impacted countries. The x-axis represents the countries, displayed in descending order of total deaths, while the y-axis shows the cumulative deaths, measured in millions. Each orange bar corresponds to a country's reported total fatalities due to COVID-19, providing a stark visual comparison of the pandemic's human toll.

The United States leads significantly, with a bar towering at approximately 1.2 million deaths, nearly double the next country. Brazil follows, reporting around 700,000 deaths, and India ranks third with approximately 525,000 fatalities. Together, these three countries account for a significant proportion of the global COVID-19 death toll.

The remaining countries show a more gradual decline. The Russian Federation reports about 400,000 deaths, while Mexico stands in the middle range with approximately 330,000 fatalities. European nations, including the United Kingdom, Italy, Germany, and France, cluster toward the lower end of the top 10, each reporting between 150,000 to 225,000 deaths.

Figure 7 provides a sobering perspective on the pandemic's impact, highlighting how factors such as population size, healthcare capacity, and reporting accuracy may contribute to the disparities in mortality across nations. It underscores the disproportionate burden carried by some of the world's largest and most populous countries.

Limitations

While this analysis provides important insights into global COVID-19 trends, it is crucial to acknowledge several limitations that may impact the interpretation of our findings. First, the accuracy of the data varies significantly across regions and countries due to differences in testing availability, reporting standards, and healthcare infrastructure. For example, underreporting of cases and deaths, particularly in regions with limited resources such as Africa, may have led to an underestimation of the pandemic's true impact. This discrepancy complicates direct comparisons between regions and highlights the need for more consistent global reporting mechanisms.

Second, vaccination rates and death rates are influenced by numerous factors not captured in this analysis. Variables such as population demographics, the spread of new variants, and the effectiveness of non-pharmaceutical interventions like lockdowns and mask mandates all play critical roles in shaping pandemic outcomes. As a result, it is challenging to isolate the direct impact of vaccination efforts on reducing cases and fatalities.

Additionally, the political context in some countries may have influenced the accuracy of reported data. In certain cases, governments may have downplayed deaths or overstated vaccination rates to project a more favorable image internationally or domestically. These political motivations can distort the data and obscure the true effectiveness of public health measures. Moreover, regional averages often mask significant within-country disparities. For example, urban areas may have better access to vaccines and healthcare services compared to rural regions, but these differences are not reflected in aggregated regional data, making it difficult to fully capture local challenges.

Lastly, this study focuses on quantitative metrics like vaccination rates, cumulative cases, and fatalities, while excluding other qualitative factors such as economic instability, political decisions, and societal trust in public health systems. These elements likely influenced how countries responded to the pandemic and their overall outcomes. Including such dimensions in future studies would provide a more comprehensive understanding of the pandemic's global impact.

Despite these limitations, the analysis sheds light on critical patterns and disparities in COVID-19 trends, offering valuable insights for improving preparedness and equity in future global health crises. Addressing these challenges will require more robust data collection, enhanced transparency, and a commitment to fair resource distribution worldwide.

Conclusion

Our analysis highlights how the global response to COVID-19 was shaped by three critical factors: the success of vaccination campaigns, the effectiveness of public health policies, and the resilience of healthcare systems. By analyzing data on case numbers, vaccination rates, and mortality rates, we revealed stark disparities. Regions like Europe and the Western Pacific achieved high vaccination coverage and consequently experienced lower death rates. In contrast, regions such as Africa and the Eastern Mediterranean faced significant challenges due to limited vaccine access, weaker healthcare infrastructure, and disparities in public health preparedness, resulting in higher mortality rates and prolonged outbreaks.

These findings underscore a troubling reality: pre-existing global inequalities significantly influenced the pandemic's impact. Disparities in vaccine distribution, testing capabilities, and healthcare access created a two-tiered response, where wealthier regions could protect their populations while others struggled to secure even basic medical resources. This inequity highlights a critical need for more inclusive global health strategies, as the pandemic did not just expose healthcare weaknesses but also systemic inequities that must be addressed.

Looking forward, the lessons from this analysis are clear. First, building robust and equitable healthcare systems is essential to ensure that all regions can respond effectively to health emergencies. Investments in healthcare infrastructure, particularly in low-resource regions, must be prioritized. Second, the creation of fair and efficient global mechanisms for distributing medical resources during crises is imperative. Vaccine equity, rapid response plans, and transparent data sharing are key components of future preparedness. Lastly, fostering stronger international cooperation will be vital. Collaborative efforts among nations, guided by shared goals of equity and resilience, can help mitigate the impact of future pandemics.

While COVID-19 exposed critical vulnerabilities in our global health system, it also provided a roadmap for building a more resilient and equitable system. The challenge now is to act on these lessons, ensuring that no region is left behind in the fight against future health crises. Only through sustained commitment to global health equity and collaboration can we hope to protect all populations and create a healthier, more equitable world.

Appendix

Source for data about COVID cases, deaths, and vaccinations:
https://data.who.int/dashboards/covid19/data