COVID-19 Pandemic Analysis

Caroline Song

2024-12-18

Abstract

The COVID-19 pandemic has had a profound impact on public health systems globally, marked by distinct waves of cases, hospitalisations, and deaths. This study analysed U.S. state-level data from January 2020 to December 2023 to identify key pandemic trends, assess state-level mortality variations, and examine the trend of COVID-19 virulence. The pandemic was segmented into five waves, each characterised by unique characteristics in cases, hospitalisations, and deaths. Wave 1, driven by limited preparedness and overwhelmed healthcare systems, exhibited the highest case-fatality rate (CFR). Subsequent waves, particularly Waves 4 and 5 dominated by Omicron variants, featured significant reductions in CFR and hospitalisation-to-death ratios, largely due to widespread vaccination rollout and improved treatment protocols.

State-level analysis revealed significant disparities in death rates, influenced by factors such as population density, availability of healthcare infrastructure, and effectiveness of public health policies. States with higher vaccination coverage consistently exhibit lower death rates, even during periods of high viral transmission. The analysis also demonstrated a decline in COVID-19 virulence over time, highlighting the critical role of vaccination in reducing severe outcomes.

These findings emphasise the importance of equitable vaccine distribution, robust public health infrastructure, and adaptive policies to address future pandemics and public health crisis. This study offers valuable lessons for enhancing pandemic preparedness and reducing health disparities in the face of evolving global health threats.

Introduction

The COVID-19 pandemic has had significant and long-lasting effect on the economy, public heath, and healthcare systems worldwide. Ever since the emergence of the spread of the SARS-CoV-2 virus in late 2019, it has caused significant morbidity and mortality, with over

6.5 million deaths globally by the end of 2023(Centers for Disease Control and Prevention 2023b), (Roser, M., Ritchie, H., Ortiz-Ospina, E., and Hasell, J. 2023). The rapid spread of the SARS-CoV-2 virus has led to unprecedented challenges, such as overloading the healthcare systems, disrupting the economy, as well as extensive public health interventions such as social distancing, mask mandates, and lockdowns. Therefore, it is crucial to understand the progression of the pandemic, and evaluate the effectiveness of interventions such as vaccination campaigns to inform future decision making when facing public health crisis.

One of the most prominent features of the COVID-19 pandemic was its distinct waves of infection, hospitalisation, and deaths. These waves were driven by various factors, including the widespread of new viral variants, seasonal effects, and changes in the public behaviours. For example, the Alpha, Delta, and Omicron variants were seen to be associated with significant increases in cases due to their enhanced transmissibility (Gupta, A., Sharma, R., and Kumar, S. 2023). In addition, the rollout of vaccination efforts in late 2020 also led to an alteration of the trajectory of the pandemic. Specifically, with the target of achieving herd immunity, over 70% of the global population had received at least one dose of a COVID-19 vaccine by the end of 2023, which significantly contributed to severe outcomes and mortality (Kaiser Family Foundation 2023).

Despite the vaccination efforts, disparities in vaccination coverage and effectiveness were evident across states in the United States. Variable vaccination rates resulted from vaccination hesitancy, logistical challenges, vaccination side-effects, and socioeconomic inequalities in turn impacted COVID-19 outcomes (Bar, K., et al. 2021). Furthermore, the mutations of SARS-CoV-19 virus led to improvements of its transmissibility among the population and immune escape capabilities, and sometimes severity, which all posed challenges to maintaining high levels of immunity in the population.

This project aims to analyse the COVID-19 trends in the Unites States from January 2020 to December 2023, with a specific focus on identifying distinct waves of the pandemic, comparing death rates by states, and characterising the virulence of SARS-CoV-19 mutants over time. Data on COVID-19 cases, hospitalisation, and deaths were obtained from the Centres for Disease Control and Prevention (CDC) (Centers for Disease Control and Prevention 2023a), and the data on state population between 2020 to 2024 were obtained from United States Census Bureau (U.S. Census Bureau 2021).

The main aims of this project are to answer the following questions:

1. What are the main waves of the pandemic and what are the key characteristics of each wave regarding confirmed case, hospitalisation rates, and death rates?
2. How did COVID-19 death rates vary by state during these waves, and what factors influenced these variations?
3. How has the COVID-19 virulence evolved over time, and how did vaccination coverage impact this trend?

By addressing these questions, this project aims to provide a comprehensive overview of the progression of COVID-19 in the United States from 2020 to 2023, giving insights into the interplay between public health measures, virus mutations, and population outcomes. The results will contribute to reflecting the effectiveness of public health initiatives employed to combat the pandemic, hence offering insights into managing future public health emergencies.

Methods

Data Sources

The datasets used to analyse COVID-19 trends across the United States between January 25th 2020 and May 20th 2023 were obtained from publicly available datasets. The data sources are:

1. U.S. Census Bureau Data; State-level population estimates for the year 2020 to 2023 were obtained from census.gov using their published CSV file. These data were crucial for normalising cases, hospitalisation, and death counts by state population to compute rates per 100, 000 individuals.
2. CDC;

- COVID-19 case data: weekly new cases by state were accessed via the CDC's COVID-19 dataset API.
- Death Data: Weekly state-level deaths due to COVID-19 were obtained from the CDC's provisional death count dataset.
- Hospitalisation Data: Weekly state-level COVID-19 hospitalizations were retrieved from the CDC's hospitalisation metrics dataset.
- Vaccination Data: Weekly vaccination rates, including series completions and booster administrations, were accessed through the CDC vaccination API.

3. Date Information: A synthetic dataset was created for weekly dates across the study period to align all other datasets temporally.

All data sources were integrated and matched by state, week, and year to create a comprehensive dataset for analysis. Data cleaning and preprocessing were conducted using R (version 4.4.0), primarily using packages including tidyverse, ggplot2, janitor, dplyr, stringr, and httr2.

Data wrangling and cleaning

1. Population Data: Population datasets were formatted to ensure consistency in column names and variable types. For example, year and population columns were transformed into numeric variables, and state abbreviations were added using state.abb and state.name variables.

- For 2020 and 2021, population data were fetched through the Census API.
- For 2022 and 2023, population data were fetched through the Census API, then cleaned and merged with the earlier years using bind_rows().

2. **COVID-19 Metrics**:

- Cases: Data were filtered to retain relevant columns (state, end_date, new_cases) and cleaned to parse dates and numeric values accurately. Weekly metrics were calculated using epiweek() and epiyear().
- Deaths, vaccination, and hospitalisations rates: Similar cleaning steps were applied, ensuring numeric parsing and alignment with weeks and years. Rates were calculated as percentages relative to state populations (rate per 100,000 individuals).

3. **Date Alignment**: A comprehensive date dataset was created using seq() to ensure all metrics were aligned temporally. This alignment facilitated left joins between the cases, hospitalisation, death, and vaccination datasets.
4. **Data Merging**: All datasets were merged using left_join() by state, week, and year to create a unified dataset. Special attention was paid to ensuring consistency in state identifiers (i.e., abbreviations and full names).

Data Analytics

1. **Segmentation of Waves**: The pandemic period was divided into five waves based on visual patterns in cases, deaths, and hospitalisations. These waves were defined using case_when() logic, with start and end dates derived from inflection points in the time-series visualisations.
2. **Calculation of Rates**:

- **Case, Death, and Hospitalisation Rates**: Per 100,000 population, calculated for each state and wave.
- **Case-Fatality Ratios (CFR)**: Calculated as deaths divided by cases (%), with safeguards to avoid division by small numbers (using filter() to remove observations less than 10).

- **Hospitalisation-to-Death Ratios (HDR)**: Calculated as hospitalisations divided by deaths, presented on a logarithmic scale due to large variability.

3. **Visualisation Techniques**:

- Time-series plots were used to identify and justify wave segmentation.
- Heatmaps highlighted state-level variations in death rates and HDR across waves.
- Scatter plots examined relationships between vaccination coverage and death rates.
- Rolling averages (7-day) were computed to smooth fluctuations in cases, deaths, and hospitalisations over time.

Assumptions and Limitations

1. Data completeness: The analysis assumes completeness and accuracy of CDC-reported cases, deaths, and hospitalisations. However, there might be issues of underreporting, especially during time periods of high transmission or resource constraints.
2. Population stability: Population estimates were assumed stable within each calendar year, despite possible mid-year migrations or demographic changes.
3. Temporal resolution: Weekly interval may mask shorter-term fluctuations.
4. Vaccination effectiveness: Analysis did not account for the differences in vaccine types, rollout strategies, or decreased immunity over time.
5. Wave definitions: The segmentation of COVID-19 waves is based on subjective interpretation of visual patterns in the data. Alternative definitions may yield slightly different results.

Tools and Reproducibility

All analyses were conducted in R, using library such as tidyverse, janitor, dplyr, stringr, ggplot2, and httr2 for data wrangling, visualisation, and statistical calculations. The use of APIs ensures that the datasets can be re-extracted for validation or future research. The final integrated dataset serves as the foundation for all subsequent analyses.

Results

Main Waves of the Pandemic and Key Characteristics

The COVID-19 pandemic was segmented into five distinct waves, each characterised by different rates of cases, deaths, and hospitalisations (Figure 1). Specifically, these waves are:

- **Wave 1** (Jan 2020–Sep 2020) marked the initial global outbreak, showing a sharp increase in death rates due to limited healthcare preparedness and no vaccines. *Supplementary Figure 1* provides detailed trends in cases, deaths, and hospitalisations for all U.S. states over time, further illustrating the sharp rise in Wave 1.
- **Wave 2** (Oct 2020–Jun 2021) corresponded with the fall and winter surges, and Alpha variant's spread, and its heightened severity led to high hospitalisation rates.
- **Wave 3** (Jul 2021–Nov 2021), a distinct resurgence in all metrics driven by the Delta variant, exhibited the highest hospitalisation rates but lower case fatality compared to Wave 1.
- **Wave 4** (Dec 2021–Apr 2022) showed a spike in cases due to the Omicron BA.1variant, which was highly transmissible but less severe, hence death and hospitalisation rates were relatively lower, reflecting reduced virulence and higher vaccination coverage. *Supplementary Figure 4* highlights the 7-day rolling averages, smoothing out fluctuations to show the clear peak in cases during Wave 4 with lower corresponding severe outcomes.
- **Wave 5** (May 2022 to May 2023) represented smaller peaks associated with Omicron subvariants such as BA.2 and BA.5, demonstrating improved public immunity and milder disease progression.

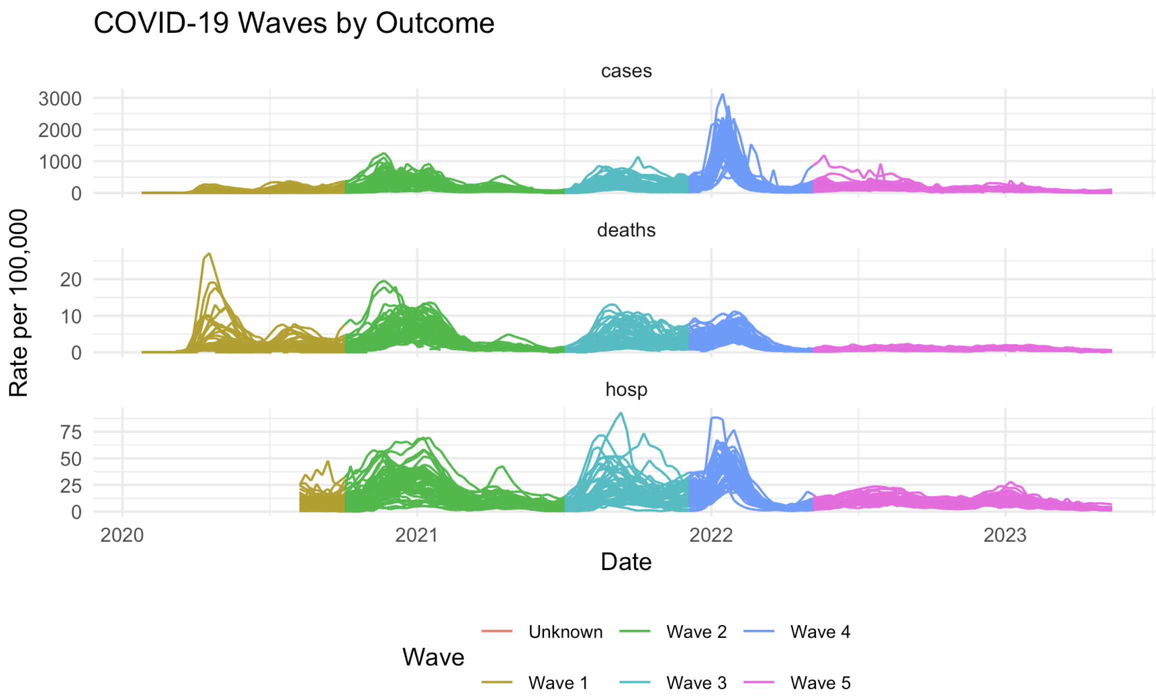

Figure 1. COVID-19 Waves by Outcome: This figure shows the weekly rates of confirmed cases, hospitalisations, and deaths per 100,000 population across five distinct waves of the COVID-19 pandemic. Each wave reflects changes in variant dynamics, public health measures, and vaccination rates.

Variation in COVID-19 Death Rates by State

Death rates varied widely across states during different pandemic waves (Figure 2).

- States like **New Jersey** and **Connecticut** exhibited extremely high death rates during Wave 1, reflecting early pandemic challenges. This observation is complemented by *Supplementary Figure 3*'s state-level comparison of the HDR across all waves. Southern and western states, such as **Alabama** and **Arizona** had relatively low death rates, indicating a slower spread.
- In Wave 2, most states featured relatively high death rates, especially in the Midwest and Southern regions, such as **South Dakota and Oklahoma** having the top death rates. **Hawaii**, **Vermont and Massachusetts** showed notable reductions in death rates, reflecting the role of geographical location, and better control measures.
- In Wave 3, Southern and Midwestern states such as **Florida**, **Texas** and **Arkansas** had higher death rates, likely due to the emergence of the Delta variant. While northeastern

states like **Vermont** and **Maine** continued to perform better, showing relatively low death rates.

- Then in Wave 4, death rates surge in several states, particularly in the South and Midwest, including **Florida** and **Oklahoma** continued to feature a relatively high death rate since Wave 2, whereas **Puerto Rico** and **Hawaii** had relatively low death rates. Northeastern states such as **Connecticut** maintained relatively low death rates.
- Lastly, in Wave 5, most states had death rates lower than 50 per 100, 000 individuals, reflecting widespread availability of vaccines and treatments. Minor variations can still be observed, with states like **West Virginia** and **Kentucky** showing slightly higher death rates compared to others.

Some states consistently had lower death rates across waves, such as Northeastern states like Vermont, Maine and Connecticut, likely due to strong healthcare infrastructure and public health measures. Southern states such as Mississippi, Florida, and Texas had higher death rates in multiple waves.

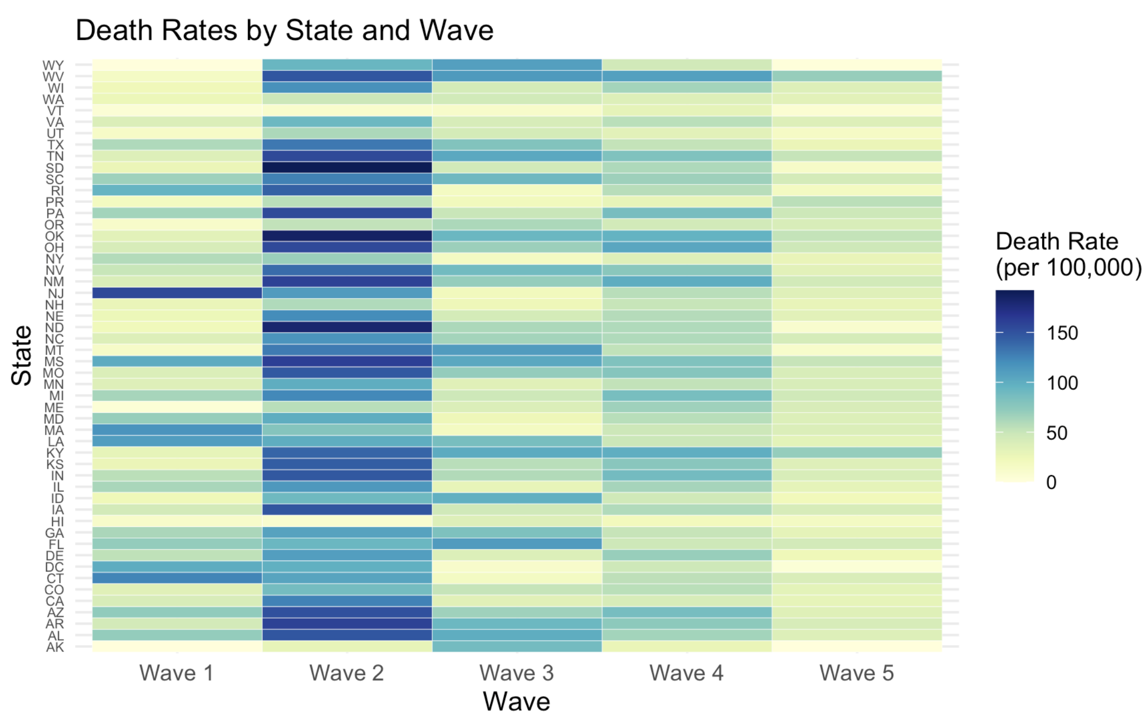

Figure 2. State-Level Death Rates by Wave Heatmap: This heatmap displays the death rates (per 100,000) for each state during the five waves. Darker shades represent higher death rates, highlighting disparities in healthcare access, public health measures, and vaccination coverage.

Evolution of COVID-19 Virulence and Impact of Vaccination

COVID-19 virulence, measured through CFR, showed a clear decline across waves (Figure 3). During **Wave 1**, the CFR was highest due to limited treatment options and overwhelmed healthcare systems. *Supplementary Figure 1* provides detailed trends for all states, showing how specific regions experienced sustained higher rates of deaths and hospitalisations during Wave 3. By **Wave 5**, the CFR dropped significantly, reflecting the combined effects of vaccination, improved healthcare capacity, and the emergence of less severe variants. Additionally, states with higher vaccination coverage had lower death rates across all waves (Figure 4), underscoring the critical role of vaccines in reducing disease severity. *Supplementary Figure 4* further provided evidence for vaccination's role by presenting rolling averages of deaths and cases, smoothing temporal trends to emphasise how vaccination coverage influenced outcomes during each wave.

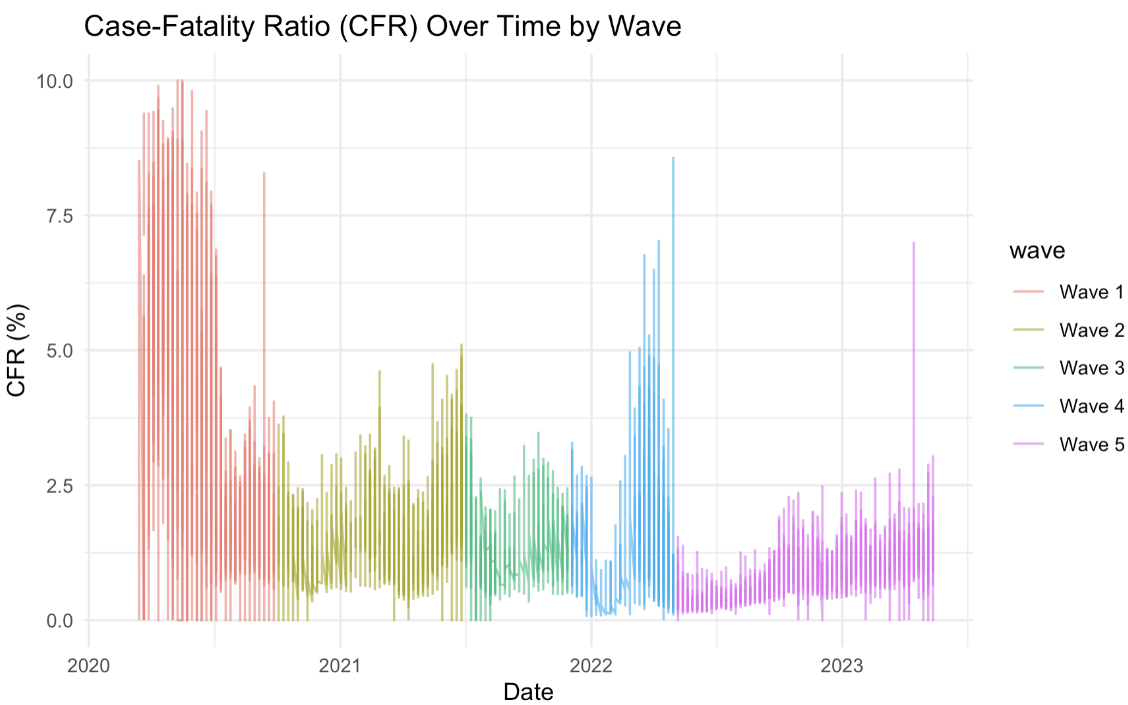

Figure 3. Case-Fatality Ratio (CFR) Over Time by Wave. A decreasing CFR implies that the disease became less virulent. Peaks in CFR during specific waves provide insight into the challenges faced in healthcare response at those times.

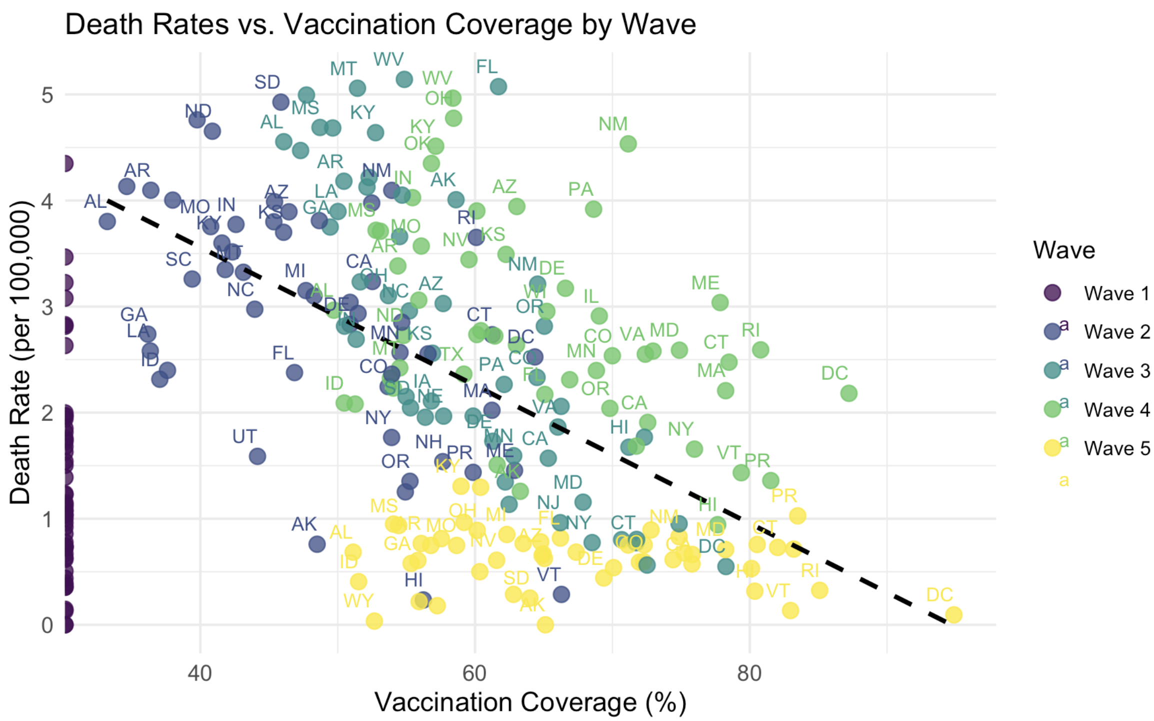

Figure 4. Vaccination Coverage vs. Death Rates by Wave. This scatter plot shows the relationship between vaccination coverage (%) and death rates (per 100,000) across states during different waves. A strong negative correlation underscores the critical role of vaccination in reducing mortality.

Discussion

Interpretation of Main Findings

The segmentation of the pandemic into five distinct waves revealed important trends in cases, hospitalisations, and deaths. Wave 1, marked by the initial outbreak, saw relatively high CFR due to limited preparedness, unavailability of vaccines, and overwhelmed healthcare systems (World Health Organization 2023), (Bedford et al. 2020). Later waves, particularly Waves 4 and 5 dominated by the Omicron variants, showed a significant reduction in CFR and HDR. This decline is likely tied to widespread vaccination campaigns and the availability of better treatment protocols (Polack et al. 2020). However, Wave 5's modest resurgence in hospitalisation rates highlighted the virus's ability to adapt and transmit under eased public health measures (Centers for Disease Control and Prevention 2023a), (Viana et al. 2022).

State-level variations in death rates revealed significant disparities that underscore the role of localised factors such as population density, healthcare infrastructure, and policy responses. States like **New York** and **New Jersey** experienced high death rates during Wave 1, likely due to population density, overwhelmed healthcare systems, and delayed public health interventions. In contrast, states with lower population densities or stricter early measures, such as **Vermont** and **Wyoming**, consistently experienced lower death rates throughout the pandemic. These findings reflect the interplay of socio-economic factors, healthcare infrastructure, and public health strategies (Hale et al. 2021).

The analysis of virulence trends through the CFR and HDR revealed an encouraging trend: **the virus became less lethal** over time. Early waves were characterised by high CFRs due to a lack of effective treatments and overburdened healthcare systems. With the rollout of vaccines during Wave 3 and beyond, the CFRs dropped significantly, demonstrating the efficacy of vaccination in preventing severe outcomes. Moreover, the negative correlation between vaccination coverage and death rates across states, even during the highly transmissible Omicron wave, further emphasised the critical role of vaccines in mitigating mortality.

Implications and Context

The results have several implications for public health policy and pandemic preparedness. The clear decline in virulence as vaccination rates increased reinforces the importance of achieving high vaccine coverage in the population. Future pandemic planning must prioritise equitable vaccine distribution to address disparities in healthcare access and outcomes, as evidenced by the variation in death rates among states.

The resurgence of cases and hospitalisations during Wave 5, despite high vaccination rates, highlights the need for ongoing surveillance and adaptive public health measures. Variants with increased transmissibility, even if less severe, can still strain healthcare systems, particularly in states with limited capacity. This underscores the importance of maintaining robust testing, contact tracing, and targeted non-pharmaceutical interventions (NPIs), especially in the face of emerging variants (Bo et al. 2021).

Limitations and Future Directions

While the study provides valuable insights, it is not without limitations. First, the reliance on state-level aggregate data may obscure intra-state disparities, such as those between rural and urban areas or among different demographic groups. Future studies could explore granular datasets to provide a more nuanced understanding of disparities in COVID-19 outcomes. Second, the analysis assumes data accuracy, yet underreporting of cases and deaths, particularly during early waves, remains a concern. Further investigation into the completeness and consistency of data across states is warranted.

Additionally, this study focused on quantifiable metrics like cases, hospitalisations, and deaths, but qualitative factors, such as public compliance with health measures and vaccine hesitancy, also play critical roles in shaping pandemic outcomes. Future research could incorporate behavioural and sociological data to complement these findings. Finally, while this analysis explored the impact of vaccination, it did not examine the effects of booster or waning immunity, both are critical to understanding long-term pandemic dynamics. (Levin et al. 2021).

Conclusion

This study highlights the complex interplay of viral evolution, public health interventions, and socio-economic factors in shaping the trajectory of the COVID-19 pandemic. The segmentation of waves provides a framework for understanding the dynamic nature of pandemics, while the analysis of state-level disparities and vaccination impacts underscores the importance of equitable healthcare policies. As the world transitions to the endemic phase of COVID-19, these findings offer valuable lessons for managing future pandemics. By prioritising vaccination, addressing healthcare inequities, and maintaining vigilance against emerging variants, public health systems can better prepare for and mitigate the impact of future global health crises.

References

Bar, K., et al. 2021. "Disparities in COVID-19 Vaccination Rates and the Impact on Health Outcomes." *Journal of Health Disparities Research* 15 (4): 289–305. https://example.com/article.

Bedford, Juliet, Delia Enria, Johan Giesecke, David L. Heymann, Chikwe Ihekweazu, Gary Kobinger, Helen C. Lane, et al. 2020. "COVID-19: Towards Controlling of a Pandemic." *Nature* 395: 1015–25. https://doi.org/10.1038/s41586-020-2008-3.

Bo, Yacong, Cui Guo, Changqing Lin, Yiqian Zeng, Hao Bi Li, Yumiao Zhang, Md Shakhaoat Hossain, et al. 2021. "Effectiveness of Non-Pharmaceutical Interventions on COVID-19 Transmission in 190 Countries from 23 January to 13 April 2020." *International Journal of Infectious Diseases* 102 (January): 247–53. https://doi.org/10.1016/j.ijid.2020.10.066.

Centers for Disease Control and Prevention. 2023a. "COVID Data Tracker." https://covid.cdc.gov/covid-data-tracker/#datatracker-home.

———. 2023b. "COVID-19 Timeline." https://www.cdc.gov/museum/timeline/covid19.html.

Gupta, A., Sharma, R., and Kumar, S. 2023. *COVID-19 and Its Impact on Society: Waves of the Pandemic*. Springer Nature. https://link.springer.com/chapter/10.1007/978-3-031-38941-2_19.

Hale, Thomas, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, et al. 2021. "A Global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker)." *Nature Human Behaviour* 5: 529–38. https://doi.org/10.1038/s41562-021-01079-8.

Kaiser Family Foundation. 2023. "Global COVID-19 Tracker and Maps." https://www.kff.org/coronavirus-covid-19/issue-brief/global-covid-19-tracker/.

Levin, Einav G., Yaniv Lustig, Carmit Cohen, Ronen Fluss, Victoria Indenbaum, Sharon Amit, Ram Doolman, et al. 2021. "Waning Immune Humoral Response to BNT162b2 Covid-19 Vaccine over 6 Months." *New England Journal of Medicine* 385 (24): e84. https://doi.org/10.1056/NEJMoa2114583.

Polack, Fernando P., Stephen J. Thomas, Nicholas Kitchin, Judith Absalon, Alejandra Gurtman, Stephen Lockhart, Kathryn Perez, et al. 2020. "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine." *New England Journal of Medicine* 383: 2603–15. https://doi.org/10.1056/NEJMoa2034577.

Roser, M., Ritchie, H., Ortiz-Ospina, E., and Hasell, J. 2023. "Coronavirus Pandemic (COVID-19) Deaths." https://ourworldindata.org/covid-deaths.

U.S. Census Bureau. 2021. "Population Estimates: 2021." https://api.census.gov/data/2021/pep/population.

Viana, Ricardo, Sikhulile Moyo, Daniel G. Amoako, Houriiyah Tegally, Cathrine Scheepers, Christian L. Althaus, Ulrich J. Anyaneji, et al. 2022. "Rapid Epidemic Expansion of the SARS-CoV-2 Omicron Variant in Southern Africa." *Nature* 603: 679–86. https://doi.org/10.1038/s41586-022-04411-y.

World Health Organization. 2023. "World Health Organization COVID-19 Dashboard." https://covid19.who.int/.