Final Project

COVID-19 Pandemic in the US

BY: Annie Song, Niloufar Saririan, Rithika Uppalapati

Abstract

Introduction: The final project investigated changes in the virulence of the COVID-19 virus during the pandemic. As a secondary analysis, the project examined whether excess mortality during the pandemic could be explained by COVID-19 mortality. We used publicly available data from the CDC and the U.S. Census. We analyzed the 50 states, Washington D.C., and Puerto Rico. Methods: For virulence, we divided the pandemic into seven waves and compared total cases, hospitalization, and mortality rates to identify periods of high and low virulence. For excess mortality, we predicted expected mortality using a linear model of all-cause mortality from 2014 to 2019 by state and week. We compared the difference in actual vs expected mortality to recorded COVID-19 deaths. Results: We found that virulence peaked during September 2020-February 2021 before decreasing. We found that deaths due to COVID-19 do not entirely explain the difference in actual vs predicted mortality. Discussion: Our analysis showed that COVID-19 virulence peaked during September 2020-February 2021 before decreasing, potentially related to viral evolution and public health adaptations by state. We also found that COVID-19 deaths do not represent the entirety of excess mortality, suggesting that the pandemic may have caused indirect deaths through other mechanisms.

Introduction

The Coronavirus Disease 2019 (COVID-19) has had a devastating impact on global health. To date, the virus has led to 776,973,432 cases and been responsible for 7,077,127 deaths worldwide, with 103,436,829 cases and 1,209,009 deaths in the U.S. [1]. Short-term symptoms of this infectious disease can include dry coughs, shortness of breath, loss of taste or smell, fatigue, and cold-like symptoms. While most people recover fully, some experience symptoms for months after the initial infection [2]. Although symptoms may vary, older individuals and those with underlying medical conditions are at higher risk of severe outcomes such as hospitalization, admission to the ICU, intubation/medical ventilation, or death [3]. After its emergence in late 2019, the World Health Organization (WHO) officially declared the COVID-19 outbreak a global pandemic on March 11, 2020 [4]. Since then, testing, quarantines, and the eventual development and distribution of vaccinations have been used to control the spread of the virus. The introduction of vaccines was critical in controlling the pandemic's impact by reducing the severity of cases and preventing further fatalities [5]. However, challenges remain, especially with the development of new variants [6]. Understanding the pandemic's evolution is valuable for several reasons. Estimating excess mortality can aid in contextualizing the virus's true impact. Highlighting the number of deaths that occurred outside of expected historical trends allows for a greater understanding of the consequences of the pandemic. Additionally, tracking the waves of the virus provides insights into the effectiveness of vaccination and the changing virulence of COVID-19. Knowledge of whether COVID-19 has a distinct seasonal wave can increase preparedness for future outbreaks, helping to refine public health responses during peak periods. Analyzing state-level death rates and excess mortality can reveal gaps and disparities in pandemic response and health outcomes. Identifying problems in the COVID-19 response can help lead to a better understanding of where improvements are needed and lead to clearer public health strategies for future pandemics. This analysis aims to examine the pandemic in the U.S. over specific waves of time, compute death rates and excess mortality by state, and investigate changes in COVID-19 virulence over time. We expect that while early waves, especially in 2020, saw higher mortality due to limited medical knowledge and lack of vaccines, subsequent waves would show a decline in death rates despite increased case numbers, due to improved treatments, widespread vaccination, and better public health measures. We also expect that new variants may cause spikes in fatality, but will likely be short-lived, as healthcare systems should be better equipped to manage the new cases. Through the use of statistical analysis and data visualization, we will identify the patterns that exist across geographic locations and periods, helping to reveal how the virus evolved and how different states were impacted by it. By examining these trends, this project intends to provide a clearer understanding of how the COVID-19 virus evolved and offer insights that can aid in improving public health planning and response in future pandemics.

Methods

Population data was obtained from the U.S. Census website for each state from 2020-2023. The U.S. population census data for 2024 was not available at this time, so it was extrapolated from 2023 population data. The national population estimate was published by the U.S. Census for 2024. The percentage of each state's population of the national population on July 1, 2023 was used to determine the proportion of each state's population in the national population. This proportion was then multiplied by the national estimate of the 2024 U.S. population from the U.S. Census to obtain population estimates by state for 2024. Since population data for 2024 was extrapolated, the analysis is limited in its accuracy for 2024. The contribution of each state to the U.S. population was assumed to be the same in 2024 as it was in 2023 on July 1.

The population data was cleaned and wrangled to create a year variable and corresponding population data. Puerto Rico (PR) and the District of Columbia (DC) were also added as their own states. COVID-19 data was pulled from the CDC website on December 17, 2024. Data on cases, hospitalizations, and deaths from COVID-19 were pulled from the CDC and used for this analysis. These data sets were also cleaned and wrangled to add Puerto Rico and the District of Columbia as states. Variables for weeks and years were created to follow the Morbidity and Mortality Weekly Report (MMWR) dates. Variables were also transformed into their correct format of numeric, date, and character as necessary. A dates data set was created consisting of dates from January 1, 2020, to December 1, 2024, and were set to their corresponding MMWR weeks and years. The population, dates, hospitalizations, cases, and deaths datasets were joined together to create a data set that was ready to use for visualization. Waves of the pandemic were determined through data visualization. Plots for hospitalization rates, case rates, and death rates per state were created through the pandemic. Based on this plot, a COVID wave is determined to be when there is a peak in both deaths and cases. It was therefore decided that the waves are January 2020-May 2020, June 2020-August 2020, September 2020-February 2021, March 2021-June 2021, July 2021-October 2021, November 2021-March 2022, and April 2022-December 2024. These waves were encoded into the dataset as waves 1-7, respectively. To determine the states' death rates by wave, the data sets were grouped by wave and state. Death rates were determined from summed deaths over the wave for each state divided by the maximum population of the state during that wave and then multiplied by 100,000 to get the death rate per 100,000 people. The same method was applied to determine hospitalization and case rates for each state during each wave. The ten states with the highest death rates and lowest death rates during each wave of the pandemic were selected to make observations about death rates during the pandemic. These were plotted as bar plots with state on the x-axis, death rate per 100,000 people on the y-axis, and colored by state. To learn more about virulence, plots were created for hospitalization, case, and death rates. The rate per 100,000 people was on the y-axis for each respective measure, and the wave was on the x-axis. Points were used to show the corresponding rate by state on the plot and were also colored by state.

Excess mortality from COVID-19 was determined through linear modeling. First, we built a linear model of weekly, state-level all-cause mortality using CDC mortality data and Census population estimates from 2014 through 2019. The fitted mortality rates were used to predict expected weekly all-cause deaths from 2020 through 2024 using state population estimates. We calculated excess deaths during the pandemic by subtracting expected deaths from actual deaths. We then subtracted deaths from COVID-19 from the excess death amounts to identify excess non-COVID-19 mortality. We visualized the trend in excess non-COVID-19 mortality using a GAM smooth from the R package ggplot. Analysis of excess mortality rates was done using the same method outlined above for death rates.

Results

Table 1 shows the ten states with the highest death rates, by wave. The states with the highest death rates in wave 1 are CT, DC, DE, LA, MA, MI, NJ, NY, PA, and RI. As the pandemic progresses, these states drop in and out of the top death rate states. CT, DC, DE, MA, NY, and RI do not make it back into the states with the top 10 death rates for any subsequent wave of the pandemic. NJ does not appear in the top ten death rate states in wave 2 or 3, but it does reappear in wave 4 before disappearing for the rest of the waves. MI and PA are in the top 10 states with the highest death rates from COVID in waves 1, 4, 6, and 7. LA is in the top 10 states for waves 1 and 2. The states with highest death rates in the last/ongoing wave of the pandemic, wave 7, are FL, KY, MI, MS, OH, OK, PA, PR, TN, and WV.

Table 1

Table 1: Top Ten States with Highest Death Rates

rank	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7
1	NJ	MS	SD	MI	FL	WV	KY
2	CT	AZ	ND	NJ	MS	NM	WV
3	MA	FL	OK	PA	AL	OH	MS
4	DC	TX	AR	FL	TN	KY	PR
5	RI	SC	MS	TN	KY	AZ	OK
6	LA	AL	NM	WV	WV	MI	TN
7	MI	LA	IA	MD	AR	OK	PA
8	NY	GA	TN	MS	MT	PA	OH
9	PA	NV	OH	KY	WY	IN	FL
10	DE	AR	KS	SC	SC	TN	MI

The states with the lowest death rates in wave 1 of the pandemic are seen in Table 2 and are AR, ID, OR, PR, UT, and WV. AK, HI, MT, and WY have death rates of 0, but this might be because these states were not collecting death data yet. The states with the lowest death rates in the last/ongoing wave of the pandemic, wave 7, are DC, HI, MT, ND, RI, SD, UT, VT, WY.

Table 2

Table 2: Top Ten States with Lowest Death Rates

rank	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7
1	AK	AK	VT	VT	VT	HI	AK
2	HI	VT	HI	WY	RI	PR	WY
3	MT	WY	AK	AK	DC	VT	DC
4	WY	ND	WA	NH	NY	NY	ND
5	ID	HI	OR	SD	CT	UT	VT
6	UT	NY	ME	ND	MA	WA	SD
7	WV	MT	PR	UT	NH	AK	MT
8	PR	WV	NY	MT	NJ	CA	UT
9	OR	OR	UT	NE	MD	DC	RI
10	AR	MI	NH	WI	PR	FL	HI

Figure 1 demonstrates the change in hospitalization, death, and case rates from COVID-19 throughout the 7 waves of the pandemic. We see that wave 2 had the lowest rates for all three measures, and wave three had the highest rates for all three measures. All three rates decreased in wave 4, but then increased again in wave 5. Case and death rates increased for wave 6 and then decreased for wave 7. Hospitalizations decreased for wave 6 and then remained relatively constant for wave 7.

Figure 1

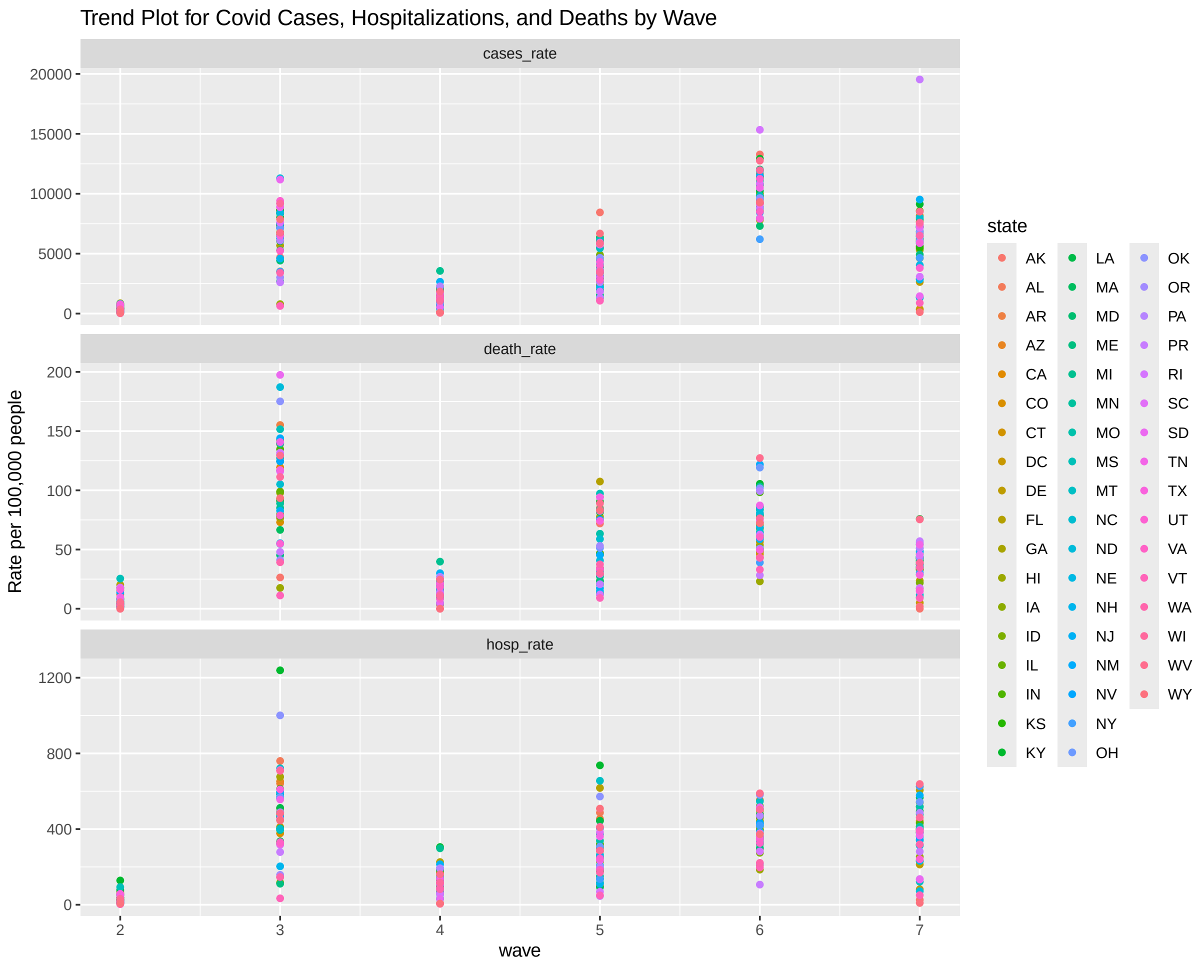


Figure 2 shows how many states had total actual deaths that were higher than the total expected deaths over the course of the pandemic, in all of the seven waves we identified. However, even after removing COVID-19 deaths from this difference, there is still a notable trend in excess non-COVID-19 mortality throughout the pandemic.

Figure 2

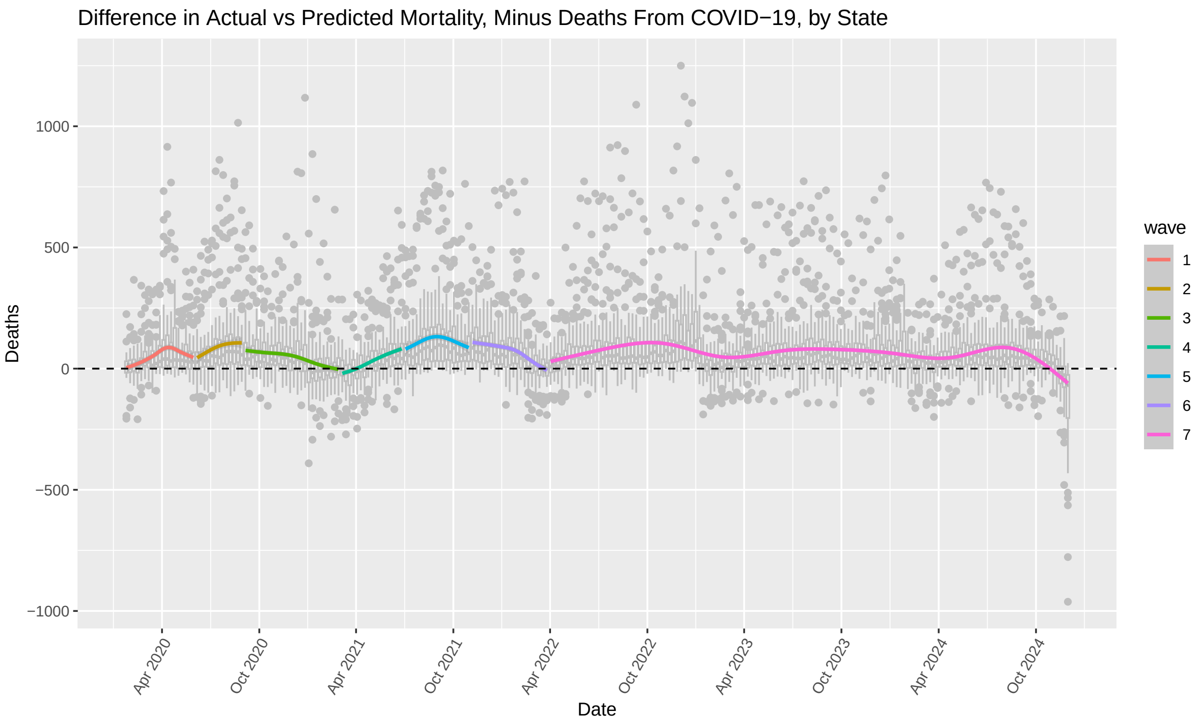


Figure 3 shows that the states with the highest excess mortality rates are CT, DC, DE, IL, LA, MA, MD, MI, NJ, and RI. These 10 states align with those found in the analysis of death rate with the only differences being IL and MD instead of NY and PA. As the pandemic progresses CT, DE, IL, MD, MA, NJ, and RI drop out of the top 10 for all subsequent waves. DC remains in wave 4, LA remains in waves 2, 4, and 5, and MI remains in waves 4 and 6. Wave 7, with the highest mortality rate, includes AK, AR, AZ, ME, MS, NM, OR, PR, WV, and WY.

Figure 3

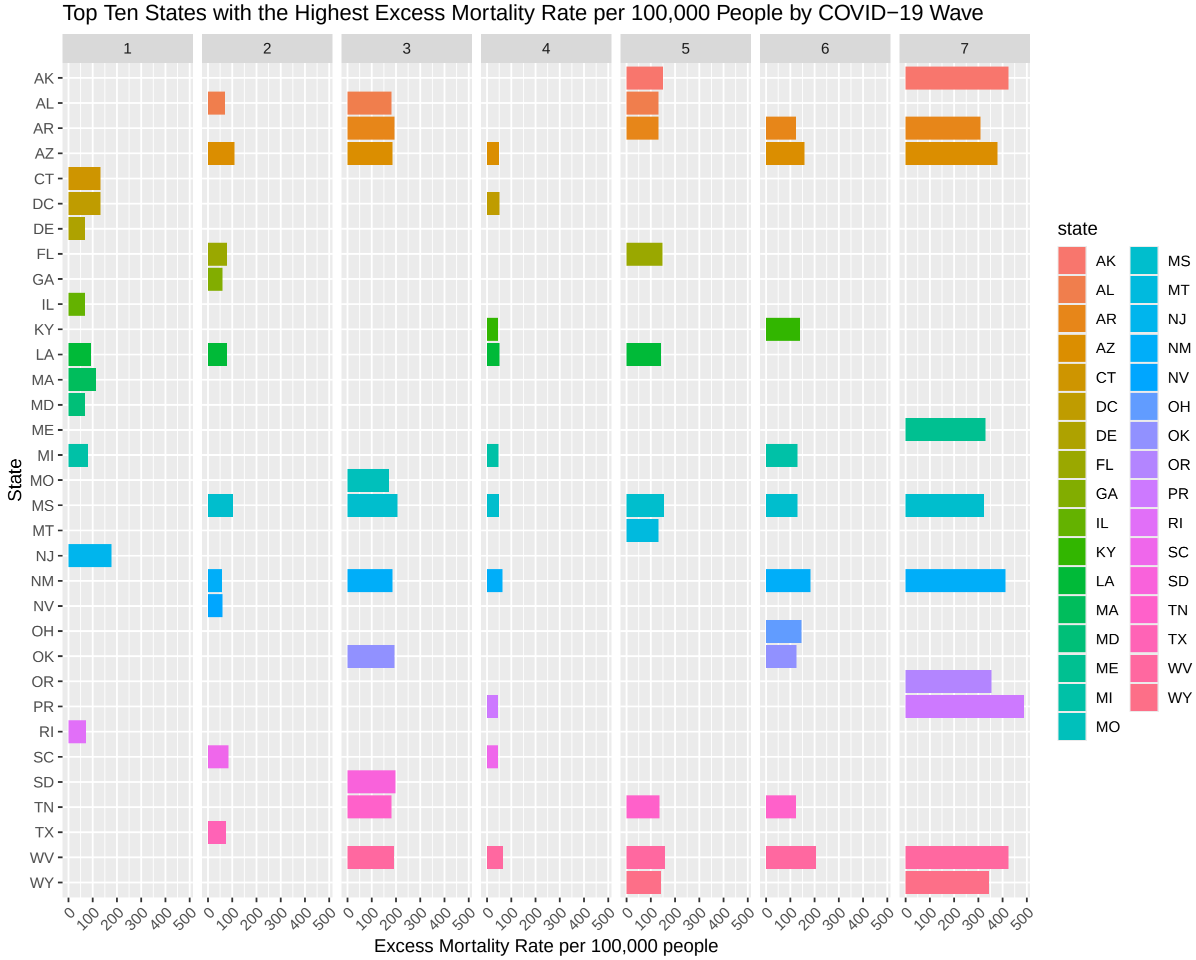
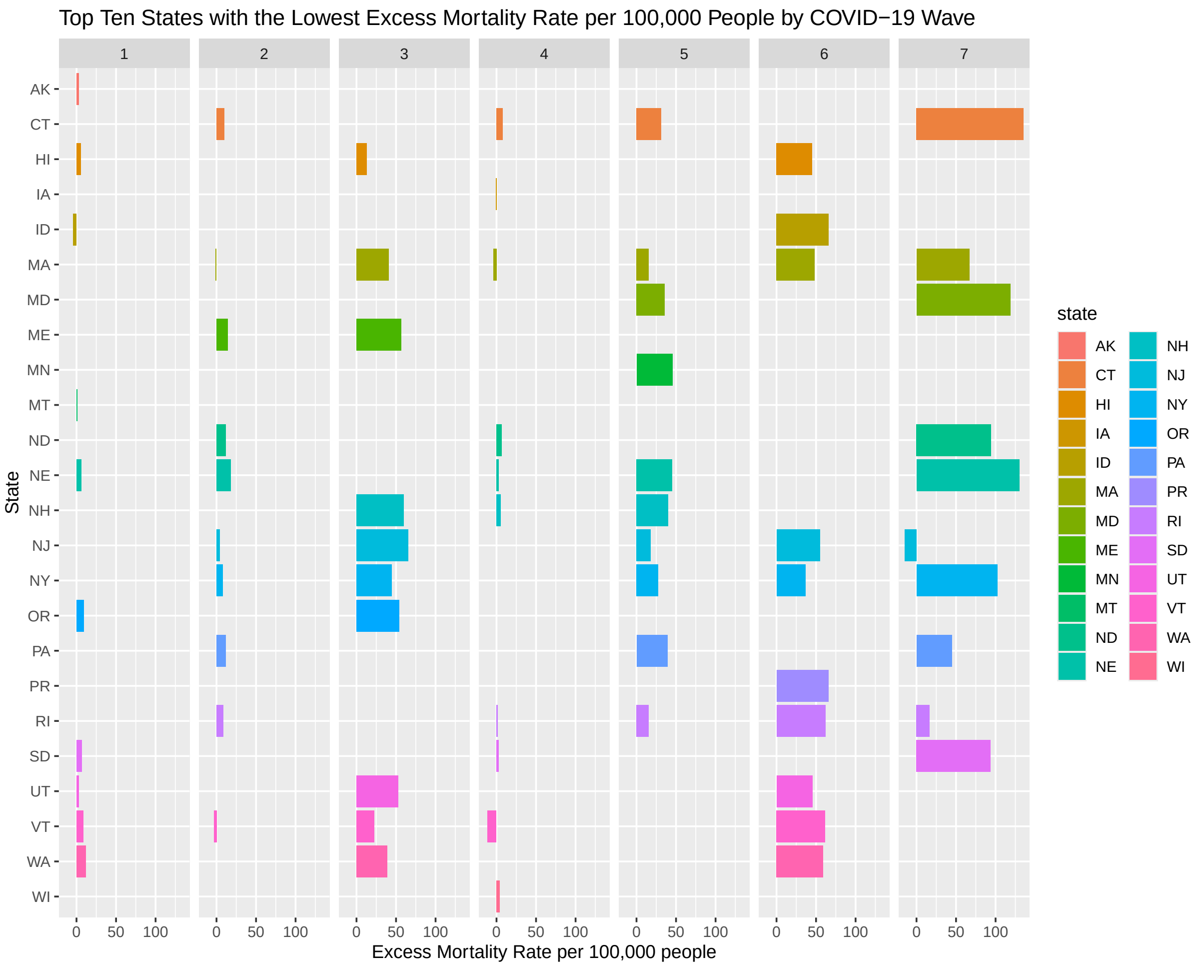


Figure 4 shows that the states with the lowest death rates in the first wave of the pandemic are AK, HI, ID, MT, NE, OR, SD, UT, VT, and WA. All states except for AK and MT remain in at least one subsequent wave, with NE and VT remaining in 4 subsequent waves. The states with the lowest excess mortality rates in the last/ongoing wave of the pandemic (wave 7) are CT, MA, MD, ND, NE, NJ, NY, PA, RI, and SD.

Figure 4



Discussion

The analysis of death rates across the U.S. reveals interesting patterns, particularly with states fluctuating between high and low death rates across different pandemic waves. For example, Rhode Island is among the states with highest death rates in wave 1, while also appearing among those with lowest death rates in waves 5 and 7. Also interesting is that Florida appears in the highest death rates for waves 2, 4, 5, and 7, but has one of the lowest death rates during wave 6. COVID-19 deaths alone do not explain the excess mortality during the pandemic. Based on our linear modeling for expected mortality, we found that many states had total deaths in excess of expected deaths for that time period, absent COVID-19. However, even when considering deaths from COVID-19, we found there is still a notable increase in actual deaths above expected, which remains unexplained. It is likely these deaths were driven by indirect factors of the pandemic. For example, lack of healthcare access, stress, and physical deterioration all could have contributed to an increase in deaths that would have been recorded as non-COVID-19 mortality. Interestingly, in wave 3, excess mortality was notably high in midwestern and southern states, while many northeastern states had low excess mortality in wave 7. Additionally, some states, such as Connecticut, exhibited both a high excess mortality rate in wave 1 and a low excess mortality rate in subsequent waves. These findings emphasize the importance of examining both the successes of states that showed improvement and the challenges faced by states with persistently high death/excess mortality rates, since there are valuable lessons to be learned from both. States that adapted to the virus could serve as models for pandemic preparedness, while those that struggled can provide insight into potential difficulties such as poor healthcare infrastructure or delayed interventions.

The CDC defines virulence as "the proportion of persons with clinical disease who, after becoming infected, become severely ill or die" [7]. From this definition, to see if virulence has increased, we are interested in an increase in cases, hospitalizations, and/or deaths. We see that from wave 2 to wave 3, COVID-19 became more virulent according to this definition as there is an increase in cases, hospitalizations, and deaths across states. These measures decrease in wave 4 but increase in wave 5, and then increase slightly in wave 6, besides hospitalization, indicating an increase in virulence. Virulence then decreases in the transition to wave 7, as case and death rates decrease and hospitalizations stay constant. COVID-19 appeared the most virulent during wave 3 of the pandemic. The findings emphasize the need to analyze not only death rates but also broader trends of the pandemic to gain insight into potential seasonal fluctuations of the virus, the effectiveness of preventative measures/vaccines, and the emergence of new variants. One limitation of this project is that the data may not fully capture indirect consequences of the pandemic, such as disturbances to mental health, economic crises, or an overburdened healthcare system, which all likely contributed to excess mortality. Future research could investigate these factors by looking into the extent of their contribution to excess mortality. Specifically, researching socioeconomic factors, such as income inequality and access to healthcare could provide a better comprehensive understanding of mortality rate trends. Our project is also limited by its geographic scope. In addition to the lack of generalizability to other countries outside the U.S., our analysis is confined to the state level. Future research could conduct similar studies at the county or zip-code level, since there is often significant variation in COVID-19 impacts within states. The results of this project have important implications for public health policy, especially for pandemic preparedness and response strategies. States that experienced high excess mortality rates early in the pandemic, such as New Jersey and Massachusetts, may have developed better public health systems for pandemic management over time. These types of states can serve as valuable models for future pandemic responses. For states with consistently high mortality rates such as Michigan and Florida, policymakers may need to focus on identifying and addressing potential underlying health disparities and vulnerabilities in healthcare systems for future responses. Understanding the fluctuations of death and excess mortality rates is essential for improving pandemic preparedness and developing strategies that minimize the negative outcomes of future public health crises.

References

1. Cumulative confirmed covid-19 cases and deaths. Our World in Data. (n.d.). https://ourworldindata.org/grapher/cumulative-deaths-and-cases-covid-19?tab=table
2. Mayo Foundation for Medical Education and Research. (2024, September 20). Coronavirus disease 2019 (covid-19). Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/coronavirus/symptoms-causes/syc-20479963
3. Centers for Disease Control and Prevention. (2024, July 30). Underlying conditions and the higher risk for severe COVID-19. Centers for Disease Control and Prevention. https://www.cdc.gov/covid/hcp/clinical-care/underlying-conditions.html
4. Cucinotta, D., & Vanelli, M. (2020). WHO Declares COVID-19 a Pandemic. Acta Biomedica Atenei Parmensis, 91(1), 157-160. https://doi.org/10.23750/abm.v91i1.9397
5. Centers for Disease Control and Prevention. (2024, March 1). Respiratory virus guidance. Centers for Disease Control and Prevention. https://www.cdc.gov/respiratory-viruses/guidance/index.html
6. Zhao, Y., Huang, J., Zhang, L., Chen, S., Gao, J., & Jiao, H. (2022). The global transmission of new coronavirus variants. Environmental research, 206, 112240. https://doi.org/10.1016/j.envres.2021.112240
7. Centers for Disease Control and Prevention. (2024b, May 15). Epidemiology glossary. Centers for Disease Control and Prevention. https://www.cdc.gov/reproductive-health/glossary/index.html