Final Report

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom

Attaching package: 'janitor'


The following objects are masked from 'package:stats':

    chisq.test, fisher.test



Attaching package: 'jsonlite'


The following object is masked from 'package:purrr':

    flatten



Attaching package: 'gridExtra'


The following object is masked from 'package:dplyr':
```

```
    combine


Attaching package: 'cowplot'


The following object is masked from 'package:patchwork':

    align_plots


The following object is masked from 'package:lubridate':

    stamp
```

Abstract

Coronavirus disease (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and its global spread has caused large number of deaths worldwide. Given the high transmissibility and virulence of SARS-CoV-2, it is crucial to understand its spread pattern from an epidemiological perspective. In this study, I analyzed the morbidity and mortality of COVID-19 in the United States from 2020 to 2023 based on the publicly available data provided by Center for Disease Control and Prevention (CDC). This investigation offers a way to examine the spread pattern of COVID-19 and how it differs between each state in the United States. Through this analysis, we found that COVID-19 pandemic can be divided into three waves, with each characterized by peaks in new cases and deaths. Comparing the data for each state, we found that.

Introduction

Coronavirus disease (COVID-19) is a contagious disease that has caused large number of deaths in the world over the past few years. It is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a pathogenic strain of coronaviruses characterized by the spike proteins projected from the viral surface. This virus infects the upper respiratory tract and causes symptoms such as fever, sore throat, and shortness of breath. SARS-CoV-2 is transmitted through through respiratory droplets and aerosols, which contributes to its high transmissibility. COVID-19 is capable of rapidly spreading among people, and due to this reason it has become a devastating global pandemic since 2019. In the United States, COVID-19 has caused millions of deaths in total over the past few years. It has caused significant public health burden and widespread social and economic disruptions. Hence, it has become increasingly urgent for public health professionals and governmental agencies to develop effective strategies for the prevention and treatment of COVID-19.

Given the widespread of COVID-19, it is crucial to understand its spread pattern through epidemiological analysis. Since the first outbreak in 2019, SARS-CoV-2 has undergone continuous evolution, leading to the emergence of various variants with distinct virulence and microbiological characteristics, such as Alpha, Beta, Delta, and Omicron. Their differing characteristics have contributed to the constantly shifting epidemiological trends. In the meantime, to address this global pandemic, scientists and pharmaceutical companies have developed vaccines and medications to prevent widespread transmission and alleviate the symptoms of infected patients. This has led to the decline in new COVID-19 positive cases as well as the mitigation of symptoms and improved prognosis in severely-infected patients. The competition between pathogenic coronavirus and human innovations in viral research and therapeutic development has resulted in constantly changing epidemiological trends of COVID-19, which is the focus of this study.

In this study, to better understand the epidemiological trends of COVID-19, I analyzed the COVID-19 data in the United States from 2020 to 2023 to examine the morbidity and mortality patterns across different time periods. The morbidity pattern was investigated using the weekly data on new clinically confirmed COVID-19 cases for each state, which provided insights into the weekly fluctuations in COVID-19 spread and helped identify potential trends in disease transmission over time. The mortality pattern was studied using the weekly number of deaths caused by COVID-19 for each state, and this official data could reflect the virulence of the virus and its impact on public health over time. This longitudinal analysis of morbidity and mortality allowed me to observe the epidemiological waves of COVID-19 in the United States and identify the time periods during which COVID-19 was more prevalent and rampant. In this way, we could gain a better understanding of the transmission and fatality of COVID-19 over time.

In addition to the examination of general trends, I also explored the COVID-19 disease pattern for each state in the United States. The goal was to compare the state-level data and assess the public health performance of each state. In fact, different states have proposed different policies for managing the pandemic, such as social distancing, mask mandates, and testing protocols. State governments also implemented different measures to manage healthcare institutions, coordinate resources, and ensure the availability of medical supplies and personnel to handle the demands of the pandemic effectively. This has resulted in the difference in each state's performance in disease prevention and treatment during the pandemic period. To evaluate each state's response to COVID-19 pandemic, I compared the average number of COVID-19 cases per day and mortality rate for each state over the course of each pandemic wave. This analysis could inform us which states responded to COVID-19 most effectively and which states were most severely impacted by the pandemic.

Method

Data for COVID-19 Analysis

In this study, two major types of data were used for COVID-19 analysis: population data and COVID-19 disease data. The population data was obtained from the official dataset ‘’Annual Estimates of the Resident Population for the United States, Regions, States, District of Columbia and Puerto Rico: April 1, 2020 to July 1, 2023 (NST-EST2023-POP)’’. This is a publicly available dataset released by United States Census Beureu. Each year, the U.S. Census Bureau’s Population Estimates Program computes and releases the estimates of the population for the nation as well as for each state, and this estimated data is produced based on the cohort-component method that combines the data from the last decennial census as well as the data related to births, deaths, and migration. The 2023 dataset used in this study is obtained from the calculations that integrate the 2020 Census and the recent data for births, deaths, and migration. In this study, the population of each state, for each vintage year, was extracted from the 2023 dataset, and this population data was used for the calculation of morbidity and mortality rates over the COVID-19 pandemic periods.

In addition to the U.S. population data, COVID-19 cases and deaths statistics from Center for Disease Control and Prevention (CDC) were used for analysis. COVID-19 cases were obtained from the dataset ‘’Weekly United States COVID-19 Cases and Deaths by State‘‘, which includes the number of new COVID-19 cases per week for each state from January 16, 2020 to May 10, 2023. On May 11, 2023, the U.S. government ended the public health emergency for COVID-19, and since then CDC has stopped collecting COVID-19 cases data. COVID-19 mortality data was obtained from the dataset‘‘Provisional COVID-19 Death Counts by Week Ending Date and State’’, which reports the weekly number of deaths caused by COVID-19 for each state in the United States. The reported data related to COVID-19 from these two datasets were combined to produce a comprehensive data frame that lists the number of new COVID-19 cases and deaths per state per week.

Identification of COVID-19 Pandemic Waves

COVID-19 morbidity and mortality rates were computed from the raw data mentioned in the previous section. Morbidity rate was measured as new COVID-19 cases per 100000 people, and this rate was computed for each state over the entire 2020 to 2023 pandemic period during which the data were available. More specifically, for each state, this was calculated by dividing the weekly new confirmed COVID-19 cases by state population, multiplied by 100000. Similarly, mortality rate was measured as weekly deaths caused by COVID-19, and it was computed by dividing the weekly deaths by state population, multiplied by 100000. Morbidity and mortality rates were then separately plotted against date for each state to visualize the epidemiological patterns of COVID-19 from 2020 to 2023. This allowed for the visualization of both the general pattern of COVID-19 and the specific trends for each state.

Comparison of State-level Mortality Rates

The mortality rate for each pandemic wave was then determined for each state in the United States. In this case, for each pandemic wave, the average value of deaths over the entire pandemic period was first calculated, and this average value was multiplied by the number of weeks recorded. This calculation aimed to find out the total deaths for each state over the whole period of each pandemic wave. Next, this total death value was divided by the state population, and the result was the number of deaths per day. In this calculation, the state population was selected to be the first recorded value of the population. Finally, the result was multiplied by 100000, which produced the mortality rate per 100000 people used for data visualization. For each pandemic wave previously identified, a bar plot was constructed to compare the mortality rate for each state. This comparison made it possible to determine which states responded more effectively to the COVID-19 pandemic and which states were more severely impacted by the disease.

Temporal trends in the virulence of COVID-19

The virulence of SARS-CoV-2 can be assessed by computing the mortality-to-morbidity ratio, which reflects the proportion of clinically confirmed COVID-19 positive patients who died because of the severe symptoms caused by this virus. A large mortality-to-morbidity ratio means the disease is more fatal, thereby leading to large percentage of deaths among infected patients. In this study, the mortality-to-morbidity ratio was computed by dividing the total number of deaths over each COVID-19 pandemic wave by the total number of COVID-19 cases reported within the same time period. For data visualization purpose, the mortality-to-morbidity ratio was calculated for each region, which integrates the data from all the states within each region. The analysis of virulence can then be performed by comparing the mortality-to-morbidity ratio for each COVID-19 pandemic wave.

Results

General trends in COVID-19 morbidity and mortality

The general trends in morbidity and mortality of COVID-19 were analyzed, and the line graphs are shown below in Figure 1. The first plot displays the morbidity rate for each state, defined as the number of COVID-19 cases per 100,000 people, from January 16, 2020, to May 10, 2023. The second plot shows the mortality rate for each state, defined as the number of COVID-19 deaths per 100,000 people, from 2020 to 2023. For the purpose of data visualization, the lines are colored according to the regions defined by the United States Public Health Service to which they belong. Figure 1 shows the general pattern of COVID-19 transmission and mortality over the past few years.

According to this graph, three pandemic waves can be identified based on the peaks in morbidity and mortality rates. The first pandemic wave, spanning from April 4, 2020, to June 20, 2020, is characterized by a slight peak in detected COVID-19 cases accompanied by a significant peak in mortality. For some states in the United States, such as New Jersy, Connecticut, and Massachusetts, large number of deaths caused by COVID-19 were reported. After several months, the United States experienced the second pandemic wave, which lasted from September 12, 2020 to March 20, 2021. This time period is characterized by a relatively high peak in the morbidity rate and a notable peak in the mortality rate. Then, in 2021, the third pandemic wave began, lasting from July 3, 2021 to April 23, 2022. This relatively long time period is characterized by a remarkably high peak in the number of new cases reported and a slightly lower peak in mortality rate.

```
Warning: Removed 128 rows containing missing values or values outside the scale range
(`geom_line()`).
```

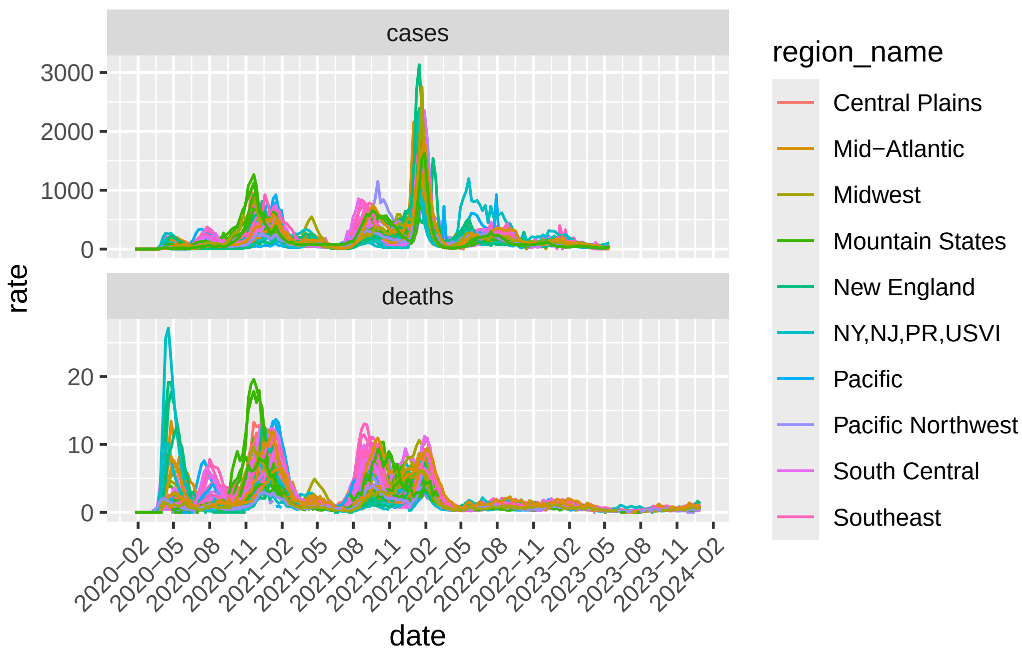
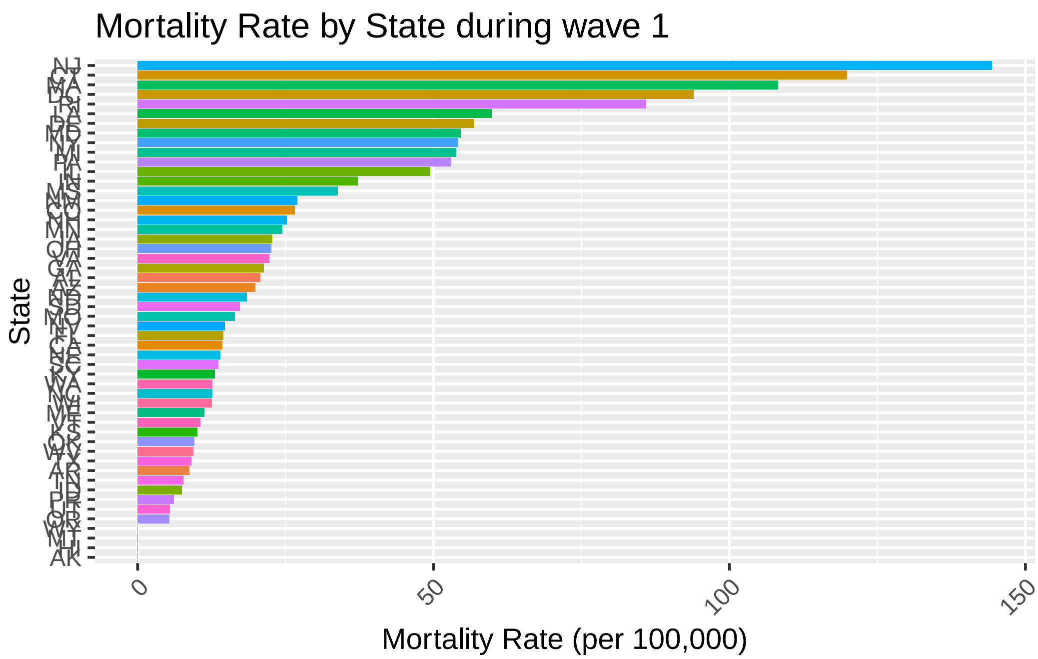
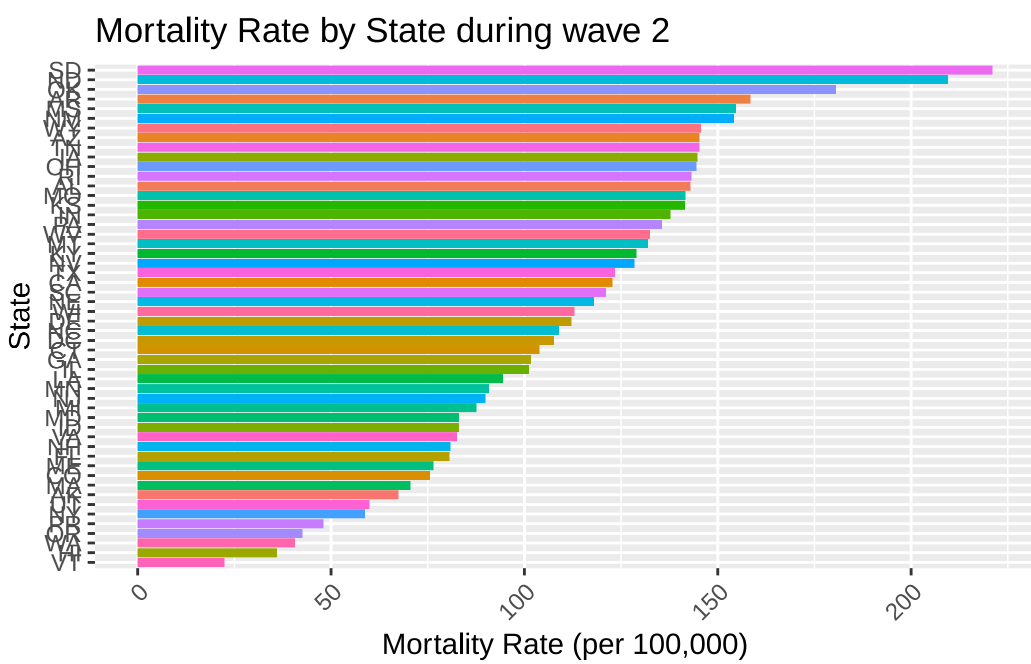


Figure 1. General trends in morbidity and mortality of COVID-19 in the United States from 2020 to 2023

Comparison of Mortality Rate Among States in the U.S.

State-level mortality data were then analyzed to evaluate each state's response to the COVID-19 pandemic. Figure 2 below shows the mortality rate of each state during the three COVID-19 pandemic waves identified previously. According to this figure, there are huge differences in mortality rates among these states in the United States. During the first pandemic wave, New Jersy, Connecticut, and Massachusetts were severely impacted by the pandemic, which is reflected by their high mortality rates. In comparison, Oregon, Utah, and Puerto Rico had the lowest mortality rates among all the states listed. Due to the missingness of reported data, the mortality rates of Wyoming, Montana, Alaska, and Hawaii are not shown. After a few months, during the second pandemic wave, South Dakota, North Dakota, and Oklahoma had the highest morality rates. In comparison, Vermont, Hawaii, and Washington had the lowest mortality rates. During the third pandemic wave, West Virginia, Kentucky, and Wyoming ranked as the top three states with the highest mortality rates. New York, Massachusetts, and Puerto Rico had the lowest mortality rates compared with other states.





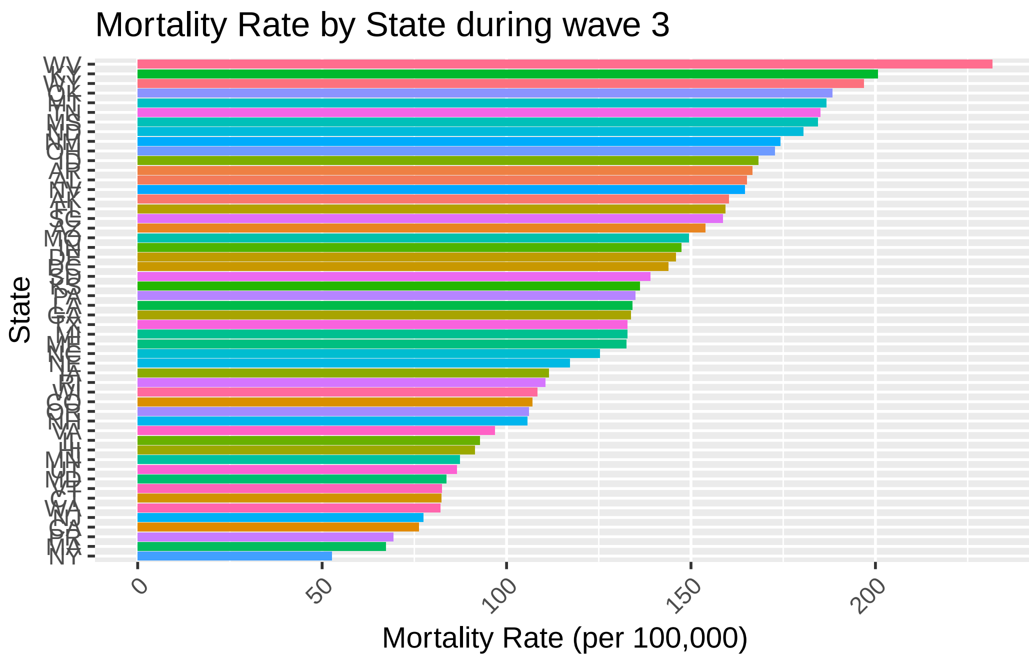


Figure 2. Mortality rate of each state during pandemic waves

Analysis of the Virulence of SARS-CoV-2 Over Time

The virulence of SARS-CoV-2 was examined by computing the mortarlity-to-morbidity ratio for each state over time. The goal was to explore whether SARS-CoV-2 has become more or less pathogenic and fatal from 2020 to 2023. Figure 3 below displays the mortality-to-morbidity ratio of COVID-19 for ten geographical regions over the three pandemic wave periods identified. This figure shows the decreasing trend in morbidity-mortality-ratio for all ten geographical regions from 2020 to 2023. During the first pandemic wave, COVID-19 caused remarkably high mortality-to-morbidity ratio, with the highest value being 0.08 for New England region. As time progressed, the mortality-to-morbidity ratio of COVID-19 significantly decreased to less than half during the second pandemic wave and continued to decline to an even lower value during the third pandemic wave.

SARS-CoV-2 gradually became less virulent

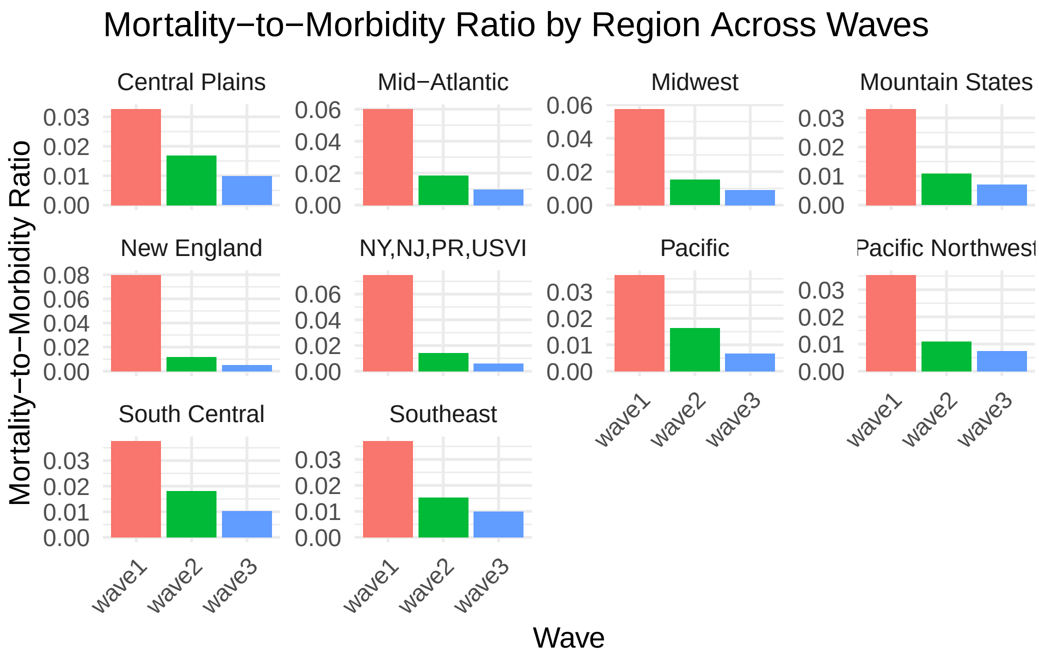


Figure 3. Comparison of the COVID-19 morbidity-to-mortality ratio across geographical regions over the three pandemic wave periods.

Discussion

In this study, the prevalence and mortality of COVID-19 in the United States from 2020 to 2023 was analyzed using the official data from U.S. Census and CDC. From 2020 to 2023, three pandemic waves were identified based on the peaks in new cases detected and deaths reported. The first pandemic wave lasted from April 4, 2020 to June 20, 2020, and this time period was characterized by the substantially high mortality rate. In fact, this pandemic wave corresponds to the emergence and spread of the initial alpha and beta variants of SARS-CoV-2. During this outbreak, SARS-CoV-2 was a newly discovered pathogenic virus, and healthcare institutions and clinical facilities were unprepared to respond to rapid transmission of COVID-19. After a few months, the whole country experienced the second pandemic wave, which lasted from September 12, 2020 to March 20, 2021. This pandemic wave was characterized by the relatively high morbidity and mortality rates as shown in the previous section. This time period corresponds to the discovery of the Delta variant. Compared with the previous Alpha and Beta variants, Delta variant was more contagious and lethal. The high transmissibility of the Delta variant is demonstrated by the larger number of new cases reported during this time period. Despite this fact, lower mortality rates were observed compared to the first outbreak. This could be explained by the development of new vaccines that could effectively reducing the severity of illness. Then, the third pandemic wave began on July 3, 2021 and lasted until April 23, 2022, which can be seen from the significantly high number of cases reported. However, this time period had relatively low mortality rates. These features could be explained by the emergence of the Omicron strain, a significantly more transmissible but less fatal variant of SARS-CoV-2.

Then, the comparison of the mortality-to-morbidity ratio across the three pandemic waves suggests a trend in the virulence of SARS-CoV-2 over time. A general decline in the mortality-to-morbidity ratio was observed across all geographical regions in the United States, suggesting a gradual decrease in the virulence of SARS-CoV-2 over the past few years. This observation is in accordance with the emergence of different SARS-CoV-2 strains discovered, as described above. From a biological perspective, viruses evolve to become more contagious and less virulent, allowing them to infect a larger number of hosts while sustaining the host-pathogen relationship to acquire necessary nutrients for survival. This could possibly explain the decreasing virulence from Alpha and Delta to the more recent Omicron variant. In addition to the features of the virus, the decreasing mortality-to-morbidity ratio could also be explained by the development of new vaccines and therapies, which could effectively prevent the viral spread and improve treatment outcomes for infected patients.

In addition to the general pandemic patterns of the whole country, the mortality rate data for each state in the United States were then analyzed. During the initial COVID-19 outbreak, New Jersy, Connecticut, and Massachusetts were relatively unprepared to respond to the first COVID-19 outbreak. Oregon, Utah, and Puerto Rico had the lowest mortality rates during the same period. In fact, most states on the eastern coast are characterized by a higher concentration of large metropolitan areas, which results in a higher population density. This might contribute to increased disease transmission and higher mortality rates compared to states with predominantly rural areas and smaller populations. During the second outbreak, South Dakota, North Dakota, and Oklahoma had the highest mortality rates while Vermont, Hawaii, and Washington had the lowest mortality rates. During the third outbreak, West Virginia, Kentucky, and Wyoming had the highest mortality rates whereas New York, Massachusetts, and Puerto Rico had the lowest mortality rates. Comparing the results across all three pandemic waves, we observed that most states responded differently to these COVID-19 outbreaks. States that were severely impacted during one wave may respond more effectively in subsequent outbreaks. In addition to population density mentioned above, this phenomenon could also be explained by education and accessibility to medical resources. Metropolitan areas, such as Boston and New York City, have more healthcare facilities and clinical professionals available to local residents, which might explain the low mortality rates of Massachusetts and New York during the second and third waves. Education may also be a factor associated with mortality rates, as individuals with higher levels of education are generally more aware of the importance of vaccination and better equipped to protect themselves from disease threats.

Given the methodological approach and findings of this study, there are some limitations that should be noted. This study was based on the official dataset from 2020 to 2023, which does not include more recent data for 2024. Hence, the results does not reveal the mortality and virulence patterns in recent time periods. Also, many factors play important roles in disease transmission and clinical outcomes of infection, such as local economics, accessibility to medical resources, education, and population density. However, this study does not account for these factors in the analysis, which might limit its ability to fully capture the complexities and underlying determinants of the observed patterns. These limitations point out the directions for future works. Integration of 2024 COVID-19 data can reveal the more recent patterns of COVID-19 pandemic, which might be helpful to guide future public health policies for disease prevention and treatment. In addition, a more comprehensive analysis that examines the public health, social, and political factors mentioned above can better find out the important determinants of COVID-19 disease transmission and treatment outcomes.

In summary, this study shows that the COVID-19 pandemic from 2020 to 2023 can be divided into three waves based on the morbidity and mortality data, and SARS-CoV-2 exhibited a gradual decline in virulence over time. During these waves, each state responded differently to the pandemic. These findings provide insights into the general trends in COVID-19 transmission and virulence, which offers suggestions for the government and public health agencies to develop future strategies for disease prevention and treatment.