Analysis of COVID-19 cases, hospitalizations, and deaths 2020-2024*

Charles Bray†

20 December, 2024

Abstract

Purpose: The abstract provides a concise summary of your project, including its objectives, key findings, and significance. Write this section last, after completing all other sections, to accurately reflect your project's focus and main results. Guidelines: Limit this section to 150-200 words. Briefly outline the purpose of your study, the approach you used, and the primary results and conclusions. The abstract should be clear, succinct, and give readers an immediate understanding of what your project entails.

Keywords: R, LaTeX, Quarto

*Thank you, BST 260 teaching team, for your hard work throughout the semester.

†SM Candidate, Department of Biostatistics, Harvard TH Chan School of Public Health, bray@hcp.med.harvard.edu

1 Introduction

The SARS-2 COVID-19 pandemic has been a major cause of death and disability in the United States and globally since the first months of 2020. XXX deaths are reported through time of writing, while millions are suspected to have diminished quality of life from the still poorly-understood set of conditions associated with "long COVID". [[DESCRIBE IMPACT OF PANDEMIC]]. It is popular to describe the net consequences of the pandemic but less so to assess the evolution of the virus and the ability of our public health and medical sectors to respond to the morbidity and mortality it caused.

To this end, it is necessary to provide a conceptualization of the distinct stages of the COVID-19 pandemic as it affected the US population, as well as the relative performance of different areas within the US. This paper aims to describe the changing dynamics of the novel coronavirus as such.

This is a link to Table 1:

Table 1: Your Caption

A	New	Table
left-aligned	centre-aligned	right-aligned
italics	~~strikethrough~~	**boldface**

A LaTeX equation. Black-Scholes (1) is a mathematical model used to price derivatives:

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial C^2} + rS\frac{\partial C}{\partial S} = rC \tag{1}$$

1. ordered list
2. item 2
    i) sub-item 1
        A. sub-sub-item 1

- unordered list
    – sub-item 1

For a demonstration of a line plot on a polar axis, see **?@fig-polar**.

2 Methods

Data were obtained from the Center for Disease Control and Prevention (CDC) via API, comprising hospitalizations, deaths, and cases associated with COVID-19 infection and reported weekly at the US state-level CDC [2]. States were defined as the 50 standard US states along with DC and Puerto Rico, for a total of 52 geographical units; states were assigned to regions of the US based on a classification scheme provided on the BST 260 course GitHub repository Rafael Izirarry [3].

The study proceeds in three parts. First, the COVID-19 pandemic was split broadly into different phases based on counts of cases, deaths, and hospitalizations across regions of the United States. Following this, the performance of individual US states was described along these measures within each determined wave. Finally, the nature of COVID-19 strains (their virulence and/or strain on hospitals) is determined by comparing the evolution of different measures from early to later waves.

In order to describe the trajectory of the COVID-19 pandemic, a novel approach to defining break-points between infection regimes (or "waves" of COVID-19) is developed. A sliding window of 9 weeks (or roughly 2 months) was applied to each measure (hospitalizations, cases, and deaths) within region, in which window were determined if the center value (the 5th index) of the window was a maximum or minimum value. Then, if at least half of the regions reporting data (typically, 10 regions) experienced a local minimum within 5 weeks of each other (or, half the minimum/maximum search window), the current wave was determined to have ended. Instead of over-weighting high-population regions, this approach appreciates potential regional segmentation of the pandemic when evaluating country-level phases of viral spread.

Prior to algorithm development, (1) break least squares as well as (2) Markov Regime Switching (MRS) models were considered (https://pmc.ncbi.nlm.nih.gov/articles/PMC10847870). Ultimately, the complexity of these models–and in the case of MRS, the infeasibility of constraining wave definitions to just two regimes–motivated the development of this data-driven identification strategy. Other, more simple, strategies such as that featured in Ayala et al., 2021, were also considered but determined to be too inflexible. Namely, hard thresholds for number of cases per population were used in defining transitions in infectious regimes, which were seen here as not considerate of changes in testing ubiquity and changing virulence of different generations of COVID-19.

3 Results

Waves

The COVID-19 pandemic in the US was divided into [[X]] waves according to case counts and hospitalizations. This was determined as appropriate based on the persistent lag of deaths data relative to these two metrics, and the fact that death timing relative to time of infection is known to exhibit much greater variance than symptoms or hospitalization relative to time of infection AUTHOR [1]. See Figure 1 below for a visualization of wave boundaries, where solid black dots indicate local minima in per-population rates of cases or hospitalizations by region, while black crosses indicate local maxima. A dotted red line shows the inter-regional wave cutoff (where one wave ends and another begins).

The first wave occurred DATE to DATE, which we classify as the "Originator" wave. This was succeeded by the __ and __ waves, which occurred DATE-DATE and DATE-DATE, respectively. Finally, while there continued to be local minima and maxima that could be considered to delineate further waves, it was decided that the low number of deaths along with moderation in number of hospitalizations characterized a "post-acute" phase of the pandemic, starting at DATE and extending through the end of data collection, December 1, 2024.

State performance

XXX

Virulence

XXX

4 Discussion

DISCUSSION.

References

[1] AUTHOR. "TITLE." In: *JOURNAL* 0.0 (2018), pp. 0-0.

[2] CDC. *Data — Centers for Disease Control and Prevention*. https://data.cdc.gov/. Accessed: 2024-11-1. 2024.

[3] Rafael Izirarry. *github.com/datasciencelabs*. https://github.com/datasciencelabs/2024/tree/refs/heads/main/data. Accessed: 2024-11-1. 2024.