Analyzing COVID-19 Mortality at the County Level in the U.S. Using Generalized Additive Models

Christian Testa

2022-10-07

Abstract

this is an abstract

```
## here() starts at /Users/cht180/Documents/2022/covid.gradient.estimation
```

Introduction

As we enter the third winter with novel coronavirus disease (COVID-19) in the United States (U.S.), evidence reflecting intense disparities in COVID-19 mortality rates among socially advantaged and disadvantaged populations persists. Eliminating inequities in health outcomes is a major policy goal of the Biden administration, representing an revitalized focus on health justice and underlining the importance of adequate data reporting systems that measure the prevalent health inequities. To this end, generalized additive models (GAMs) provide a flexible regression framework which can illustrate the evolving roles and relationships social conditions have with trends in COVID-19 mortality.

Background

Having passed over 1 million COVID-19 deaths in the United States earlier this year, and facing uncertain prospects for the third COVID-19 winter looming even as new iterations on the COVID-19 vaccines become available, it remains critical that inequities in COVID-19 outcomes are documented and analyzed to reckon with the unjust and unfair burden of preventable illness. Even though the first vaccines were granted emergency use authorization by the U.S. Food and Drug Administration in 2020, with the first shots going in arms in December 2020, COVID-19 is still continuing to cause hundreds of deaths a day in the U.S. in the late summer and early fall of 2022. The new bivalent vaccines released at the end of August 2022 contain messenger RNA components of the original strain as well as the BA.4 and BA.5 lineages in an effort to make the nation's immunity more up-to-date and robust against the myriad of phylogenetic directions the COVID-19 virus is exploring. Despite the updated bivalent boosters representing a significant step forward in prevention strategy, less than 4% of eligible Americans had taken the booster in the first month after it became available. As such, and with an enduring history of inequities in health care availability, it is not clear that all communities will be equally able to take advantage of the new vaccines and inequities in COVID-19 illness and mortality may persist despite the technological innovations in vaccine therapy.

Methods

Data Sources

The following variables were retrieved at the county level:

- Counts of deaths by month with underlying cause of death as COVID-19 (ICD-10 code U07.1) from CDC WONDER, Provisional Multiple Cause of Death Data

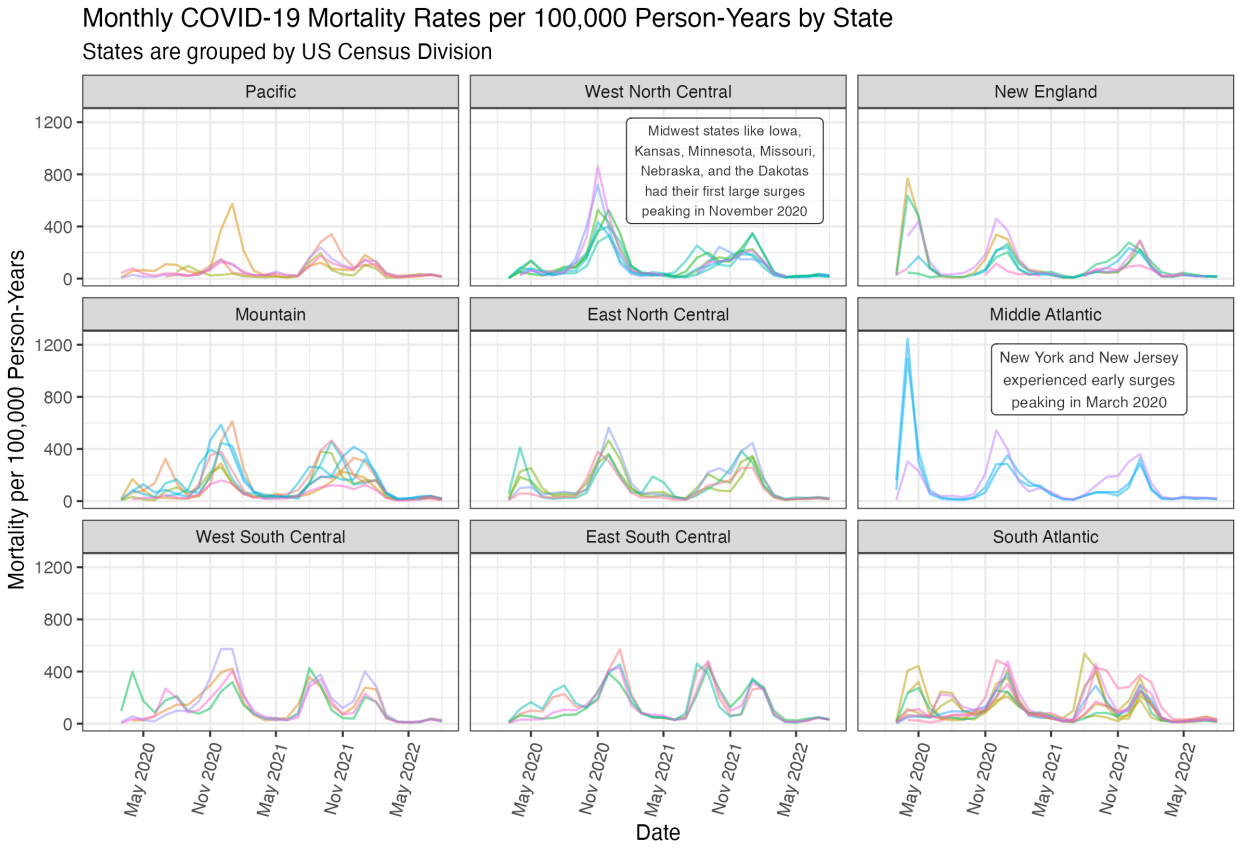

Figure 1: Monthly estimates of COVID-19 mortality per 100,000 residents by state are shown. States are in distinct colors to make each state's trajectory visible. Reference the online version TODO: XXX

- Population Size Estimates for 2020 from the U.S. Census (Redistricting File Public-Law 94-171 Dataset). *Note:* While the complete data for the 2020 U.S. Census have not yet been released, the summary tables including population counts have been released as mandated by Congress for the purposes of redistricting and updating the boundaries of the 435 voting Congressional Districts, which are what is made us of here as population size estimates.
- Percent Black, percent Hispanic, percent Asian, percent American Indian or Alaska Native, Percent below the federal poverty line, median household income, median age, percent 65 years or older, and percent 75 years or older from the 2014-2019 5-year American Community Survey. *Note:* while more recent American Community Survey 5-year estimates are available, survey data collection in 2020 was severely affected by the pandemic with the 2020 data release denoted as "experimental".

Generalized Additive Models

Results

We used GAMs to describe the following:

- the changing associations between individual area-based measures and COVID-19 mortality over time;
- the changing associations between bivariate area-based measures and COVID-19 mortality over time;
- associations with area-based measures and COVID-19 mortality after controlling for spatio-temporal autocorrelation.

Additionally, as sensitivity analyses, random forest models were fit to compare with the coefficient estimates from the GAMs.

Discussion

Limitations

- Missing data
- Ecological fallacy
- Age effects

Conclusions

References

Supplementary Materials