Approximating the Number of COVID-19 Infections in the United States Across 2021

Quinn White

2023-06-30

Abstract

Introduction

Results

State-level Estimates

For simplicity, at the state-level we focus on the implementation that does not vary by state or date. This also allows us to consider the entirety of 2021, since survey data is only available for dates after March 20, 2021. A full comparison of implementations is included in Supplementary Figure covidestim-concordance-state.

In Figure 1, we consider three distinct two-week intervals during waves of the pandemic in 2021.

Although prevalence of COVID-19 was highest during the time interval during the Omicron wave, we see that the ratio of estimated infections to observed infections is higher during the time intervals in the alpha and delta waves. This distinction is explained by the differences in testing rates during these period: the testing rate during this two-week interval during the omicron wave was 2.4 times that of the alpha wave and 4.9 times that of the delta wave.

Several states consistently have among the highest or lowest ratios of estimated to observed infections. In particular, there are 6 states with among the lowest 10 ratios of estimated infections to observed infections, and as such the highest case ascertainment rates, for more than 80% of time intervals considered. These states were Rhode Island, Massachusetts, District of Columbia, Alaska, New York, and Vermont. Meanwhile, states that had the highest ratios, and equivalently the lowest case ascertainment rates, include Mississippi, South Dakota, Oklahoma, Nebraska, and Tennessee.

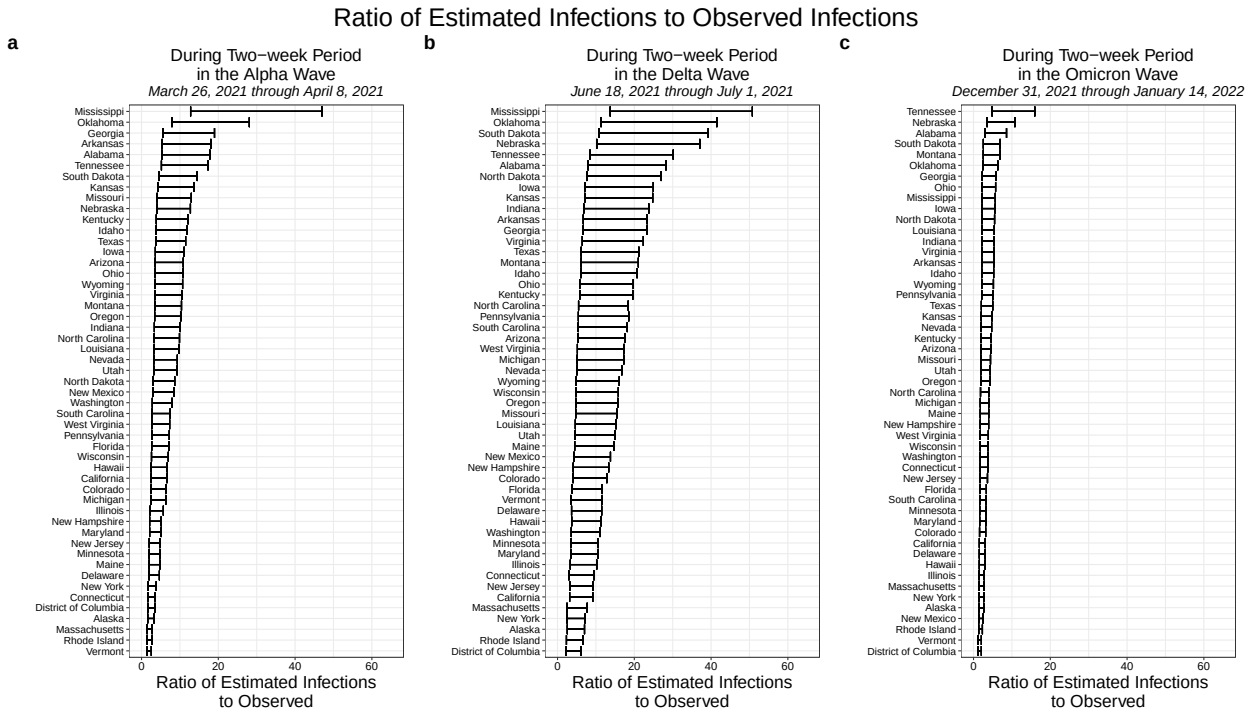


Figure 1: The ratio of estimated infections to observed infections for three time intervals of interest: one during the alpha wave, one during the delta wave, and one during the omicron wave. Although the prevalence of COVID-19 was highest during the omicron wave, the ratios of estimated to observed infections are higher for the time intervals during the alpha and delta waves, a difference that was driven by lower testing rates during these times. The trend we see in these three time intervals where Mississippi, South Dakota, Oklahoma, Nebraska, and Tennessee have among the highest ratios of estimated infections to observed infections, and as such the lowest case ascertainment rates, is consistent across the full set of time intervals considered from January of 2021 to March of 2022.

In Figure 2, we see the simulation intervals for all two-week intervals and all states.

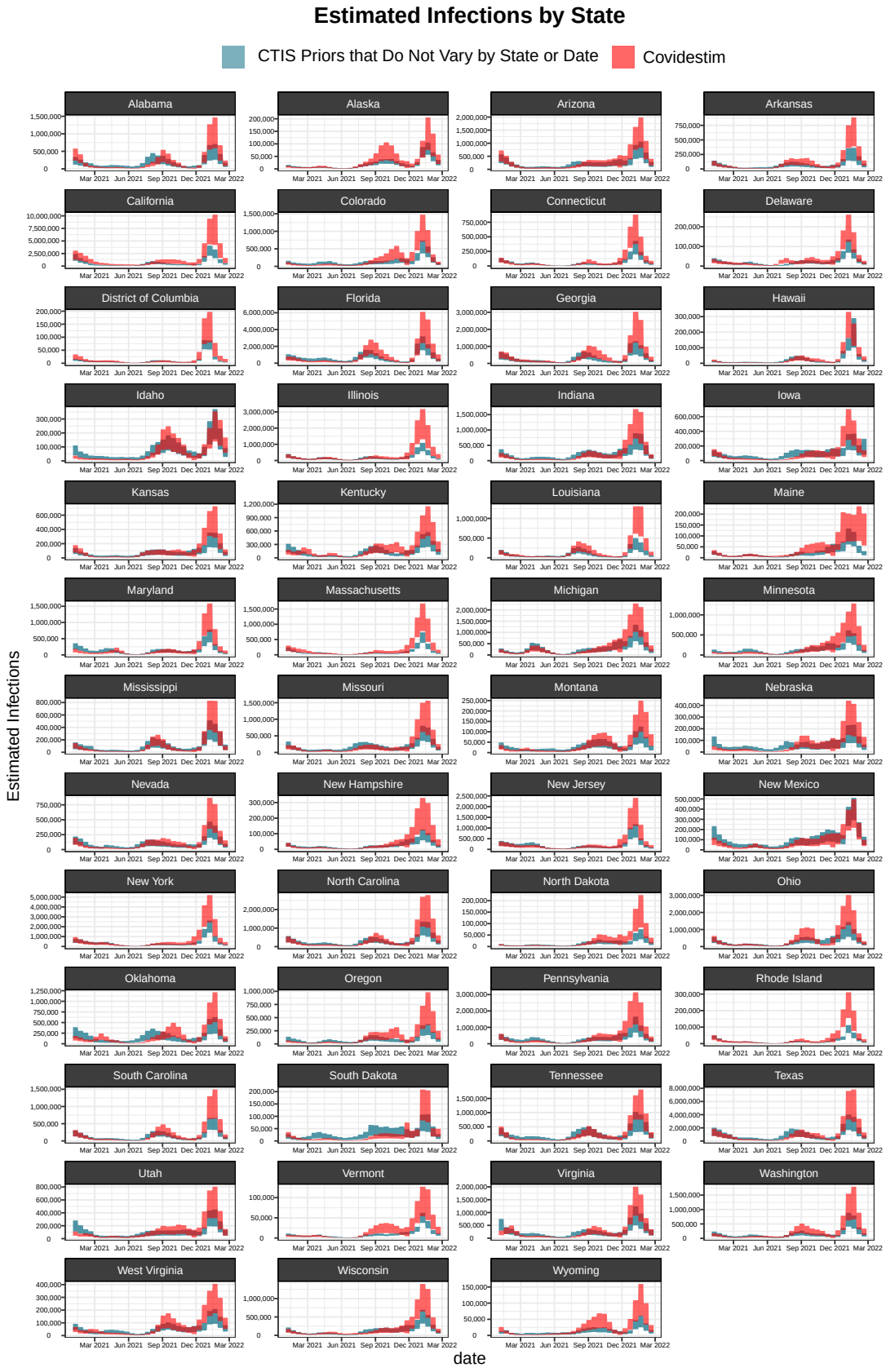


Figure 2: Simulation intervals for each 2-week interval considered, for all states. For any given state, each vertical bar shows the 2.5% percentile and 97.5% percentile for the total number of infections in that two-week interval. Covidestim intervals summed over the same two-week time-scale are shown in red. The scale on the y-axis is distinct across states to highlight differences across time within each state.

County Level Estimates in Massachusetts

Comparison to Wastewater Data and Covidestim Estimates

Because there is no established ground truth to compare to regarding the true number of infections for any time-interval, at the county level we compare our results to two distincts sources of information: wastewater data aggregated at the county-level, and results from a previously published Bayesian evidence synthesis model, Covidestim. This also allows us to compare how different implementations of probabilistic bias analysis compare to Covidestim estimates and wastewater concentrations.

Our first comparison considers the correlations between wastewater concentrations and the probabilistic bias estimates as well as between the Covidestim estimates and the probabilistic bias estimates (Figure 3).

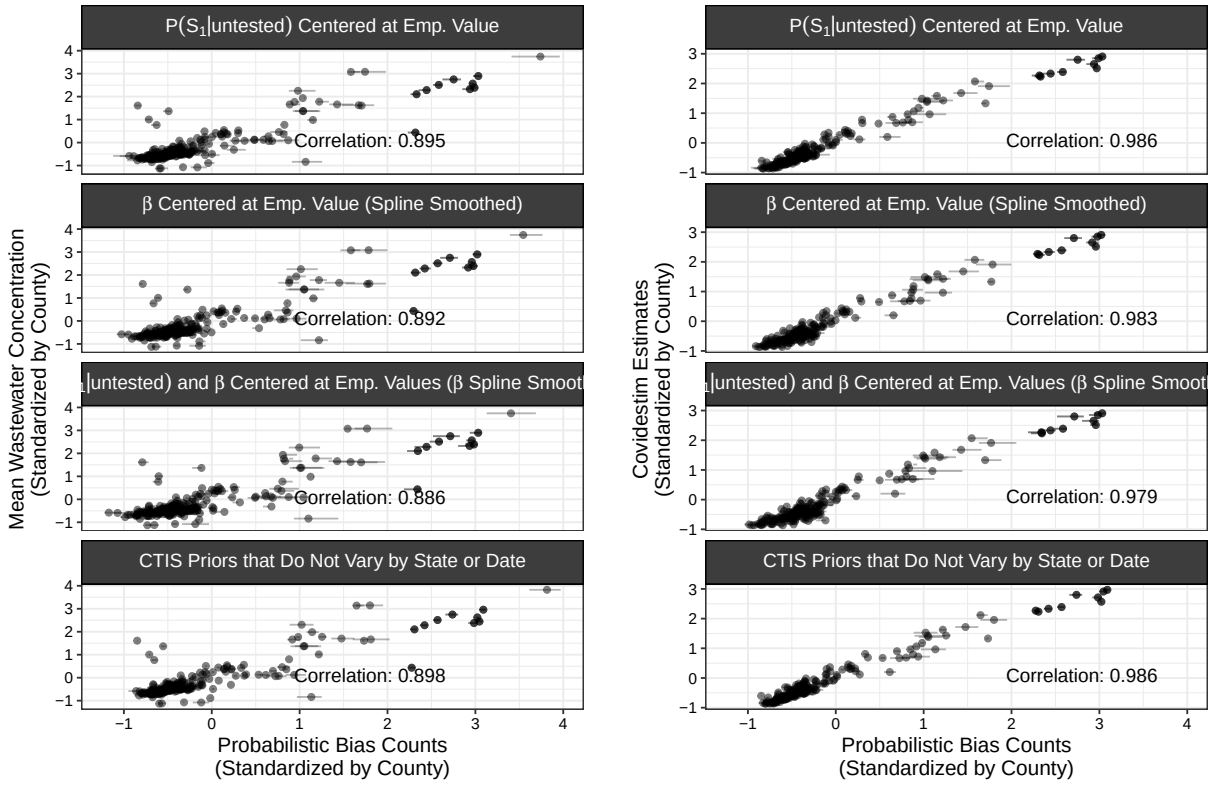
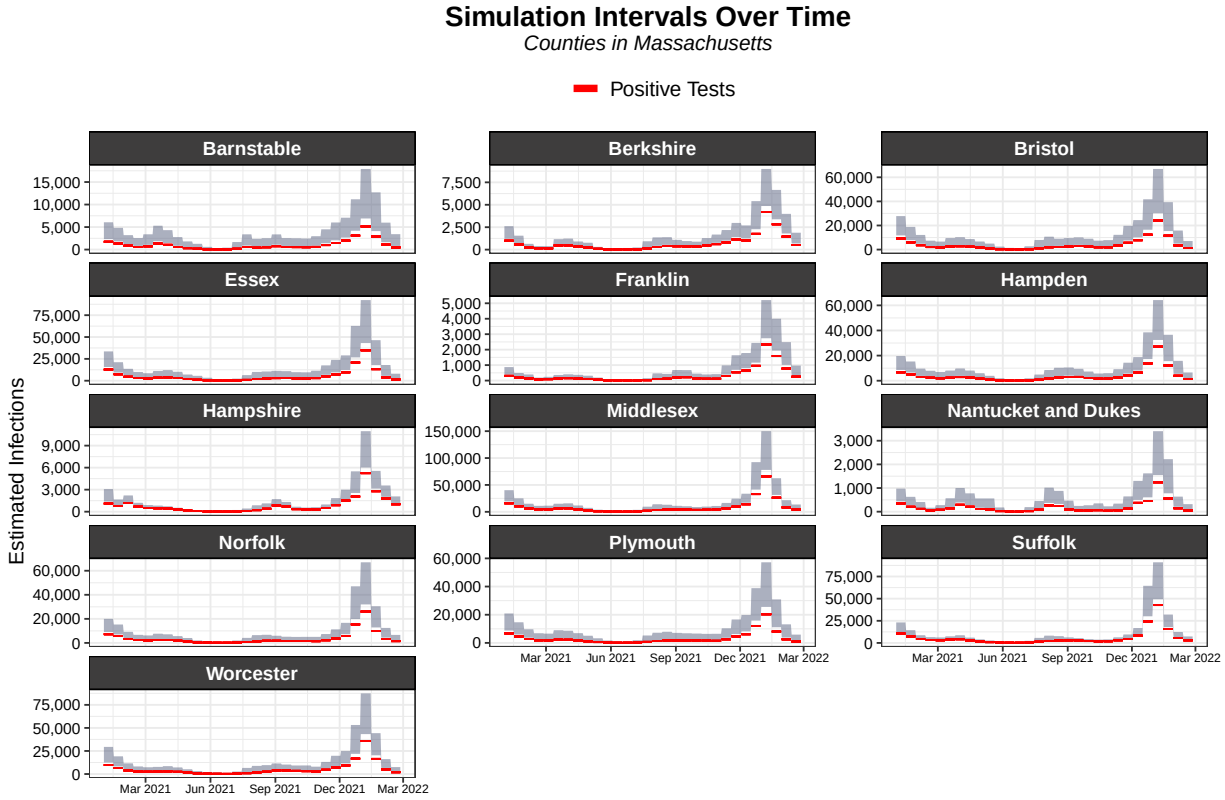


Figure 3: Considering the correlations between probablistic bias estimates and wastewater concentrations (a) and between probablistic bias estimates and Covidestim estimates (b). We see that all implementations considered are highly correlated with both wastewater concentrations and Covidestim estimates. The implementation that does not allow priors to vary by state or date has the highest correlations.

NOTE add (a), (b) to this figure



Methods

Data

Massachusetts County Level

State Level

Survey Data

The COVID-19 Trends and Impact Survey was run in collaboration by . . .

Wastewater Data

Biobot analytics . . .

Statistical Methods

Probabilistic bias analysis

Bayesian Melding

Specification of Priors