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

Quinn White

2023-07-09

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.

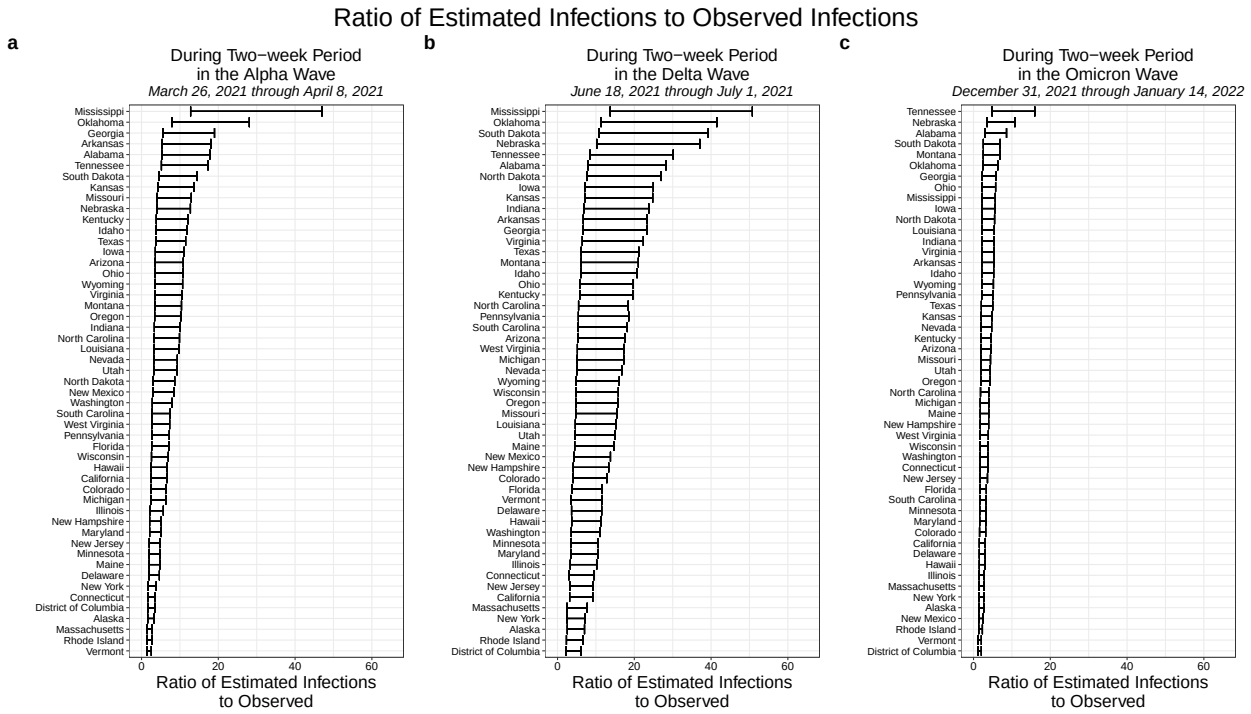

Figure 1: The ratio of estimated infections to observed infections for three time intervals of interest: one during the alpha wave, one during the delta wave, and one during the omicron wave. Although the prevalence of COVID-19 was highest during the omicron wave, the ratios of estimated to observed 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.

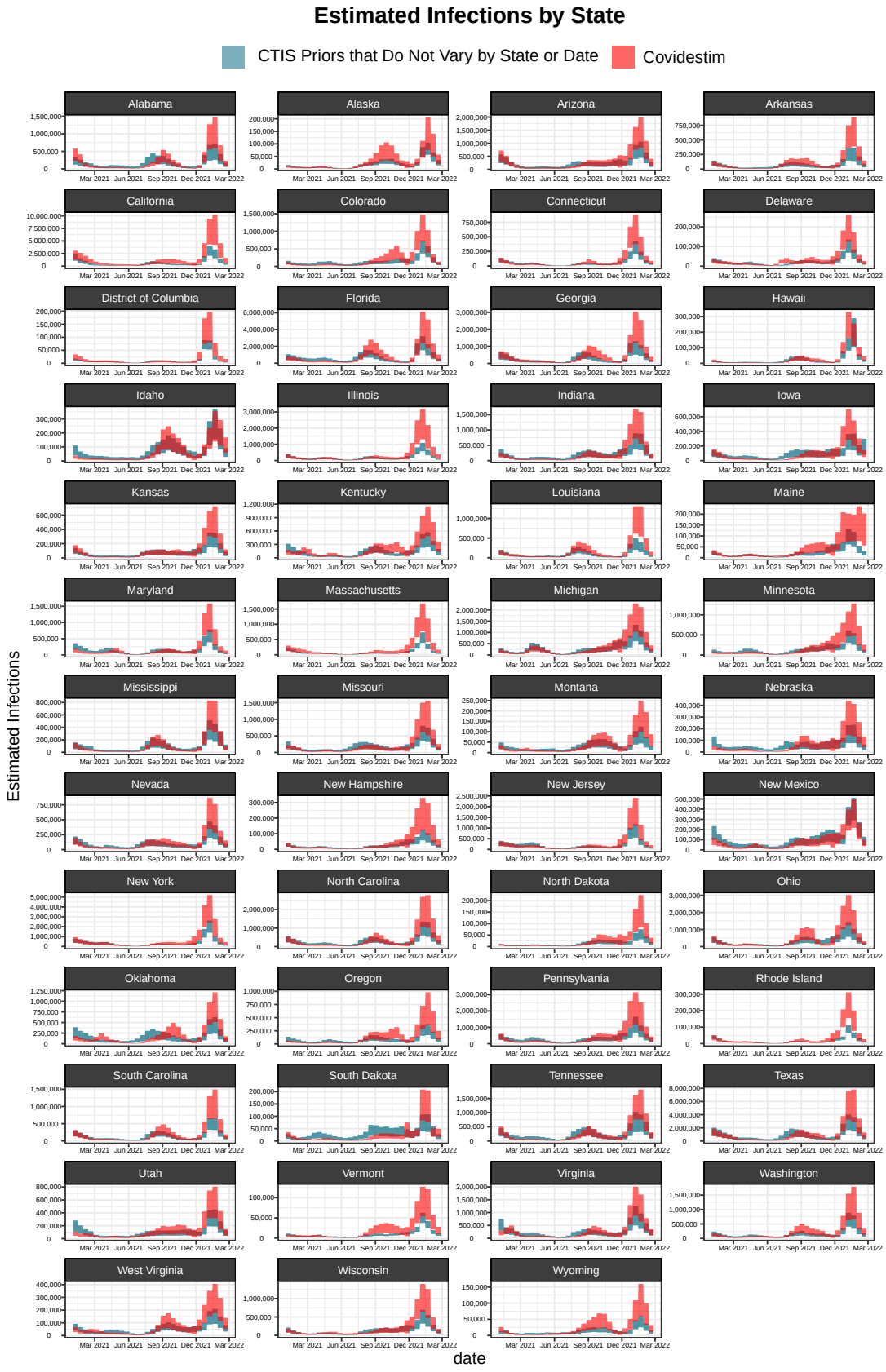

Figure 2: Simulation intervals for each 2-week interval considered, for all states. For any given state, each vertical bar shows the 2.5% percentile and 97.5% percentile for the total number of infections in that two-week interval. Covidestim intervals summed over the same two-week time-scale are shown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

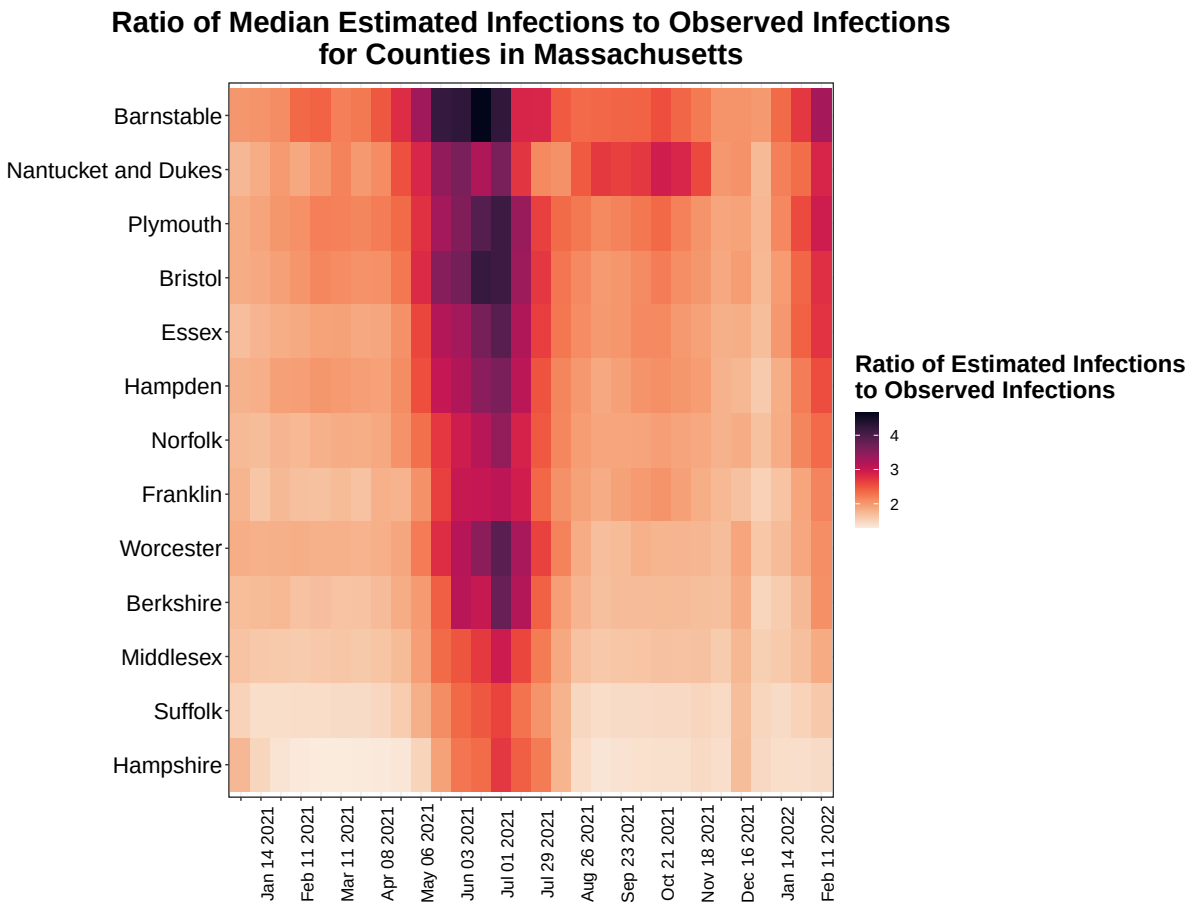

Figure 3: The ratio of estimated to observed infections across time for counties in Massachusetts. Counties are ordered by the median ratio across time intervals, from the highest ratio (Barnstable) to the lowest (Hampshire). Similar to what we see at the state level, the highest ratios were during the summer of 2021 during the Delta wave – a period of decreased testing. The span of time with the highest ratio of estimated to observed infections was July 2, 2021 through July 30, 2021.

County Level Estimates in Massachusetts

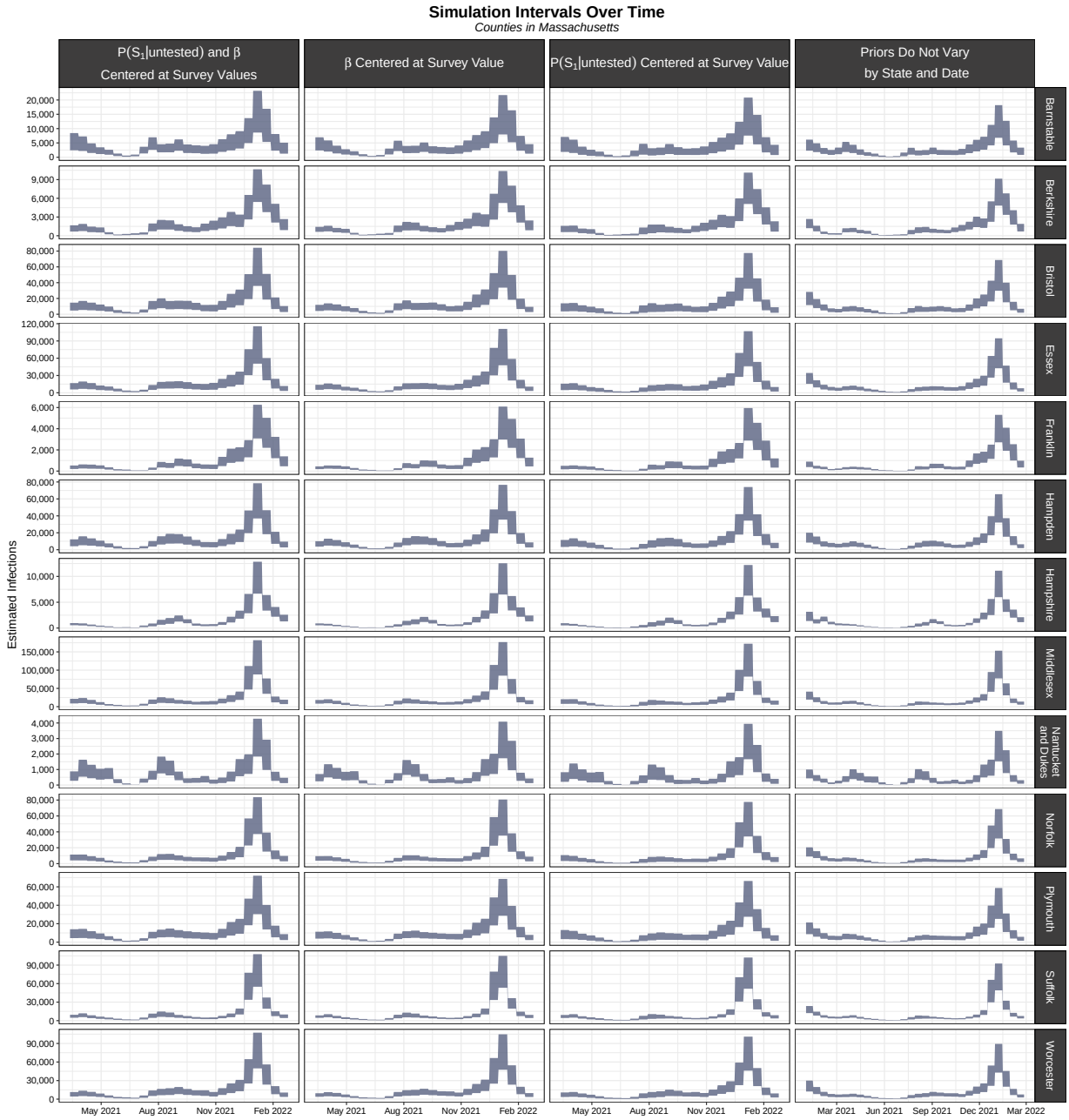

Figure 4: Simulation intervals for counties in Massachusetts. Each interval corresponds to a 95% simulation interval for the total number of estimated infections for that county in that two-week time interval. The columns represent different implementations of the probabilistic bias analysis. The first column corresponds to the implementation where we specify priors that are the same for all dates. For the implementation in the second column, we center the distribution of β at the ratio of the screening test positivity to the overall test positivity from the survey. For the third column, we center the distribution of $\Pr(S_1|\text{untested})$ at the percentage of the population experiencing COVID-19-like illness from the survey. The fourth column centers both β and $\Pr(S_1|\text{untested})$ at the aforementioned values.

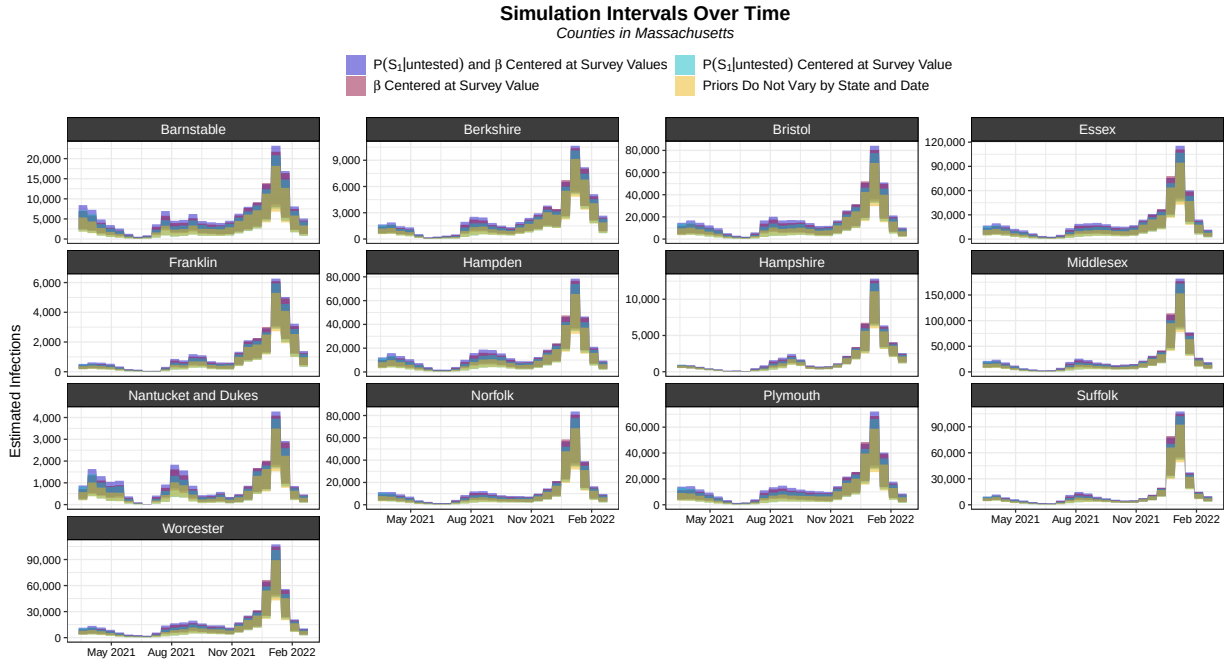

Figure 5: Simulation intervals for counties in Massachusetts, colored by the implementation of probabilistic bias analysis. The implementation that centers both $\Pr(S_1|\text{untested})$ and β at their empirical values is consistently the highest among the implementations, and the version only centering $\Pr(S_1|\text{untested})$ at the survey value is highly concordant with the version where the priors do not vary by date.

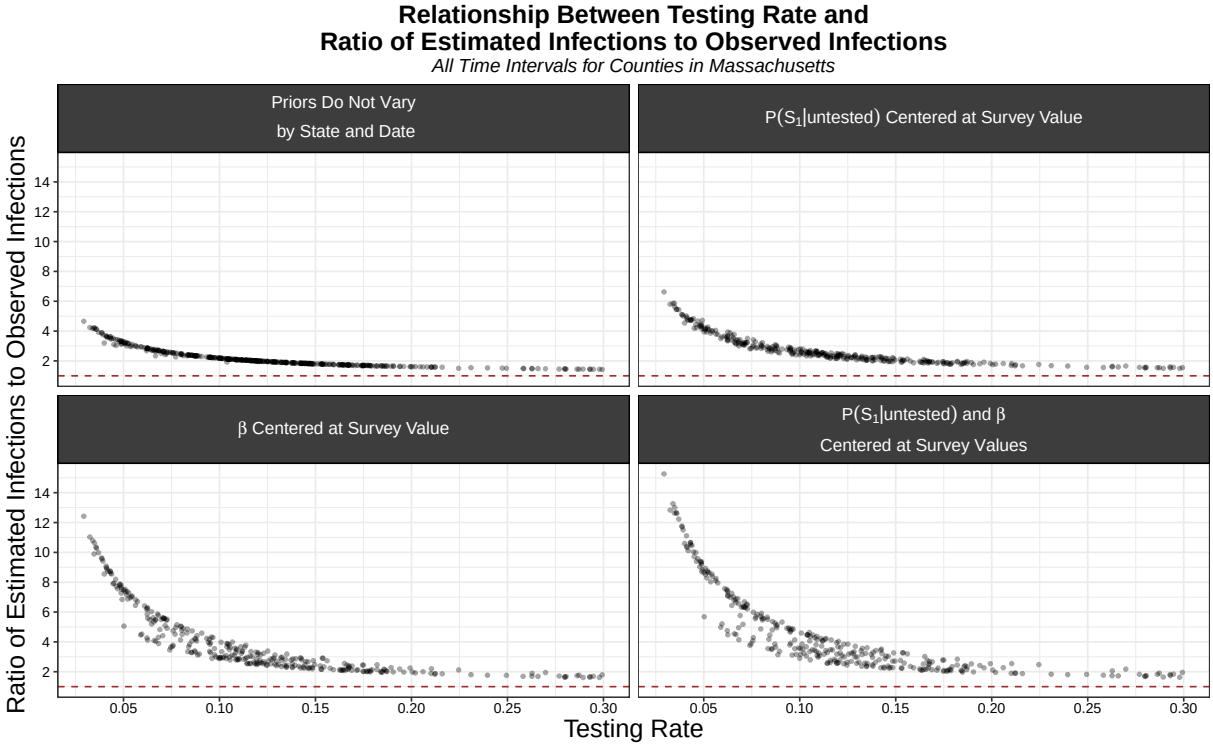

Figure 6: The ratio of the median estimated infections to observed infections plotted against the testing rate, where the testing rate is calculated as the total number tested in a two-week interval over the population size. When the priors are the same for all time intervals, there is minimal variability relationship between the testing rate and the ratio of estimated to observed infections, since the correction for incomplete testing and diagnostic test inaccuracy is identical for each time-interval. However, when we allow β or $\Pr(S_1|\text{untested})$ to vary over time, there is more variability in the relationship. A horizontal line in red at 1 is included to reference; a ratio of exactly 1 would indicate no infections went unobserved.

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 7).

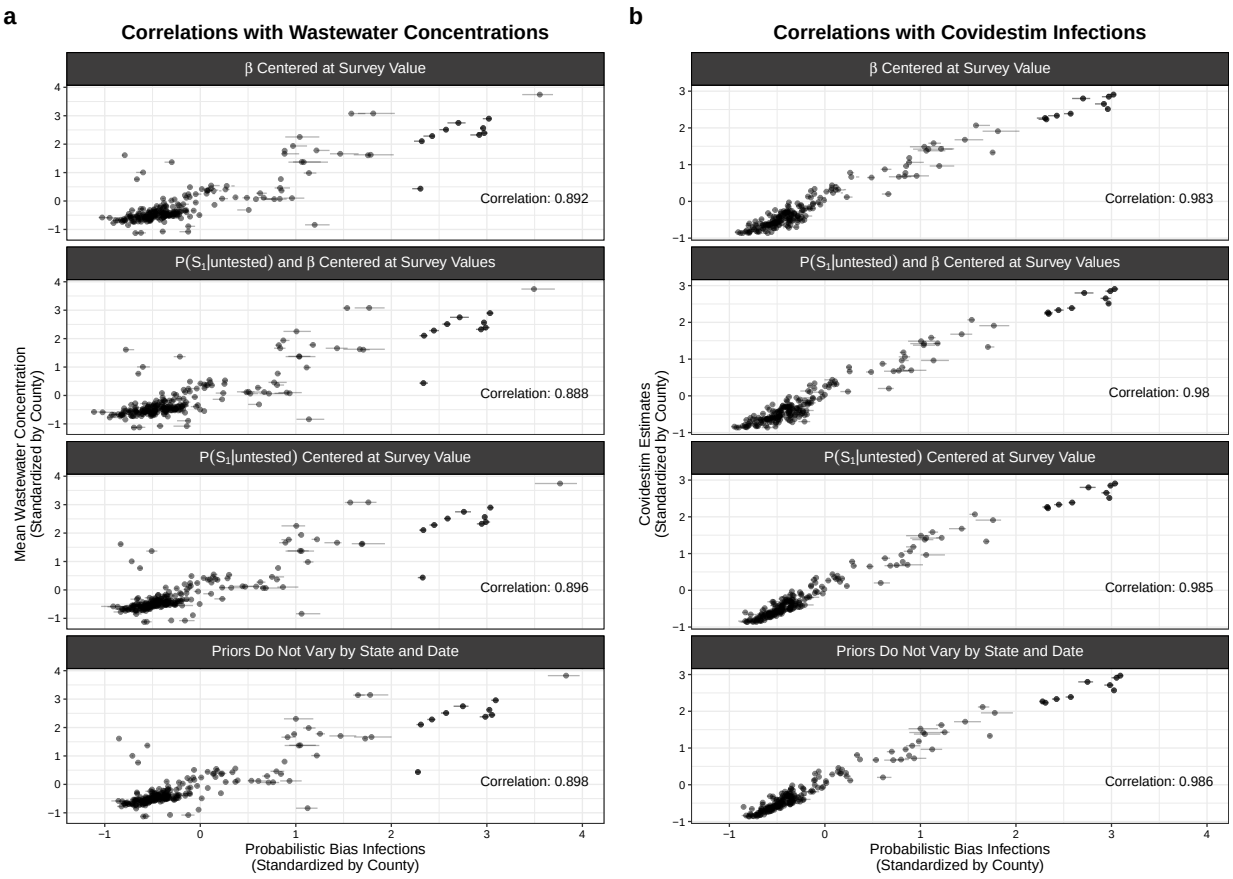

Figure 7: Considering the correlations between probablistic bias estimates and wastewater concentrations (a) and between probablistic bias estimates and Covidestim estimates (b). We see that all implementations considered are highly correlated with both wastewater concentrations and Covidestim estimates. The implementation that does not allow priors to vary by state or date has the highest correlations.

Table 1 is

Table 1: Coverage of Covidestim Medians at the County and State Levels

Implementation	Percent Below Interval	Percent Contained in Interval	Percent Above Interval
County			
$P(S_1$|untested) Centered at Survey Value	1.282	87.500	11.218
Priors Do Not Vary by State or Date	1.075	74.194	24.731
β Centered at Survey Value	29.487	63.782	6.731
$P(S_1$|untested) and β Centered at Survey Values	34.615	62.179	3.205
State			
$P(S_1$|untested) Centered at Survey Value	0.889	76.000	23.111
Priors Do Not Vary by State or Date	5.817	71.634	22.549
β Centered at Survey Value	13.481	69.630	16.889
$P(S_1$|untested) and β Centered at Survey Values	23.259	62.370	14.370

The percent of simulation intervals where the Covidestim median falls below, within, or above the interval, when considering all simulation intervals for that implementation and geographic scale.

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

Sample from priors on $\Pr(S_1|\text{untested}), \alpha, \beta$

$\downarrow$

Constrain priors with Bayesian melding to obtain constrained distributions

for $\Pr(S_1|\text{untested}), \alpha, \beta,$ and $\Pr(S_0|\text{test}_+, \text{untested})$

$\downarrow$

For each geographic unit, use sampled α and β to calculate:

$$\Pr(\text{test}_+|S_1, \text{untested}) = \alpha \ \Pr(\text{test}_+|\text{tested})$$
$$\Pr(\text{test}_+|S_0, \text{untested}) = \beta \ \Pr(\text{test}_+|\text{tested})$$

$\downarrow$

$$N^*_{\text{untested},S_0} = N_{\text{untested}}(1 - \Pr(S_1|\text{untested}))\Pr(\text{test}_+|S_0, \text{untested})$$
$$N^*_{\text{untested},S_1} = N_{\text{untested}}(\Pr(S_1|\text{untested}))\Pr(\text{test}_+|S_1, \text{untested})$$

$\downarrow$

Estimate total unobserved positive tests as

$$N^*_{\text{untested}} = N^*_{\text{untested},S_0} + N^*_{\text{untested},S_1}$$

$\downarrow$

Take the sum to acquire total positive tests

$$N^* = N^*_{\text{untested}} + N^*_{\text{tested}}$$

$\downarrow$

Correct for diagnostic test inaccuracy

$$N^+ = \frac{(N^* - (1 - S_p)N)}{(S_e + S_p - 1)}$$

Probabilistic bias analysis

Bayesian Melding

Specification of Priors

$\Pr(S_0|\textbf{test}_+, \textbf{untested})$

There is substantial heterogeneity in estimates of the percent of infections that are asymptomatic. This is in part due to distinct study populations and selection criteria. In particular, estimates from screening studies may be better estimates of the asymptomatic rate among the untested population: estimates from studies where the population was not screened, and as such was comprised of individuals that sought out a PCR test, may include a higher proportion of symptomatic individuals, biasing estimates of the asymptomatic rate downwards.

[1] conducted a meta-analysis including studies across the globe as of February 4, 2021, and estimated the pooled percent of asymptomatic infections among confirmed infections to be 40.50% (95% CI 33.50%-47.50%). This analysis did not restrict to screening studies. Another meta-analysis, when restricting to screening studies, found the pooled asymptomatic percentage to be 47.3% (95% CI, 34.0 - 61.0%)[2]. Both meta-analyses noted the substantial amount of heterogeneity in the percent of asymptomatic infections.

[3] conducted a large screening study consisting of individuals arriving from overseas and found the asymptomatic rate was 76.8%, while a screening study among children admitted to a pediatric emergency department between May 2020 and January 2021 found it to be 51.7%[4]. Several studies were conducted among university students. Among students at the University of Arizona in the fall semester of 2020, including students who sought testing and who were required to test, the asymptomatic rate of infection was 79.2%. A study at the University of Notre Dame distinguished between presymptomatic infection and asymptomatic infection, and found found 32% to be asymptomatic throughout the entire course of infection, 27.0% to be presymptomatic, and 40.5% to be symptomatic. The asymptomatic rate among nonresidential students participating in the surveillance testing system at Clemson University was 69%.

Vaccine coverage also may influence the asymptomatic rate. In a study in Israel on the effectiveness of the Pfizer–BioNTech mRNA COVID-19 vaccine BNT162b2, 55.7% (49,138 out of 88203) of infections were asymptomatic in the unvaccinated group, and 68.2% of infections (3632 out of 5324) of infections were asymptomatic in the vaccinated group.

Numerous additional factors contribute to the heterogeneity we see among estimates of the percent of infections that are asymptomatic, including community prevalence, the study population, and the time period when the study population was tested. The use of different definitions also may contribute. This includes the definition of a symptomatic infection, since our understanding of the clinical presentation of COVID-19 has evolved over time[2], as well as variation in making a distinction between presymptomatic cases, where people that had no symptoms upon testing positive but may have went on to develop symptoms at a later date, and truly asymptomatic cases, where an infected individual never goes on to develop symptoms.

Because of the heterogeneity in estimates of the percent of infections that are asymptomatic, for this prior we specified a beta distribution with the majority of the density between 0.3 and 0.8, with a mean of 0.55 and standard deviation of 0.12.

Limitations

References

1. Ma, Q. *et al.* Global Percentage of Asymptomatic SARS-CoV-2 Infections Among the Tested Population and Individuals With Confirmed COVID-19 Diagnosis: A Systematic Review and Meta-analysis. *JAMA Netw Open* **4**, e2137257 (2021).
2. Sah, P. *et al.* Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2109229118 (2021).
3. Fang, L.-L. *et al.* PCR combined with serologic testing improves the yield and efficiency of SARS-CoV-2 infection hunting: A study in 40,689 consecutive overseas arrivals. *Front. Public Health* **11**, 1077075 (2023).
4. Ford, J. S. *et al.* Use of an Asymptomatic COVID-19 Testing Protocol in a Pediatric Emergency Department. *The Journal of Emergency Medicine* **63**, 332–338 (2022).