

## Article

# Bayesian Inference of State-Level COVID-19 Basic Reproduction Numbers across the United States

Abhishek Mallela <sup>1</sup>, Jacob Neumann <sup>2,†</sup>, Ely F. Miller <sup>2</sup>, Ye Chen <sup>3</sup>, Richard G. Posner <sup>2</sup>, Yen Ting Lin <sup>4</sup> and William S. Hlavacek <sup>5,\*</sup>

<sup>1</sup> Department of Mathematics, University of California, Davis, CA 95616, USA; amallela@ucdavis.edu

<sup>2</sup> Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA; jn382@cornell.edu (J.N.); efm46@nau.edu (E.F.M.); richard.posner@nau.edu (R.G.P.)

<sup>3</sup> Department of Mathematics and Statistics, Northern Arizona University, Flagstaff, AZ 86011, USA; ye.chen@nau.edu

<sup>4</sup> Los Alamos National Laboratory, Information Sciences Group, Computer, Computational and Statistical Sciences Division, Los Alamos, NM 87545, USA; yentingl@lanl.gov

<sup>5</sup> Los Alamos National Laboratory, Theoretical Biology and Biophysics Group, Theoretical Division, Los Alamos, NM 87545, USA

\* Correspondence: wish@lanl.gov

† Current address: Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA.

**Abstract:** Although many persons in the United States have acquired immunity to COVID-19, either through vaccination or infection with SARS-CoV-2, COVID-19 will pose an ongoing threat to non-immune persons so long as disease transmission continues. We can estimate when sustained disease transmission will end in a population by calculating the population-specific basic reproduction number  $\mathcal{R}_0$ , the expected number of secondary cases generated by an infected person in the absence of any interventions. The value of  $\mathcal{R}_0$  relates to a herd immunity threshold (HIT), which is given by  $1 - 1/\mathcal{R}_0$ . When the immune fraction of a population exceeds this threshold, sustained disease transmission becomes exponentially unlikely (barring mutations allowing SARS-CoV-2 to escape immunity). Here, we report state-level  $\mathcal{R}_0$  estimates obtained using Bayesian inference. Maximum a posteriori estimates range from 7.1 for New Jersey to 2.3 for Wyoming, indicating that disease transmission varies considerably across states and that reaching herd immunity will be more difficult in some states than others.  $\mathcal{R}_0$  estimates were obtained from compartmental models via the next-generation matrix approach after each model was parameterized using regional daily confirmed case reports of COVID-19 from 21 January 2020 to 21 June 2020. Our  $\mathcal{R}_0$  estimates characterize the infectiousness of ancestral strains, but they can be used to determine HITs for a distinct, currently dominant circulating strain, such as SARS-CoV-2 variant Delta (lineage B.1.617.2), if the relative infectiousness of the strain can be ascertained. On the basis of Delta-adjusted HITs, vaccination data, and seroprevalence survey data, we found that no state had achieved herd immunity as of 20 September 2021.

**Keywords:** mathematical model; coronavirus disease 2019 (COVID-19); basic reproduction number; herd immunity; Bayesian inference



**Citation:** Mallela, A.; Neumann, J.; Miller, E.F.; Chen, Y.; Posner, R.G.; Lin, Y.T.; Hlavacek, W.S. Bayesian Inference of State-Level COVID-19 Basic Reproduction Numbers across the United States. *Viruses* **2022**, *14*, 157. <https://doi.org/10.3390/v14010157>

Academic Editors: Tomasz Lipniacki, Marek Kočańczyk and Marek Bodnar

Received: 30 November 2021

Accepted: 12 January 2022

Published: 15 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Vaccines to protect against coronavirus disease 2019 (COVID-19) became available in the United States (US) in December 2020 [1]. In the US, as of 20 September 2021, 181,728,072 persons have been fully vaccinated (55% of the total population), an additional 30,307,256 persons have been partially vaccinated, and an uncertain number of persons have acquired immunity through infection [2]. The entire US population does not need to be vaccinated to end sustained COVID-19 transmission because of the phenomenon of

herd immunity [3], which is reached when a critical fraction of the population becomes immune. This fraction is called the herd immunity threshold (HIT).

The HIT for a population relates to the basic reproduction number,  $\mathcal{R}_0$ , as follows [3]:  $\text{HIT} = 1 - 1/\mathcal{R}_0$ .  $\mathcal{R}_0$  is defined as the expected number of secondary infections arising from a primary case in the absence of any immunity or intervention. As is well known,  $\mathcal{R}_0$  and HIT are population-specific [4,5], which means that the effort required to control the local COVID-19 epidemic may vary from community to community. However, knowledge of the HIT for a given region is insufficient to determine when disease transmission within the region will end. One also needs to know the fraction of the population that has immunity. Estimating the immune fraction is difficult, because we cannot simply count the number of persons who have been vaccinated or the number of persons detected to be infected. Immunity is acquired not only through vaccination but also through infection [6], and case detection is imperfect. Insight into the immune fraction can be obtained from seroprevalence surveys, which use blood tests to identify persons who have antibodies against the SARS-CoV-2 virus (acquired through vaccination or infection).

Various estimates of  $\mathcal{R}_0$  for transmission of COVID-19 have been provided in the literature [7]. The estimates that have received the most attention are those given for China and Italy [8–12], which were among the first regions to be impacted by COVID-19. However, the relevance of these estimates for populations within the US (or elsewhere outside of China and Italy) is unclear. Several studies have estimated  $\mathcal{R}_0$  for the US at the national level [13–15], the state level [16–18], and the county level [19,20]. The usefulness of a national estimate is unclear given the heterogeneity of the US, and none of the county-level estimates are comprehensive. Some state-level estimates are also incomplete [16,18]. Because responses to COVID-19 within the US have been and continue to be driven mainly by governors of US states [21], we undertook a study to generate comprehensive state-level  $\mathcal{R}_0$  estimates through Bayesian inference. With this approach, we were able to quantify uncertainty in each estimate through a parameter posterior distribution.

In earlier work, we developed a compartmental model for COVID-19 transmission dynamics that reproduces surveillance data and generates accurate forecasts for the 15 most populous metropolitan statistical areas (MSAs) in the US [22]. Here, for each of the 50 states, we found a state-specific parameter posterior conditioned on this model from state-level COVID-19 surveillance data available from 21 January to 21 June 2020 [23]. From these parameter posteriors, we then obtained region-specific  $\mathcal{R}_0$  and HIT posteriors and maximum a posteriori (MAP) estimates. The MAP estimates for HITs together with other data—vaccination tracking data [24], serological survey data [25,26], and quantitative estimates of the increased transmissibility of the recently introduced SARS-CoV-2 variant Delta (lineage B.1.617.2) [27,28]—provide insight into the progress of each state toward herd immunity.

## 2. Materials and Methods

### 2.1. Model

To obtain regional  $\mathcal{R}_0$  and HIT estimates, we used a compartmental model developed previously for the purpose of forecasting COVID-19 disease incidence [22]. This model, which is capable of making accurate forecasts [22], is a COVID-19-specific elaboration of the classic SEIR model [29] that accounts for effects of nonpharmaceutical interventions, including social distancing. Consideration of nonpharmaceutical interventions is important because the widespread adoption of such interventions began in the US around 13 March 2020 [30], a time roughly coincident with the start of sustained community transmission of COVID-19 in many parts of the US (see below). We found region-specific parameterizations that allow the model to reproduce surveillance data (daily reports of new confirmed COVID-19 cases) available for each region of interest over a defined period (e.g., 21 January to 21 June 2020). The model is able to account for a variable number of social-distancing periods. We considered versions of the model accounting for one, two, and three social-distancing periods. The number of social-distancing periods deemed best (i.e., to provide

the most parsimonious explanation of the data) for a given time period was determined using the model selection procedure described by Lin et al. [22]. As in the study of Lin et al. [22], the model has 14 parameters with universal fixed values (applicable to all regions). The model also has  $3(n + 1) + 3$  parameters with region-specific adjustable values determined through Bayesian inference, where  $n + 1$  denotes the number of social-distancing periods. In this study, for a given region, we censored case-reporting data whenever the cumulative reported case count was less than 10 cases. We also specified the onset time of the first social-distancing period,  $\sigma$ , as the earliest day on which the cumulative reported case count was 200 cases or more. A full description of model parameters is given in Lin et al. [22].

## 2.2. Simulations

Each region-specific model consists of a coupled system of ordinary differential equations (ODEs), which are given by Lin et al. [22]. The ODEs were numerically integrated using the SciPy [31] interface to LSODA [32] and the BioNetGen [33] interface to CVODE [34]. Python code was converted to machine code using Numba [35]. The initial conditions were determined as in Lin et al. [22].

## 2.3. Calculation of Epidemic Parameters $\mathcal{R}_0$ and $\lambda$

To find the basic reproduction number  $\mathcal{R}_0$ , we considered a reduced form of the model of Lin et al. [22], which is given in Equations (1)–(8) of the Supplementary Materials Text S1. The reduced model omits consideration of interventions, including social distancing, quarantine, and self-isolation, which are all considered in the full model. From the reduced model, we derived an expression for  $\mathcal{R}_0$  by applying the next-generation matrix method [36]. In this procedure,  $\mathcal{R}_0$  is determined as the spectral radius of the so-called next-generation matrix. Denoting this matrix as  $\mathcal{N}$ , the  $(i, j)$  entry of  $\mathcal{N}$  is the expected number of new infections in the  $i$ th compartment produced by persons initially in the  $j$ th compartment. The expression for  $\mathcal{R}_0$  given in the Results section below was obtained using Mathematica [37]. The matrix  $\mathcal{N}$  was obtained using Mathematica's LinearSolve function, and  $\mathcal{R}_0$  was computed as the dominant eigenvalue of  $\mathcal{N}$ .

To characterize the initial rate of exponential growth for a local epidemic within a given region, we computed the epidemic growth rate  $\lambda$  as the dominant eigenvalue of the Jacobian of the reduced model linearized at the disease-free equilibrium [38]. The derivation of  $\lambda$  is provided in the SI.

## 2.4. Bayesian Inference

To infer region-specific values of adjustable model parameters (and  $\mathcal{R}_0$  and HIT estimates), we followed the Bayesian inference approach of Lin et al. [22]. In inferences, we used all region-relevant confirmed COVID-19 case-count data available in the GitHub repository maintained by *The New York Times* newspaper [23] for the period starting on 21 January 2020 and ending on 21 May 2020, 21 June 2020, or 21 July 2020 (inclusive dates). The first case in the US was reported on 21 January 2020 [39]. We focused on early surveillance data (vs. all available surveillance data) so as to characterize COVID-19 transmission within populations that are nearly wholly susceptible. Markov Chain Monte Carlo (MCMC) sampling was performed using the Python code of Lin et al. [22] and a new release of PyBioNetFit [40], version 1.1.9, which includes an implementation of the adaptive MCMC method used in the study of Lin et al. [22]. Inference job setup files for PyBioNetFit, including data files, are provided for each of the 50 states online (<https://github.com/lanl/PyBNF/tree/master/examples/Mallela2021States> (accessed on 19 September 2021)). Results from both methods were found to be consistent (Supplementary Materials Figure S1). To ensure that the MCMC sampling procedures converged, we visually inspected trace plots for log-likelihood (Supplementary Materials Figure S2), parameters (Supplementary Materials Figure S3) and pairs plots (Supplementary Materials Figure S4). We also performed simulations using maximum likelihood estimates (MLEs)

for parameter values to assess the agreement of the simulations with the training data (Supplementary Materials Figure S5).

The maximum a posteriori (MAP) estimate of a parameter is the value of the parameter corresponding to the peak of its marginal posterior distribution, where probability density is highest. Because we assumed a proper uniform prior distribution for each of the adjustable parameters, as in the study of Lin et al. [22], the MAP estimates are MLEs.

### 3. Results

#### 3.1. Bayesian Uncertainty Quantification

Following the Bayesian inference approach of Lin et al. [22], we quantified uncertainty in the predicted trajectories of confirmed case counts for all 50 states, using data from 21 January to 21 June 2020. As illustrated in Figure 1 for the states of New Jersey, Wyoming, Florida, and Alaska, we find that each region-specific model parameterized on the basis of our MCMC sampling procedure reproduces the corresponding surveillance data over the period of interest. Results for the remaining states are shown in Supplementary Materials Figure S5. At the end of each MCMC sampling procedure, we obtained a marginal posterior distribution for  $\beta$  (the rate constant in the model for disease transmission) which provides a probabilistic characterization of region-specific SARS-CoV-2 transmissibility. If the marginal posterior was narrow, we have high confidence in the MAP estimate of  $\beta$ ; if it is wide, we had less confidence in its value. Each state-specific marginal posterior yielded a MAP estimate for  $\beta$ .

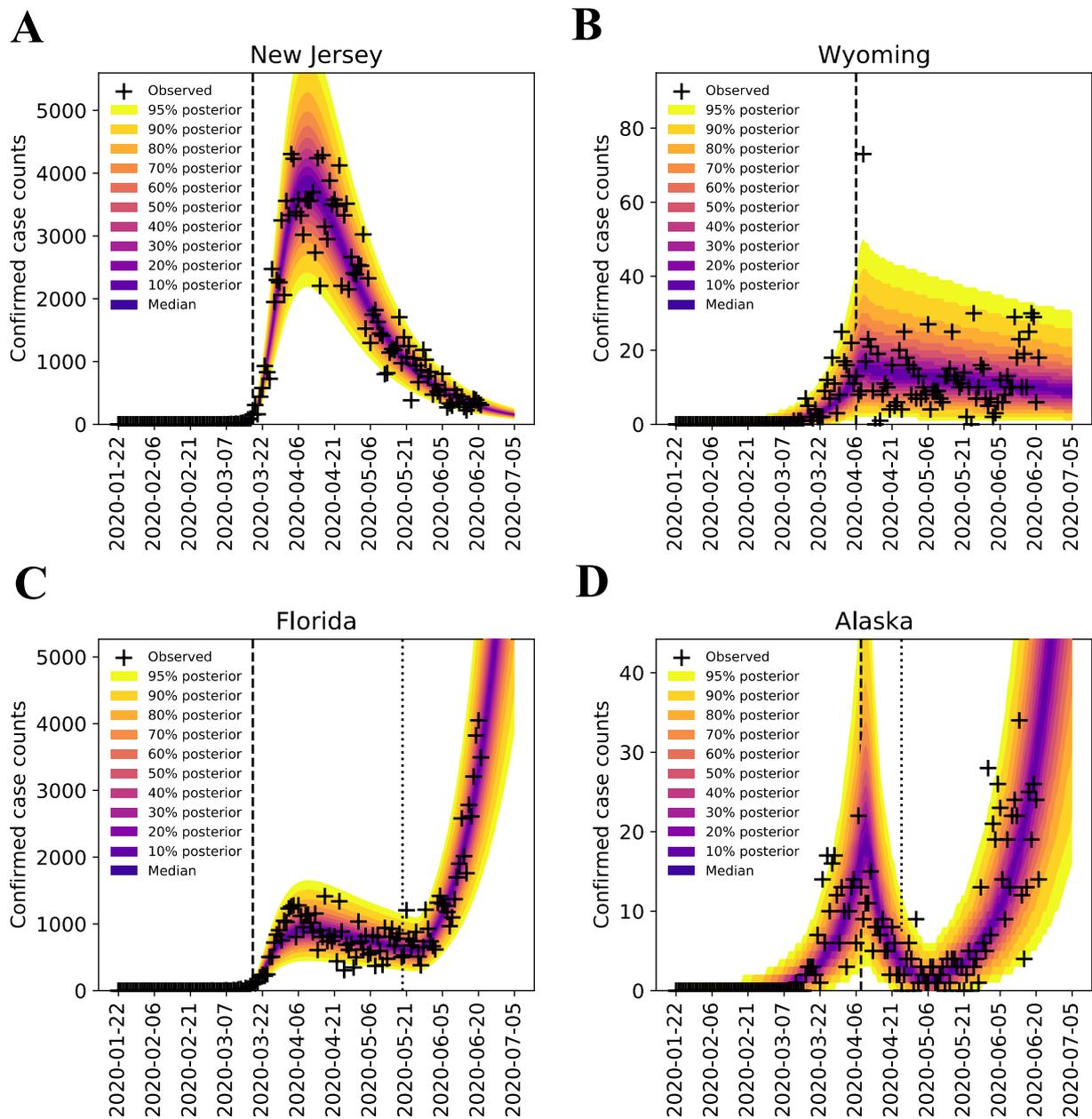
We can propagate the uncertainty in  $\beta$  into uncertainty in  $\mathcal{R}_0$  and HIT estimates, using the formula for  $\mathcal{R}_0$  given below and  $\text{HIT} = 1 - 1/\mathcal{R}_0$ . In Figure 2, we show marginal posterior distributions for  $\mathcal{R}_0$  and HIT for the states of New Jersey, Wyoming, Florida, and Alaska. We provide MAP estimates of the model parameters for all states in Supplementary Materials Table S1. The model parameters were found to be identifiable in practice (we had no proof of identifiability). MAP estimates for  $\mathcal{R}_0$  and HIT for all 50 states are provided in Supplementary Materials Table S2. These tables also provide 95% credible intervals. These estimates characterize the infectiousness of SARS-CoV-2 ancestral strains in each region of interest.

#### 3.2. Region-Specific Basic Reproduction Numbers and Herd Immunity Thresholds

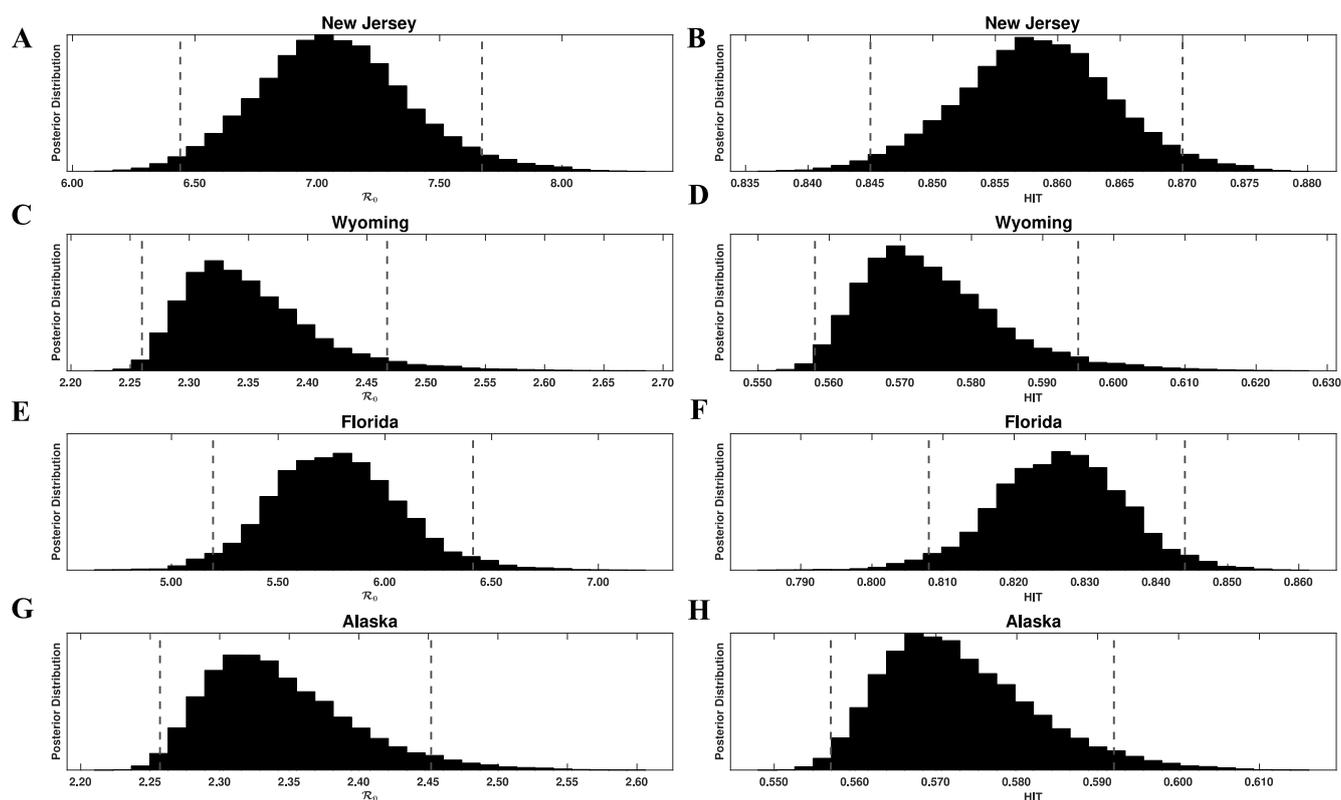
To calculate the herd immunity threshold (HIT) for a specific region, we need to know the corresponding region-specific value of the basic reproduction number  $\mathcal{R}_0$ , which is given by the following formula (obtained as described in Materials and Methods and Supplementary Materials Text S1):

$$\mathcal{R}_0 = \beta \times \left( \frac{1 - f_A}{c_I} + \frac{f_A \rho_A}{c_A} + \frac{(m - 1) \rho_E}{k_L} \right) \quad (1)$$

where  $\beta$  characterizes the rate of transmission attributable to contacts between persons who are not protected by social distancing,  $f_A$  denotes the fraction of infected persons who never develop symptoms (i.e., the fraction of asymptomatic cases),  $c_A$  characterizes the rate at which asymptomatic persons recover during the immune clearance phase of infection,  $c_I$  characterizes the rate at which symptomatic persons with mild disease recover or progress to severe disease,  $\rho_E$  is a constant characterizing the relative infectiousness of presymptomatic persons compared to symptomatic persons (with the same behaviors),  $\rho_A$  is a constant characterizing the relative infectiousness of asymptomatic persons compared to symptomatic persons (with the same behaviors),  $m$  denotes the number of stages in the incubation period, and  $k_L$  characterizes disease progression from one stage of the incubation period to the next and ultimately to an immune clearance phase. The value of  $\mathcal{R}_0$  depends on one inferred region-specific parameter,  $\beta$ , and seven fixed parameters, which have values taken to be applicable for all regions (i.e.,  $f_A, c_A, c_I, \rho_E, \rho_A, k_L$ , and  $m$ ). Estimates of these fixed parameters were taken from Lin et al. [22].



**Figure 1.** Bayesian predictive inferences for daily confirmed case counts of COVID-19 in (A) New Jersey (B) Wyoming (C) Florida (D) Alaska, from 21 January to 21 June 2020 (inclusive dates). The compartmental model [22] accounts for an initial social distancing period followed by  $n$  additional periods. We considered  $n = 0, 1$  and  $2$  and selected the best  $n$  using the model selection procedure of Lin et al. [22]. Plus signs indicate daily case reports. The shaded region indicates the prediction uncertainty and inferred noise in the detection of new cases. The color-coded bands within the shaded region indicate the median and different credible intervals (e.g., the dark purple band corresponds to the median, the band with lightest shade of yellow corresponds to the 95% credible interval, and gradations of color between these two extremes correspond to different credible intervals, as indicated in the legend). In each panel, the vertical broken line indicates the onset time of the first social-distancing period. For states with  $n = 1$  (Alaska and Florida), there is an additional broken line, which indicates the onset time of the second social-distancing period. The model was used to make forecasts of new case detection for 14 days after 21 June 2020. The last prediction date was 5 July 2020.



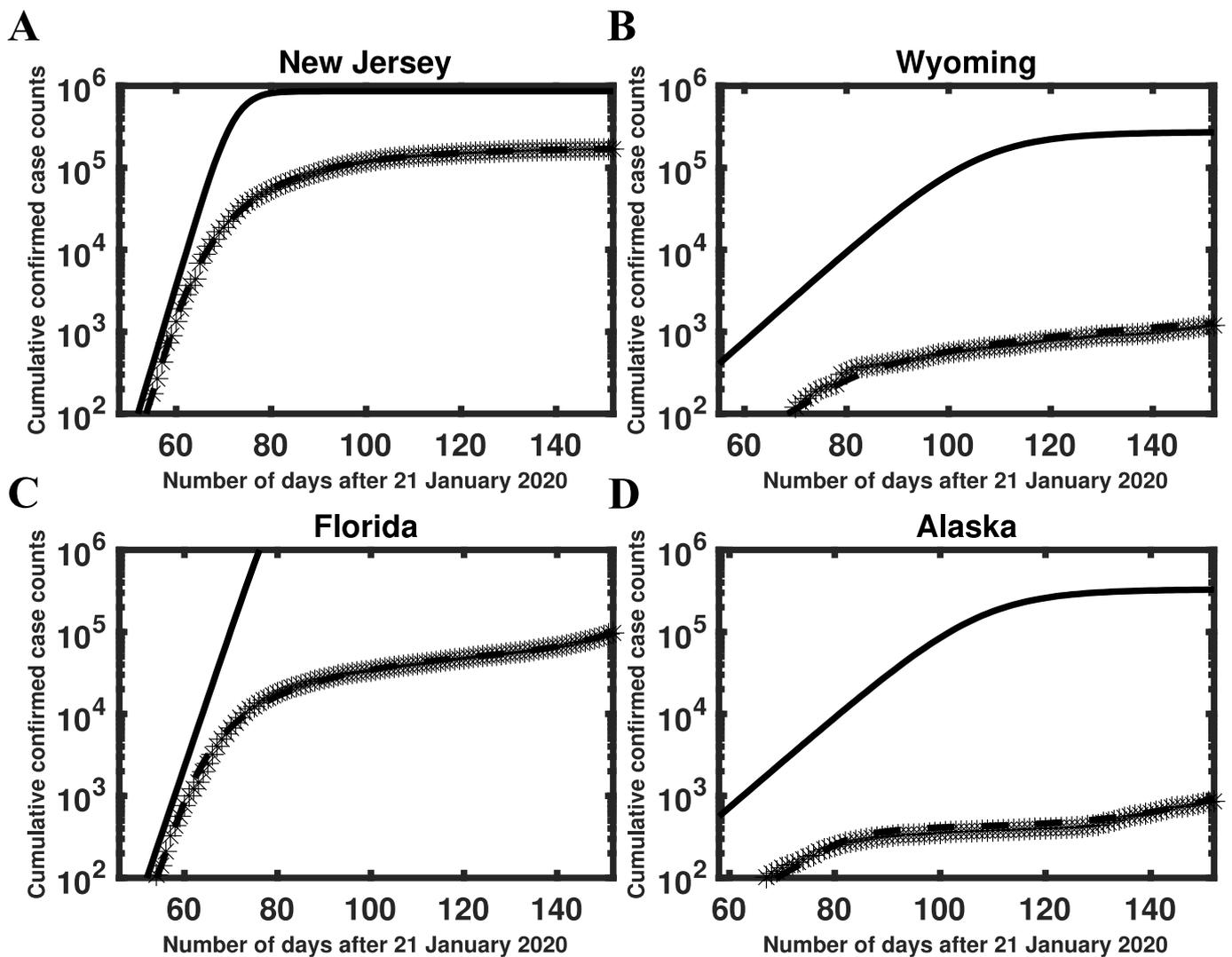
**Figure 2.** Marginal posterior distributions of  $\mathcal{R}_0$  (left panels) and HIT (right panels) for ancestral strains of SARS-CoV-2 in four US states: (A,B) New Jersey, (C,D) Wyoming, (E,F) Florida, and (G,H) Alaska. Inferences are based on daily reports of new cases from 21 January to 21 June 2020. Each  $\mathcal{R}_0$  posterior was obtained from the corresponding marginal posterior for  $\beta$  and Equation (1). Each HIT posterior was obtained from the relation  $\text{HIT} = 1 - 1/\mathcal{R}_0$  and the corresponding marginal posterior for  $\mathcal{R}_0$ . The 95% credible intervals for  $\mathcal{R}_0$  are as follows: (6.44, 7.67) for New Jersey, (2.26, 2.47) for Wyoming, (5.20, 6.41) for Florida, and (2.26, 2.45) for Alaska. The 95% credible intervals for the HIT estimates are as follows: (0.84, 0.87) for New Jersey, (0.56, 0.59) for Wyoming, (0.81, 0.84) for Florida, and (0.56, 0.59) for Alaska. For each panel, the endpoints of the corresponding credible interval are indicated with vertical broken lines.

The SARS-CoV-2 variant Delta (lineage B.1.617.2) has been estimated to be 1.64 times more infectious than variant Alpha (lineage B.1.1.7) [28], which has been estimated to be 1.50 times more infectious than ancestral strains [27]. Assuming that Delta is the dominant circulating SARS-CoV-2 strain throughout the US (as of 20 September 2021) and that  $\beta$  for Delta is  $1.64 \times 1.50 = 2.46$  times greater than  $\beta$  for ancestral strains (with other parameters in Equation (1) remaining the same), the MAP estimate of the Delta-adjusted  $\mathcal{R}_0$  ranges from 5.6 for Wyoming to 18 for New Jersey (from the multiplier given above and Supplementary Materials Table S2). The population-weighted Delta-adjusted  $\mathcal{R}_0$  for the US is 12. These estimates indicate that the herd immunity threshold (HIT) for the Delta variant of SARS-CoV-2 ranges from 82% to 94%.

### 3.3. Estimates of Initial Region-Specific Epidemic Growth Rates

HIT estimates were directly determined by estimates of the basic reproduction number, which were related to the initial growth rate of the epidemic in a given region. Here, our  $\mathcal{R}_0$  estimates were conditioned on a compartmental model that has been parameterized to reproduce case-reporting data available for each region over a five-month period (21 January to 21 June 2020). We can use parameter estimates obtained for each region to calculate the initial epidemic growth rate  $\lambda$ , which is directly comparable to early surveillance data (Figure 3 and Supplementary Materials Figure S6). We provide MAP estimates and

95% credible intervals for  $\lambda$ ,  $\mathcal{R}_0$ , and HIT for selected states in Table 1. MAP estimates and 95% credible intervals for  $\lambda$ ,  $\mathcal{R}_0$ , and HIT for all states are provided in Supplementary Materials Table S2. These estimates are based on the state-specific marginal posteriors for the parameter  $\beta$  of our compartmental model. State-specific MAP estimates and 95% credible intervals for  $\beta$  (and other adjustable model parameters) are given in Supplementary Materials Table S1. As can be seen (e.g., in Figure 3), our  $\lambda$  estimates are consistent with early case reporting data during the exponential takeoff phase of disease transmission.



**Figure 3.** Consistency of model-derived  $\lambda$  estimates with empirical growth rates during initial exponential increase in disease incidence in (A) New Jersey, (B) Wyoming, (C) Florida, and (D) Alaska. In each panel, the initial slope of the solid curve corresponds to  $\lambda$  (calculated as described in Materials and Methods), the crosses indicate empirical cumulative case counts, and the broken line is the model prediction based on MAP estimates for adjustable parameters. The solid curve is derived from the reduced model (Equations (1)–(8) in the Supplementary Materials Text S1). This curve shows cumulative case counts had there not been any interventions to limit disease transmission. As can be seen, the initial slopes of the solid and broken curves are comparable. We selected  $n = 0$  for New Jersey and Wyoming and  $n = 1$  for Florida and Alaska. Among 35 states with  $n = 0$ , New Jersey had the largest inferred  $\lambda$  value (0.45) and Wyoming had the smallest inferred  $\lambda$  value (0.13). Among 15 states with  $n = 1$ , Florida had the largest inferred value of  $\lambda$  (0.39) and Alaska had the smallest inferred value of  $\lambda$  (0.13). It should be noted that, in contrast with Figure 1, the y-axis here indicates cumulative (vs. daily) number of cases on a logarithmic (vs. linear) scale.

**Table 1.** Maximum a posteriori (MAP) estimates and 95% credible intervals for epidemic parameters ( $\beta$ ,  $\lambda$ ,  $\mathcal{R}_0$ , HIT, and Delta-adjusted HIT) for the states of New Jersey, Wyoming, Florida, and Alaska.

State	$\beta$ ( $d^{-1}$ )	$\lambda$ ( $d^{-1}$ ) *	$\mathcal{R}_0$ **	HIT ***	Delta-Adjusted HIT ****
New Jersey	0.65 (0.59–0.71)	0.45 (0.41–0.48)	7.1 (6.4–7.7)	0.86 (0.84–0.87)	0.94 (0.94–0.95)
Wyoming	0.21 (0.21–0.23)	0.13 (0.13–0.15)	2.3 (2.3–2.5)	0.56 (0.56–0.59)	0.82 (0.82–0.84)
Florida	0.55 (0.48–0.59)	0.39 (0.34–0.41)	6.0 (5.2–6.4)	0.83 (0.81–0.84)	0.93 (0.92–0.94)
Alaska	0.21 (0.21–0.23)	0.13 (0.13–0.14)	2.3 (2.3–2.5)	0.56 (0.56–0.59)	0.82 (0.82–0.84)

In this analysis, we used surveillance data (daily reports of new cases) available from 21 January 2020 to 21 June 2020 (inclusive dates) to estimate parameter values through Bayesian inference. \* Computed as described in SI. \*\* Calculated using Equation (1). \*\*\* Obtained through the relation  $HIT = 1 - 1/\mathcal{R}_0$ . \*\*\*\* Based on Delta being 2.46 times more infectious than ancestral strains.

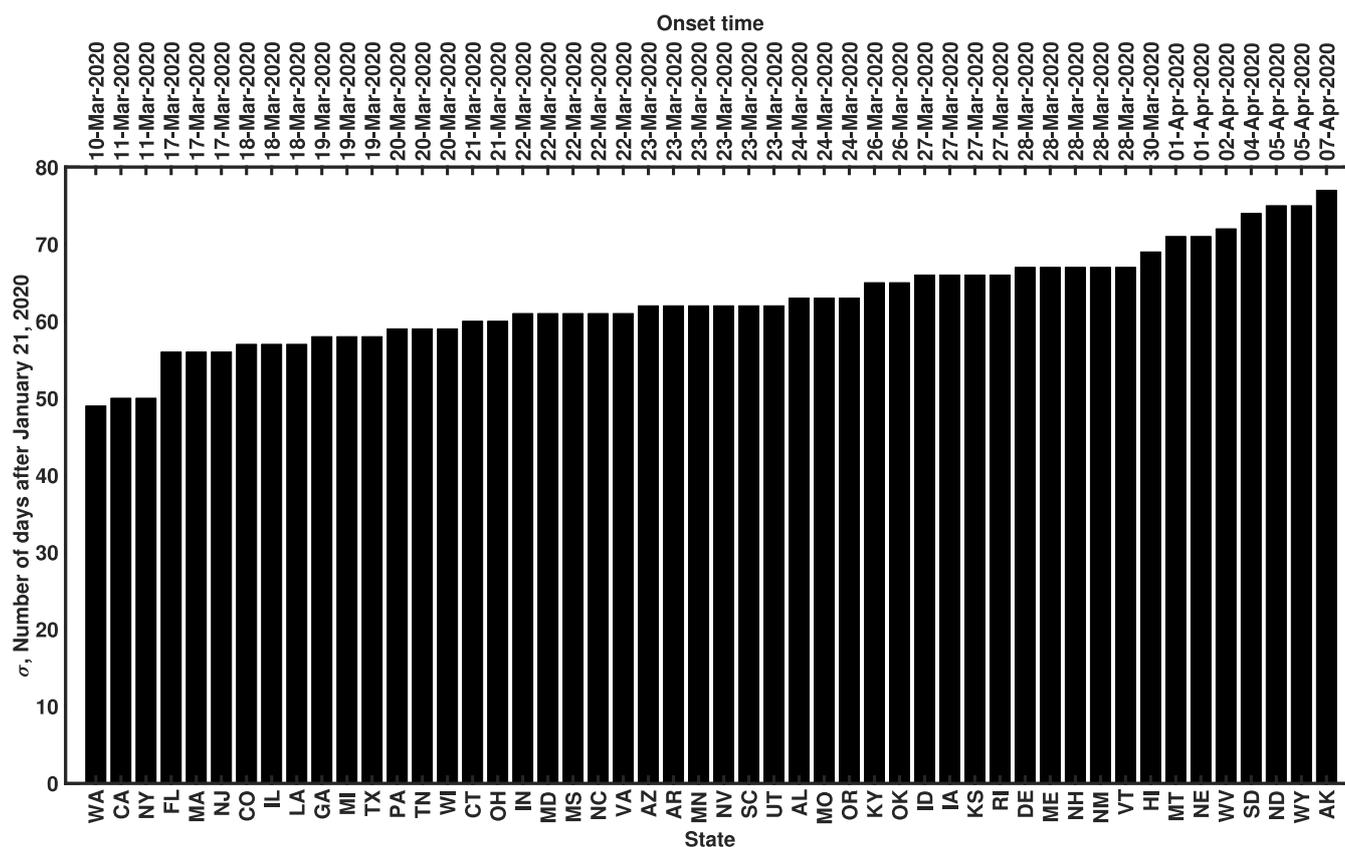
### 3.4. Sensitivity of $\beta$ to the Surveillance Data Used in Inference

For each state, we used surveillance data available from 21 January to 21 June 2020 to infer the MAP estimate of  $\beta$  (and the values of the other region-specific adjustable model parameters). This time window encompasses the onset time  $\sigma$  for all 50 states (Figure 4), which ranged from 10 March to 7 April 2020. Recall that  $\sigma$  is a region-specific parameter of the model of Lin et al. [22], which we take as the first time at which the cumulative confirmed case count for a given state was 200 or more. The value of  $\sigma$  provides a rough estimate of the start of sustained community transmission. To check the robustness of the MAP estimates for  $\beta$  to variations in training data, we performed a sensitivity analysis wherein we inferred  $\beta$  using data collected over three distinct periods in 2020, namely: (1) 21 January to 21 May, (2) 21 January to 21 June, and (3) 21 January to 21 July 2020. By visualizing our estimates with a rank order plot (Figure 5) and conducting pairwise two-sample Kolmogorov–Smirnov tests [41], we found that the 4-, 5-, and 6-month training datasets yielded estimates for  $\beta$  that were not statistically significantly different from each other. The MAP estimates for  $\beta$  obtained using the 4-, 5-, and 6-month datasets are listed in Supplementary Materials Table S3. We assessed sensitivity by computing the relative error between the  $\beta$  estimates obtained from the 5-month dataset and the average  $\beta$  estimate over all datasets considered. We found that none of the state-level MAP estimates for  $\beta$  showed sensitivity (i.e., a relative error exceeding 100% in magnitude) to variations in the training data (Supplementary Materials Table S4). The largest relative error was 12% (for Kansas).

### 3.5. Global Asymptotic Stability of the Disease-Free Equilibrium

The model of Lin et al. [22] has a globally asymptotically stable disease-free equilibrium (DFE) if  $\mathcal{R}_0 < 1$ , which can be deduced by following the approach of Shuai and van den Driessche [42]. As a consequence, the model predicts that the epidemic will be extinguished as the system dynamics are attracted to the DFE.

To confirm that the model behaves as expected around the HIT, we conducted a perturbation analysis for the states of New York (Figure 6A,B) and Washington (Figure 6C,D). We simulated disease dynamics starting from an arbitrarily chosen initial condition near the HIT number of persons,  $S_h$ , given by the following formula:  $S_h = HIT \times S_0$ , where  $S_0$  denotes the population size of the region considered. We defined the size of our perturbation as  $\varepsilon = 0.2 \times S_h$  for Figure 6A,C and as  $\varepsilon = -0.2 \times S_h$  for Figure 6B,D. The initial condition was  $S_0 - S_h - 1 + \varepsilon$  susceptible persons, 1 infected person, and  $S_h - \varepsilon$  recovered persons. As expected, for  $S_h < HIT \times S_0$  (Figure 6A,C), the number of infectious persons grows over time, whereas for  $S_h > HIT \times S_0$  (Figure 6B,D), the number of infectious persons decays over time.



**Figure 4.** Onset times of COVID-19 disease transmission for ancestral strains of SARS-CoV-2 in all 50 US states. The onset time  $\sigma$  is defined as the first day on which the cumulative reported case count was 200 cases or more. Dates corresponding to  $\sigma$  values on the vertical axis are indicated above each bar. States are indicated using two-letter US postal service state abbreviations (<https://about.usps.com/who-we-are/postal-history/state-abbreviations.pdf> (accessed on 19 September 2021)).

In the two scenarios considered above (i.e., introduction of an infected person into a disease-free population with or without herd immunity), the rate at which disease burden changes is sensitive to different factors (Figure 7). As illustrated in Figure 7A, the rate at which disease burden decreases in a population with herd immunity (as in the scenario considered in Figure 6B,D) depends sensitively on the duration of the incubation period. As illustrated in Figure 7B, the rate at which disease burden increases (as in the scenario considered in Figure 6A,C) depends sensitively on the size of the subpopulation of susceptible persons.

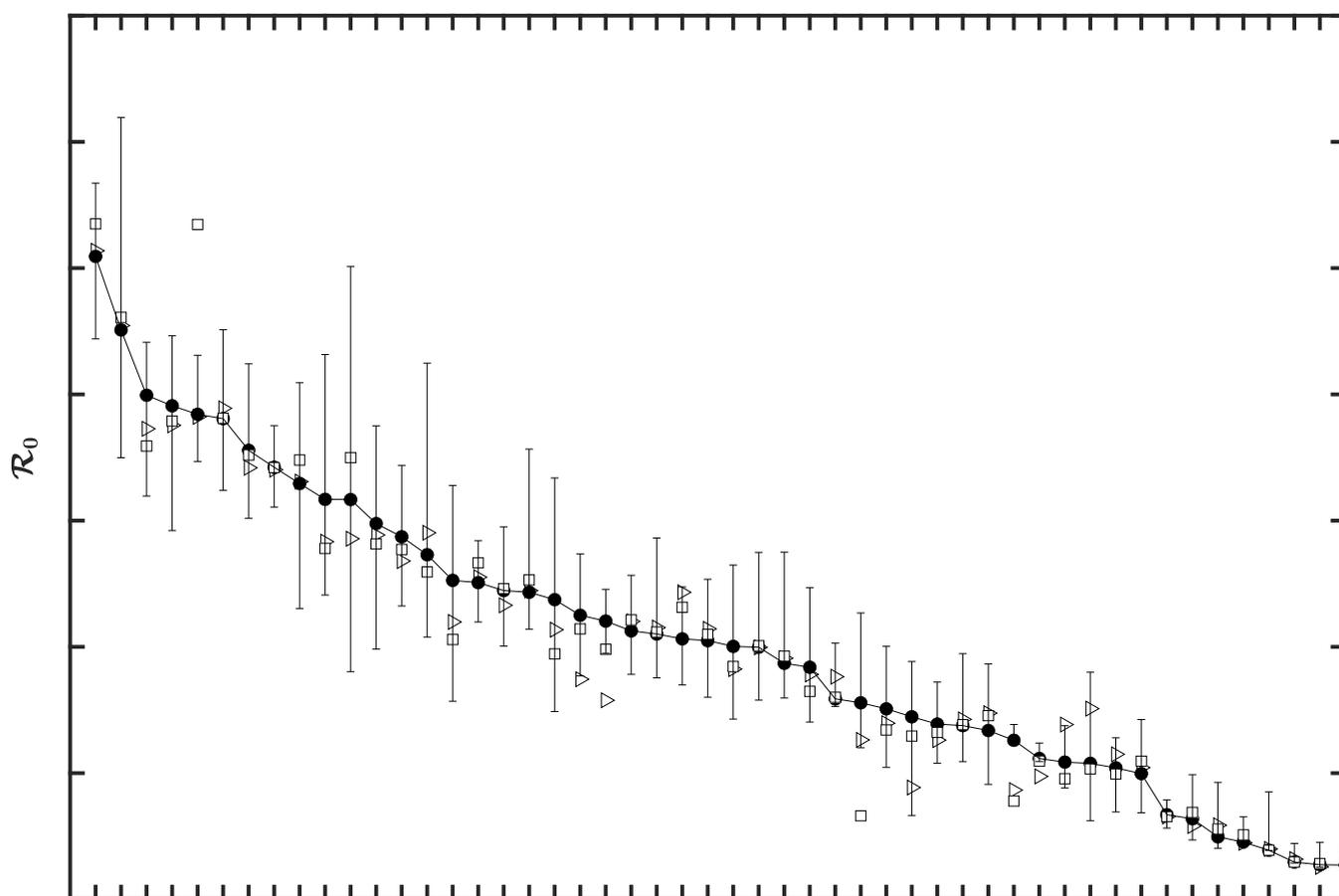
### 3.6. Progress toward Herd Immunity

From our state-specific HIT estimates and other information (discussed below), we were able to calculate percent progress toward herd immunity for each state (Figure 8, Supplementary Materials Table S5). We estimated the percent progress of each state's population toward herd immunity,  $\mathcal{P} \in [0\%, 100\%]$ , using the following equation (the derivation of which is given in the Supplementary Materials Text S1):

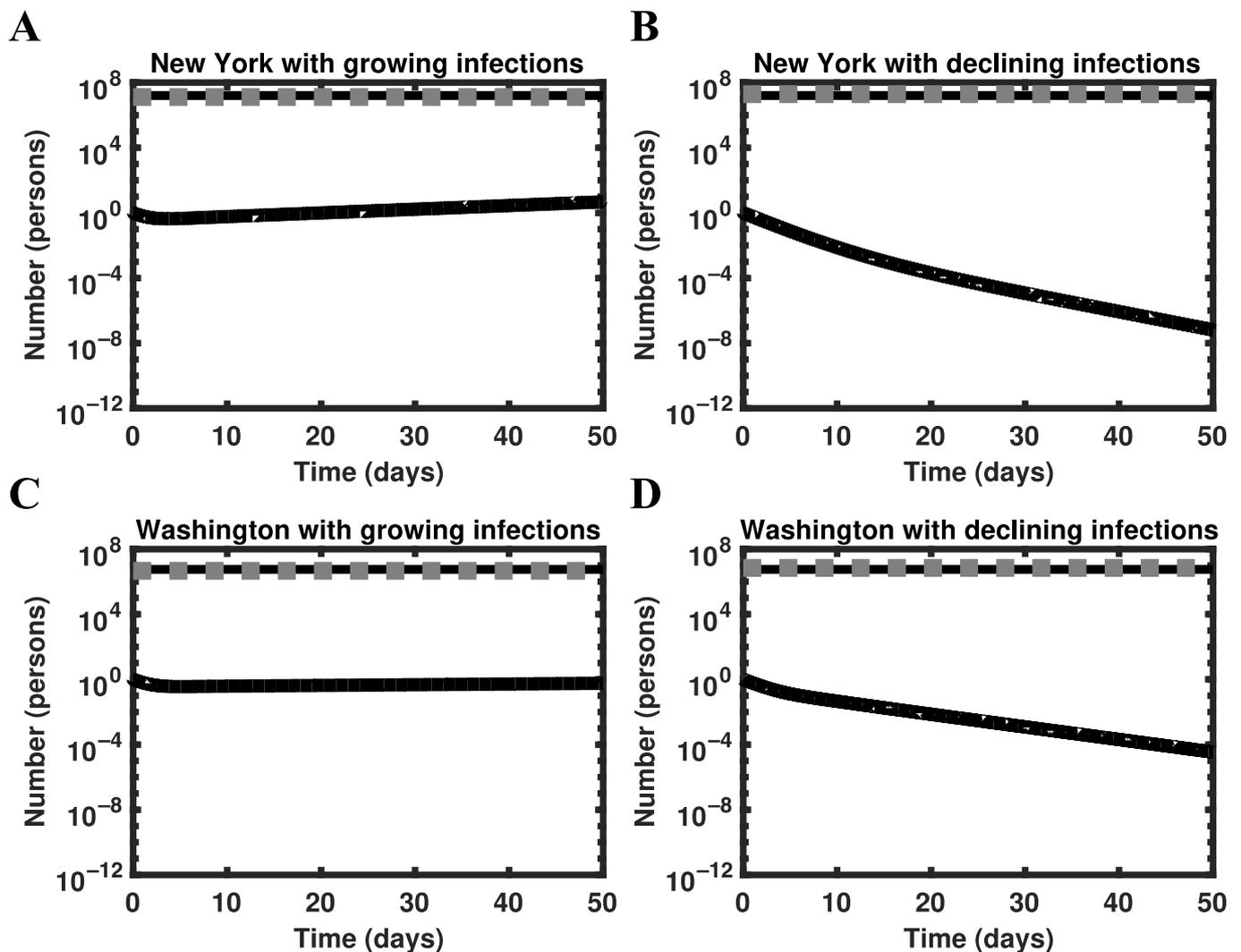
$$\mathcal{P} \equiv (\varepsilon_v(1 - f_r)f_v + \varepsilon_r f_r) \left(1 - \frac{1}{Y_{Delta} \mathcal{R}_0}\right)^{-1} \times 100\% \quad (2)$$

where  $\mathcal{R}_0$  is the population-specific basic reproduction number that we estimated for ancestral strains (Supplementary Materials Table S2),  $Y_{Delta}$  is a multiplier that accounts for the increased transmissibility of SARS-CoV-2 variant Delta,  $f_r$  denotes the fraction of the population with immunity acquired through infection,  $f_v$  is the fraction of the population

that has been vaccinated [24],  $\varepsilon_r$  is the fraction of infected persons who are protected against productive infection (i.e., an infection that can be transmitted to others), and  $\varepsilon_v$  is the fraction of vaccinated persons who are protected against productive infection. Recall that we use  $Y_{Delta} = 2.46$  [27,28]. We estimate that  $\varepsilon_r = 1.0$  [43] and  $\varepsilon_v = 0.66$  [44]. We obtain four different estimates for  $f_r$  as follows. In the first case, we obtain  $f_r$  as the cumulative number of detected cases within a population divided by the population size. In the second case, we adjust our previous estimate for  $f_r$  by a multiplier of 5.8 [45]. In other words, we assume that the true disease burden is 5.8 times higher than the detected number of cases. In the third case, we obtain  $f_r$  as the fraction of the population that has been infected according to the latest serological survey results reported online at Ref. [25]. In the fourth case, we assume  $f_r = f_{r,0}/(1 - f_A)$ , where  $f_{r,0}$  denotes the estimate of seroprevalence in a given region and  $f_A$  denotes the fraction of all cases that are asymptomatic. With this approach, we are assuming that asymptomatic cases are not detected in serological testing [46]. We adopt the estimate of Lin et al. [22] that  $f_A = 0.44$ .

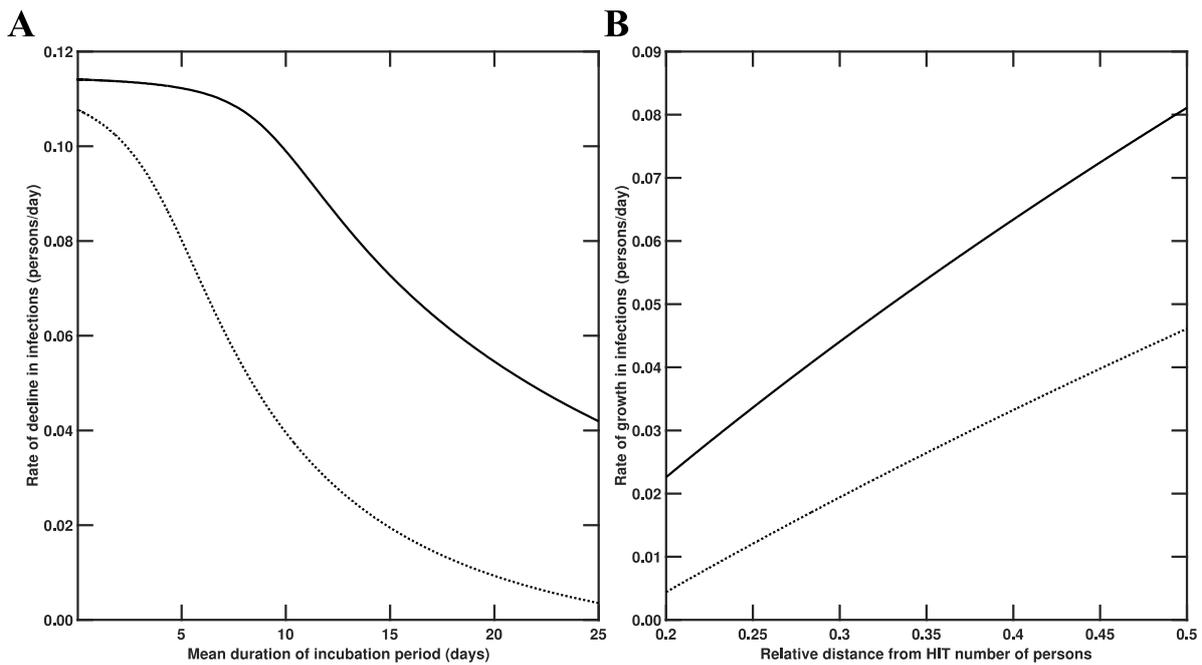


**Figure 5.** MAP estimates of the basic reproduction number  $\mathcal{R}_0$  for ancestral strains of SARS-CoV-2 in all 50 US states. The different symbols refer to different training datasets used to estimate  $\mathcal{R}_0$ . Open triangles correspond to surveillance data collected from 21 January to 21 May 2020, filled circles correspond to surveillance data collected from 21 January to 21 June 2020, and open squares correspond to surveillance data collected from 21 January to 21 July 2020. Estimates of  $\mathcal{R}_0$  are sorted by state from largest to smallest values according to the  $\mathcal{R}_0$  estimates derived from the surveillance data collected for 21 January to 21 June 2020. The whiskers associated with each filled circle indicate the 95% credible interval (inferred from the 5-month dataset). States are indicated using two-letter US postal service state abbreviations (<https://about.usps.com/who-we-are/postal-history/state-abbreviations.pdf> (accessed on 19 September 2021)).

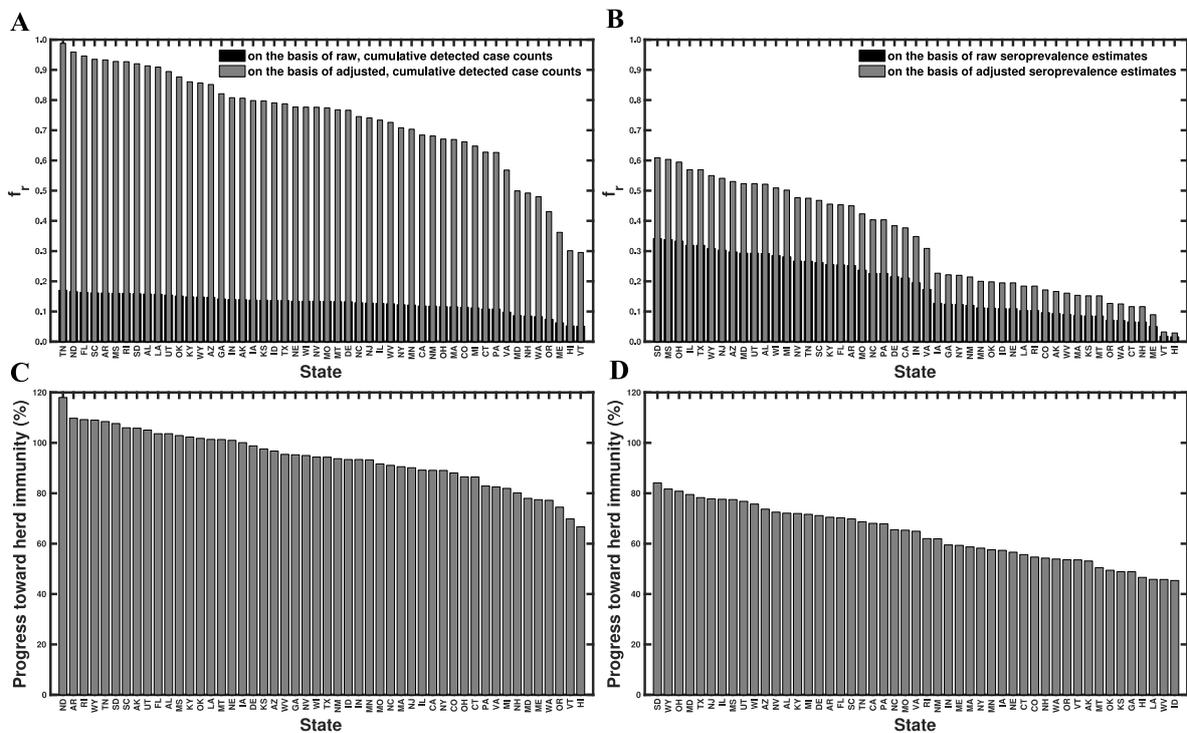


**Figure 6.** Perturbation analysis using the full model of Lin et al. [22] for the states of New York (panels (A,B)) and Washington (panels (C,D)). In each panel, the black solid line represents the number of infectious persons (initially 1), the black broken line represents the threshold number of persons required for herd immunity (i.e.,  $S_h$ ), and the gray broken line represents the number of recovered persons (initially  $S_h - \epsilon$ , obtained as described in Results). Simulations are based on MAP estimates for model parameters obtained using surveillance data collected from 21 January to 21 June 2020.

As can be seen in Figure 8C, which is based on case reporting data, 18 of the 50 states have reached herd immunity. However, in Figure 8D, which is based on serological survey data, none of the states have reached herd immunity. South Dakota is closest to herd immunity, with the 84% of immune persons required for herd immunity. Idaho is furthest from herd immunity, with 45% of the immune persons required for herd immunity. The mean (median) progress toward herd immunity, across all states, is 63% (63%).



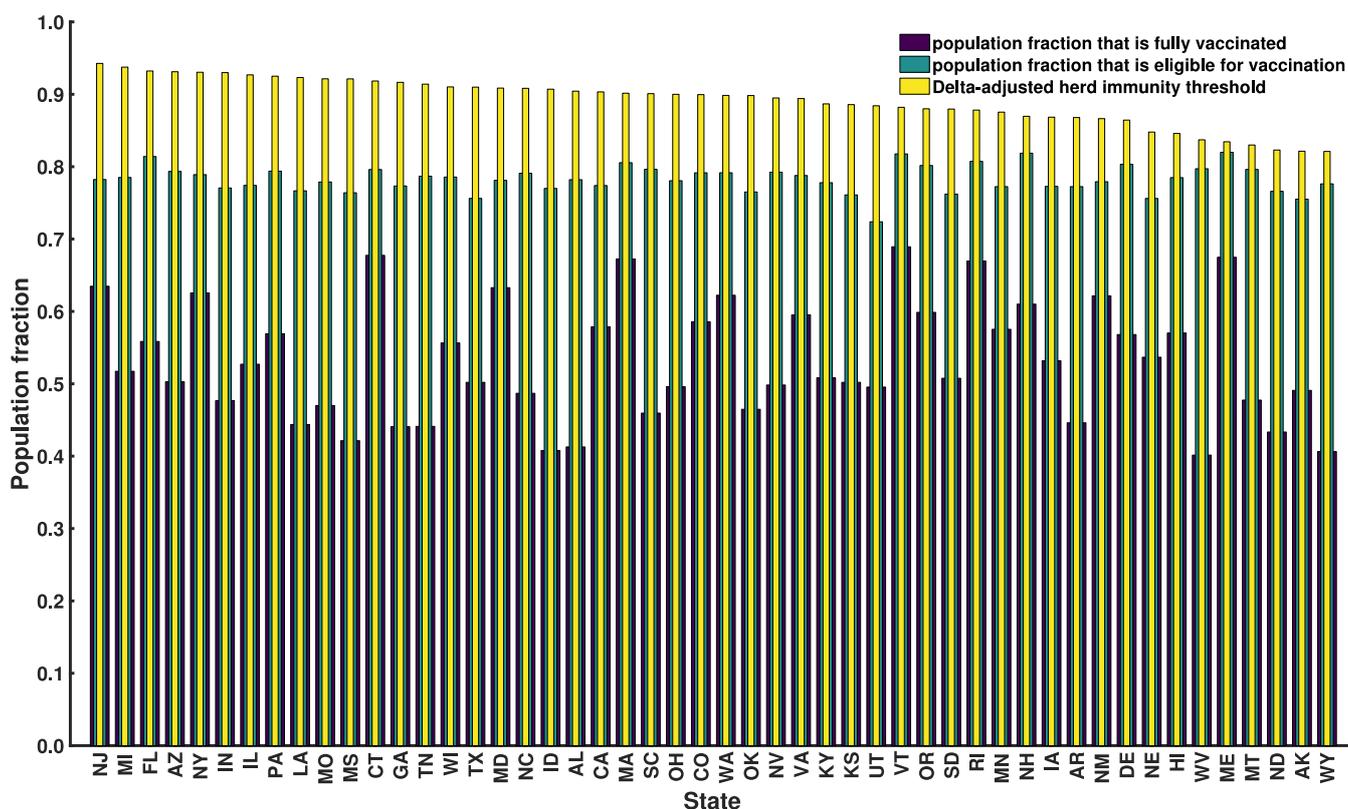
**Figure 7.** Dependence of disease burden on key model parameters for the states of New York and Washington. In each panel, the solid line corresponds to New York and the broken line corresponds to Washington. In Panel (A), rate of decline in infections is plotted as a function of the mean duration of the incubation period (in days), which is obtained as  $m/k_L$ , where  $m$  is the number of stages in the incubation period and  $k_L$  characterizes disease progression from one stage to the next. We take  $m = 5$  as in the study of Lin et al. [22]. In Panel (B), the rate of growth in infections is plotted as a function of the relative distance from the herd immunity threshold number of persons, which is defined as  $\epsilon/S_H$ .



**Figure 8.** Percent progress toward herd immunity in each of the 50 US states. Percent progress  $\mathcal{P}$  indicates the fraction of immune persons required for herd immunity.  $\mathcal{P}$  was calculated using

Equation (2). Black bars (Panel (A)) correspond to the first scenario (i.e.,  $f_r$  estimated as the number of detected cases divided by population size), gray bars (Panels (A,C)) correspond to the second scenario (i.e.,  $f_r$  estimated as the number of detected cases within a population divided by the population size, adjusted for lack of detection of undiagnosed SARS-CoV-2 infections), black bars (Panel (B)) correspond to the third scenario (i.e.,  $f_r$  given by seroprevalence survey results), and gray bars (Panels (B,D)) correspond to the fourth scenario (i.e.,  $f_r$  given by seroprevalence survey results adjusted for lack of detection of asymptomatic cases). Estimates for  $\mathcal{P}$  are sorted by state from largest to smallest values according to the second scenario (Panels (A,C)) and the fourth scenario (Panels (B,D)). North Dakota was omitted from Panels (B,D) because a recent estimate of seroprevalence was not available at Ref. [25]. States are indicated using two-letter US postal service state abbreviations (<https://about.usps.com/who-we-are/postal-history/state-abbreviations.pdf> (accessed on 19 September 2021)).

In Figure 9, we show the fraction of each state's population that has been vaccinated and the fraction that is eligible for vaccination based on data available as of 20 September 2021. Vaccination data was taken from Ref. [24]. We assumed that only persons 18 years or older were eligible for vaccination. Age data were taken from Ref. [47]. In Figure 9, we also show Delta-adjusted HITs from Supplementary Materials Table S2. As can be seen, vaccine coverage is below that required for herd immunity (in the face of Delta) in all cases, even if we take vaccines to provide sterilizing immunity in 100% of cases. For example, vaccine coverage for New Jersey (Wyoming) is 63% (41%) (Figure 9) and the corresponding Delta-adjusted HIT is 94% (82%) (Figure 9, Table 1, Supplementary Materials Table S2). It seems that herd immunity cannot be reached through vaccination alone.



**Figure 9.** Vaccine eligibility and vaccine coverage in each of the 50 US states on 20 September 2021. Purple bars correspond to vaccine coverage, i.e., the population fraction that is fully vaccinated [24]. Teal bars correspond to vaccine eligibility, i.e., the population fraction that is eligible for vaccination. We estimated the eligible population fraction as the adult fraction of the population [47], i.e., the

population fraction 18 years or older. Yellow bars correspond to Delta-adjusted HIT estimates from Supplementary Materials Table S2. States are indicated using two-letter US postal service state abbreviations (<https://about.usps.com/who-we-are/postal-history/state-abbreviations.pdf> (accessed on 19 September 2021)).

#### 4. Discussion

One of our most important findings is a quantification of how COVID-19 transmissibility, in terms of the basic reproduction number  $\mathcal{R}_0$ , varies across the 50 US states. The MAP value of  $\mathcal{R}_0$  for ancestral strains of SARS-CoV-2 ranges from 2.3 for Wyoming to 7.1 for New Jersey. The population-weighted mean for the US is 4.7. These estimates indicate that the herd immunity threshold (HIT) for the Delta variant of SARS-CoV-2 ranges from 82% to 94%, assuming that Delta is 2.46 times more transmissible than ancestral strains. The uncertainty in each  $\mathcal{R}_0$  estimate was quantified: 95% credible intervals are indicated in Figure 5. The 95% credible intervals for ancestral HIT estimates are given in Supplementary Materials Table S2. Because we can estimate the relative effort required to reach herd immunity across the US (in terms of HIT), resources for vaccination campaigns can be targeted to those areas where it is more difficult to achieve herd immunity.

Our  $\mathcal{R}_0$  and HIT estimates differ from estimates given in previous studies. For example, various researchers derived point estimates for  $\mathcal{R}_0$  from data using tools from time-series analysis, without assuming an underlying mechanistic model [13,15]. These tools depend on slope estimation and thus can be expected to depend sensitively on noise and errors in early case-reporting data. Ives and Bozzuto [16] provided state-level estimates for  $\mathcal{R}_0$  (in 36 states), and Fellows et al. [17] used a Bayesian framework to obtain state-level estimates for  $\mathcal{R}_0$  (in all 50 states). For the 30 states that are considered in Ives and Bozzuto [16], Fellows et al. [17], Milicevic et al. [18], and the present study, our estimates for  $\mathcal{R}_0$  were most similar to those of Milicevic et al. [18] (Supplementary Materials Table S6). Milicevic et al. [18] provided state-level  $\mathcal{R}_0$  point estimates (for 45 states) that are statistically consistent with our MAP estimates of  $\mathcal{R}_0$  for ancestral strains of SARS-CoV-2. The main points of difference between these earlier studies and the present study are as follows. Our  $\mathcal{R}_0$  and HIT estimates were obtained from a model consistent with new case-reporting data, as illustrated in Figs 1 and 3. We were able to provide estimates for all 50 states (Figure 5, Supplementary Materials Table S2), and we were able to obtain a Bayesian quantification of the uncertainty in each estimate (Figure 5, Supplementary Materials Table S2).

In the face of Delta, the estimates of Figure 8C (based on case reporting data) suggest that a majority of states have yet to achieve herd immunity, and the estimates of Figure 8D (based on serological survey results) suggest that no state in the US has achieved herd immunity as of 20 September 2021. In either case, persons in the US lacking immunity are still at risk [48]. The perspective provided by Figure 8D is consistent with the study of Moghadas et al. [49] indicating that only 62% of persons in the US had some form of immunity as of 15 July 2021 (either through infection or vaccination). Given that the percentage of immune persons required for herd immunity according to Figure 8D ranges from 84% for South Dakota to 45% for Idaho (Figure 8D) ~20 months (counting from January 2020) into the COVID-19 pandemic and ~9 months after vaccines became widely available, it seems that this situation will persist for months, if not years.

How can the US accelerate the approach to herd immunity (if herd immunity is even possible)? Policies that encourage infection of children and vaccinated persons who have healthy immune systems may be rationalized because such persons seem to be well-protected against severe (but not mild) disease [50] and infected persons seem to have greater protection against productive infection [43]. However, this approach has obvious drawbacks, starting with the risks of infection. Another is that non-immune persons may not be able to self-identify as such. Unfortunately, it seems that we cannot rely on currently available vaccines to stop community transmission. Delta-adjusted HITs are mathematically impossible to achieve through vaccination alone because these HITs are close to 1 (Supplementary Materials Table S2) and vaccine protection against productive infection is imperfect (i.e.,  $\varepsilon_v$  is significantly less than 1) [44]. This situation is exacerbated

by the emergence of the SARS-CoV-2 variant Omicron (lineage B.1.1.529) [51], which has been estimated to be roughly 2 to 4 times more transmissible than Delta [52–54]. Other factors influencing the feasibility of herd immunity are waning immunity [55–57], limited vaccine uptake, and vaccine eligibility (Figure 9). Thus, use of variant-targeted vaccines may be needed to achieve herd immunity and to minimize COVID-19 impacts.

As is well-known, population features, not just pathogen features, affect the value of  $\mathcal{R}_0$  [58]. These features potentially include numerous biological, sociobehavioral, and environmental factors, such as age, physical fitness, social network structure, population density, and aspects of the built environment. Variations in these features across regions can give rise to spatial heterogeneity in  $\beta$  and  $\mathcal{R}_0$ , although not in immediately obvious ways. One benefit of our comprehensive state-level  $\mathcal{R}_0$  estimates is that they quantify how differences in population features across the US influence the spread of an aerosol-transmitted virus [59,60]. This information, by identifying the regions in the US where transmission is likely to be highest, could be useful in preparing for and responding to future pandemics caused by viruses similar to SARS-CoV-2. Disease transmission can be reduced through nonpharmaceutical interventions, such as early detection and isolation of infected persons [61–63].

Our study has several notable limitations. Our HIT estimates are potentially biased downward because of general awareness within the US of the impacts of COVID-19 in other countries (e.g., China and Italy), which could have resulted in a fraction of the US population changing their behaviors to protect themselves from COVID-19 before the start of the local epidemic. In addition, our estimation of percent progress toward herd immunity crucially depends on the seroprevalence estimates of the true disease burden. These estimates are associated with some uncertainty [64–66]. As illustrated in Figure 8, percent progress toward herd immunity is underestimated if serological tests fail to detect all cases of infection. The reader must also be cautioned that our analysis depends on a number of assumptions. For example, we considered a compartmental model in which populations are taken to be well-mixed and to lack age structure. This is clearly a simplification. More refined estimates could be obtained by making the model more realistic, but this would have the drawback of increasing the complexity of inference, which at some point would make inference impracticable.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v14010157/s1>, Text S1: supplementary text and equations, Table S1: estimates of state-specific model parameters, Table S2: state-specific MAP estimates and 95% credible intervals for epidemic parameters ( $\lambda$ ,  $\mathcal{R}_0$ , HIT, and Delta-adjusted HIT), Table S3: state-specific MAP estimates of the rate constant  $\beta$  obtained using different training data, Table S4: sensitivity of the MAP estimate of  $\beta$  to training data, Table S5: percent progress  $\mathcal{P}$  toward herd immunity for each of the 50 states, Table S6: comparison of estimates of  $\mathcal{R}_0$  made in different studies, Figure S1: consistency of results obtained from different codes used to perform Markov chain Monte Carlo sampling, Figure S2: log-likelihood trace plots for each of the 50 US states, Figure S3, parameter trace plots for each of the 50 US states, Figure S4, matrix of 1- and 2- dimensional marginalizations of the parameter posterior samples obtained for each of the 50 US states, Figure S5, posterior predictive checking, Figure S6, consistency of model-derived  $\lambda$  estimates with empirical growth rates during initial exponential increase in disease incidence.

**Author Contributions:** A.M., R.G.P., Y.T.L. and W.S.H. designed research; A.M., J.N., E.F.M., Y.C., R.G.P., Y.T.L. and W.S.H. performed research; A.M., J.N., Y.T.L. and W.S.H. analyzed data; and A.M. and W.S.H. wrote the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** A.M. was supported by the 2020 Mathematical Sciences Graduate Internship program, which is sponsored by the Division of Mathematical Sciences of the National Science Foundation. E.F.M., J.N., Y.C., R.G.P. and W.S.H. were supported by grant R01GM111510 from the National Institute of General Medical Sciences of the National Institutes of Health. Y.T.L. was supported by the Laboratory Directed Research and Development program at Los Alamos National Laboratory. Computational resources used in this study included the FARM cluster at the University of California,

Davis and the Monsoon cluster at Northern Arizona University, which is funded by Arizona's Technology and Research Initiative Fund.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Inferences were performed using problem-specific code. The functionality of the code has been added to a freely available open-source software package (PyBioNetFit, version 1.1.9). We have confirmed that the results of the problem-specific code are reproduced by PyBioNetFit. PyBioNetFit is available online (<https://github.com/lanl/pybnf> (accessed on 29 November 2021).) along with inference job setup files (<https://github.com/lanl/PyBNF/tree/master/examples/Mallela2021States> (accessed on 29 November 2021)). The EXP files contain the data used in inference.

**Acknowledgments:** We thank R.M. Ribeiro for helpful discussions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Gee, J.; Marquez, P.; Su, J.; Calvert, G.M.; Liu, R.; Myers, T.; Nair, N.; Martin, S.; Clark, T.; Markowitz, L.; et al. First month of COVID-19 vaccine safety monitoring—United States, 14 December 2020–13 January 2021. *MMWR Morb. Mortal Wkly Rep.* **2020**, *70*, 283–288. [\[CrossRef\]](#)
2. National Center for Immunization and Respiratory Diseases (NCIRD), Data from Centers for Disease Control and Prevention (CDC). Available online: <https://data.cdc.gov/Vaccinations/\protect\unhbox\voidb@x\hbox{COVID-19}-Vaccinations-in-the-United-States-Jurisd/uns-k-b7fc> (accessed on 20 September 2021).
3. Fine, P.; Eames, K.; Heymann, D.L. Herd immunity: A rough guide. *Clin. Infect. Dis.* **2011**, *7*, 911–916. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Ridenhour, B.; Kowalik, J.M.; Shay, D.K. Unraveling  $\mathcal{R}_0$ : Considerations for public health applications. *Am. J. Public Health* **2018**, *108*, S445–S454. [\[CrossRef\]](#)
5. Temime, L.; Gustin, M.-P.; Duval, A.; Buetti, N.; Crépey, P.; Guillemot, D.; Thiébaud, R.; Vanhems, P.; Zahar, J.-R.; Smith, D.R.M.; et al. A conceptual discussion about the basic reproduction number of severe acute respiratory syndrome coronavirus 2 in healthcare settings. *Clin. Infect. Dis.* **2021**, *72*, 141–143. [\[CrossRef\]](#)
6. Dan, J.M.; Mateus, J.; Kato, Y.; Hastie, K.M.; Yu, E.D.; Faliti, C.E.; Grifoni, A.; Ramirez, S.I.; Haupt, S.; Frazier, A.; et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* **2021**, *371*, eabf4063. [\[CrossRef\]](#)
7. Yu, C.-J.; Wang, Z.-X.; Xu, Y.; Hu, M.-X.; Chen, K.; Qin, G. Assessment of basic reproductive number for COVID-19 at global level: A meta-analysis. *Medicine* **2021**, *100*, e25837. [\[CrossRef\]](#)
8. Kucharski, A.J.; Russell, T.W.; Diamond, C.; Liu, Y.; Edmunds, J.; Funk, S.; Eggo, R.M. Early dynamics of transmission and control of COVID-19: A mathematical modelling study. *Lancet* **2020**, *20*, 553–558. [\[CrossRef\]](#)
9. Li, R.; Pei, S.; Chen, B.; Song, Y.; Zhang, T.; Yang, W.; Shaman, J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **2020**, *368*, 489–493. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Ferretti, L.; Wymant, C.; Kendall, M.; Zhao, L.; Nurtay, A.; Abeler-Dörner, L.; Parker, M.; Bonsall, D.; Fraser, C. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **2020**, *368*, eabb6936. [\[CrossRef\]](#) [\[PubMed\]](#)
11. D'Arienza, M.; Coniglio, A. Assessment of the SARS-CoV-2 basic reproduction number,  $R_0$ , based on the early phase of COVID-19 outbreak in Italy. *Biosaf. Health* **2020**, *2*, 57–59. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Sanche, S.; Lin, Y.T.; Xu, C.; Romero-Severson, E.; Hengartner, N.W.; Ke, R. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. *Emerg. Infect. Dis.* **2020**, *26*, 1470–1477. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Romero-Severson, O.E.; Hengartner, N.; Meadors, G.; Ke, R. Change in global transmission rates of COVID-19 through May 6 2020. *PLoS ONE* **2020**, *15*, e0236776. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Ke, R.; Romero-Severson, E.O.; Sanche, S.; Hengartner, N. Estimating the reproductive number  $R_0$  of SARS-CoV-2 in the United States and eight European countries and implications for vaccination. *J. Theor. Biol.* **2021**, *517*, 110621. [\[CrossRef\]](#)
15. Kong, D.J.; Tekwa, E.W.; Gignoux-Wolfsohn, S.A. Social, economic, and environmental factors influencing the basic reproduction number of COVID-19 across countries. *PLoS ONE* **2021**, *16*, e0252373. [\[CrossRef\]](#)
16. Ives, C.A.; Bozzuto, R. State-by-State estimates of  $R_0$  at the start of COVID-19 outbreaks in the USA. *medRxiv* **2020**. Available online: <https://www.medrxiv.org/content/10.1101/2020.05.17.20104653v3> (accessed on 4 September 2021).
17. Fellows, I.E.; Slayton, R.B.; Hakim, A.J. The COVID-19 pandemic, community mobility and the effectiveness of non-pharmaceutical interventions: The United States of America, February to May 2020. *arXiv* **2020**. Available online: <https://arxiv.org/abs/2007.12644> (accessed on 8 September 2021).
18. Milicevic, O.; Salom, I.; Rodic, A.; Markovic, S.; Tumbas, M.; Zigic, D.; Djordjevic, M.; Djordjevic, M.  $PM_{2.5}$  as a major predictor of COVID-19 basic reproduction number in the USA. *Environ. Res.* **2021**, *201*, 111526. [\[CrossRef\]](#)

19. Ives, A.R.; Bozzuto, C. Estimating and explaining the spread of COVID-19 at the county level in the USA. *Commun. Biol.* **2021**, *4*, 1–9. [CrossRef]
20. Sy, K.T.; White, L.F.; Nichols, B.E. Population density and basic reproductive number of COVID-19 across United States counties. *PLoS ONE* **2021**, *16*, e0249271. [CrossRef] [PubMed]
21. Weissert, C.S.; Uttermark, M.J.; Mackie, K.R.; Artilles, A. Governors in control: Executive orders, state-local preemption, and the COVID-19 pandemic. *Publius* **2021**, *51*, 396–428. [CrossRef]
22. Lin, Y.T.; Neumann, J.; Miller, E.F.; Posner, R.G.; Mallela, A.; Safta, C.; Ray, J.; Thakur, G.; Chintavali, S.; Hlavacek, W.S. Daily forecasting of regional epidemics of coronavirus disease with bayesian uncertainty quantification. *Emerg. Infect. Dis.* **2021**, *27*, 767–778. [CrossRef] [PubMed]
23. The New York Times COVID-19 Data Team. Data from The New York Times. Available online: <https://github.com/nytimes/covid-19-data> (accessed on 20 September 2021).
24. The Covid Act Now COVID-19 Data Team. Data from Covid Act Now. Available online: <https://covidactnow.org/data-api> (accessed on 20 September 2021).
25. Surveillance Review and Response Group, Data from Centers for Disease Control and Prevention (CDC). Available online: <https://covid.cdc.gov/covid-data-tracker/#national-lab> (accessed on 20 September 2021).
26. Bajema, K.L.; Wiegand, R.E.; Cuffe, K.; Patel, S.V.; Iachan, R.; Lim, T.; Lee, A.; Moyses, D.; Havers, F.P.; Harding, L.; et al. Estimated SARS-CoV-2 Seroprevalence in the US as of September 2020. *JAMA* **2021**, *181*, 450–460. [CrossRef]
27. Fort, H. A very simple model to account for the rapid rise of the alpha variant of SARS-CoV-2 in several countries and the world. *Virus Res.* **2021**, *304*, 198531. [CrossRef] [PubMed]
28. Allen, H.; Vusirikala, A.; Flannagan, J.; Twohig, K.A.; Zaidi, A.; Chudasama, D.; Lamagni, T.; Groves, N.; Turner, C.; Rawlinson, C.; et al. Increased Household Transmission of COVID-19 Cases Associated with SARS-CoV-2 Variant of Concern B.1.617.2: A National Case-Control Study. 2021. Available online: <https://khub.net/documents/135939561/405676950/Increased+Household+Transmission+of+%5Cprotect%5Cunhbox%5Cvoidb%5Cx%5Chbox%5C%5BCOVID-19%5D+Cases+-+national+case+study.pdf/7f7764fbcfb0-da31-77b3-b1a8ef7be9aa> (accessed on 9 July 2021).
29. Hethcote, H.W. The mathematics of infectious diseases. *SIAM Rev. Soc. Ind. Appl. Math.* **2000**, *42*, 599–653. [CrossRef]
30. White House. Proclamation on Declaring a National Emergency Concerning the Novel Coronavirus Disease (COVID-19) Outbreak. Available online: <https://trumpwhitehouse.archives.gov/presidential-actions/proclamation-declaring-national-emergency-concerning-novel-coronavirus-disease-covid-19-outbreak/> (accessed on 29 November 2021).
31. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef] [PubMed]
32. Petzold, L. Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM J. Sci. Comput.* **1983**, *4*, 136–148. [CrossRef]
33. Blinov, M.L.; Faeder, J.R.; Goldstein, B.; Hlavacek, W.S. BioNetGen: Software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics* **2004**, *20*, 3289–3291. [CrossRef] [PubMed]
34. Cohen, S.D. CVODE. A stiff/nonstiff ODE solver in C. *Comput. Phys.* **1996**, *10*, 138–143. [CrossRef]
35. Lam, S.K.; Pitrou, A.; Seibert, S. Numba: A LLVM-based Python JIT compiler. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*; Association for Computing Machinery: New York, NY, USA, 2015; pp. 1–6.
36. Diekmann, O.; Heesterbeek, J.A.; Roberts, M.G. The construction of next-generation matrices for compartmental epidemic models. *J. R. Soc. Interface* **2010**, *7*, 873–875. [CrossRef]
37. Wolfram, S. *Mathematica: A System for Doing Mathematics by Computer*; Addison Wesley Longman Publishing Co., Inc.: Boston, MA, USA, 1991.
38. Holshue, M.L.; DeBolt, C.; Lindquist, S.; Lofy, K.H.; Wiesman, J.; Bruce, H.; Spitters, C.; Ericson, K.; Wilkerson, S.; Tural, A.; et al. First case of 2019 novel coronavirus in the United States. *N. Engl. J. Med.* **2020**, *382*, 929–936. [CrossRef]
39. Wearing, H.J.; Rohani, P.J.M. Keeling. Appropriate models for the management of infectious diseases. *PLoS Med.* **2005**, *7*, e174.
40. Mitra, E.D.; Suderman, R.; Colvin, J.; Ionkov, A.; Hu, A.; Sauro, H.M.; Posner, R.G.; Hlavacek, W.S. PyBioNetFit and the Biological Property Specification Language. *iScience* **2019**, *19*, 1012–1036. [CrossRef]
41. Massey, F.J. The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **1951**, *46*, 68–78. [CrossRef]
42. Shuai, Z.; van den Driessche, P. Global stability of infectious disease models using Lyapunov functions. *SIAM J. Appl. Math.* **2013**, *73*, 1513–1532. [CrossRef]
43. Dorigatti, I.; Lavezzo, E.; Manuto, L.; Ciavarella, C.; Pacenti, M.; Boldrin, C.; Cattai, M.; Saluzzo, F.; Franchin, E.; Del Vecchio, C.; et al. SARS-CoV-2 antibody dynamics and transmission from community-wide serological testing in the Italian municipality of Vo'. *Nat. Commun.* **2021**, *12*, 1–11.
44. Fowlkes, A.; Gaglani, M.; Groover, K.; Thiese, M.S.; Tyner, H.; Ellingson, K. Effectiveness of COVID-19 vaccines in preventing SARS-CoV-2 infection among frontline workers before and during B.1.617.2 (Delta) variant predominance—Eight US locations, December 2020–August 2021. *MMWR Morb. Mortal Wkly Rep.* **2021**, *70*, 1167–1169. [CrossRef]
45. Kalish, H.; Klumpp-Thomas, C.; Hunsberger, S.; Baus, H.A.; Fay, M.P.; Siripong, N.; Wang, J.; Hicks, J.; Mehalko, J.; Travers, J.; et al. Undiagnosed SARS-CoV-2 seropositivity during the first 6 months of the COVID-19 pandemic in the United States. *Sci. Transl. Med.* **2021**, *13*, eabh3826. [CrossRef]

46. Takahashi, S.; Greenhouse, B.; Rodríguez-Barraquer, I. Are seroprevalence estimates for severe acute respiratory syndrome coronavirus 2 biased? *J. Infect. Dis.* **2020**, *222*, 1772–1775. [[CrossRef](#)]
47. U.S. Census Bureau, Population Division. 2020. Available online: <https://www2.census.gov/programs-surveys/popest/tables/2010-2020/national/asrh/sc-est2020-18+pop-res.xlsx> (accessed on 29 November 2021).
48. Randolph, H.E.; Barreiro, L.B. Herd immunity: Understanding COVID-19. *Immunity* **2020**, *5*, 737–741. [[CrossRef](#)]
49. Moghadas, S.M.; Sah, P.; Shoukat, A.; Meyers, L.A.; Galvani, A.P. Population immunity against COVID-19 in the United States. *Ann. Intern. Med.* **2021**, *174*, 1586–1591. [[CrossRef](#)]
50. Science Brief: COVID-19 Vaccines and Vaccination. 2021. Available online: <https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/fully-vaccinated-people.html> (accessed on 8 September 2021).
51. Callaway, E.; Ledford, H. How bad is Omicron? What scientists know so far. *Nature* **2021**, *600*, 197–199. [[CrossRef](#)] [[PubMed](#)]
52. Nishiura, H.; Ito, K.; Anzai, A.; Kobayashi, T.; Piantham, C.; Rodríguez-Morales, A.J. Relative reproduction number of SARS-CoV-2 Omicron (B.1.1.529) compared with Delta variant in South Africa. *J. Clin. Med.* **2022**, *11*, 30. [[CrossRef](#)] [[PubMed](#)]
53. Ito, K.; Piantham, C.; Nishiura, H. Relative instantaneous reproduction number of Omicron SARS-CoV-2 variant with respect to the Delta variant in Denmark. *J. Med. Virol.* **2021**. [[CrossRef](#)]
54. Sofonea, M.T.; Roquebert, B.; Foulongne, V.; Verdurme, L.; Trombert-Paolantoni, S.; Roussel, M.; Haim-Boukobza, S.; Alizon, S. From Delta to Omicron: Analysing the SARS-CoV-2 epidemic in France using variant-specific screening tests (September 1 to December 18, 2021). *medRxiv* **2022**. Available online: <https://www.medrxiv.org/content/10.1101/2021.12.31.21268583v1> (accessed on 29 November 2021).
55. Tartof, S.Y.; Slezak, J.M.; Fischer, H.; Hong, V.; Ackerson, B.K.; Ranasinghe, O.N.; Frankland, T.B.; Ogun, O.A.; Zamparo, J.M.; Gray, S.; et al. Six-month effectiveness of BNT162B2 mRNA COVID-19 vaccine in a large US Integrated health system: A retrospective cohort study. *Lancet* **2021**, *398*, 1407–1416. [[CrossRef](#)]
56. Chemaitelly, H.; Tang, P.; Hasan, M.R.; AlMukdad, S.; Yassine, H.M.; Benslimane, F.M.; Al Khatib, H.A.; Coyle, P.; Ayoub, H.H.; Al Kanaani, Z.; et al. Waning of BNT162b2 vaccine protection against SARS-CoV-2 infection in Qatar. *N. Engl. J. Med.* **2021**, *385*, e83. [[CrossRef](#)]
57. UK Health Security Agency, Technical Briefing 33. Available online: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1043807/technical-briefing-33.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1043807/technical-briefing-33.pdf) (accessed on 29 November 2021).
58. Delamater, P.L.; Street, E.J.; Leslie, T.F.; Yang, Y.T.; Jacobsen, K.H. Complexity of the basic reproduction number (R0). *Emerg. Infect. Dis.* **2019**, *25*, 1–4. [[CrossRef](#)]
59. Stadnytskyi, V.; Bax, C.E.; Bax, A.; Anfinrud, P. The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 11875–11877. [[CrossRef](#)]
60. Echternach, M.; Gantner, S.; Peters, G.; Westphalen, C.; Benthous, T.; Jakubaß, B.; Kuranova, L.; Döllinger, M.; Kniesburges, S. Impulse dispersion of aerosols during singing and speaking: A potential COVID-19 transmission pathway. *Am. J. Respir. Crit. Care Med.* **2020**, *202*, 1584–1587. [[CrossRef](#)]
61. Pascarella, G.; Strumia, A.; Piliago, C.; Bruno, F.; del Buono, R.; Costa, F.; Scalata, S.; Agrò, F.E. COVID-19 diagnosis and management: A comprehensive review. *J. Intern. Med.* **2020**, *288*, 192–206. [[CrossRef](#)]
62. Falzone, L.; Gattuso, G.; Tsatsakis, A.; Spandidos, D.A.; Libra, M. Current and innovative methods for the diagnosis of COVID-19 infection. *Int. J. Mol. Med.* **2021**, *47*, 1–23. [[CrossRef](#)] [[PubMed](#)]
63. Van der Toorn, W.; Oh, D.Y.; von Kleist, M. COVID StrategyCalculator: A software to assess testing and quarantine strategies for incoming travelers, contact management, and de-isolation. *Patterns* **2021**, *2*, 100262. [[CrossRef](#)] [[PubMed](#)]
64. Larremore, D.B.; Fosdick, B.K.; Zhang, S.; Grad, Y.H. Jointly modeling prevalence, sensitivity and specificity for optimal sample allocation. *bioRxiv* **2020**. Available online: <https://www.biorxiv.org/content/10.1101/2020.05.23.112649v1> (accessed on 8 September 2021).
65. Gelman, A.; Carpenter, B. Bayesian analysis of tests with unknown specificity and sensitivity. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2020**, *69*, 1269–1283. [[CrossRef](#)]
66. Bendavid, E.; Mulaney, B.; Sood, N.; Shah, S.; Bromley-Dulfano, R.; Lai, C.; Weissberg, Z.; Saavedra-Walker, R.; Tedrow, J.; Bogan, A.; et al. COVID-19 antibody seroprevalence in Santa Clara County, California. *Int. J. Epidemiol.* **2021**, *50*, 410–419. [[CrossRef](#)]