State-level tracking of COVID-19 in the United States

H. Juliette T. Unwin[1,*], Swapnil Mishra[1,*], Valerie C. Bradley[2,*], Axel Gandy[3,*], Thomas A. Mellan[1], Helen Coupland[1], Jonathan Ish-Horowicz[3], Michaela A. C. Vollmer[1], Charles Whittaker[1], Sarah L. Filippi[3], Xiaoyue Xi[3], Mélodie Monod[3], Oliver Ratmann[3], Michael Hutchinson[2], Fabian Valka , Harrison Zhu[3], Iwona Hawryluk[1], Philip Milton[1], Kylie E. C. Ainslie[1], Marc Baguelin[1,4], Adhiratha Boonyasiri[5], Nick F. Brazeau[1], Lorenzo Cattarino[1], Zulma Cucunuba[1], Gina Cuomo-Dannenburg[1], Ilaria Dorigatti[1], Oliver D. Eales[1], Jeffrey W. Eaton[6], Sabine L. van Elsland[1], Richard G. FitzJohn[1], Katy A. M. Gaythorpe[1], William Green[1], Wes Hinsley[1], Benjamin Jeffrey[1], Edward Knock[1], Daniel J. Laydon[1], John Lees[1], Gemma Nedjati-Gilani[1], Pierre Nouvellet[1,7], Lucy Okell[1], Kris V. Parag[1], Igor Siveroni[1], Hayley A. Thompson[1], Patrick Walker[1], Caroline E. Walters[1], Oliver J. Watson[8,1], Lilith K. Whittles[1], Azra C. Ghani[1], Neil M. Ferguson[1], Steven Riley[1], Christl A. Donnelly[1,2], Samir Bhatt[1,*], and Seth Flaxman[3,*,†]

[1]MRC Centre for Global Infectious Disease Analysis, Abdul Latif Jameel Institute for Disease and Emergency Analytics (J-IDEA), Imperial College, London, UK

[2]Department of Statistics, University of Oxford, Oxford, UK

[3]Department of Mathematics, Imperial College, London, UK

[4]Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, UK

[5]NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, Imperial College London, UK

[6]MRC Centre for Global Infectious Disease Analysis, Imperial College, London, UK

[7]School of Life Sciences, University of Sussex, Brighton, UK

[8]Department of Laboratory Medicine and Pathology, Brown University, Providence, RI, USA

[*]Contributed equally

†Corresponding author: s.flaxman@imperial.ac.uk

Abstract

As of 1st June 2020, the US Centers for Disease Control and Prevention reported 104,232 confirmed or probable COVID-19-related deaths in the US. This was more than twice the number of deaths reported in the next most severely impacted country. We jointly modelled the US epidemic at the state-level, using publicly available death data within a Bayesian hierarchical semi-mechanistic framework. For each state, we estimate the number of individuals that have been infected, the number of individuals that are currently infectious and the time-varying reproduction number (the average number of secondary infections caused by an infected person). We used changes in mobility to capture the impact that non-pharmaceutical interventions and other behaviour changes have on the rate of transmission of SARS-CoV-2. On 1st June, we estimated that R_t was only below one in 23 states. We also estimated that 3.7% [3.4%-4.0%] of the total population of the US had been infected, with wide variation between states, and approximately 0.01% of the population was infectious. We demonstrate good 3 week model forecasts of deaths with low error and good coverage of our credible intervals.

1 Introduction

The first death caused by COVID-19 in the United States is currently believed to have occurred in Santa Clara County, California on the 6th February [1]. Throughout March 2020, US state governments implemented a variety of non-pharmaceutical interventions (NPIs), such as school closures and stay-at-home orders, to limit the spread of SARS-CoV-2 and ensure the number of severe COVID-19 cases did not exceed the capacity of the health system. In April 2020, the number of deaths attributed to COVID-19 in the United States (US) surpassed that of Italy [2]. Courtemanche *et al.* [3] used an event-study model to determine that such NPIs were successful in reducing the growth rate of COVID-19 cases across US counties. We similarly seek to estimate the impact of NPIs on COVID-19 transmission, but through a semi-mechanistic Bayesian model that reflects the underlying process of disease transmission and relies on mobility data released by companies such as Google [4].

Mobility measures revealed stark changes in behaviour following the large-scale government interventions in the first stage of the epidemic, with individuals spending more time at home and correspondingly less time at work, at leisure centres, shopping, and on public transit [4, 5]. As states continued to ease the stringency of their NPIs in the end of June, policy decisions relied on the interaction between mobility and NPIs and their subsequent impact on transmission, alongside other measures to track and curtail SARS-CoV-2 transmission.

We introduced a new Bayesian statistical framework for estimating the rate of transmission and attack rates for COVID-19 in Flaxman *et al.* [6]. In that paper, we inferred the time-varying reproduction number, R_t, or the average number of people an infected person will infect over time. We calculated the number of new infections through combining previous infections with the generation interval (the distribution of times between infections) and chose the number of deaths to be a function of the number of infections and the infection fatality ratio (IFR). We estimated the posterior probability of our parameters given the observed data, while incorporating prior uncertainty. This made our approach empirically driven, whilst incorporating uncertainty. This approach has also been implemented for Italy [7] and Brazil [8].

In this paper, we extend the Flaxman *et al.* [6] framework to model transmission in the US at the state level and include reported cases in our model. We parameterise R_t as a function of several mobility types and include an autoregressive term to capture changes in transmission that are decoupled from mobility, for example hand-washing, social distancing and changes in transmission that are decoupled from mobility. We utilise partial pooling of parameters, where information is shared across all states to leverage as much signal as possible, but individual effects are also included for state and region-specific idiosyncrasies. In this paper, we infer plausible upper and lower bounds (Bayesian credible interval summaries of our posterior distribution) of the total population that had been infected by COVID-19 on 01 June 2020 (also called the cumulative attack rate or attack rate) and estimate the effective number of individuals currently infectious given our generation distribution. We also present effect sizes of the mobility covariates and make short term forecasts, which we compare with reality throughout June. Details of the data sources and a technical description of our model are found in Sections 4 and 5 respectively. General limitations of our approach are presented in the conclusions.

2 Results

2.1 Infections

The percentage of the total population across the US infected by COVID-19 was 3.7% [3.4%-4.0%] on 01 June 2020. However, this low national average masked a stark heterogeneity across the states (Table 1). New York and New Jersey had the highest estimated cumulative attack rates, of 15.9% [12.4%-19.9%] and 14.8% [11.2%-18.2%] respectively, and Connecticut and Massachusetts both had cumulative attack rates over 10%. Conversely, other states that have drawn attention for early outbreaks, such as California, Washington, and Florida, only had cumulative attack rates of around 2% and states that were only in the early stages of their epidemics, like Maine, had estimated cumulative attack rates of less than 1%.

Figure 1 shows the effective number of infectious individuals and the number of newly infected individuals on any given day up until 01 June 2020 for each of the 8 regions in our model, which are based on US census regions (see Appendix A for further descriptions of our groupings). The effective number of infectious individuals is calculated using the generation time distribution, where individuals are weighted by how infectious they are over time, see Section 5.2 for more information. The fully infectious average includes asymptotic and symptomatic individuals. On 01 June 2020, we estimate that there were 41,100 [34,500-46,800] infectious individuals across the US, which corresponds to 0.01% of the population. Table 1 shows estimates of the number of new infections across each states on 01 June 2020. By this date, the estimated number infections were beginning to increase in the Pacific (Alaska, California, Hawaii, Oregon and Washington) and Mountain (Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah and Wyoming) regions.

Our model includes a state-level parameter for the infection ascertainment ratio, IAR, which we define as the number of reported cases divided by the true number of infections (including asymptomatic infections). We only estimate this parameter from 11 May 2020 when more than 375,000 tests are done each day, see Appendix B for further information. Column 3 of Table 1 shows the value of the infection ascertainment ratio in our model (see Section 5) and varies significantly between state. We would not expect the infection ascertainment to be 100% because our model includes asymptomatic individuals who may not know they have COVID-19. The mean value of this ratio varies between 43% (Missouri) to 74% (Kansas and Tennessee), which suggests that states are doing very different levels of testing.

2.2 Reproduction number

The mean estimate for R_t was below one in 23 states[1] on 01 June 2020 and the 95% credible intervals did not exclude one in any state (see Appendix C for R_ts by state). Figure 2 depicts the geographical variation in the posterior probability that R_t was less than 1. The closer a value is to 100%, the more certain we were that the reproduction number was below 1, indicating that new infections were not increasing. The probability was less than 40% that $R_t < 1$ in 20 states. There was substantial geographical clustering; most states in the Midwest and the South had reproduction numbers that suggested that the epidemic was not under control. We include figures of R_t, infections and deaths over time for each state in Appendix D.

[1] 45.10% of states

Table 1: Posterior model estimates of percentage of total population ever infected, mean new infections per day over week ending 01 June 2020, and infection ascertainment ratio as of 01 June 2020. We present the mean and the 95% credible intervals in square brackets.

State	**% of total population infected**	**Estimated mean new infections per day over week ending 01 June 2020**	**Infection ascertainment ratio**
Alabama	1.8% [1.4%-2.3%]	1,065 [300-2,400]	59% [35%-80%]
Alaska	0.1% [0.0%-0.2%]	20 [0-100]	69% [46%-88%]
Arizona	1.6% [1.2%-2.0%]	1,003 [400-1,800]	55% [35%-81%]
Arkansas	0.7% [0.5%-1.0%]	451 [100-900]	66% [45%-86%]
California	1.5% [1.1%-1.9%]	4,863 [2,100-10,600]	59% [37%-80%]
Colorado	3.2% [2.6%-4.1%]	674 [200-1,400]	54% [33%-79%]
Connecticut	11.4% [9.1%-14.5%]	520 [200-1,200]	53% [32%-78%]
Delaware	4.4% [3.4%-5.6%]	153 [0-300]	68% [45%-87%]
District of Columbia	9.7% [7.6%-12.3%]	134 [0-300]	60% [39%-83%]
Florida	1.2% [0.9%-1.5%]	1,350 [600-2,700]	61% [39%-83%]
Georgia	2.7% [2.1%-3.4%]	1,528 [600-3,600]	46% [25%-70%]
Hawaii	0.1% [0.0%-0.3%]	2 [0-100]	69% [49%-89%]
Idaho	0.6% [0.4%-0.9%]	47 [0-100]	70% [48%-88%]
Illinois	5.2% [4.1%-6.5%]	2,198 [800-4,500]	63% [40%-84%]
Indiana	3.8% [3.1%-4.9%]	779 [300-1,700]	61% [36%-82%]
Iowa	2.3% [1.7%-2.8%]	542 [200-1,100]	58% [36%-81%]
Kansas	1.1% [0.8%-1.4%]	189 [0-400]	74% [58%-91%]
Kentucky	1.2% [0.9%-1.6%]	359 [100-800]	58% [36%-81%]
Louisiana	7.1% [5.7%-9.0%]	660 [300-1,400]	63% [38%-86%]
Maine	0.7% [0.5%-1.0%]	78 [0-200]	64% [42%-85%]
Maryland	5.5% [4.3%-6.7%]	1,675 [600-3,200]	60% [38%-83%]
Massachusetts	11.2% [9.0%-14.0%]	3,387 [1,300-7,000]	43% [23%-68%]
Michigan	5.8% [4.5%-7.2%]	641 [200-1,500]	54% [30%-76%]
Minnesota	2.6% [1.9%-3.2%]	1,110 [400-2,400]	57% [36%-80%]
Mississippi	3.1% [2.5%-4.1%]	687 [300-1,600]	48% [27%-73%]
Missouri	1.5% [1.1%-1.9%]	504 [200-1,100]	43% [24%-69%]
Montana	0.2% [0.0%-0.3%]	11 [0-100]	71% [47%-87%]
Nebraska	1.5% [1.2%-2.0%]	379 [100-900]	73% [53%-90%]
Nevada	1.8% [1.4%-2.3%]	197 [0-400]	62% [40%-84%]
New Hampshire	2.0% [1.5%-2.6%]	152 [0-400]	54% [30%-78%]
New Jersey	14.8% [11.2%-18.2%]	1,493 [500-3,200]	52% [31%-79%]
New Mexico	2.0% [1.6%-2.6%]	176 [0-400]	61% [36%-81%]
New York	15.9% [12.4%-19.9%]	2,056 [800-4,200]	59% [37%-81%]
North Carolina	1.3% [1.0%-1.7%]	1,859 [800-4,100]	56% [34%-78%]
North Dakota	1.2% [0.8%-1.7%]	49 [0-200]	71% [47%-88%]
Ohio	2.1% [1.7%-2.7%]	1,141 [400-2,700]	48% [28%-75%]
Oklahoma	1.1% [0.8%-1.4%]	117 [0-300]	66% [43%-85%]
Oregon	0.4% [0.3%-0.6%]	85 [0-200]	72% [50%-89%]
Pennsylvania	4.4% [3.4%-5.5%]	1,310 [400-2,600]	51% [28%-78%]
Rhode Island	7.5% [5.8%-9.4%]	246 [0-700]	51% [27%-74%]
South Carolina	1.3% [0.9%-1.8%]	743 [200-1,400]	51% [30%-78%]
South Dakota	1.1% [0.7%-1.5%]	110 [0-300]	69% [48%-87%]
Tennessee	0.8% [0.6%-1.1%]	406 [100-800]	74% [54%-90%]
Texas	0.9% [0.7%-1.2%]	2,208 [1,000-4,400]	65% [44%-86%]
Utah	0.8% [0.6%-1.1%]	420 [100-800]	66% [45%-86%]
Vermont	0.9% [0.5%-1.3%]	6 [0-100]	69% [46%-87%]
Virginia	2.3% [1.8%-2.9%]	1,879 [800-3,900]	62% [40%-83%]
Washington	1.8% [1.4%-2.3%]	533 [200-1,200]	62% [38%-83%]
West Virginia	0.5% [0.3%-0.7%]	70 [0-200]	65% [43%-86%]
Wisconsin	1.3% [1.0%-1.7%]	846 [300-1,900]	61% [37%-80%]
Wyoming	0.3% [0.1%-0.6%]	15 [0-100]	65% [39%-85%]
National	3.7% [3.4%-4.0%]	41,100 [34,500-46,800]	

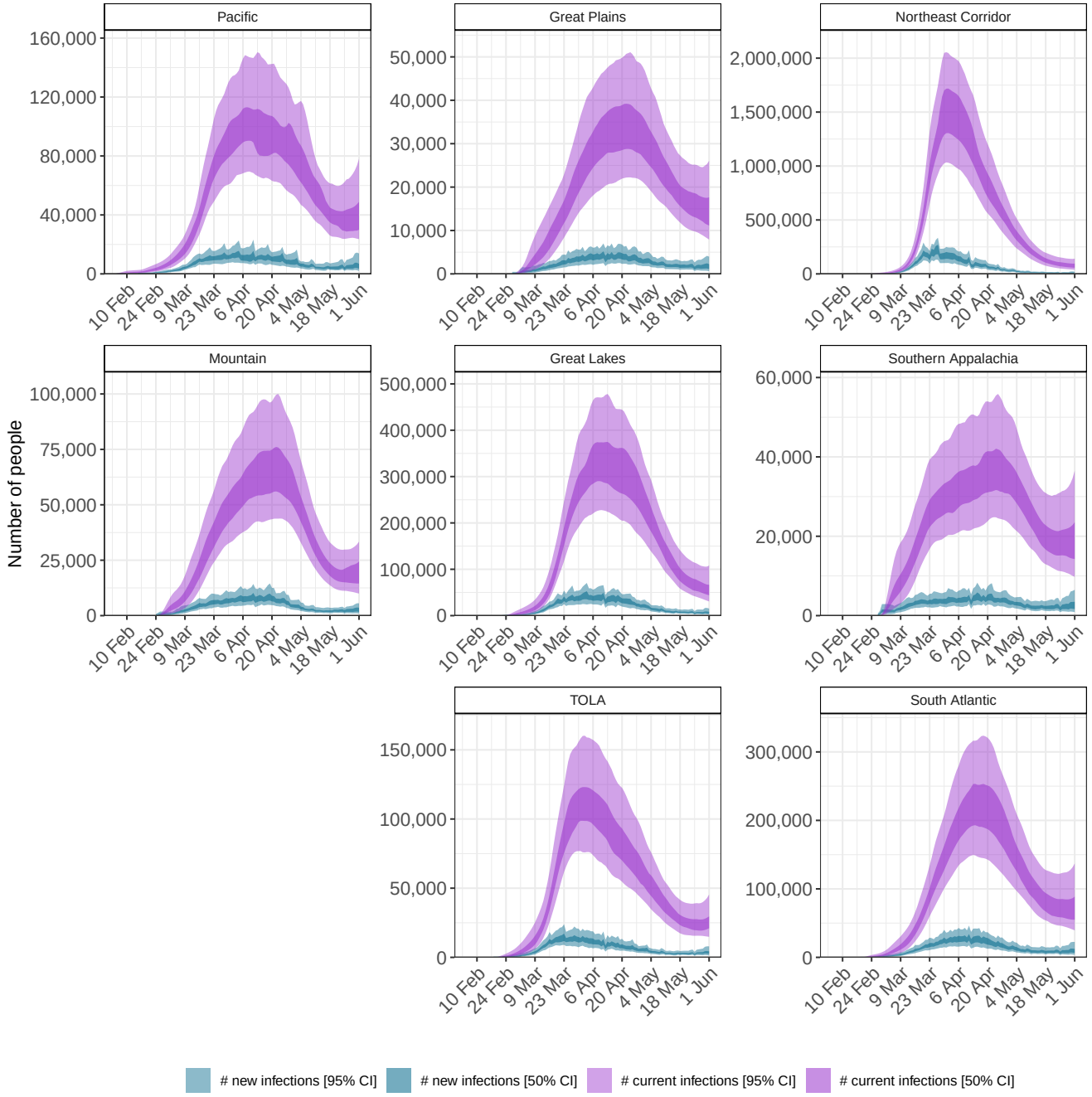


Figure 1: Daily estimates of the number of infectious (those still able to transmit) individuals (light purple, 95% credible interval(CI), dark purple 50% CI) and newly infected individuals (light blue, 95% CI, dark blue: 50% CI).

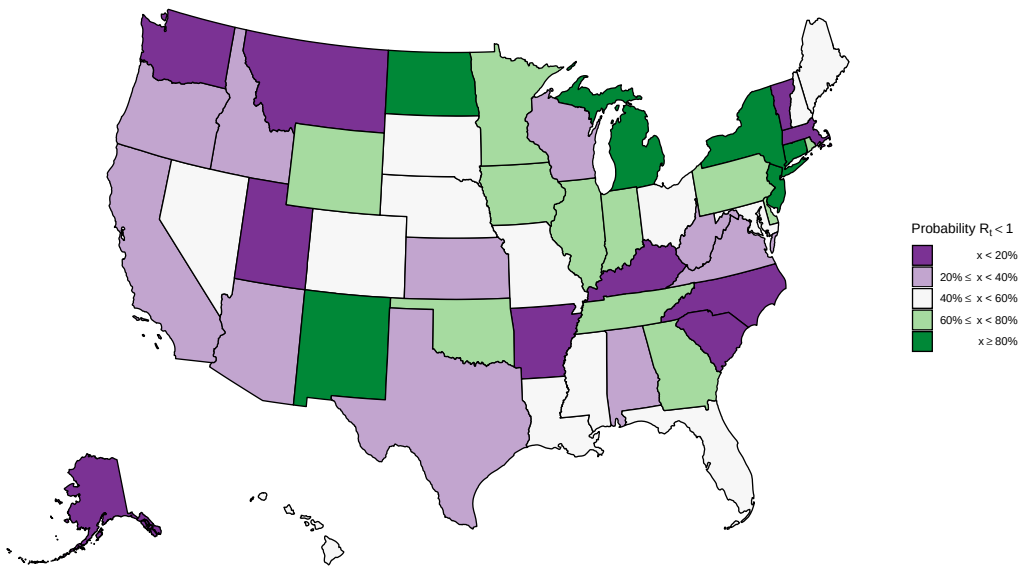


Figure 2: Estimates of the probability that the time-varying reproduction number R_t is less than one in each state. These values are an average over the week ending 01 June 2020. This plot shows the certainty that the rate of transmission is under control.

2.3 Model effect sizes

We find that decreases in the overall average number of visits to different places had a significant effect on reducing transmission. If mobility stopped entirely (100% reduction in average mobility) then R_t would be reduced by 55.1% [26.5% - 77.0%]. The country effect size estimates are given in Figure 3, with regional and state-level effects given in Appendix E. However, in the US, the average mobility covariate never approached a 100% reduction, and only about half the states had reductions below 50% of the baseline. We define the baseline as the pre-epidemic mobility for each state [4]. As an example, consider the largest reduction observed, -62% of the baseline (Minnesota on 12 April 2020). The effect on R_t was a reduction of 37% [16% - 56%] from the country level effect.

Increased time spent in residences also reduced transmission; if time spent in residences increased to 100% of the baseline, R_t would be reduced by 15.3% [-27.5% - 54.6%]. Time spent in residences increased by 20% or more from the baseline in 36 states. As an example, consider the largest reduction observed, a 33% increase from the baseline (New Jersey on 10 April 2020). The effect on R_t from this was a reduction of 5% [-10% - 20%] in New Jersey from the country level effect.

Average mobility and residential mobility are no doubt correlated—when people spend less time in public spaces, captured by our average mobility metric, they conversely spend more time at home. Due to this collinearity, our model is unable to distinguish between the independent contributions of these covariates, with most of the effect assigned to the average mobility coefficient, due to its greater explanatory power. As a check that our overall findings were not biased by this collinearity, we verified that the posterior estimates of these coefficients was not correlated.

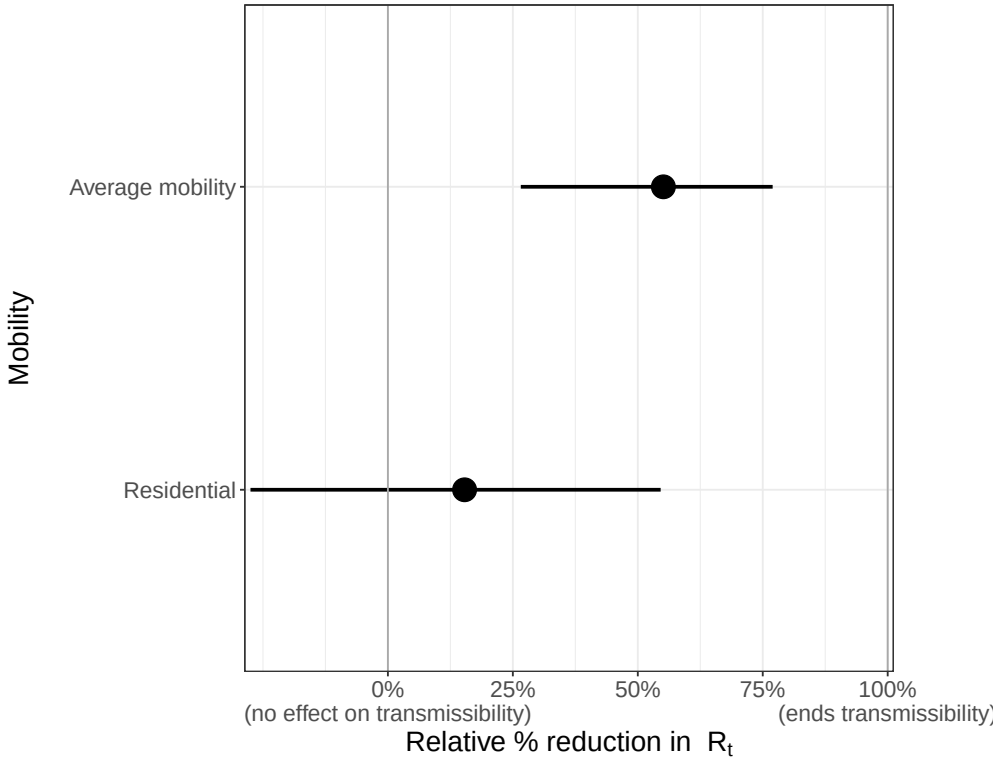


Figure 3: Country level covariate effect sizes assuming mobility stopped entirely (100% reduction in average mobility) and residential mobility was increased fully (100% increase in residential mobility): Average mobility combines "retail & recreation", "grocery & pharmacy", "workplaces".

2.4 Short term forecasts

We used our model to produce short term death forecasts. Figure 4 compares our forecasts for the three weeks after 01 June 2020 (blue line with shaded uncertainty intervals) with the recorded daily number of deaths during this period (coral bars). As expected from our R_t values, deaths were noticeably declining in the Northeastern Corridor, where $R_t > 1$, with particularly low error between our forecasts and reality in New York and Connecticut. In the South, we forecast a flattening or slight increase of deaths, especially in Arkansas, Texas and Florida.

We investigated the numerical accuracy of our forecast using three metrics: mean absolute error, continuous ranked probability score (CRPS) and coverage of credible intervals. We fitted our model to three end points: 1 May, 15 May and 1 June and performed three week forecasts from each end point. We compared the metric scores with a log-linear "null" model fit to 31 days of data prior to the three specified end points (see Appendix F for further information). We find our model performs similarly to the null model (1 June) or better (15 May), however our model fit to 1 May is worse than the null model because we only include cases after 11 May in our models. This suggests that including cases improves the forecasting ability of our model and further justifies our inclusion of them. The coverage of our credible intervals is good for all models, in particular our model and the null model fit to 1 June.

2.5 Model selection and sensitivity

Mobility data provided a proxy for the behavioural changes that occur in response to non-pharmaceutical interventions. Appendix H shows the mobility trends for the 50 states and the District of Columbia up until 01 June 2020 (see Section 4 for a description of the mobility dimensions). The median correlation between the observed average

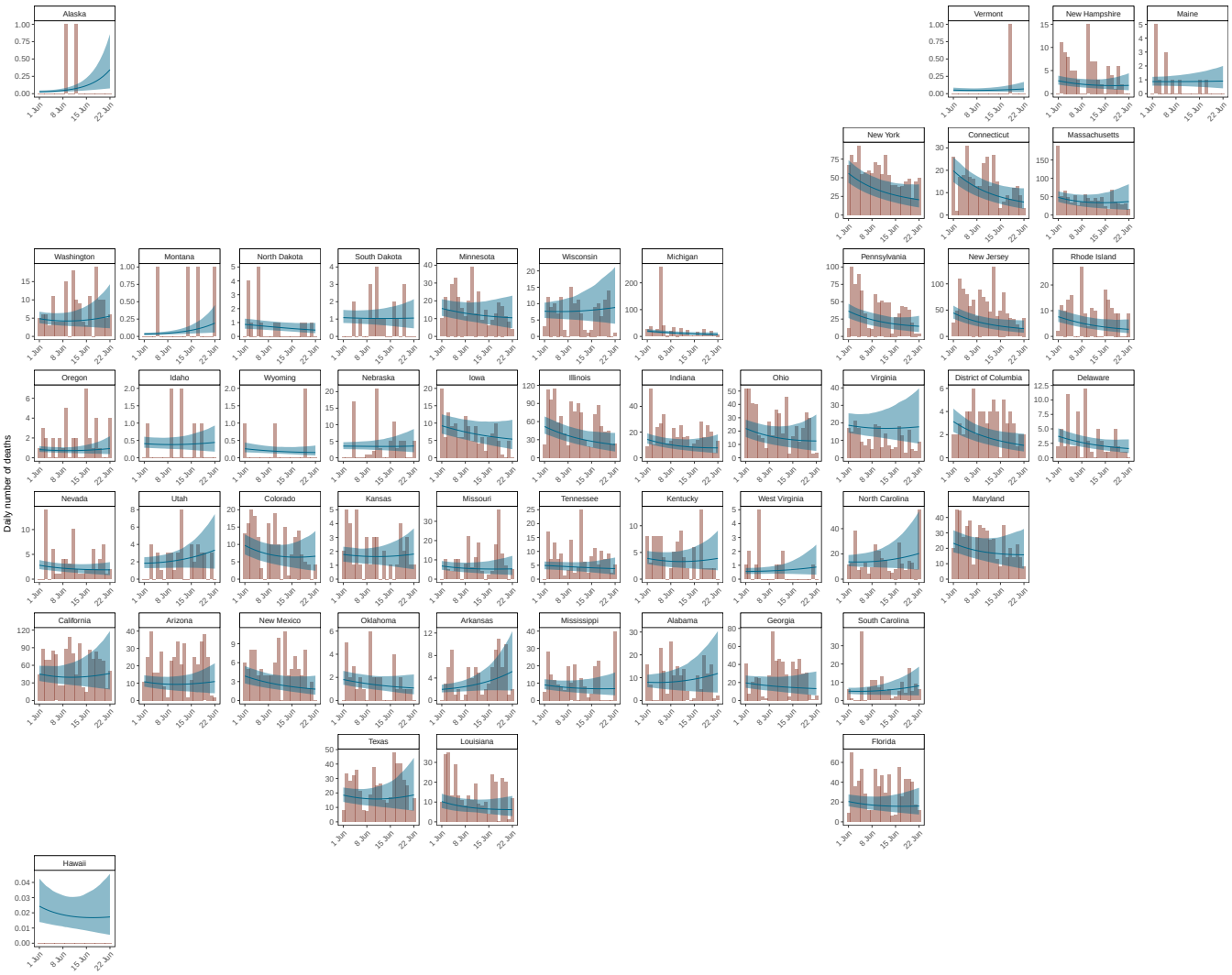


Figure 4: Three week forecasts for model fitted up until 01 June 2020. The coral bars show the reported number of deaths for the three weeks after 01 June 2020, and the blue line and ribbon show the mean and 95% CI for our forecast estimates.

mobility and the timing of the introduction of major NPIs (represented as step functions) was approximately 86% (see Appendix I). We make no explicit causal link between NPIs and mobility because this relationship is plausibly causally linked by other factors. The mobility trends data suggests that substantial early outbreak in New York state may have led to substantial changes in mobility in nearby states, like Connecticut, prior to any mandated interventions in those states, which supports including regions in our model. Including both mobility trends and the timing of imposition and lifting of "stay-at-home" orders did not affect the estimated cumulative attack rates (see Appendix J).

Mobility alone cannot fully capture how transmission evolves over time. In particular, it cannot capture the impact of case-based interventions (such as testing and tracing) or behaviour changes (such as mask wearing or hand washing). We use a second-order, weekly, autoregressive process to allow our changes in transmission to be decoupled from mobility. This autoregressive process is an additional term in our parametric equation for R_t and accounts for residual effects by capturing a correlation structure where current R_t is correlated with previous weeks R_t. This means that our forecasts were equally good whatever combination of mobility covariates were used because this term could capture the unexplained behaviour. The learnt random effects from this process are shown in Appendix K for all states. We show the contributions to R_t from the mobility and autoregressive terms in Figure 19 for three example states. The autoregressive term increases R_t before lockdown in New York, which could be explained by behaviour such as panic buying. In contrast, the autoregressive term reduces R_t in Montana and could reflect behavioural changes such as hand-washing and self isolation, which can reduce transmission with maintained mobility levels. The autoregressive term remains mostly constant in Washington and suggests that mobility is sufficient to capture the behaviour there.

3 Discussion

We developed a Bayesian semi-mechanistic modelling approach to investigate the impact of NPIs on the spread of SARS-CoV-2 in the United States through changes in mobility. Our model relies on death data from the start of the epidemic and recently reported case data to inform our predictions. This enabled us to estimate a realistic infection ascertainment ratio for the three weeks before 01 June 2020 for each state, which could help inform policy as to where testing may be lacking. The mean value of this ratio varies between 43% (Missouri) to 74% (Kansas and Tennessee). Our epidemiological grounded mechanistic model links unobserved infections to reported cases and deaths, all within a principled Bayesian statistical framework. This is a significant advancement over curve-fitting models fit directly to reported cases.

Our model suggests that although initial reductions in the daily infections had plateaued in most states by 01 June 2020, the reservoir of infectious individuals still remained large with approximately 0.01% of the population being infectious on that date. Despite this, the cumulative attack rate across the US still remained low. We found our attack rate for New York was in line with those from recent serological studies [9]. There is now evidence that mild infection is able to lead to robust immunity (via T cells) but potentially not induce antibody production, which are detected in serosurveys [10]. Therefore, serosurveys might underestimate exposure, particularly in mild cases, and our model may provide an alternative way to measure population exposure. Our cumulative attack rates are, however, sensitive to the assumed values of infection fatality rate (IFR). We account for each individual state's age structure, and further adjust for contact mixing patterns [11], but age specific modelling may be necessary to capture

potential changes in the demographics of cases in states such as Texas, Florida and South Carolina where there is evidence that younger people than were infected at the start of the epidemic are being infected [12, 13].

We estimated that 23 states had a posterior mean reproduction number R_t below one on 01 June 2020 and in no states were we more than 95% confident that R_t was below one. We compared our estimates with predictions made by rt.live [14] who use a method that fits the most likely R_t curve to the daily new daily cases (see Appendix L). Overall, our estimates were weakly correlated ($\rho = 0.42$) with both of us estimating $R_t > 1$ in 23 states (red points) including Montana and Alaska. However, the rt.live estimates are slightly more pessimistic because they predict $R_t > 1$ in 10 states where we predict $R_t < 1$ (blue points). In contrast we predict $R_t > 1$ in 5 states where they predict $R_t < 1$ (green points). Both sets of reproduction numbers strongly implied that the US epidemic was not under control in many states, and that in the presence of continued migration and the loosening of interventions seen in June, increased infections were to be expected with high probability. We found that state with high reproduction numbers on 01 June 2020 were geographically clustered in the west and south US, whilst the states that had suffered high COVID-19 mortality (such as the Northeast Corridor) in the early phase of the epidemic had lower reproduction numbers. After the period covered by this study, reported cases began to increase in the US, and 7 states (Arizona, Arkansas, California, North Carolina, South Carolina, Tennessee and Texas) had recorded higher levels of hospitalisations in early July than before [15, 16]. This suggests our estimates that R_t was not less than one were accurate. More recent estimates of R_t, the number of infections, and the number of people currently infectious are presented on our website[2].

Our three week forecasts of daily deaths were highly accurate, confirming the predictive validity of our modelling approach, despite our having kept mobility constant during our forecasts. These forecasts, alongside our R_t values, show that the epidemic was not under control at the end of May. The accuracy of our forecasts varied during the epidemic and could be due to our assumption that mobility is kept constant over these three weeks. Our forecast would perform worse in weeks where mobility was significantly different to the last week of our model fit. When we include cases in our model, we are able to get similar results to a simple "null" model whilst also being about to estimate effect sizes of different mobility trends. We also compared our cumulative death forecasts with those presented by Friedman et al. [17]. Friedman et al. compared the median absolute percentage error (MAPE) for SEIR and dynamic growth rate types of models for models fit to some point in June. Unlike those models, we find the MAPE of our cumulative death forecasts did not increase significantly over time and our 3 week median cumulative death MAPE across all states (9.9%) was similar to the US estimate from Friedman et al. (4.1-8.6), see Appendix G for more information.

Our model uses mobility to predict SARS-COV2 transmission. We find that the timings that non-pharmaceutical interventions were implemented was strongly correlated to changes in mobility. This is similar to findings in Abouk and Heydari [5] who find that statewide stay-at-home orders had the strongest causal impact on reducing social interaction and that these orders significantly increase the presence of individuals at home by about six fold (our "residential mobility trend"). This supports our choice of using mobility instead of the timings of NPIs in this study instead of the times of interventions as in Flaxman et al. [6]. We find that magnitude of the reductions in average mobility, and the resulting increases in residential mobility, are important in determining the size of reduction in R_t. This agrees with Wang et al. [18] who use a stochastic age- and risk-structured susceptible-exposed-asymptomatic-symptomatic-hospitalized-recovered (SEAYHR) model to considered the effect of various levels of social distancing.

[2] https://mrc-ide.github.io/covid19usa/

They found that social distancing measures, which reduced non-household contacts by <50%, would not prevent a healthcare crisis and that only their 75% and 90% contact reduction scenarios were projected to enable metropolitan areas to remain within health care levels.

While mobility, or social distancing measures, will explain a large amount of the trend in R_t, there is likely to be substantial residual variation from other behavioural changes such as mask wearing and hand-washing. We accounted for this residual variation through a second order, weekly, autoregressive process. This stochastic process captures changes in R_t reflected in the data, but is unable to attribute these changes to other determinants of transmission or interventions. We pool parameters in our model to leverage as much signal in our data as possible and to reflect the conjoined nature of some states, in particular in the Northeastern Corridor. While this sharing can potentially lead to over or under estimation of effect sizes, it also means that a consistent signal for all states can be estimated before that signal is presented in an individual state with little data, such as Alaska and Hawaii. Pooling also increases the robustness of our models to under reporting and time lags [6, 7, 8].

4 Data

Our model uses daily real-time state-level aggregated data published by New York Times (NYT) [19] for New York State and John Hopkins University (JHU) [2] for the remaining states. We include 105,006 deaths in our model up until 1 June and 479,422 cases from 11 May to 1 June. Age specific population counts were drawn from the U.S. Census Bureau in 2018 [20] to estimate state-specific infection fatality ratio reflective of the population age structure. The timing of NPIs were collated by the University of Washington [21]. We used Google's COVID-19 Community Mobility Report [4], which provides data on movement in the US by states and highlights the percent change in visits to:

- Grocery & pharmacy: Mobility trends for places like grocery markets, food warehouses, farmers markets, speciality food shops, drug stores, and pharmacies.
- Parks: Mobility trends for places like local parks, national parks, public beaches, marinas, dog parks, plazas, and public gardens.
- Transit stations: Mobility trends for places like public transport hubs such as subway, bus, and train stations.
- Retail & recreation: Mobility trends for places like restaurants, cafes, shopping centres, theme parks, museums, libraries, and movie theatres.
- Residential:Mobility trends for places of residence.
- Workplaces: Mobility trends for places of work.

The residential data includes length of stay at different places compared to a baseline, whereas the other mobility trends are based on number of visits to a certain place. These trends are therefore relative, i.e. mobility of -20% means that, compared to normal circumstances individuals are engaging in a given activity 20% less.

5 Methods

Flaxman *et al.* [6] introduced a Bayesian model for estimating the transmission intensity and attack rate (percentage of the population that has been infected) from COVID-19 from the reported number of deaths. This framework used the time-varying reproduction number R_t to inform a latent function for infections, and then these infections, together with probabilistic lags, were calibrated against observed deaths. Observed deaths, while still susceptible to under reporting and delays, comprise a more consistent time series than the reported number of confirmed cases, which are susceptible to changes in the probability of ascertainment over the course of the epidemic as testing strategies changed. Our model code is available on GitHub[3].

We adapted the original Bayesian semi-mechanistic model of the infection cycle to all the states in the US and the District of Columbia to infer the reproduction number over time (R_t), plausible upper and lower bounds (95% Bayesian credible intervals) of the total populations infected (attack rates) and the number of people currently infected on 01 June 2020. In this paper, we also include the reported number of cases after 11 May 2020, see Appendix B. This reflects the point in time when over 375,000 tests were being done each day across the US. We include this in our likelihood but do not use them to calculate transmission directly. We parametrise R_t as a function of Google mobility data and include an autoregressive term to capture non-mobility driven behaviour. We fit our model jointly to COVID-19 data from all states to assess the attack rates and number of people who were currently infected. Finally, we use our model to forecast for 3 weeks from 01 June 2020 and compare our estimates of deaths to those recorded in the US for each state. We assume mobility remains constant at the previous value of mobility on the same day the previous week in our forecasts and the autoregressive term remains constant.

5.1 Model specifics

The true number of infected individuals, i, is modelled using a discrete renewal process. We specify a generation distribution g with density $g(\tau)$ as:

$$g \sim \text{Gamma}(6.5, 0.62). \tag{1}$$

Given the generation distribution, the number of infections $i_{t,m}$ on a given day t, and state m, is given by the following discrete convolution function:

$$\begin{aligned} i_{t,m} &= S_{t,m} R_{t,m} \sum_{\tau=0}^{t-1} i_{\tau,m} g_{t-\tau}, \\ S_{t,m} &= 1 - \frac{\sum_{j=0}^{t-1} i_{j,m}}{N_m}, \end{aligned} \tag{2}$$

where the generation distribution is discretised by $g_s = \int_{s-0.5}^{s+0.5} g(\tau)d\tau$ for $s = 2, 3, ...,$ and $g_1 = \int_0^{1.5} g(\tau)d\tau$. The population of state m is denoted by N_m. We include the adjustment factor $S_{t,m}$ to account for the number of susceptible individuals left in the population.

Both deaths and cases are observed in our model. We define daily deaths, $D_{t,m}$, for days $t \in \{1, \ldots, n\}$ and states $m \in \{1, \ldots, M\}$. These daily deaths are modelled using a positive real-valued function $d_{t,m} = \mathbb{E}[D_{t,m}]$ that represents the expected number of deaths attributed to COVID-19. The daily deaths $D_{t,m}$ are assumed to follow a negative binomial distribution with mean $d_{t,m}$ and variance $d_{t,m} + \frac{d_{t,m}^2}{\psi_1}$, where ψ_1 follows a positive half normal distribution,

[3] https://github.com/ImperialCollegeLondon/covid19model

i.e.

$$D_{t,m} \sim \text{Negative Binomial}\left(d_{t,m}, d_{t,m} + \frac{d_{t,m}^2}{\psi_1}\right), \quad t = 1, \ldots, n \tag{3}$$

$$\psi_1 \sim \mathcal{N}^+(0, 5). \tag{4}$$

Here, $\mathcal{N}(\mu, \sigma)$ denotes a normal distribution with mean μ and standard deviation σ. We say that X follows a positive half normal distribution $\mathcal{N}^+(0, \sigma)$ if $X \sim |Y|$, where $Y \sim \mathcal{N}(0, \sigma)$.

We link our observed deaths mechanistically to transmission as in Flaxman et al. [6]. We use a previously estimated COVID-19 infection fatality ratio (IFR, probability of death given infection) together with a distribution of times from infection to death π. Details of this calculation can be found in [22, 23]. From the above, every region has a specific mean infection fatality ratio ifr_m (see Appendix M). To incorporate the uncertainty inherent in this estimate we allow the ifr_m for every state to have additional noise around the mean. Specifically we assume

$$\text{ifr}_m^* \sim \text{ifr}_m \cdot N(1, 0.1). \tag{5}$$

We believe a large-scale contact survey similar to polymod [11] has not been collated for the USA, so we assume the contact patterns are similar to those in the UK. We conducted a sensitivity analysis, shown in Appendix M, and found that the IFR calculated using the contact matrices of other European countries lay within the posterior of ifr_m^*.

Using estimated epidemiological information from previous studies, we assume the distribution of times from infection to death π (infection-to-death) to be the convolution of an infection-to-onset distribution (π') [23] and an onset-to-death distribution [22]:

$$\pi \sim \text{Gamma}(5.1, 0.86) + \text{Gamma}(17.8, 0.45). \tag{6}$$

The expected number of deaths $d_{t,m}$, on a given day t, for state m is given by the following discrete sum:

$$d_{t,m} = \text{ifr}_m^* \sum_{\tau=0}^{t-1} i_{\tau,m} \pi_{t-\tau}, \tag{7}$$

where $i_{\tau,m}$ is the number of new infections on day τ in state m and where, similar to the generation distribution, π is discretized via $\pi_s = \int_{s-0.5}^{s+0.5} \pi(\tau) d\tau$ for $s = 2, 3, ...$, and $\pi_1 = \int_0^{1.5} \pi(\tau) d\tau$, where $\pi(\tau)$ is the density of π.

For every state m, we also observe daily cases $C_{t,m}$ after $t_c = 11$ May 2020. Similar to daily deaths, daily cases are modelled using a positive real-valued function $\bar{c}_{t,m} = \mathbb{E}[C_{t,m}]$ that represents the expected number of symptomatic and asymptomatic cases. Again, the daily cases $C_{t,m}$ are assumed to follow a negative binomial distribution but with mean $c_{t,m}$ and variance $c_{t,m} + \frac{c_{t,m}^2}{\psi_2}$, where ψ_2 follows a positive half normal distribution, i.e.

$$C_{t,m} \sim \text{Negative Binomial}\left(c_{t,m}, c_{t,m} + \frac{c_{t,m}^2}{\psi_2}\right), \quad t = t_c, \ldots, n, \tag{8}$$

$$\psi_2 \sim \mathcal{N}^+(0, 5). \tag{9}$$

As before, we assume the distribution of times from infection to becoming a case π' (infection-to-onset) to be

$$\pi' \sim \text{Gamma}(5.1, 0.86). \tag{10}$$

We add in a new link between our observed daily cases and our estimated daily infections. We use our model to estimate an infection ascertainment ratio (iar_m) for each state m, which is defined as the number of reported cases divided by the true number of infections (including both symtomatic and asymptomatic infections). This follows a Beta distribution, specifically $u_m \sim \text{Beta}(12, 5)$.

The expected number of cases $c_{t,m}$, on a given day t, for state m is given by the following discrete sum:

$$c_{t,m} = \text{iar}_m \sum_{\tau=0}^{t-1} i_{\tau,m} \pi'_{t-\tau}, \tag{11}$$

where, again, $c_{\tau,m}$ is the number of new infections on day τ in state m and where π' is discretized via $\pi'_s = \int_{s-0.5}^{s+0.5} \pi'(\tau) d\tau$ for $s = 2, 3, ...$, and $\pi'_1 = \int_0^{1.5} \pi'(\tau) d\tau$, where $\pi'(\tau)$ is the density of π'.

We parametrise $R_{t,m}$ as a linear function of the relative change in time spent and number of visits (from a baseline)

$$R_{t,m} = R_{0,m} \cdot f\left(-\left(\sum_{k=1}^{2} X_{t,m,k} \alpha_k \right) - \sum_{l=1}^{2} Y_{t,m,l} \alpha^{\text{region}}_{r(m),l} - Z_{t,m} \alpha^{\text{state}}_m - \epsilon_{m,w_m(t)} \right), \tag{12}$$

where $f(x) = 2\exp(x)/(1+\exp(x))$ is twice the inverse logit function. $X_{t,m,k}$ are covariates that have the same effect for all states, $Y_{t,m,l}$ is a covariate that has region-specific effects, $r(m) \in \{1, \ldots, R\}$ is the region a state is in (see Figure 15), $Z_{t,m}$ is a covariate that has a state-specific effect and $\epsilon_{m,w_m(t)}$ is a weekly AR(2) process, centred around 0, that captures variation between states that is not explained by the covariates.

The prior distribution for $R_{0,m}$[24] was chosen to be

$$R_{0,m} \sim \mathcal{N}(3.28, \kappa) \text{ with } \kappa \sim \mathcal{N}^+(0, 0.5), \tag{13}$$

where κ is the same among all states.

In the analysis of this paper we chose the following covariates: $X_{t,m,1} = M^{\text{average}}_{t,m}$, $X_{t,m,2} = M^{\text{residential}}_{t,m}$, $Y_{t,m,1} = 1$ (an intercept), $Y_{t,m,2} = M^{\text{average}}_{t,m}$ and $Z_{t,m} = M^{\text{average}}_{t,m}$, where the mobility variables are from [4] and defined as follows (all are encoded so that 0 is the baseline and 1 is a full reduction of the mobility in this dimension):

- $M^{\text{average}}_{t,m}$ is an average of retail and recreation, groceries and pharmacies, and workplaces. An average is taken as these dimensions are strongly collinear.
- $M^{\text{residential}}_{t,m}$ are the mobility trends for places of residences.

We include regional, as well as state level parameters, in our model to encapsulate the connected nature of states. This was particularly important in the Northeasten corridor where residents in New Jersey and Connecticut regularly commuted into New York, the early epicenter of the US epidemic (see Appendix A for a map of the regions). Regions are based on US Census Divisions, modified to account for coordination between groups of state governments [25].

We assume that seeding of new infections begins 30 days before the day after a state has cumulatively observed 10 deaths. From this date, we seed our model with 6 sequential days of an equal number of infections: $i_{1,m} = \cdots = i_{6,m} \sim \text{Exponential}(\frac{1}{\tau})$, where $\tau \sim \text{Exponential}(0.03)$. These seed infections are inferred in our Bayesian posterior distribution.

The weekly, state-specific effect is modelled as a weekly AR(2) process, centred around 0 with stationary standard deviation σ_w that, in every state, starts on the first day of its seeding of infections, i.e. 30 days before a total of 10

cumulative deaths have been observed in this state. The AR(2) process starts with $\epsilon_{1,m} \sim \mathcal{N}(0, \sigma_w^*)$,

$$\epsilon_{w,m} \sim \mathcal{N}(\rho_1 \epsilon_{w-1,m} + \rho_2 \epsilon_{w-2,m}, \sigma_w^*) \text{ for } m = 2, 3, 4, \ldots \tag{14}$$

with independent priors on ρ_1 and ρ_2 that are normal distributions conditioned to be in $[0, 1]$; the prior for ρ_1 is a $\mathcal{N}(0.8, .05)$ distribution conditioned to be in $[0, 1]$ and the prior for ρ_2 is a $\mathcal{N}(0.1, .05)$ distribution, conditioned to be in $[0, 1]$. The prior for σ_w, the standard deviation of the stationary distribution of ϵ_w is chosen as $\sigma_w \sim \mathcal{N}^+(0, .2)$. The standard deviation of the weekly updates to achieve this standard deviation of the stationary distribution is $\sigma_w^* = \sigma_w \sqrt{1 - \rho_1^2 - \rho_2^2 - 2\rho_1^2 \rho_2 / (1 - \rho_2)}$. The conversion from days to weeks is encoded in $w_m(t)$. Every 7 days, w_m is incremented, i.e. we set $w_m(t) = \lfloor (t - t_m^{\text{start}})/7 \rfloor + 1$, where t_m^{start} is the first day of seeding. We keep the AR(2) process constant over the last 7 days of observations since this is less informed by data due to the lags and also over the forecast period.

The prior distribution for the shared coefficients were chosen to be

$$\alpha_k \sim \mathcal{N}(0, 0.5), k = 1, \ldots, 3, \tag{15}$$

and the prior distribution for the pooled coefficients were chosen to be

$$\alpha_{r,l}^{\text{region}} \sim \mathcal{N}(0, \gamma_r), r = 1, \ldots, R, l = 1, 2, \text{ with } \gamma_r \sim \mathcal{N}^+(0, 0.5), \tag{16}$$

$$\alpha_m^{\text{state}} \sim \mathcal{N}(0, \gamma_s), m = 1, \ldots, M \text{ with } \gamma_s \sim \mathcal{N}^+(0, 0.5). \tag{17}$$

We estimated parameters jointly for all states in a single hierarchical model. Fitting was done in the probabilistic programming language Stan[26] using an adaptive Hamiltonian Monte Carlo (HMC) sampler.

5.2 Generated quantities

The effective number of infectious individuals, i^*, on a given day considers how infectious a previously infected individual is on a given day and includes both asymptotic and symptomatic individuals. It is calculated by first re-scaling the generation distribution by its maximum, i.e. $g_\tau^* = \frac{g_\tau}{\max_t g_t}$. Based on (2), the number of infectious individuals is then calculated from the number of previously infected individuals, c, using the following:

$$i_{t,m}^* = \sum_{\tau=0}^{t-1} i_{\tau,m} g_{t-\tau}^*, \tag{18}$$

where $i_{t,m}$ is the number of new infections on day t in state m.

6 Acknowledgements

We would like to thank Amazon AWS and Microsoft Azure for computational credits and we would like to thank the Stan development team for their ongoing assistance. We would also like to thank David Joerg and Jacob Steinhardt for their comments through Open Review. This work was supported by Centre funding from the UK Medical Research Council under a concordat with the UK Department for International Development, the NIHR Health Protection Research Unit in Modelling Methodology and Community Jameel. HJTU is funded by Imperial College London through an Imperial College Research Fellowship grant. SB would like to acknowledge the NIHR BRC Imperial College NHS Trust Infection and COVID themes the Academy of Medical sciences Springboard award and the Bill and Melinda Gates Foundation.

7 Author Contributions

HJTU, SM, VB, AG, SB and SF conceived and designed the study. HJTU, SM, TM, MACV, HZ and PM performed mobility analysis. JI-H, SLF and XX contributed to statistical analysis and MH and IH did other analysis. HJTU, SM, AG, TM, HC, MACV, VW, MM, OR, SB and SF contributed to code development. HJTU, MACV and SF did the plotting. SM and FV created the website. HJTU, VB, SB and SF wrote the first draft of the paper. All authors discussed the results and contributed to the revision of the final manuscript.

8 Data and Code availability

All source code and data necessary for the replication of our results and figures is available at `https://github.com/ImperialCollegeLondon/covid19model`.

9 Competing Interests statement

The authors declare no competing interests.

References

1. Santa Clara County Public Health. *County of Santa Clara Identifies Three Additional Early COVID-19 Deaths* 2020. `https://www.sccgov.org/sites/covid19/Pages/press-release-04-21-20-early.aspx`.
2. Dong, E, Du, H & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. eng. *The Lancet. Infectious diseases,* 1473–3099. ISSN: 1474-4457. `https://pubmed.ncbi.nlm.nih.gov/32087114https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7159018/` (Feb. 2020).
3. Courtemanche, C, Garuccio, J, Le, A, Pinkston, J & Yelowitz, A. Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate. *Health Affairs* **39** (2020).
4. Aktay, A *et al. Google COVID-19 Community Mobility Reports: Anonymization Process Description (version 1.0)* 2020.
5. Abouk, R. & Heydari, B. The immediate effect of covid-19 policies on social distancing behavior in the united states. *Available at SSRN* (2020).
6. Flaxman, S *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* (2020).
7. Vollmer, M *et al. Report 20: Using mobility to estimate the transmission intensity of COVID-19 in Italy: A subnational analysis with future scenarios* 2020.
8. Mellan, T. A. *et al. Report 21 - Estimating COVID-19 cases and reproduction number in Brazil* 2020.
9. Kissler, S. M. *et al. Reductions in commuting mobility predict geographic differences in SARS-CoV-2 prevalence in New York City* 2020. `http://nrs.harvard.edu/urn-3:HUL.InstRepos:42665370`.

10. Long, Q.-X. *et al.* Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nature Medicine* (2020).

11. Mossong, J *et al.* Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLOS Medicine* **5,** 1–1. https://doi.org/10.1371/journal.pmed.0050074 (Mar. 2008).

12. Champagne, S. R. & Oxner, R. *Surge in coronavirus cases linked to more Texans in their 20s getting sick, officials say* Accessed on 9/7/2020. 2020. https://www.texastribune.org/2020/06/16/texas-coronavirus-spike-young-adults/.

13. NY Times. *Florida and South Carolina Again Set Records as U.S. Coronavirus Cases Surge* Accessed on 9/7/2020. 2020. https://www.nytimes.com/2020/06/20/world/coronavirus-updates.html#link-2204ff25.

14. Systrom, K., Vladek, T. & Krieger, M. *Project Title* https://github.com/rtcovidlive/covid-model. 2020.

15. Buffa, P. *Coronavirus in the U.S.: Where cases are growing and declining* Accessed on 26/6/2020. 2020. https://www.nationalgeographic.com/science/2020/05/graphic-tracking-coronavirus-infections-us/.

16. Knowles, H *et al. Seven states report highest coronavirus hospitalizations since pandemic began* Accessed on 26/6/2020. 2020. https://www.washingtonpost.com/nation/2020/06/23/coronavirus-live-updates-us/.

17. Friedman, J., Liu, P., Gakidou, E. & IHME COVID-19 Model Comparison Team. Predictive performance of international COVID-19 mortality forecasting models. *medRxiv* (2020).

18. Wang, X. *et al.* Impact of Social Distancing Measures on Coronavirus Disease Healthcare Demand, Central Texas, USA. *Emerging Infectious Diseases* **26** (2020).

19. Smith, M *et al. Coronavirus (Covid-19) Data in the United States* 2020. https://github.com/nytimes/covid-19-data.

20. reporter, C. *Census reporter* 2020. https://censusreporter.org.

21. Fullman, N *et al. State-level social distancing policies in response to COVID-19 in the US* 2020. http://www.covid19statepolicy.org.

22. Verity, R *et al.* Estimates of the severity of COVID-19 disease. *Lancet Infect Dis* (2020).

23. Walker, P *et al.* The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science.* https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/news--wuhan-coronavirus/ (2020).

24. Liu, Y, Gayle, A, Wilder-Smith, A & Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine.* ISSN: 17088305 (2020).

25. Reston, M., Sgueglia, K. & Mossburg, C. *Governors on East and West coasts form pacts to decide when to reopen economies* 2020. https://edition.cnn.com/2020/04/13/politics/states-band-together-reopening-plans/index.html.

26. Carpenter, B *et al. Stan*: A Probabilistic Programming Language. *Journal of Statistical Software* **76,** 1–32. ISSN: 1548-7660. http://www.jstatsoft.org/v76/i01/ (2017).

27. Gneiting, T. & Raftery, A. E. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* **102,** 359–378. eprint: https://doi.org/10.1198/016214506000001437. https://doi.org/10.1198/016214506000001437 (2007).

28. Marchant, R, Samia, N. I., Rosen, O, Tanner, M. A. & Cripps, S. *Learning as We Go: An Examination of the Statistical Accuracy of COVID19 Daily Death Count Predictions* 2020. `arXiv:2004.04734`.

29. IHME. *COVID-19 Projections* `https://covid19.healthdata.org/global`. 2020.

30. Gu, Y. *COVID-19 Projections Using Machine Learning* `https://github.com/youyanggu/covid19_projections`. 2020.

31. Li, M. L. *et al.* Forecasting COVID-19 and Analyzing the Effect of Government Interventions. *medRxiv* (2020).

32. Imperial College. *Future scenarios of the healthcare burden of COVID-19 in Low- or Middle-Income Countries* `https://github.com/mrc-ide/global-lmic-reports`. 2020.

33. Los Alamos National Laboratory. *LANL COVID-19 Cases and Deaths Forecasts* `https://covid-19.bsvgateway.org`. 2020.

A Model regions

We choose 8 regions in our model, see Figure 5. These regions are based on US Census Divisions, modified to account for coordination between groups of state, see Section 5 for more information.

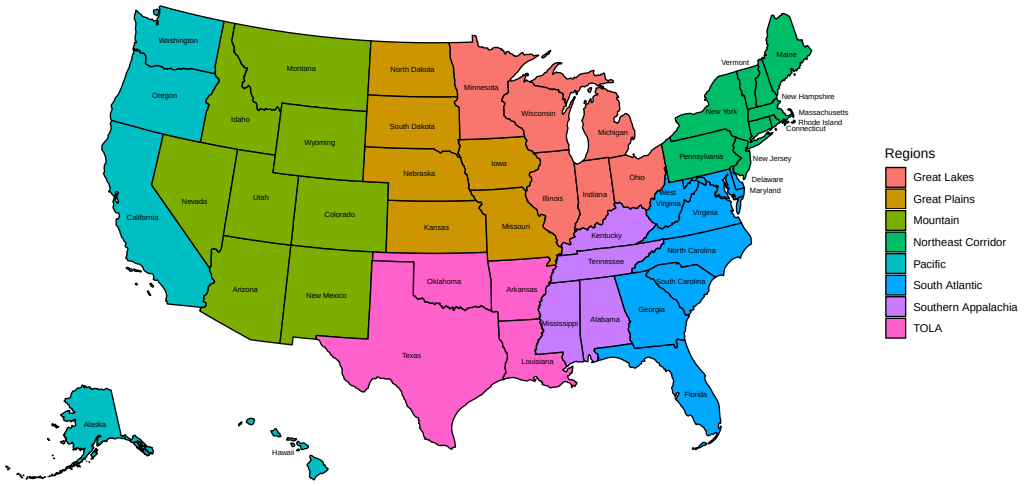


Figure 5: Map showing the 8 regions in our model. Great Lakes - Illinois Indiana, Michigan, Minnesota, Ohio and Wisconsin; Great Plains - Iowa, Kansas, Missouri, Nebraska, North Dakota and South Dakota; Mountain - Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah and Wyoming; Northeast Corridor - Connecticut, Massachusetts, Maine, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island and Vermont; Pacific - Alaska, California, Hawaii, Oregon and Washington; South Atlantic - Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia and West Virginia; Southern Appalachia - Alabama, Kentucky, Mississippi and Tennessee; TOLA - Arkansas, Louisiana, Oklahoma, Texas.

B Ratio of reported cases to estimated infections

We compare the weekly moving average of the percentage of reported cases to mean estimated infections for the time period of the model. We use the same model formulation without cases here to illustrate two points. First, the percentage is small initially, which is why we do not use cases in the model until 11 May 2020. Second, sufficient tests are being done towards the end of May so they are informative to our model and we can estimate an infection ascertainment ratio, see Section 5 for further details. Percentages may be greater than 100% when our model estimates fewer cases than were reported. This occurs in states with little data.

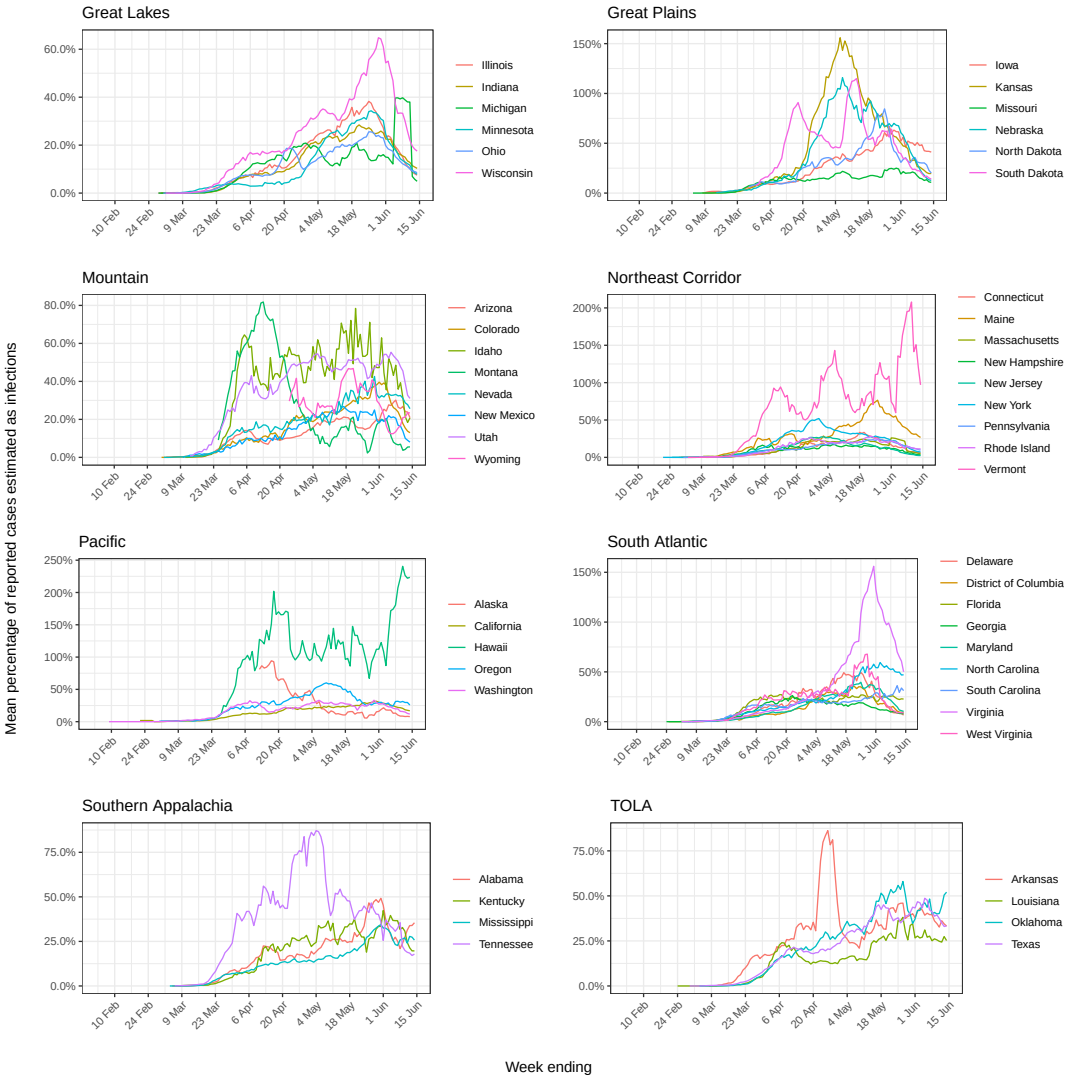


Figure 6: Weekly moving average of the percentage of reported cases to mean estimated infections for the time period of the model.

C State-level R_t values

Figure 7 shows the value of R_t at two points during the epidemic. The first time point is for the week centered on the date on which each state implemented any emergency decree order such as State of Emergency, Public Health Emergency, and Public Health Disaster declarations. The second is for the week ending on 01 June 2020.

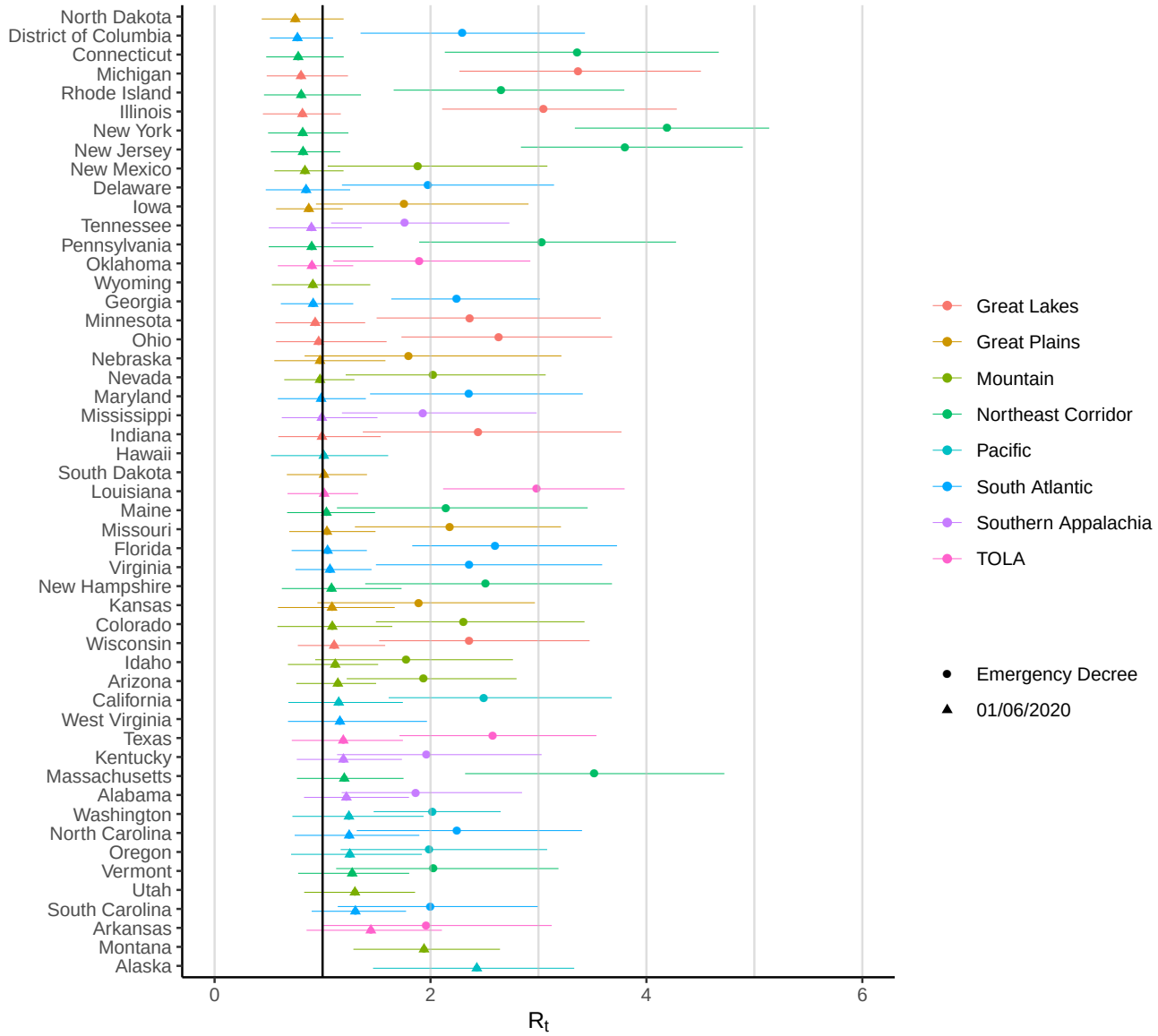
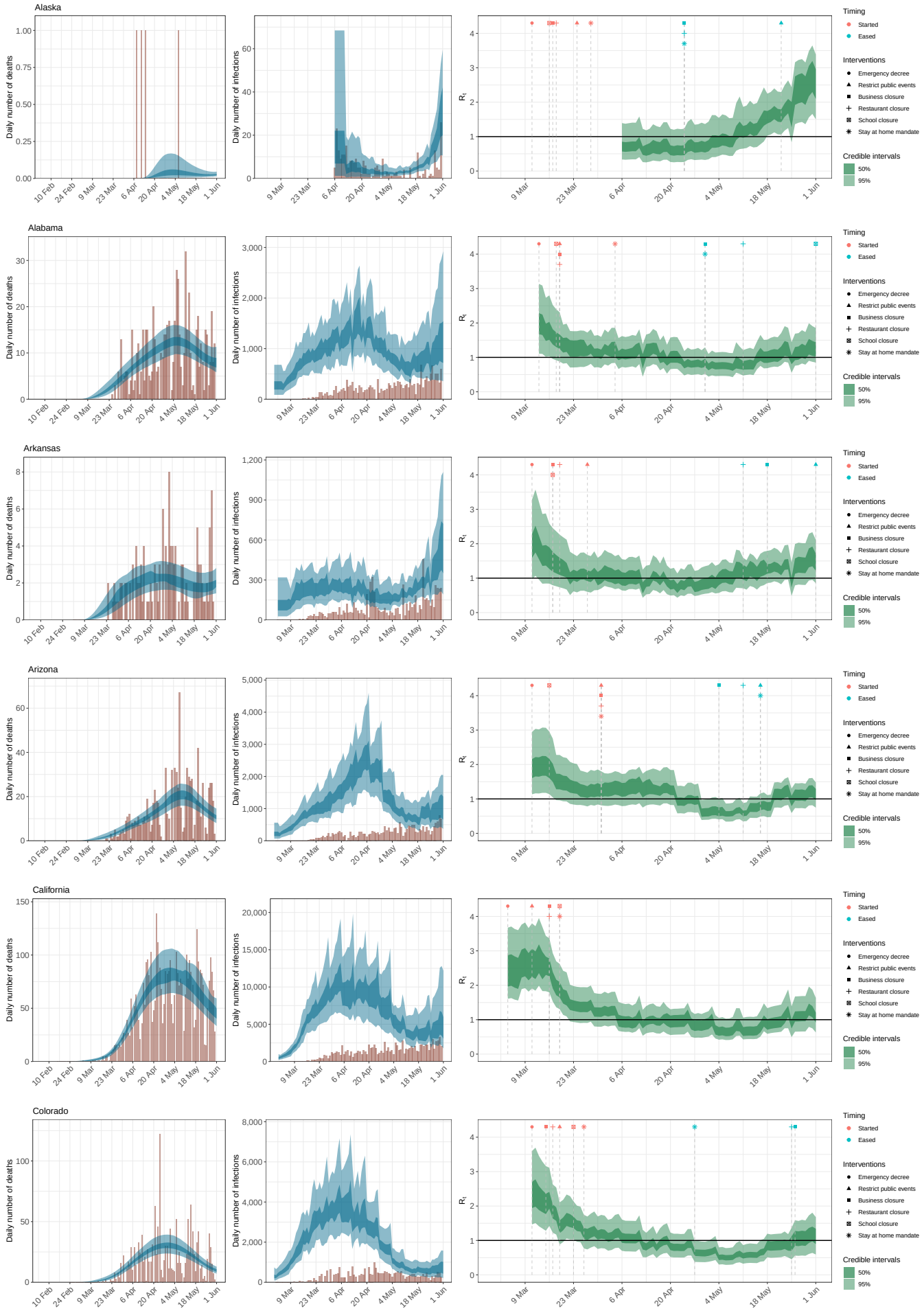
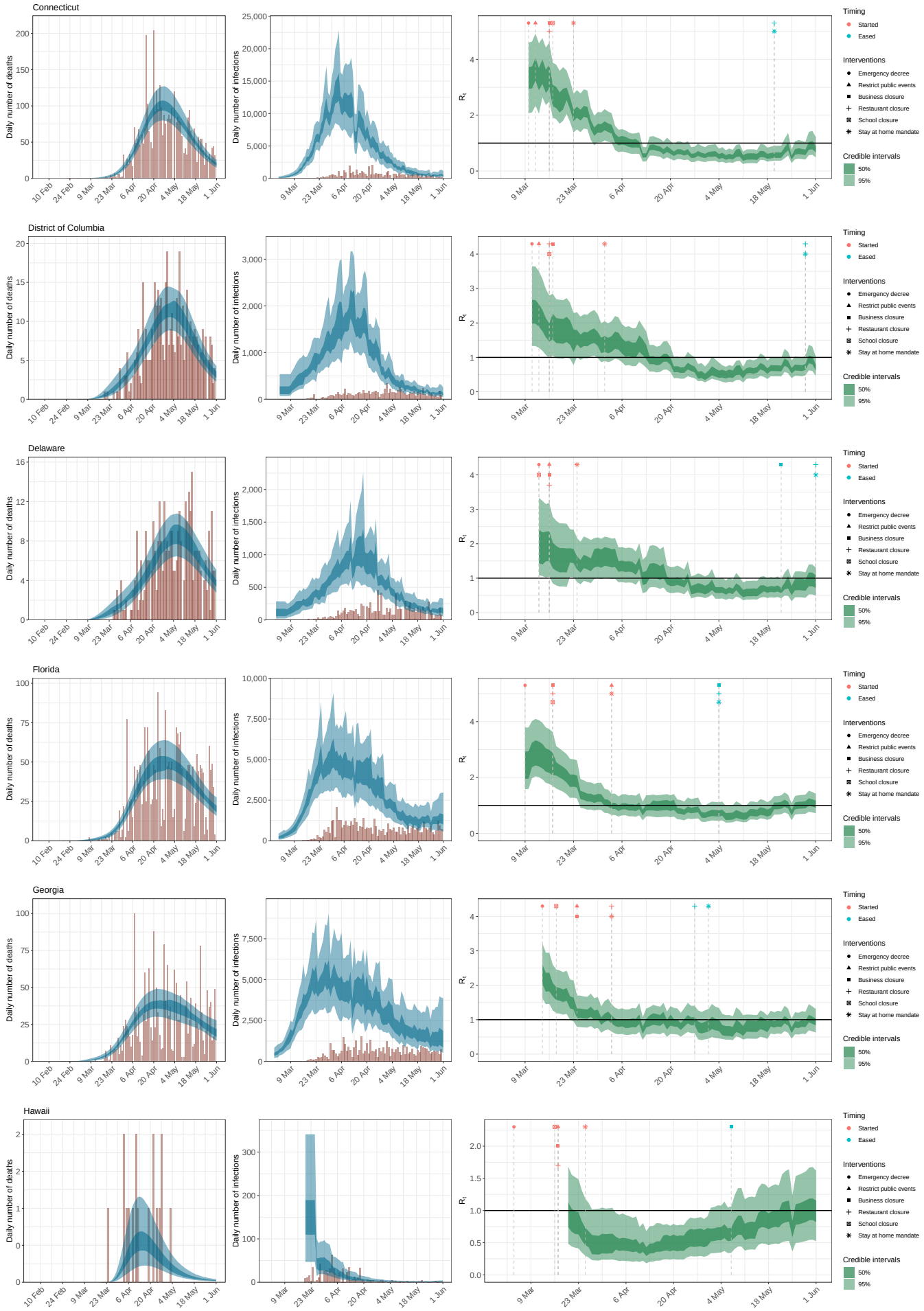


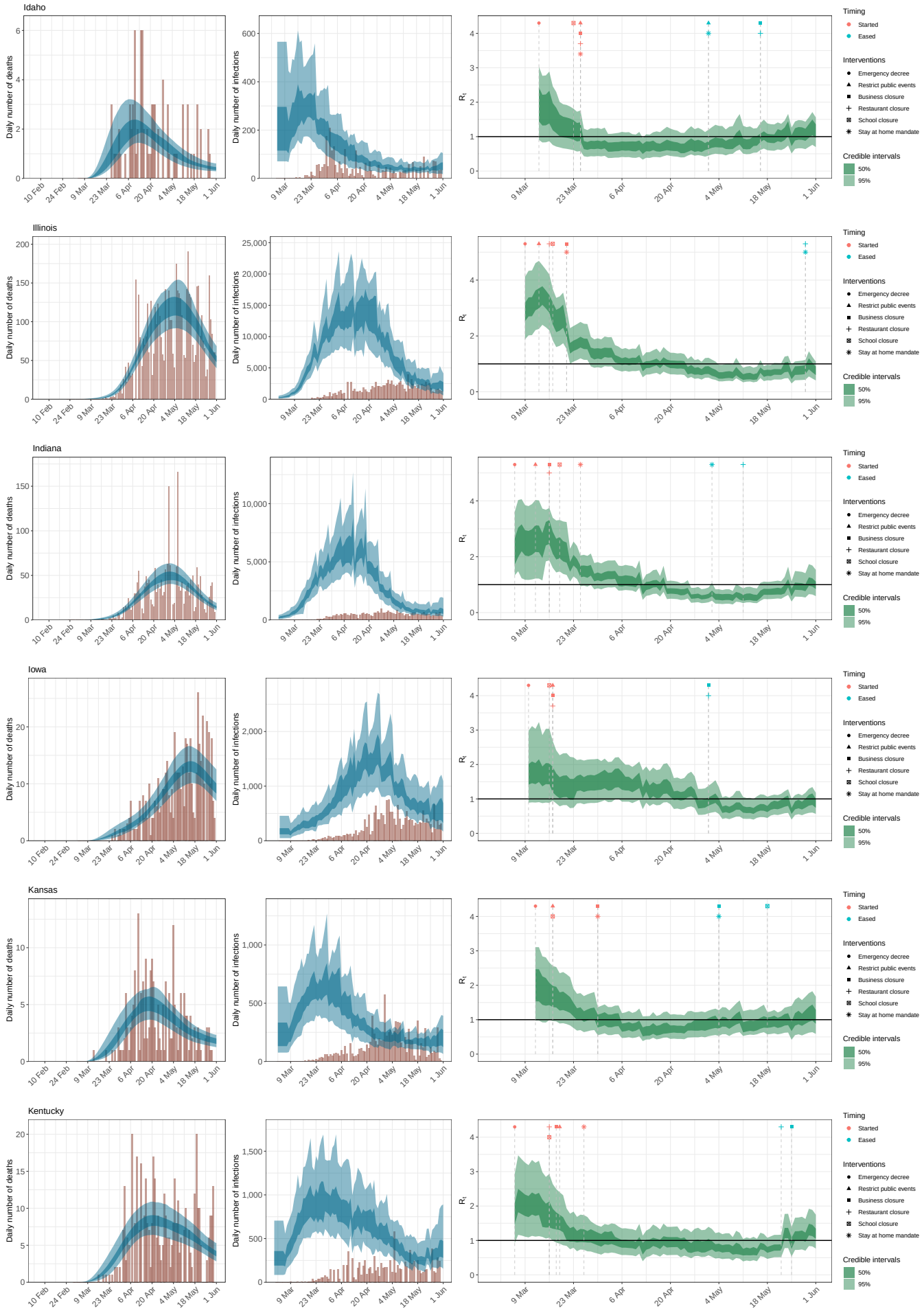
Figure 7: State-level estimates of $R_{t=EmergencyDecree}$ and the average R_t over the week ending 01 June 2020. The colours indicate regional grouping as shown in Figure 5. We do not include estimates of $R_{t=EmergencyDecree}$ for Alaska, Hawaii, Montana, North Dakota, South Dakota, Utah, West Virgina and Wyoming as Emergency Decree was declared in these states before we start modelling these states.

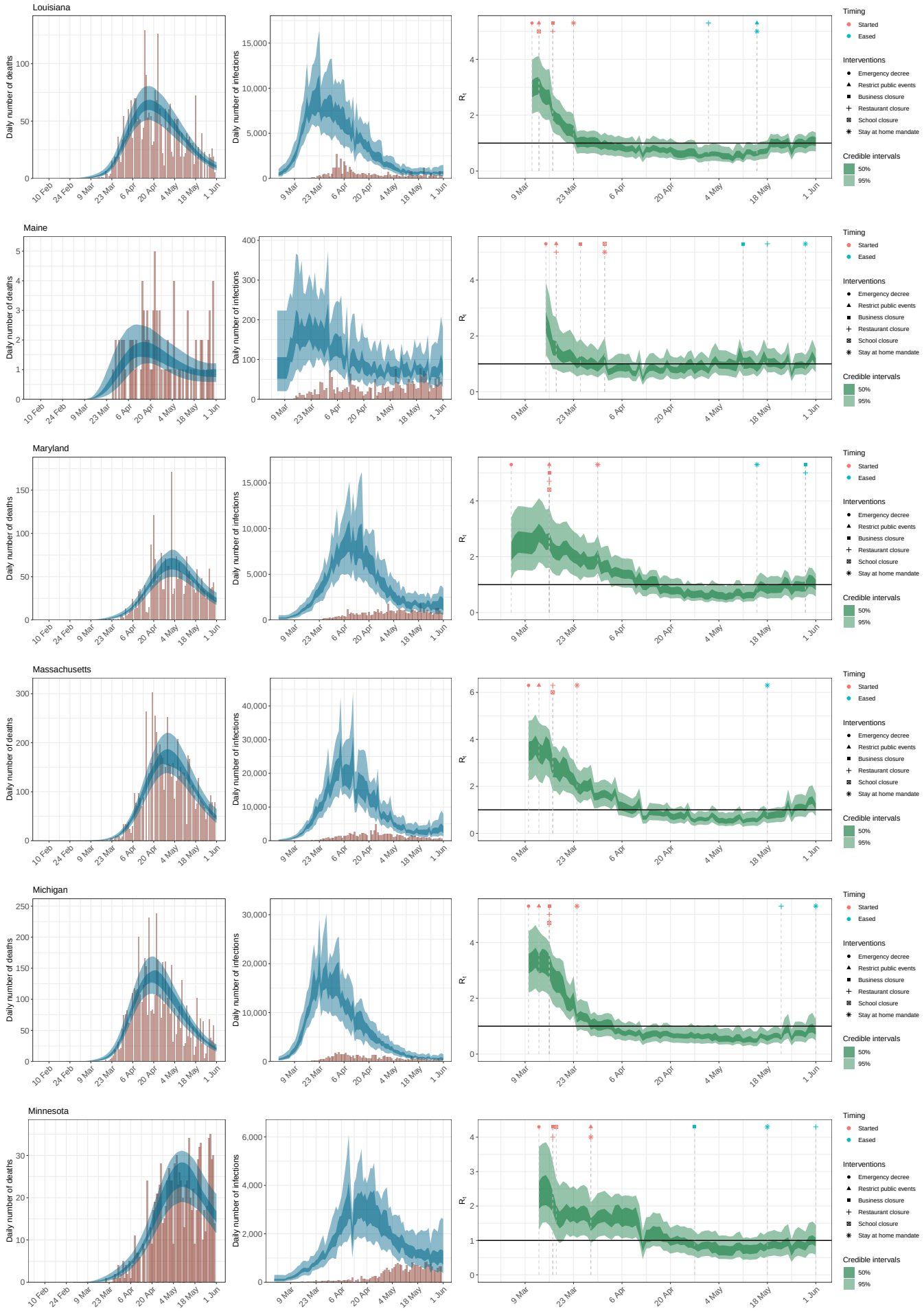
D Model predictions for all states

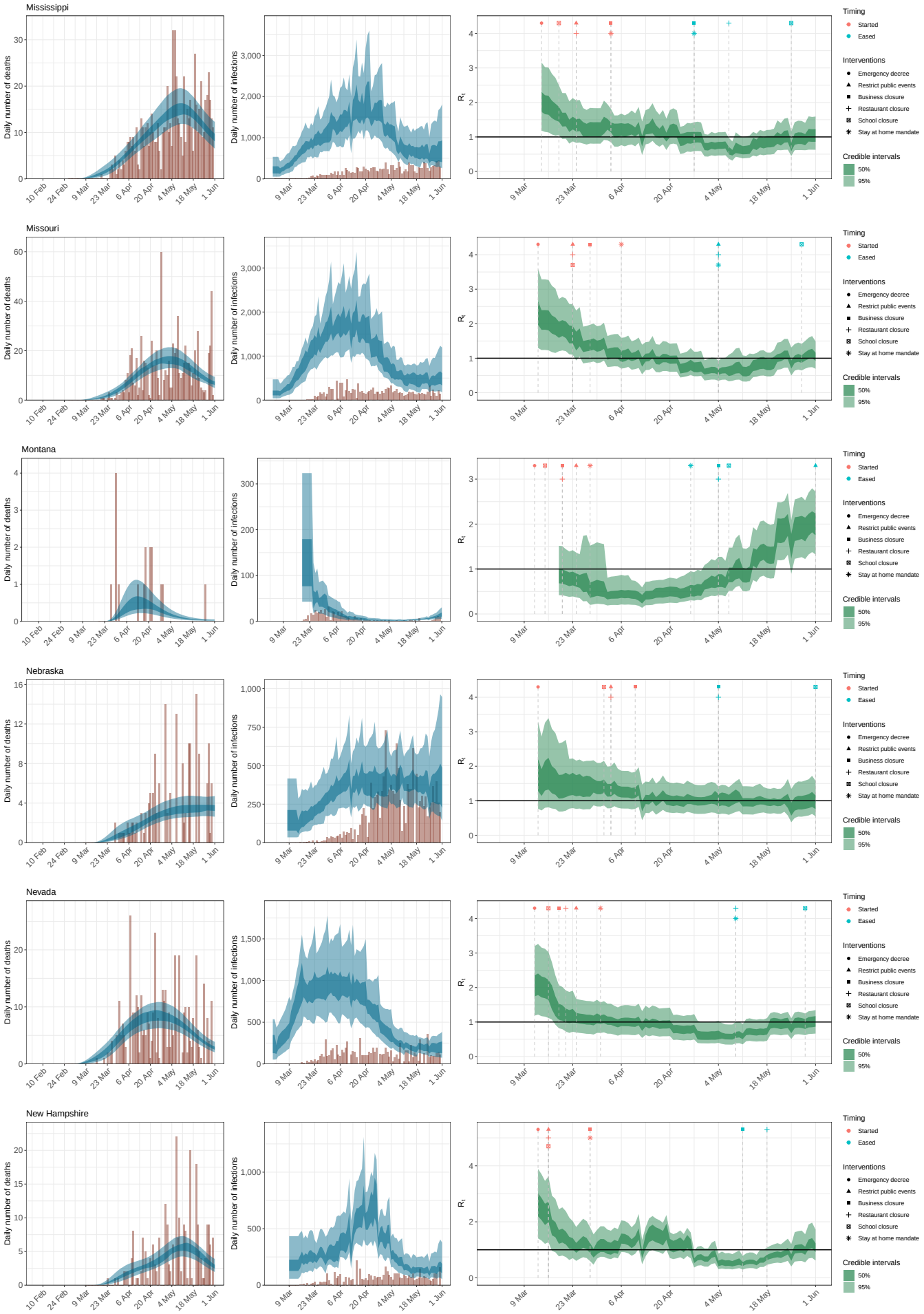
State-level estimates of infections, deaths and R_t. Left: daily number of deaths, brown bars are reported deaths, blue bands are predicted deaths, dark blue 50% credible interval (CI), light blue 95% CI. Middle: daily number of infections, brown bars are reported infections, blue bands are predicted infections, CIs are same as left. The number of daily infections estimated by our model drops immediately after an intervention, as we assume that all infected people become immediately less infectious through the intervention. Afterwards, if the R_t is above 1, the number of infections will start growing again. Right: time-varying reproduction number R_t dark green 50% CI, light green 95% CI. Icons are interventions shown at the time they occurred.

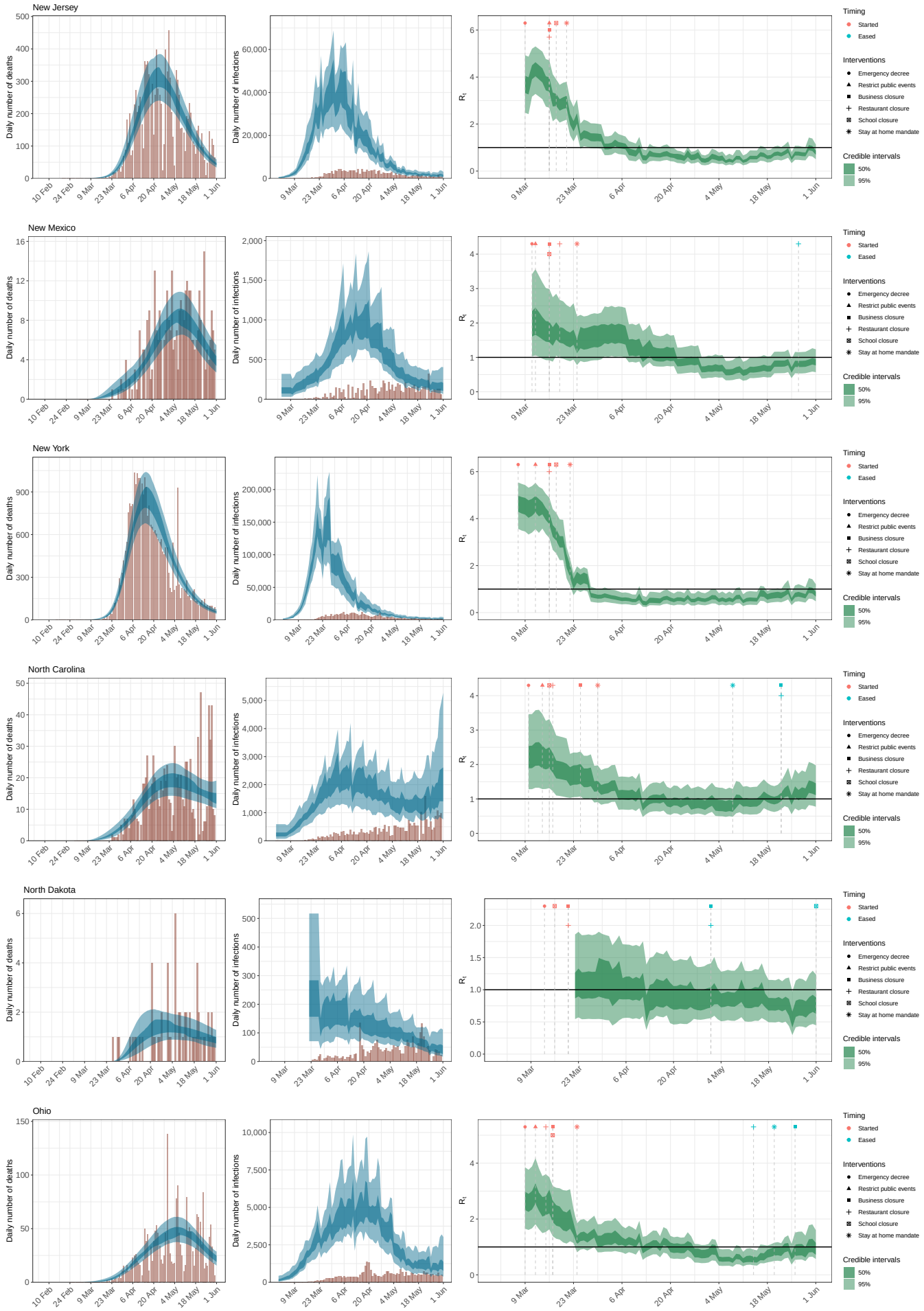

Alaska
Alabama
Arkansas
Arizona
California
Colorado
Daily number of deaths
Daily number of infections
R_t
10 Feb
24 Feb
9 Mar
23 Mar
6 Apr
20 Apr
4 May
18 May
1 Jun
Timing
Started
Eased
Interventions
Emergency decree
Restrict public events
Business closure
Restaurant closure
School closure
Stay at home mandate
Credible intervals
50%
95%

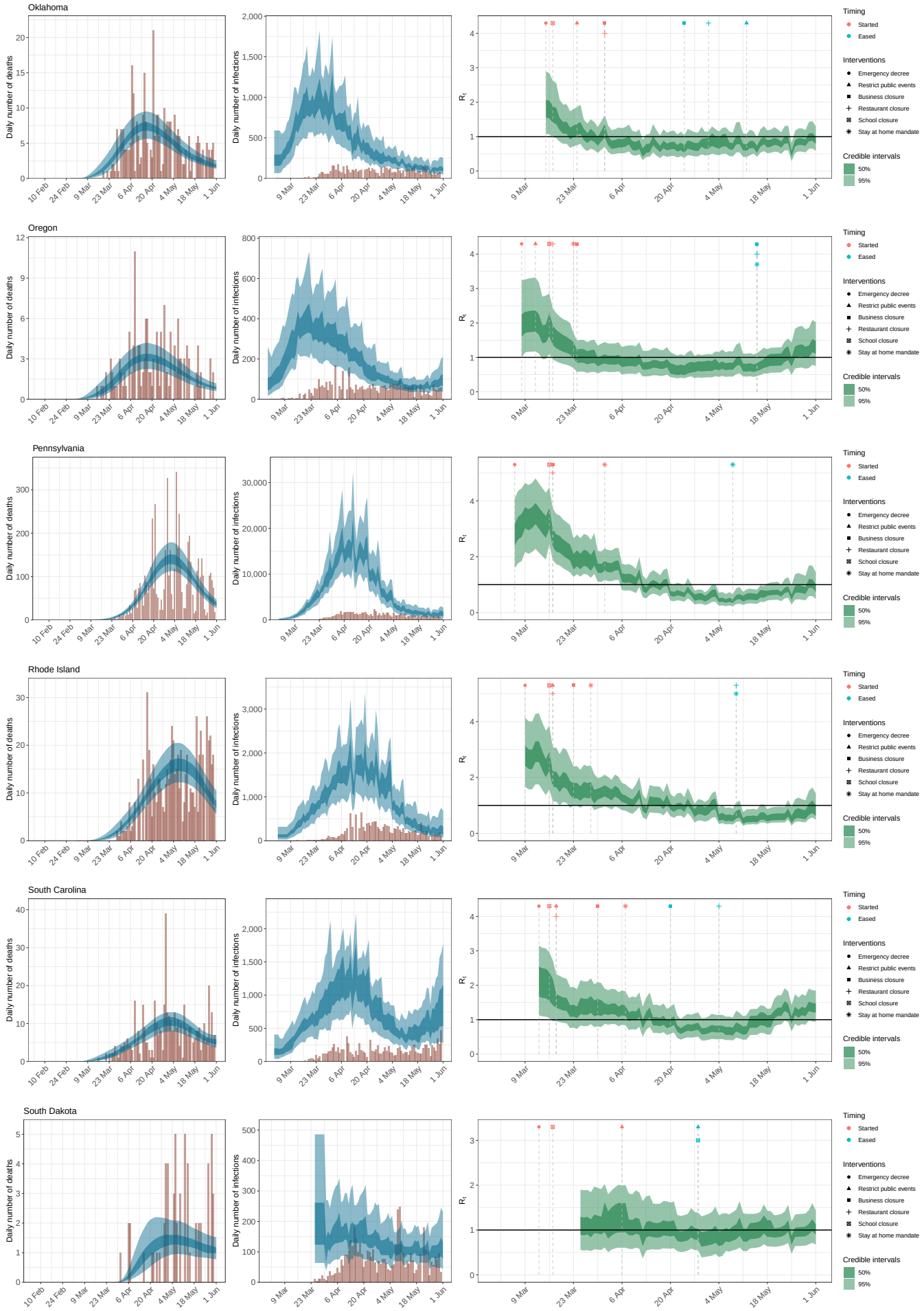

Connecticut
District of Columbia
Delaware
Florida
Georgia
Hawaii
Daily number of deaths
Daily number of infections
R_t
10 Feb
24 Feb
9 Mar
23 Mar
6 Apr
20 Apr
4 May
18 May
1 Jun
Timing
Started
Eased
Interventions
Emergency decree
Restrict public events
Business closure
Restaurant closure
School closure
Stay at home mandate
Credible intervals
50%
95%

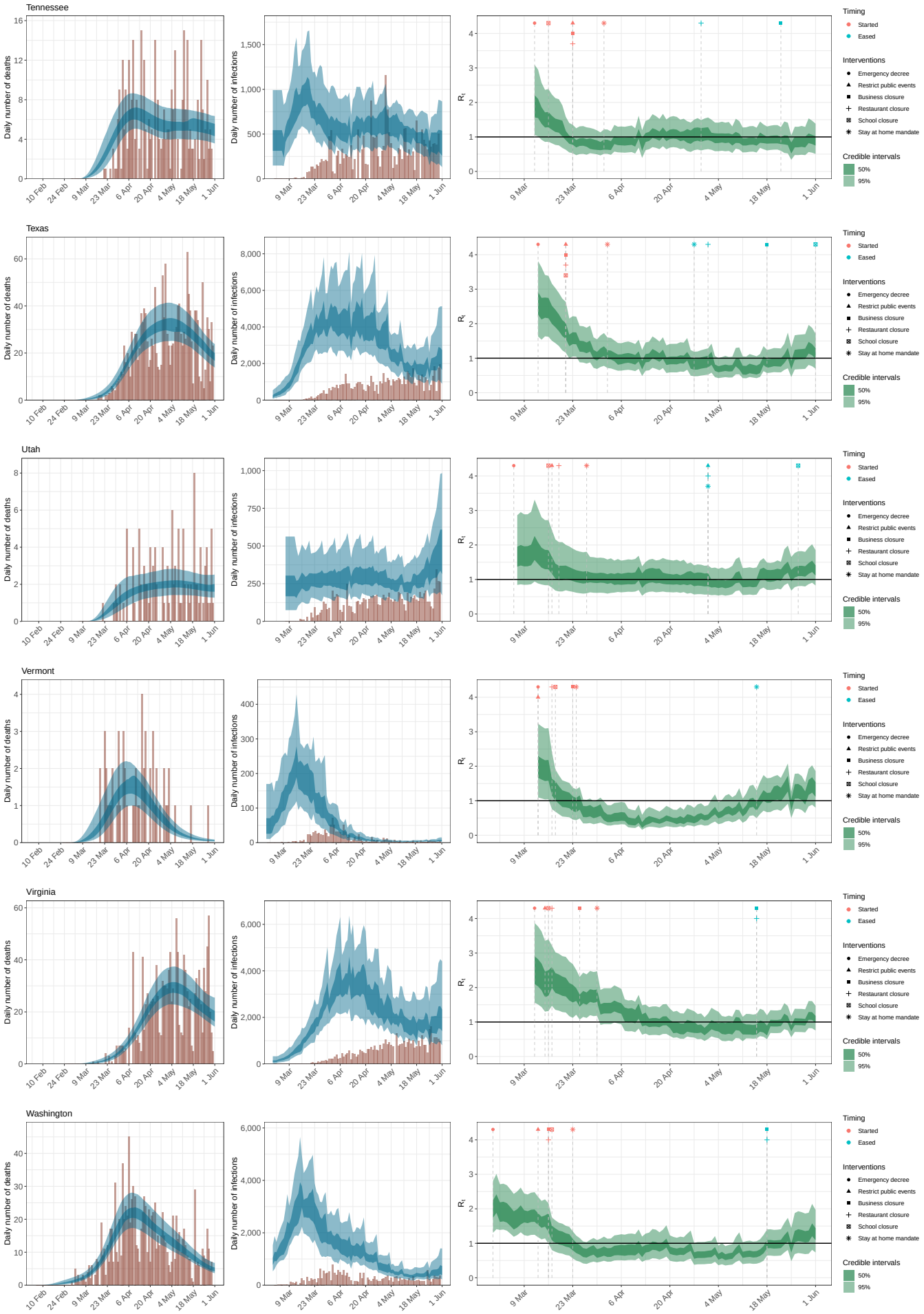

Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Daily number of deaths
Daily number of infections
R_t
10 Feb
24 Feb
9 Mar
23 Mar
6 Apr
20 Apr
4 May
18 May
1 Jun
Timing
Started
Eased
Interventions
Emergency decree
Restrict public events
Business closure
Restaurant closure
School closure
Stay at home mandate
Credible intervals
50%
95%

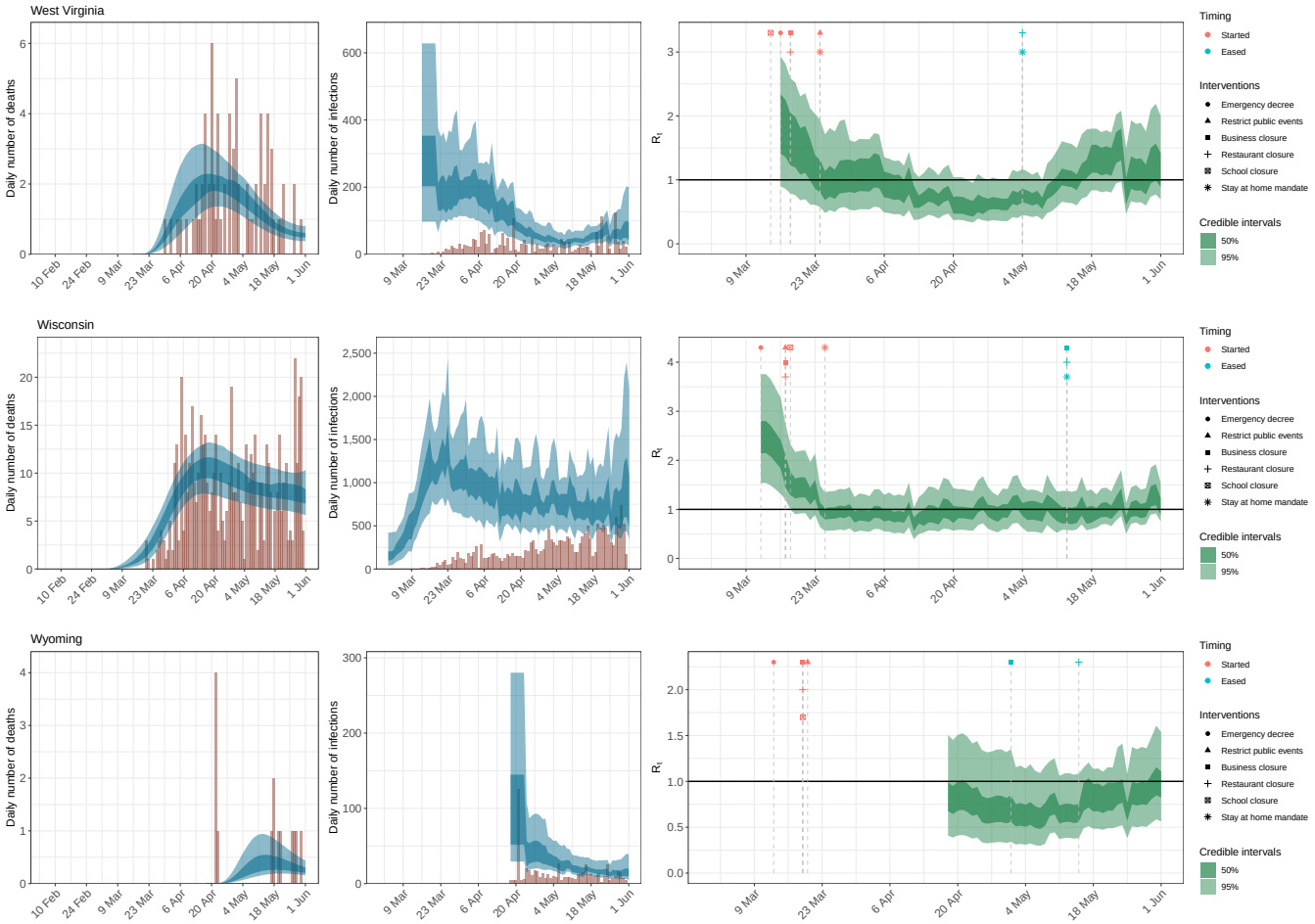

Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Daily number of deaths
Daily number of infections
R_t
10 Feb
24 Feb
9 Mar
23 Mar
6 Apr
20 Apr
4 May
18 May
1 Jun
Timing
Started
Eased
Interventions
Emergency decree
Restrict public events
Business closure
Restaurant closure
School closure
Stay at home mandate
Credible intervals
50%
95%


Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
Daily number of deaths
Daily number of infections
R_t
10 Feb
24 Feb
9 Mar
23 Mar
6 Apr
20 Apr
4 May
18 May
1 Jun
Timing
Started
Eased
Interventions
Emergency decree
Restrict public events
Business closure
Restaurant closure
School closure
Stay at home mandate
Credible intervals
50%
95%


New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Daily number of deaths
Daily number of infections
R_t
10 Feb
24 Feb
9 Mar
23 Mar
6 Apr
20 Apr
4 May
18 May
1 Jun
Timing
Started
Eased
Interventions
Emergency decree
Restrict public events
Business closure
Restaurant closure
School closure
Stay at home mandate
Credible intervals
50%
95%


Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Daily number of deaths
Daily number of infections
R_t
10 Feb
24 Feb
9 Mar
23 Mar
6 Apr
20 Apr
4 May
18 May
1 Jun
Timing
Started
Eased
Interventions
Emergency decree
Restrict public events
Business closure
Restaurant closure
School closure
Stay at home mandate
Credible intervals
50%
95%


Tennessee
Texas
Utah
Vermont
Virginia
Washington
Daily number of deaths
Daily number of infections
R_t
10 Feb
24 Feb
9 Mar
23 Mar
6 Apr
20 Apr
4 May
18 May
1 Jun
Timing
Started
Eased
Interventions
Emergency decree
Restrict public events
Business closure
Restaurant closure
School closure
Stay at home mandate
Credible intervals
50%
95%



E Effect sizes

We plot estimates of the posterior mean effect sizes and 95% credible intervals for each mobility category. The relative % reduction in R_t metric is interpreted as follows: the larger the percentage, the more R_t decreases, meaning the disease spreads less; a 100% relative reduction ends disease transmissibility entirely. The smaller the percentage, the less effect the covariate has on transmissibility. A 0% relative reduction has no effect on R_t and thus no effect on the transmissibility of the disease, while a negative percent reduction implies an increase in transmissibility.

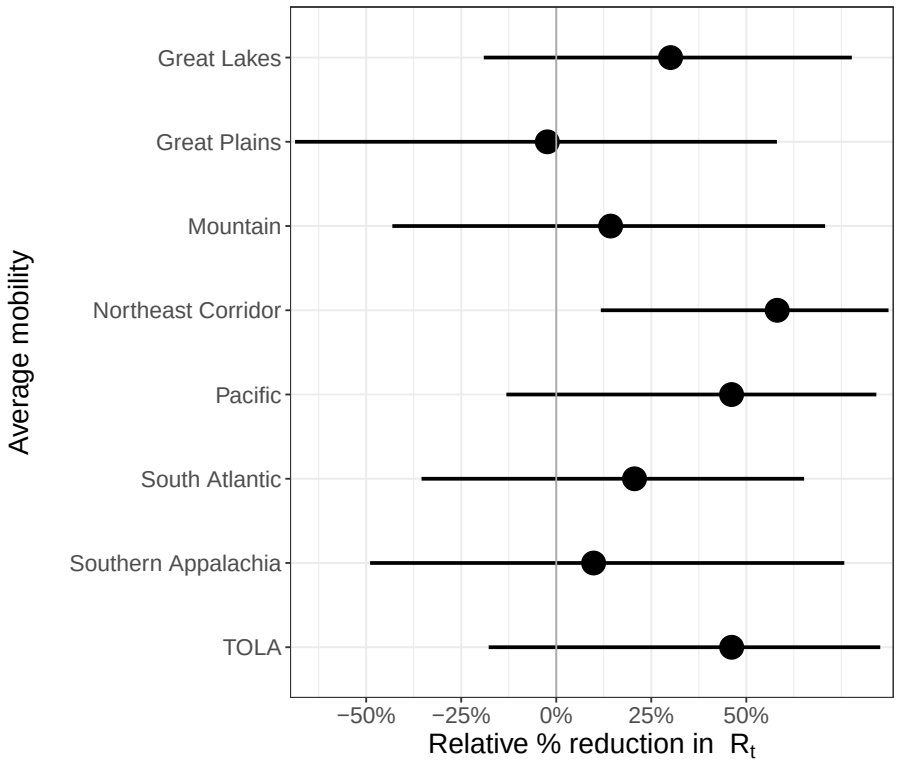


Figure 8: Regional average mobility covariate effect size plots assuming mobility stopped entirely (100% reduction in average mobility).

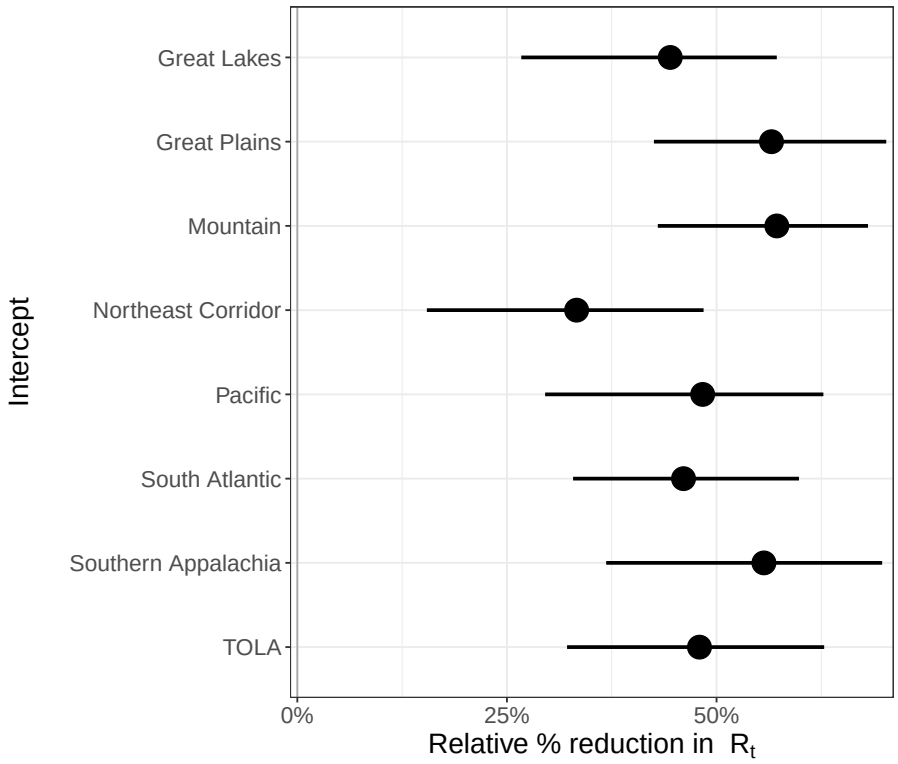


Figure 9: Regional intercept covariate effect size plots.

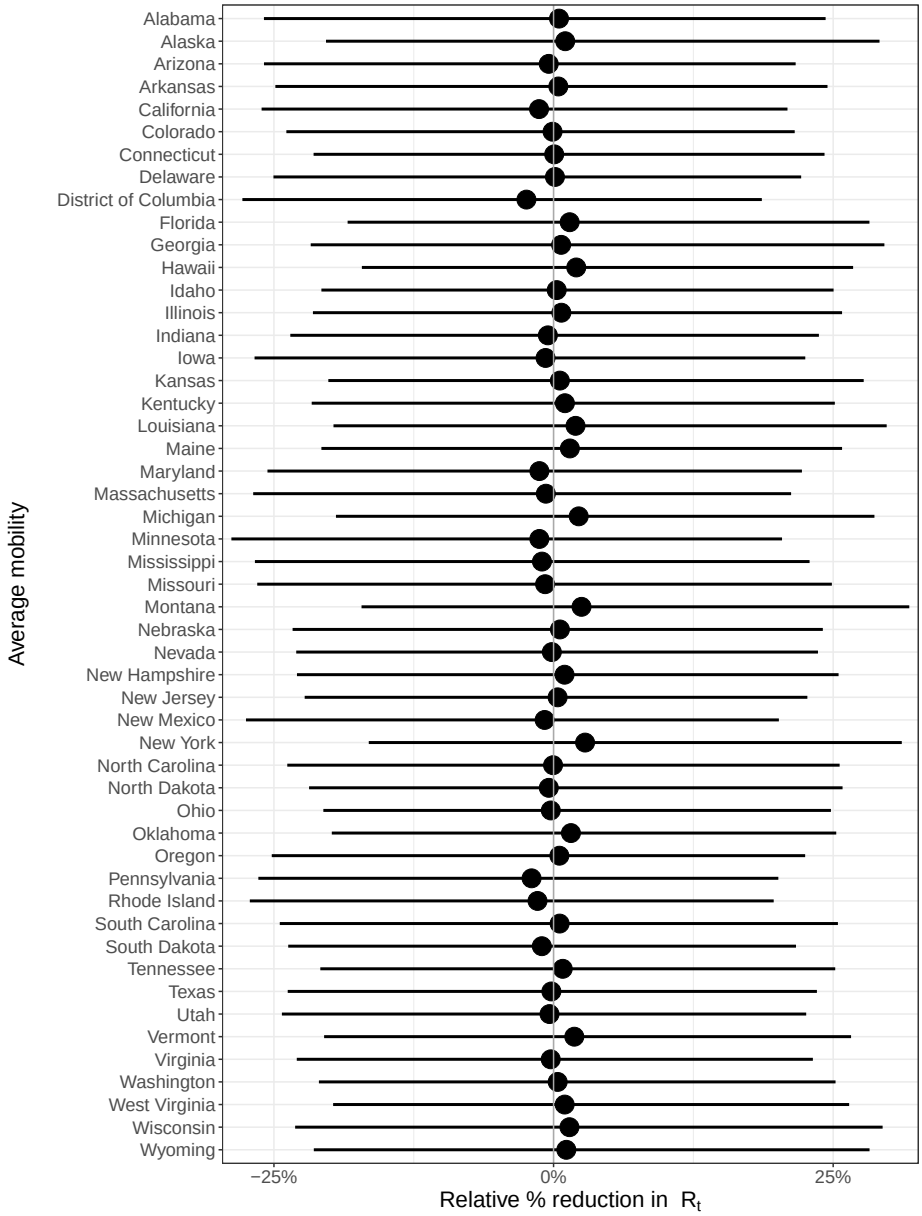


Figure 10: State-level covariate effect size plots assuming mobility stopped entirely (100% reduction in average mobility).

F Forecast evaluation

We evaluate three-week ahead model forecasts (days $t \in \{T+1, \ldots, T+21\}$) for our model using two metrics. The first metric is the mean absolute error (MAE), which is given by

$$\mathrm{MAE}_{t,m} = \frac{1}{S} \sum_{s=1}^{S} |\widehat{y}_{t,m}^{s} - y_{t,m}| \,, \qquad t = T+1, \ldots, T+21 \,, \tag{19}$$

where $\widehat{y}_{t,m}^{1}, \ldots, \widehat{y}_{t,m}^{S}$ are S posterior predictive samples of daily deaths on day t in country m and $y_{t,m}$ is the observed value of the corresponding quantity. The continuous ranked probability score (CRPS) is a generalisation of MAE to probabilistic forecasts and can be estimated using Gneiting and Raftery [27]

$$\mathrm{CRPS}_{t,m} = \frac{1}{S} \sum_{s=1}^{S} |\widehat{y}_{t,m}^{s} - y_{t,m}| - \frac{1}{2S^2} \sum_{j=1}^{S} \sum_{k=1}^{S} |\widehat{y}_{t,m}^{j} - \widehat{y}_{t,m}^{k}| \,, \qquad t = T+1, \ldots, T+21 \,. \tag{20}$$

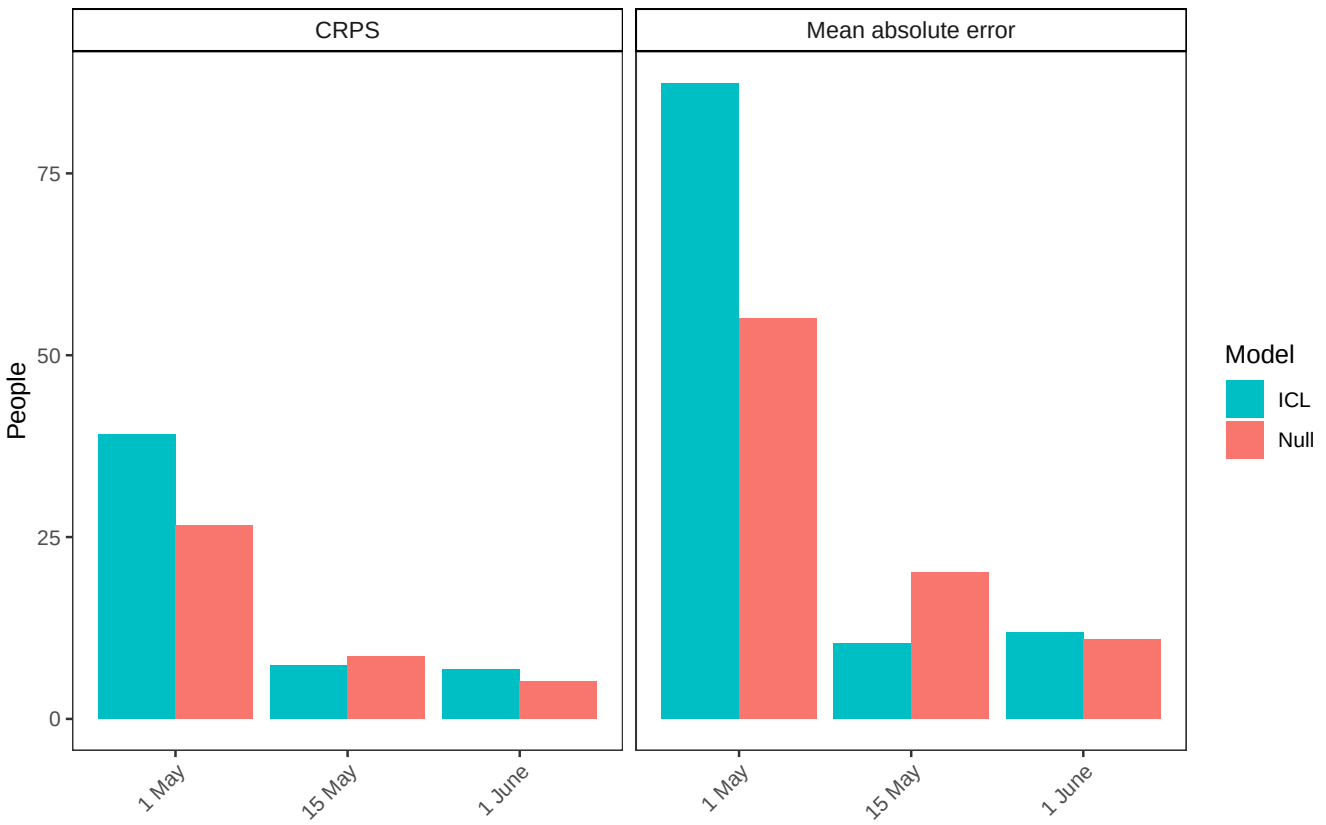


Figure 11: MAE and CRPS 21 day forecast estimates for three models fit until 1 May, 15 May and 1 June. The blue ICL model refers to our model and the coral bars refer to the null model.

Figure 11 shows the mean MAE and CRPS for three different forecast periods across all time points and states. In addition to the forecast results shown in Section 2.4, we also fit our model up to 1 May and 15 May and show the error in our 21 day forecasts. We compare these forecasts to a simple log-linear "null" model fit to the 31 days of death data prior to the end dates of our model runs (1 May, 15 May and 1 June) to evaluate our predictions. We find the MAE and CRPS error were different for models fit up to different dates. Our model fit to 1 May performs worse than the "null" model because this model only includes deaths (we only include cases in our model after 11 May, see Appendix B. Our model performs similarly (1 June) or better than the null model (15 May) in both MAE and CRPS. We include a state-level break down of MAE in Figure 12 and CRPS in Figure 13.

The forecasts were also evaluated using the mean coverage of their 95% and 50% credible intervals. If model uncertainty is well-calibrated then the observed quantities will fall outside of the 95% credible intervals 5% of the time and 50% of the time for the 50% credible intervals. Recent work has highlighted that other prominent models do not meet this criterion, which suggests that their uncertainty estimates are not well-calibrated [28]. The mean coverage of the 95% credible interval in a time period starting at time t_0 with length L is given by

$$\frac{1}{L}\sum_{t=t_0+1}^{t_0+1+L} 1\left(y_{t,m} \in \left[p_{2.5}(\{\widehat{y}^s_{t,m}\}_{s=1}^S), p_{97.5}(\{\widehat{y}^s_{t,m}\}_{s=1}^S)\right]\right), \tag{21}$$

where $1(\cdot)$ is the indicator function and $p_z(\{\widehat{y}^s_{t,m}\}_{s=1}^S)$ is the z-th percentile of the samples on day t in country m. The mean coverage of the 50% credible interval is

$$\frac{1}{L}\sum_{t=t_0+1}^{t_0+1+L} 1\left(y_{t,m} \in \left[p_{25}(\{\widehat{y}^s_{t,m}\}_{s=1}^S), p_{75}(\{\widehat{y}^s_{t,m}\}_{s=1}^S)\right]\right). \tag{22}$$

The coverage was calculated for each state individually and then the mean across all the 50 state and the District of

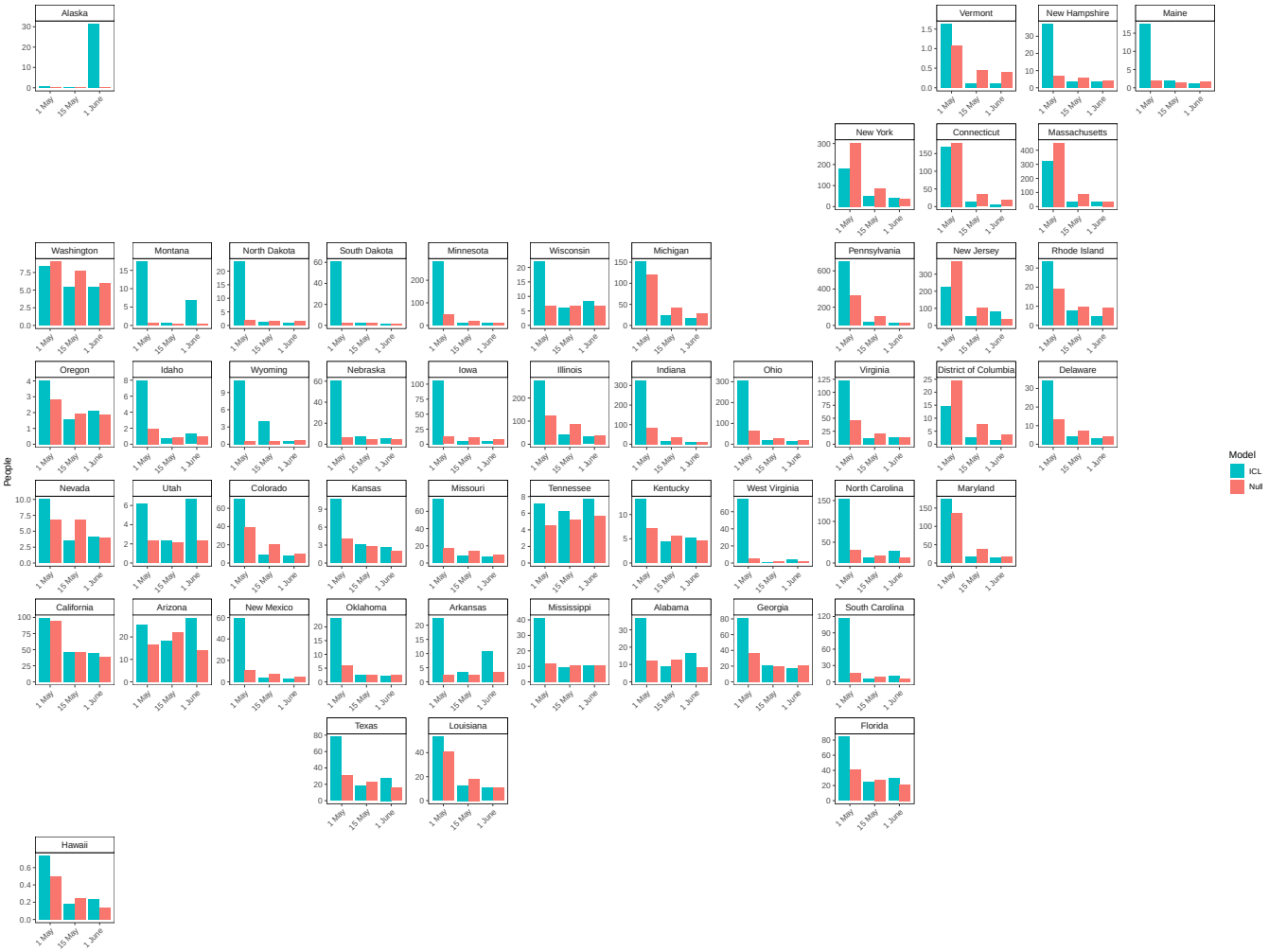


Figure 12: State level 21 day forecast mean absolute error estimates for three models fit until 1 May, 15 May and 1 June. The blue ICL model refers to our model and the coral bars refer to the null model.

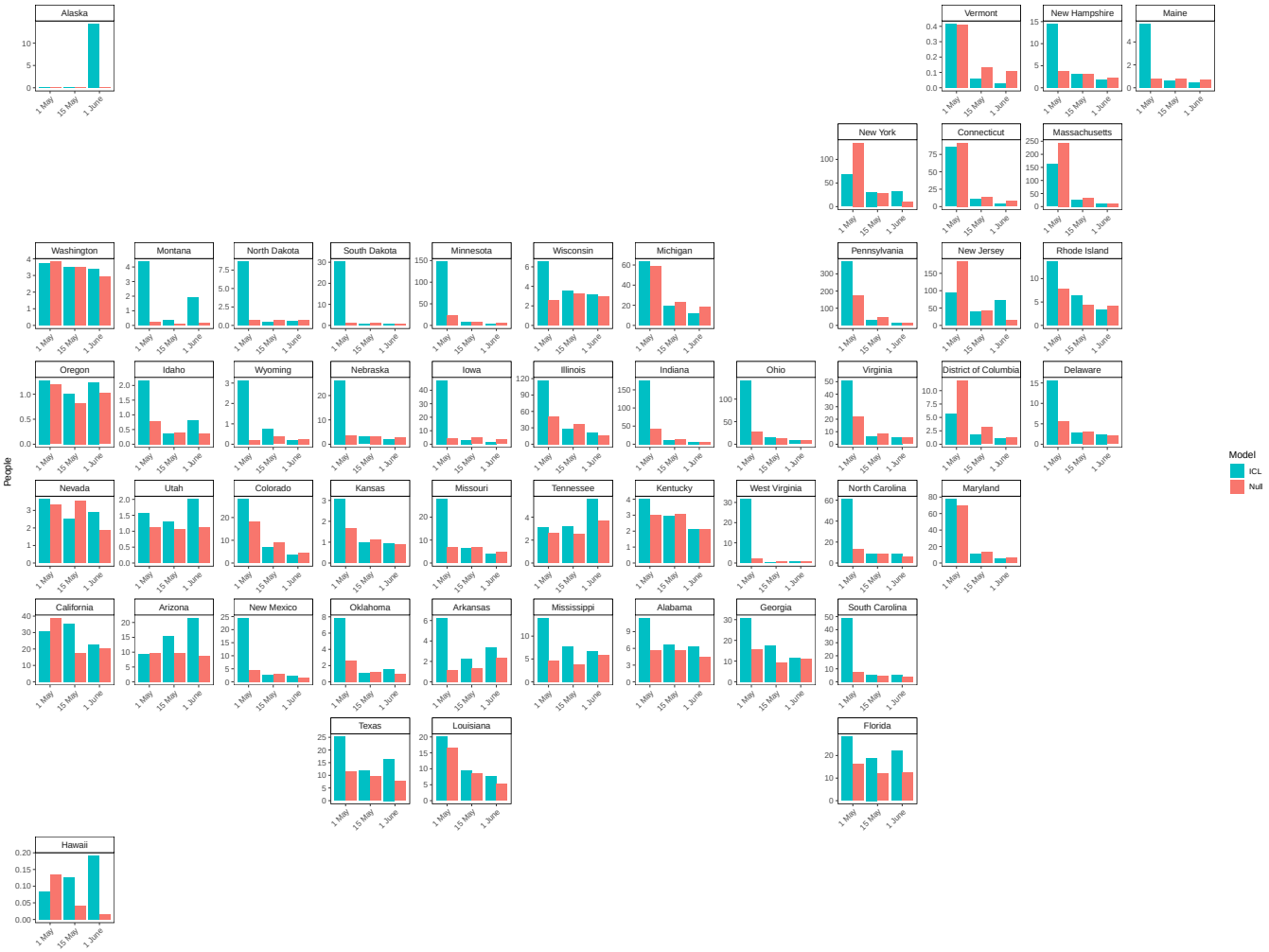


Figure 13: State level 21 day forecast CRPS estimates for three models fit until 1 May, 15 May and 1 June. The blue ICL model refers to our model and the coral bars refer to the null model.

Columbia was computed. Again these coverage probabilities were compared with the log-linear "null" model.

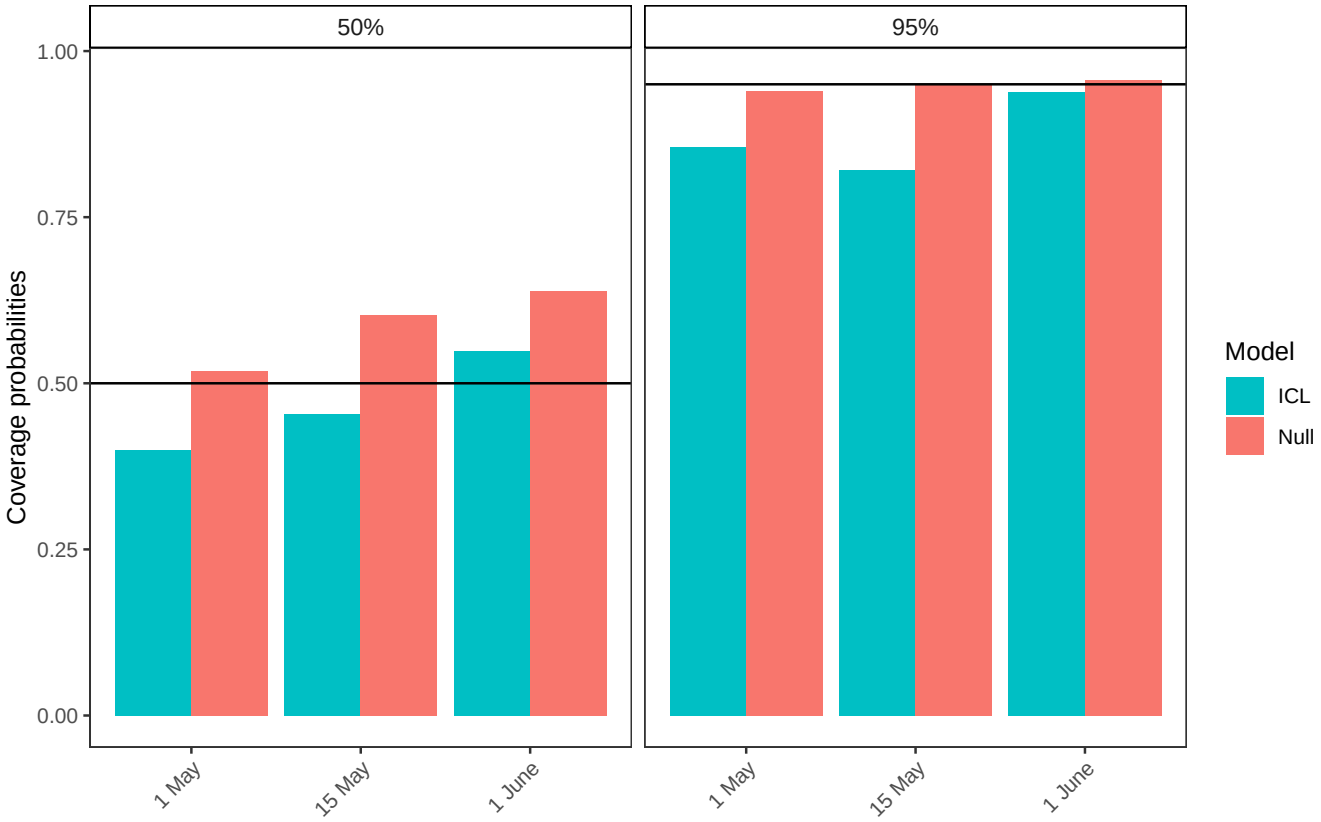


Figure 14: Coverage probabilities for 21 day forecasts for three models with data fit to 1 May, 15 May and 1 June. The blue ICL model refers to our model and the coral bars refer to the null model.

These coverage results suggest that both the 95 and 50% credible intervals are well-calibrated for all the models but best for our model and the null model fit to the 1 June. We hypothesise this is because they include more in data in the forecast.

G Median absolute percent error

We compared the magnitude of our cumulative death median absolute percent error (MAPE) estimates from our forecasts from 1 June given in Table 2 with global COVID-19 forecasts found in Friedman et al. [17]. This is not a parameter we explicitly wanted to forecast but enabled us to compare with other models in literature. Friedman et al.'s paper included both SEIR type models (IHME – MS SEIR [29], Youyang Gu [30], MIT - DELPHI [31] and Imperial-LMIC [32]) and dynamic growth rate models (LANL-Growthrate [33]). The forecasts in Friedman et al. [17] were not all fit to the same date range, but to some point in June and are only presented for the whole of the US and not to each state specifically. Unlike the models compared by Friedman et al., the MAPE of our cumulative death forecasts did not increase significantly over time. Our 3 week median cumulative death MAPE across all states (9.9%) was similar to the US estimate from Friedman et al. (4.1-8.6 excluding the Imperial-LMIC model which was not calibrated for high income settings and never meant to be used for the US), but the magnitude was slightly higher in our forecasts for some states, in particular Alaska and Wyoming, which were the last two states to reach 10 cumulative deaths and so had the least data in our models.

Table 2: State level 1 week, 2 week and 3 week median absolute percentage error (MAPE) from our cumulative death forecasts fitted until 1 June.

State	**1 week**	**2 weeks**	**3 weeks**
Alabama	7.8%	8.1%	9.1%
Alaska	75.0%	66.7%	50.0%
Arizona	9.1%	13.5%	17.4%
Arkansas	9.7%	10.4%	13.7%
California	7.4%	8.4%	10.0%
Colorado	7.1%	7.1%	7.1%
Connecticut	7.4%	7.4%	7.4%
Delaware	8.3%	8.3%	8.3%
District of Columbia	8.6%	8.3%	8.4%
Florida	8.5%	10.3%	12.9%
Georgia	6.6%	8.8%	10.2%
Hawaii	29.4%	29.4%	29.4%
Idaho	15.5%	13.6%	14.4%
Illinois	8.4%	8.9%	9.9%
Indiana	8.6%	9.4%	10.8%
Iowa	8.6%	8.2%	7.9%
Kansas	11.7%	11.4%	11.2%
Kentucky	8.1%	8.0%	7.9%
Louisiana	7.7%	7.7%	8.1%
Maine	13.1%	11.9%	11.8%
Maryland	7.6%	7.7%	7.9%
Massachusetts	7.6%	7.6%	7.3%
Michigan	7.8%	7.7%	7.7%
Minnesota	8.9%	9.8%	9.7%
Mississippi	9.3%	9.4%	11.7%
Missouri	7.6%	8.2%	10.3%
Montana	33.3%	31.6%	33.3%
Nebraska	11.1%	10.4%	10.4%
Nevada	9.0%	8.8%	8.9%
New Hampshire	11.5%	15.3%	16.2%
New Jersey	8.1%	8.0%	8.0%
New Mexico	10.5%	12.7%	14.5%
New York	7.8%	7.9%	7.9%
North Carolina	8.2%	8.0%	8.2%
North Dakota	13.9%	13.5%	13.0%
Ohio	7.5%	8.2%	8.9%
Oklahoma	8.6%	8.3%	8.1%
Oregon	9.8%	11.1%	12.0%
Pennsylvania	8.1%	8.9%	9.4%
Rhode Island	8.3%	8.8%	10.6%
South Carolina	8.4%	8.6%	9.4%
South Dakota	17.2%	13.5%	13.8%
Tennessee	8.3%	8.0%	9.4%
Texas	6.9%	7.3%	9.2%
Utah	11.6%	10.8%	11.7%
Vermont	16.1%	16.1%	15.8%
Virginia	7.5%	7.7%	8.6%
Washington	7.3%	7.6%	8.3%
West Virginia	14.3%	13.6%	15.7%
Wisconsin	8.3%	8.2%	8.4%
Wyoming	33.3%	31.2%	33.3%
Median	8.5%	8.8%	9.9%

H Mobility trends

In Figure 15 we show the Google mobility trends [4] across the 50 states of the US and the District of Columbia. In our model, we use the time spent at one's residence and the number of visits to grocery stores, pharmacies, recreation centres, and workplaces. Our regions are based on US Census Divisions, modified to account for coordination between groups of state governments [25]. These trends are relative to a state-dependent baseline, which was calculated shortly before the COVID-19 epidemic. For example, a value of -20% in the average trend means that individuals, on average, are visiting 20% less shops, recreation places and workplaces than before the epidemic. We overlay the timing of two major state-wide NPIs (stay at home and emergency decree) on Figure 15 (see [21] for details). We note intuitive changes in mobility such as the spike on 11th and 12th April for Easter.

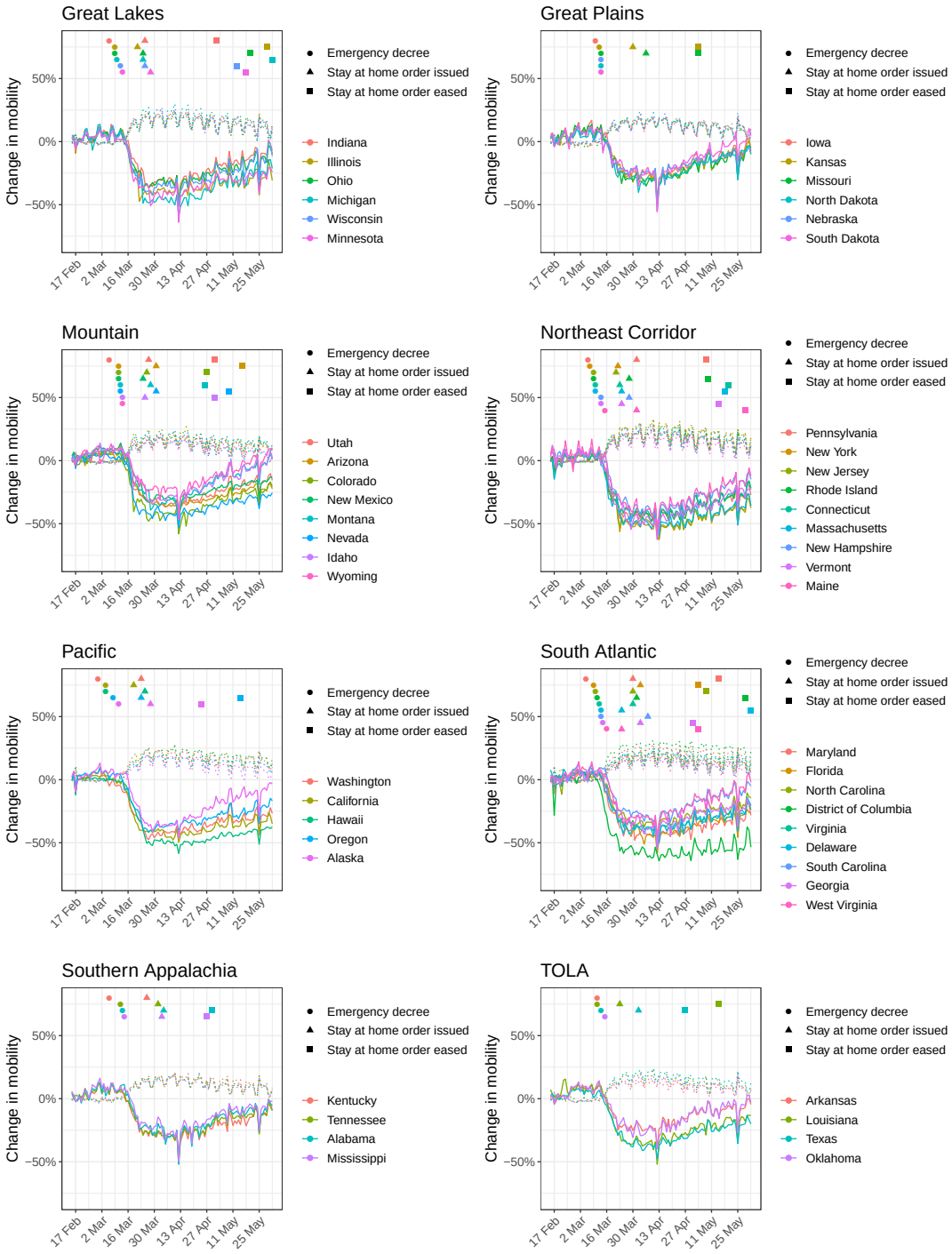


Figure 15: Comparison of mobility data from Google with government interventions for the 50 states and the District of Columbia. The solid lines show average mobility (across categories "retail & recreation", "grocery & pharmacy", "workplaces") and the dotted lines show "residential". Intervention dates are indicated by shapes as shown in the legend; see Section 4 for more information about the interventions implemented. There is a strong correlation between the onset of interventions and reductions in mobility.

I Mobility regression analysis

In Figure 16 we regressed NPIs against average mobility. We parameterised NPIs as piece-wise constant functions that are zero when the intervention has not been implemented and one when it has. We evaluated the correlation between

the predictions from the linear model and the actual average mobility. We also lagged the timing of interventions and investigate its impact on predicted correlation. We observed reduced correlation when lagging (forward and backwards) the timing of NPIs, which suggested an immediate impact on mobility.

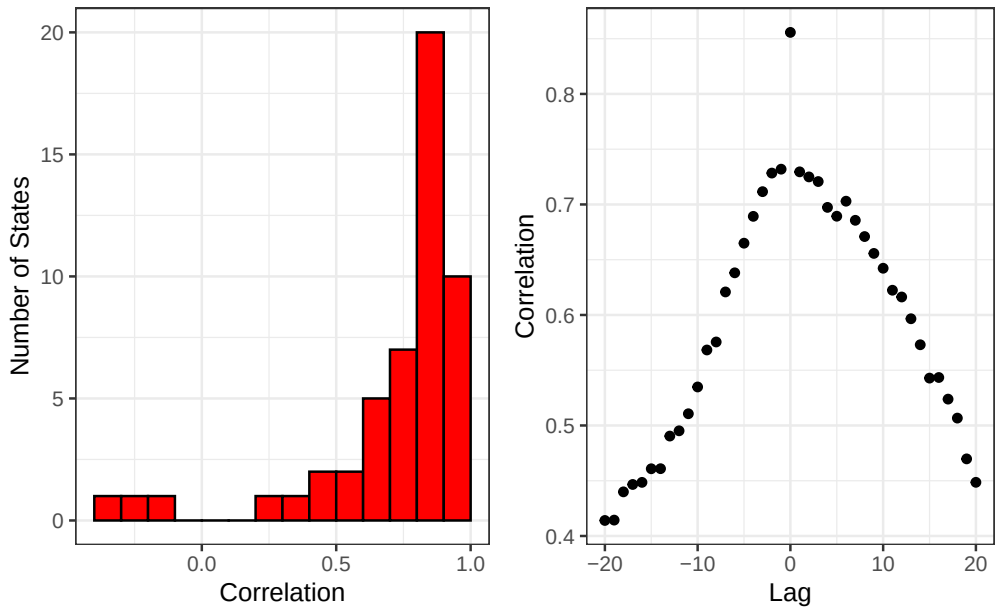


Figure 16: Mobility regression analysis.

J Comparison of model attack rates with and without interventions in the model

We show in Figure 17 that there is little difference between the attack rates for two models where we include the "Stay at Home" intervention in our model as well and when we just use mobility. We, therefore, choose the mobility only model to limit the numbers of parameters we are fitting.

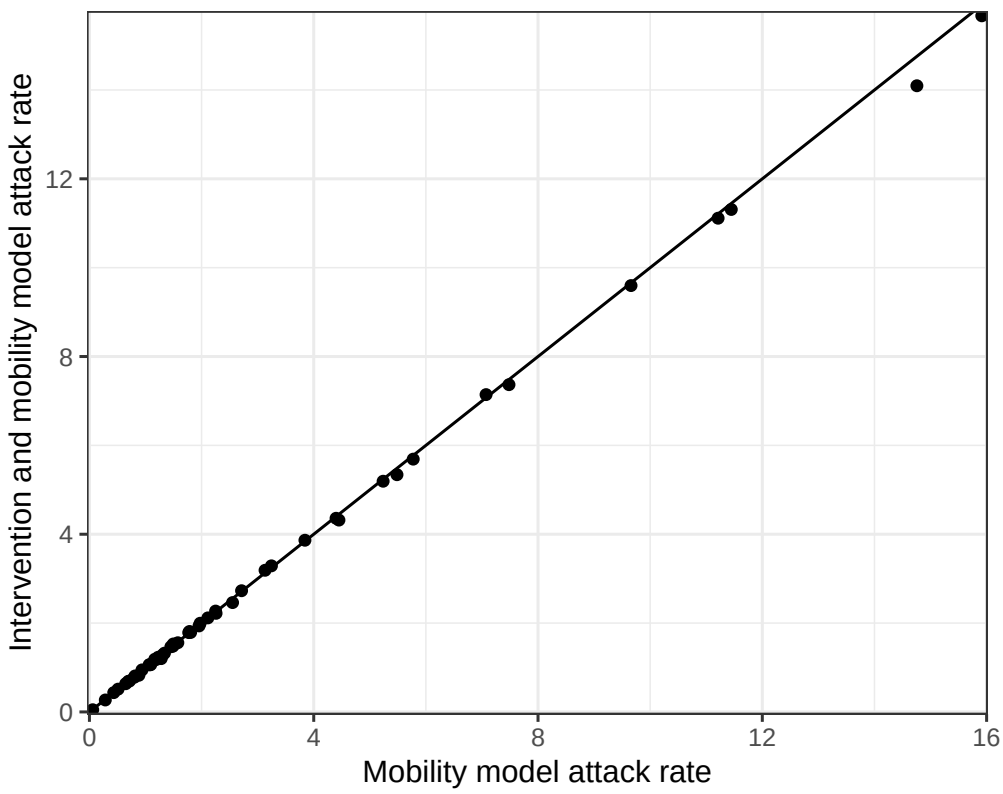


Figure 17: Mobility regression analysis.

K State-specific weekly effects

Our model includes a state-specific weekly effect $\epsilon_{w,m}$ (see equations 12, 14) for each week w in the epidemic period for a state. As described in Section 5, we assign an autoregressive process with mean 0 as prior to this effect. Figure 20 shows the posterior of this effect on the same scale as in Figure 3, that is, the percent reduction in R_t with mobility variables held constant[4]. Values above 0 have the interpretation that the state-specific weekly effect lowers the reproduction number $R_{t,m}$, i.e. transmission for week t and state m is lower than what is explained by the mobility covariates.

[4]Draws from the posterior are transformed with $1 - f(-\epsilon_{m,w_m(t)})$, where $f(x) = 2\exp(x)/(1+\exp(x))$ is twice the inverse logit function.

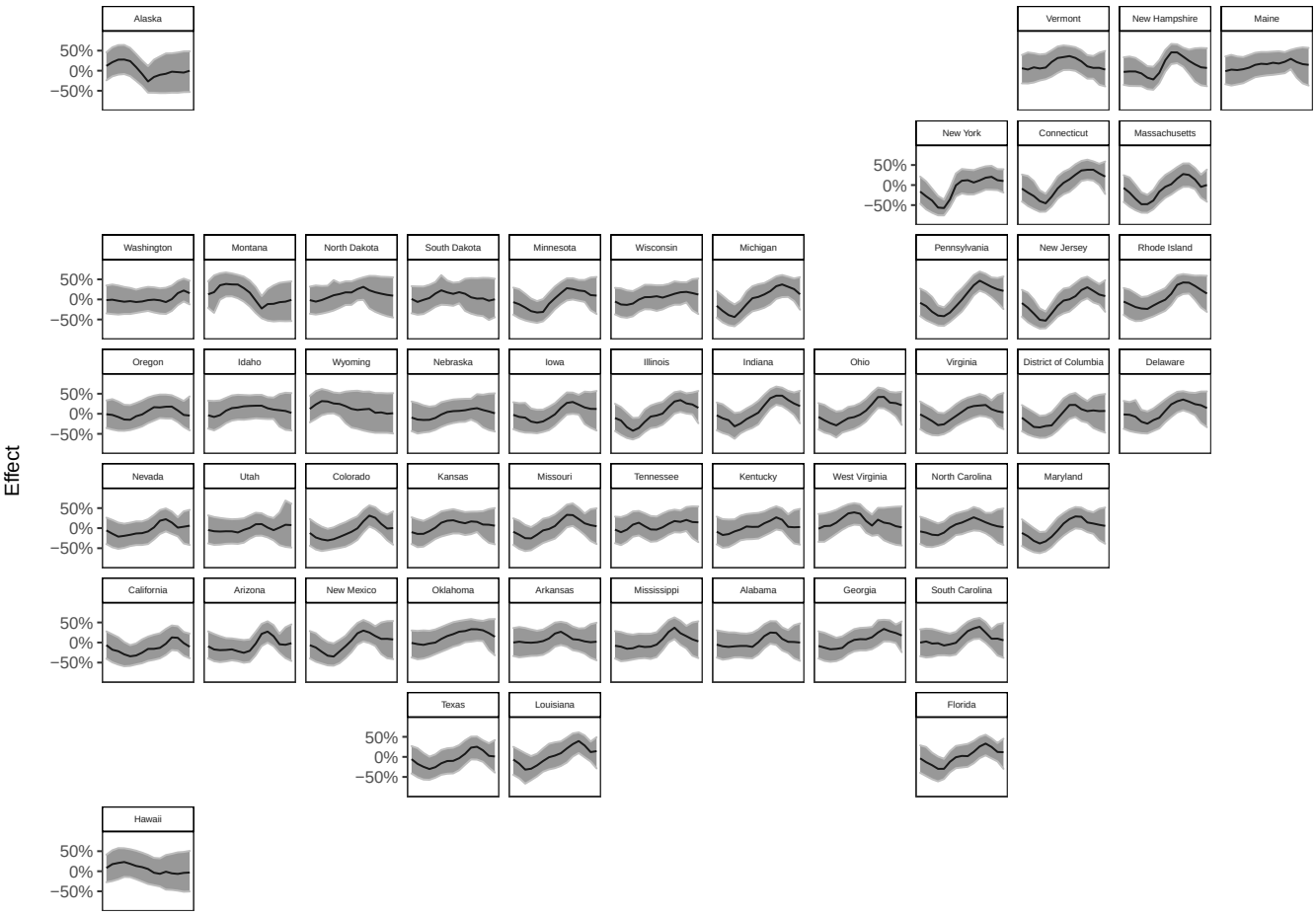


Figure 18: Percent reduction in R_t due to the weekly, state-level autoregressive effect.

In Figure 19, we compare the different terms contributing to R_t, as explained by equation (12), for three example states. We choose New York, Montana and Washington for our vignettes because the autoregressive (purple) trend behave differently in these states. In New York, the autoregressive term increases R_t before lockdown, which we hypothesise corresponds to importations and new seeding events from Europe driving transmission. Then the autoregressive term decreases as social distancing, mask wearing and handwashing are implemented along side behavioural changes. In contrast, the autoregressive term reduces R_t at the start in Montana and the mobility trends remain very flat, adding some weight to the hypothesis of importations. This could reflect the low density of that state and the correspondingly long distances between where people live, along with behavioural changes. The autoregressive term remains mostly constant in Washington, in line with the states early and effective epidemic response, and suggests that mobility is sufficient to capture the behaviour there.

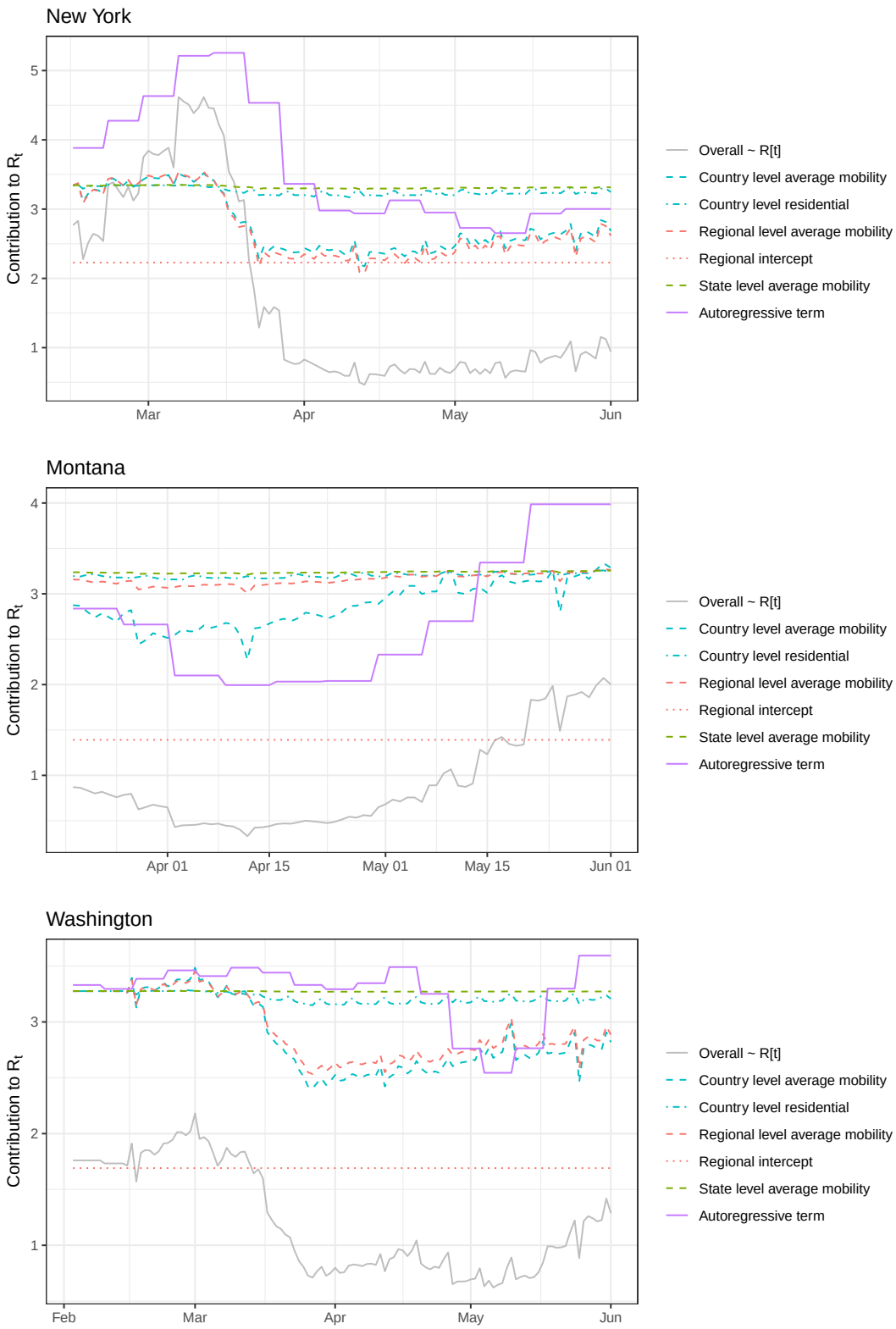


Figure 19: Comparison of different contributors to R_t for New York, Montana and Washington. Here we consider all the contributions given in equation (12) - they grey solid line shows the posterior mean estimate of R_t and the coloured lines show the contributions from the other terms. The colour denotes pooling level and the line type the mobility type where appropriate. Note the inverse logit transformation means these terms are not additive and each line, apart from the overall trend, is calculated by assuming all the other terms in equation (12) are zero.

L Comparison of R_t estimations

We compare our R_t estimates on 01 June 2020 with estimates made by rt.live [14]. Overall, our estimates were weakly correlated ($\rho = 0.42$)with both of us estimating $R_t > 1$ in 23 states (red points) including Montana and Alaska. However, the rt.live estimates are slightly more pessimistic because they predict $R_t > 1$ in 10 states where we predict $R_t < 1$ (blue points). In contrast we predict $R_t > 1$ in 5 states where they predict $R_t < 1$ (green points).

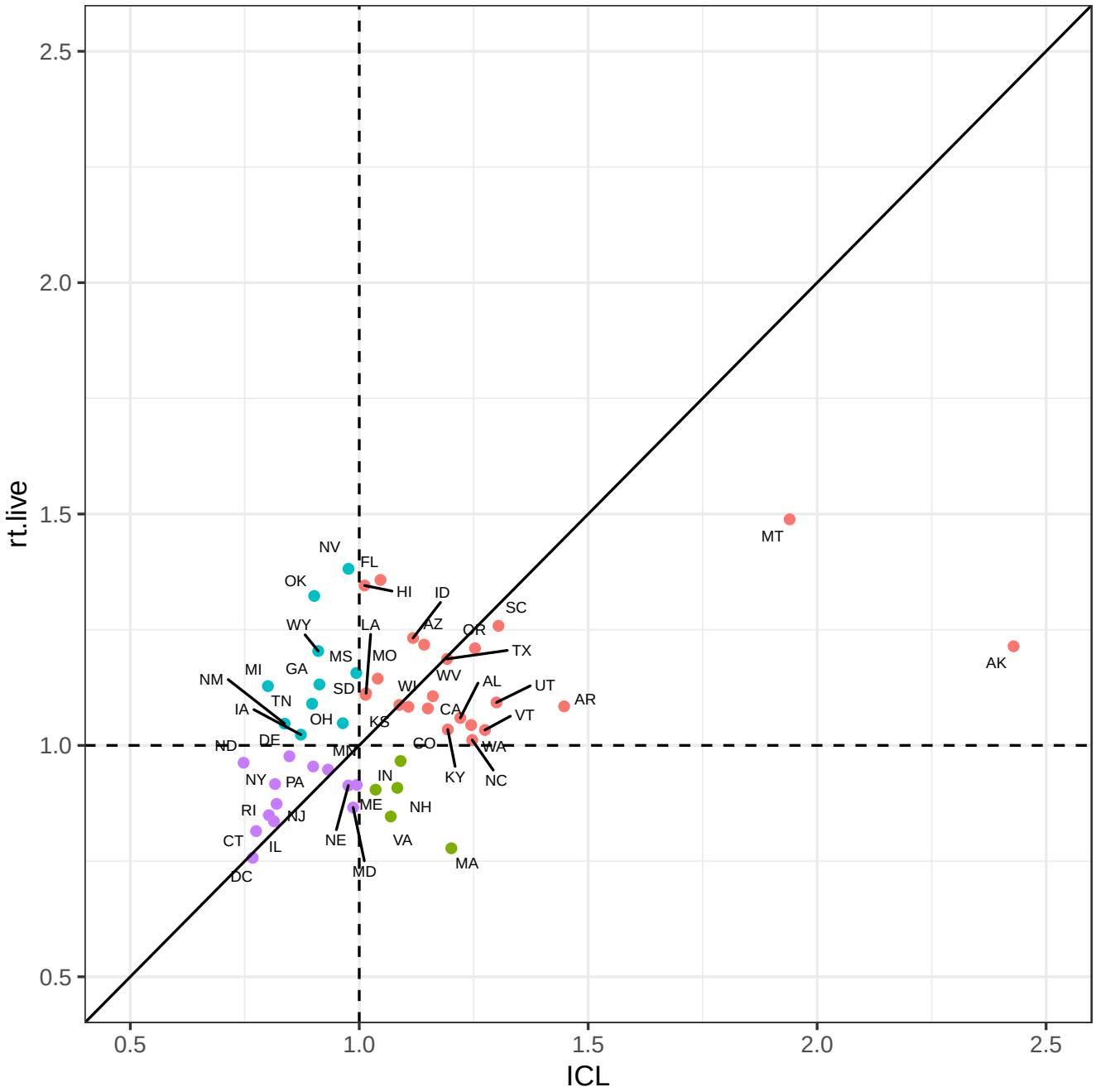


Figure 20: Comparison of our and rt.live [14] R_t estimates. Colours are used to indicate groupings if each model predicted $R_t > 1$. Standard 2 letter state codes are used as labels- Alabama AL, Alaska AK, Arizona AZ, Arkansas AR, California CA, Colorado CO, District of Columbia DC, Connecticut CT, Delaware DE, Florida FL, Georgia GA, Hawaii HI, Idaho ID, Illinois IL, Indiana IN, Iowa IA, Kansas KS, Kentucky KY, Louisiana LA, Maine ME, Maryland MD, Massachusetts MA, Michigan MI, Minnesota MN, Mississippi MS, Missouri MO, Montana MT, Nebraska NE, Nevada NV, New Hampshire NH, New Jersey NJ, New Mexico NM, New York NY, North Carolina NC, North Dakota ND, Ohio OH, Oklahoma OK, Oregon OR, Pennsylvania PA, Rhode Island RI, South Carolina SC, South Dakota SD, Tennessee TN, Texas TX, Utah UT, Vermont VT, Virginia VA, Washington WA, West Virginia WV, Wisconsin WI and Wyoming WY.

M Sensitivity analysis to infection fatality ratio

Geographic-specific contact surveys are important for calculating the weighted IFR values according to the methods in [22, 23]. There is no large-scale cross-generational contact survey, similar to the polymod survey [11], implemented in the USA. Therefore, it was important to understand if the model was robust to changes in the underlying contact survey. We calculated the IFRs using three different contact matrices: UK, France and Netherlands. We believe that the USA is culturally closest to that UK out of those countries we had contact matrices for, but also considered France where we saw the greatest mixing of the elderly and the Netherlands which showed the *average* behaviour of the European studies used in [23]. We found that the IFR, calculated for each state using the three contact matrices, lay within the posterior of IFR in our model (Figure 21). We also noted that our results remained approximately constant when using the IFR calculated from the three different contact matrices as the mean of the prior IFR in out model, see Section 5.

Since we are using the same contact matrix across all the states, the differences in IFR are due to the population demographics and not due to differential contacts. The low IFR in Texas and Utah reflects the younger population there whereas the higher IFR in Florida and Maine is due to the older population. This is a limitation of our methods.

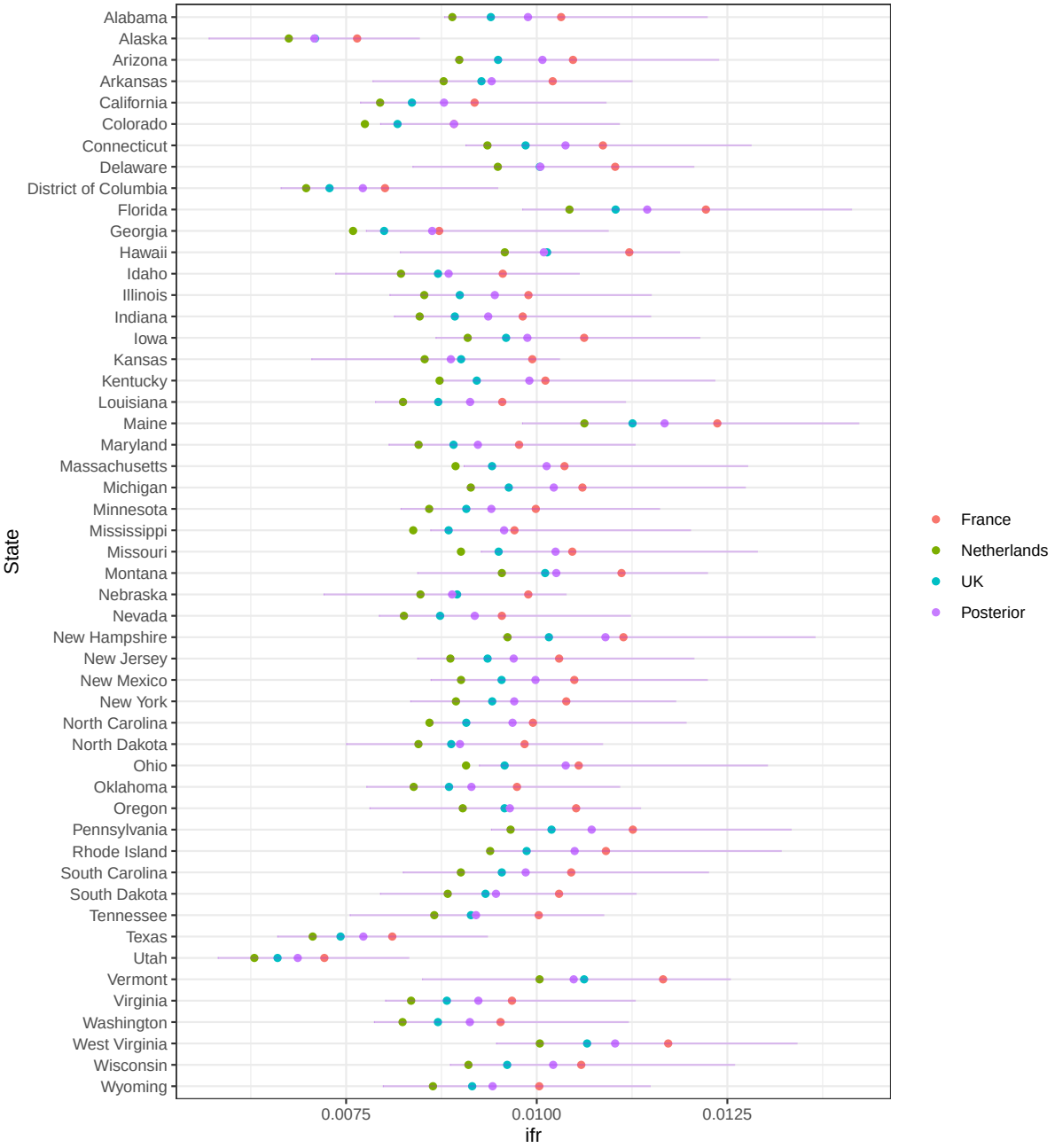


Figure 21: Sensitivity analysis for IFR. The red, green and blue dots show the IFR values calculated according to [22, 23] using the French, Dutch and and UK contact matrices respectively. The purple dot shows the mean of our posterior estimates for the IFR run using the UK contact matrix estimate and the purple error shows the 95% credible intervals of the distribution.