ARTICLES
https://doi.org/10.1038/s41562-020-00969-7

nature human behaviour

State-level needs for social distancing and contact tracing to contain COVID-19 in the United States

Weihsueh A. Chiu[1]✉, Rebecca Fischer[2] and Martial L. Ndeffo-Mbah[1,2]✉

Starting in mid-May 2020, many US states began relaxing social-distancing measures that were put in place to mitigate the spread of COVID-19. To evaluate the impact of relaxation of restrictions on COVID-19 dynamics and control, we developed a transmission dynamic model and calibrated it to US state-level COVID-19 cases and deaths. We used this model to evaluate the impact of social distancing, testing and contact tracing on the COVID-19 epidemic in each state. As of 22 July 2020, we found that only three states were on track to curtail their epidemic curve. Thirty-nine states and the District of Columbia may have to double their testing and/or tracing rates and/or rolling back reopening by 25%, while eight states require an even greater measure of combined testing, tracing and distancing. Increased testing and contact-tracing capacity is paramount for mitigating the recent large-scale increases in US cases and deaths.

The novel coronavirus pandemic (COVID-19), first detected in Wuhan, China in December 2019, has now reached pandemic status with spread to >210 countries and territories, including the United States[1]. The United States reported its first imported case of COVID-19 on 20 January 2020, arriving via an international flight from China[2]. Since then the disease has spread rapidly within the country, with every state reporting confirmed cases within 3 weeks of the first reported community transmission. As of 1 August, the United States has exceeded 4.5 million cases and 150,000 deaths, heterogeneously distributed across all states[1]. To date, states such as New York, New Jersey and California have borne the highest burden, with <420,000, 183,000 and 510,000 cases and 32,000, 15,000 and 9,000 deaths, respectively, while Alaska and Hawaii have each reported <4,000 cases and 25 deaths[1].

COVID-19 is caused by a newly described and highly transmissible SARS-like coronavirus (SARS-CoV-2). Severe clinical outcomes have been observed in approximately 20% of symptomatic cases[3,4]. There is no vaccine and no cure or approved pharmaceutical intervention for this disease, making the fight against the pandemic reliant on non-pharmaceutical interventions (NPIs). These NPIs include: (1) case-driven measures such as testing, contact tracing and isolation[5]; (2) personal preventive measures such as hand hygiene, cough etiquette, face mask use, eye protection, physical distancing and surface cleaning, which aim to reduce the risk of transmission during contact with potentially infectious individuals[6]; and (3) social-distancing measures to reduce interpersonal contact in the population. In the United States, social-distancing measures have included policies and guidelines to close schools and workplaces, cancel and restrict mass gatherings and group events, restrict travel, maintain physical separation from others (for example, keeping six feet apart) and stay-at-home orders[7].

Non-pharmaceutical interventions and other responses to COVID-19, especially stay-at-home orders, have varied widely across states, leading to spatial and temporal variation in the timing and implementation of mitigation strategies. This variation in policies and response efforts may have contributed to the observed heterogeneity in COVID-19 morbidity and mortality across states[8]. Recent studies suggest that state-wide social-distancing measures have probably contributed to reducing the spread COVID-19 epidemic in the United States[9,10]. Understanding the extent to which NPIs, such as social distancing, testing, contact tracing and self-quarantine, influence COVID-19 transmission in a local context is pivotal for predicting and better managing the future course of the epidemic on a state-by-state basis. This in turn will inform how these NPIs should be optimized to mitigate the spread and burden of COVID-19 while awaiting development of pharmaceutical interventions (for example, therapeutics and vaccines).

After several weeks of state-wide stay-at-home orders, most US states began to ease their social-distancing requirements in May/June 2020 (ref. [11]) while attempting to increase their testing and contact-tracing capacities[12]. Mathematical modelling is a unique tool to help answer these important and timely questions. Models can contribute valuable insight for public health decision-makers by providing an evaluation of the effectiveness of ongoing control strategies along with predictions of the potential impact of alternative policy scenarios[13].

To address these needs, we developed and validated a data-driven transmission dynamic model to evaluate the impact of social distancing, state reopening, testing and contact tracing on the state-level dynamics of COVID-19 infections and mortality in the United States, shown schematically in Fig. 1. Like many other COVID-19 transmission models[14–17], we used an extended susceptible, exposed, infectious, removed (SEIR) compartmental model. The model divides the population into several disease compartments and tracks movements of individuals between the compartments through different transition rates. The main model compartments include: *S*, susceptible; *E*, exposed; *A*, infectious and asymptomatic; *I*, infectious and symptomatic; *R*, recovered; and *F*, dead. In addition to disease progression stages, our model incorporates social distancing informed by several public sources of mobility data, case identification via testing, isolation of detected cases and contact tracing. This is a mean-field epidemiological modelling approach that captures the average disease dynamics behaviour within a population[18,19]. We used Bayesian inference methods

[1]Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX, USA. [2]Department of Epidemiology and Biostatistics, School of Public Health, Texas A&M University, College Station, TX, USA. ✉e-mail: wchiu@tamu.edu; m.ndeffo@tamu.edu

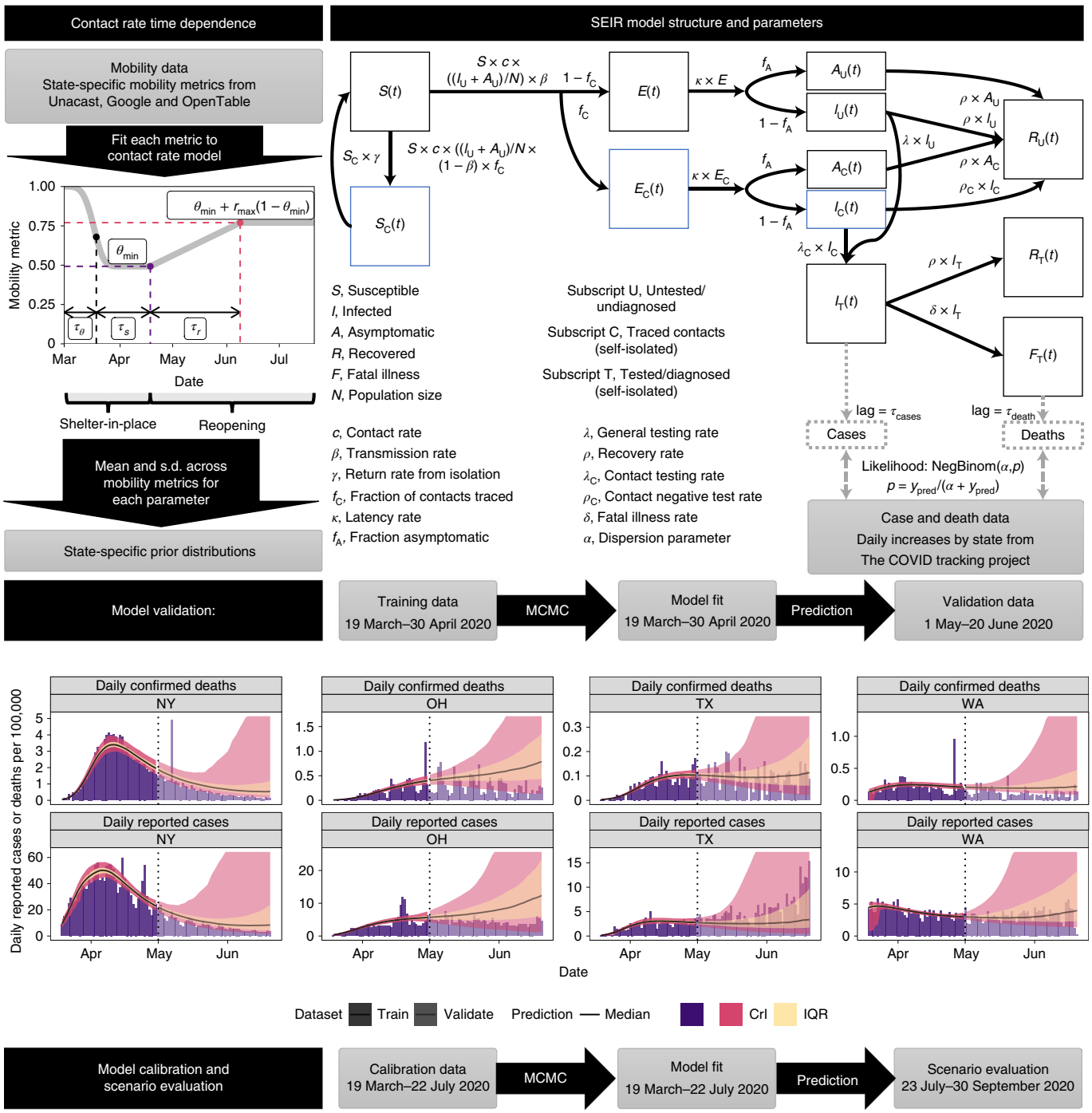


Fig. 1 | SEIR model structure, parameters, data sources and fitting/validation methods. We used mobility data to constrain the time dependence of the contact rate. We fitted the model to daily reported cases and confirmed deaths from 19 March to 30 April 2020 and validated its projections against data from 1 May to 20 June 2020. On the model projections, the black solid line is the median, the pink band is the 95% CrI and the orange band is the IQR. We show model fitting and validation for four states: New York (NY), Ohio (OH), Texas (TX) and Washington (WA).

to calibrate and validate our model prediction to state-level daily reported COVID-19 cases and fatality data. Model parameters, prior distributions and their sources are shown in Table 1. We used the calibrated model to evaluate the transmissibility of COVID-19 in each state from March 2020 to late July 2020, to estimate the state-level impact of shelter-in-place and reopening on COVID-19 transmission. Finally, we evaluated the degree to which increasing testing efforts (rate of identification of infected cases) and/or contact tracing could curtail the spread of the disease and enable greater relaxation of social-distancing restrictions while preventing a resurgence of infections and deaths. A detailed description of the model considerations, parameterization and analysis is provided in Methods.

Results

Model performance and validation. We used state-level mobility data from Unacast, Google and OpenTable to calibrate a parametric model of shelter-in-place and reopening (Supplementary Fig. 1),

Table 1 | Model inputs, parameters and prior distributions for Bayesian analysis

Symbol	Definition (units)	Calibrated parameter(s)	Prior (truncation)	Notes/refs.
N	Population size	Input (not calibrated)	Constant	40
N_{init}	Initial I_U on 29 February 2020	N_{init}	$\log_N(1{,}000, 10)$ [1, 10,000]	a
$1/\gamma$	Self-isolation time after contact tracing	$T_{isolation} = 1/\gamma$	$\log_N(14, 2)$ [7, 21]	b
$1/\kappa$	Latent period (d)	$T_{latent} = 1/\kappa$	$N(4,1)$ [2,7]	41,42
c_0	Baseline contact rate (contacts d^{-1})	c_0	$N(13, 5)$ [7, 20]	43
ρ	Recovery rate (d^{-1})	$T_{recover} = 1/\rho$	$\log_N(10, 1.5)$ [5, 30]	42,44
β_0	Transmission probability per contact (unitless)	$R_0 = c_0\beta_0/\rho$	$N(2.9, 0.78)$ [1.46, 4.5]	45–47
f_C	Fraction of contacts traced (unitless)	f_C	$\log_N(0.25, 2)$ [0.05, 1]	48
f_A	Fraction of infected asymptomatic (unitless)	f_A	$N(0.295, 0.275)$ [0.02, 0.57]	49
$T50_T$	Date of 50% of final testing rate (d)	$T50_T$	$U(60, 106)$ (1 Mar–15 Apr)	a
λ	General positive diagnosis rate (d^{-1})	$\lambda = F_{test}\,Sens_{test}\,k_{test}$	Derived	45,50,51
F_{test}	General test coverage (unitless)	F_{test}	$\beta(2,2)$	45,50,51
$Sens_{test}$	Test sensitivity (unitless)	$Sens_{test}$	$N(0.7, 0.1)$ [0.6, 0.95]	52
k_{test}	General testing rate (d^{-1})	$\tau_{test} = 1/k_{test}$	$N(7, 3)$ [2, 12]	53,54
λ_C	Contact positive diagnosis rate (d^{-1})	$\lambda_C = Sens_{test}\,k_{test,C}$	Derived	
$k_{C,test}$	Contact testing rate (d^{-1})	$\tau_{C,test} = 1/k_{C,test}$	$N(2, 1)$ [1, 3]	a
ρ_C	Rate of infected contacts testing negative (d^{-1})	$\rho_C = (1 - Sens_{test})\,k_{test,C}$	Derived	
δ	Fatal illness rate (d^{-1})	IFR[d]	$\log_N(0.01, 2)$ [0.001, 0.1]	44,55
θ_{min}	Minimum of $\theta(t)$	θ_{min}	Validation: $\beta(2,2)$ Calibration: state-specific	a c
τ_θ	Weibull scale parameter	τ_θ	Validation: $N(21, 7)$ [7, 35] Calibration: state-specific	a c
n_θ	Weibull shape parameter	n_θ	Validation: $\log_N(6, 2)$ [1,11] Calibration: state-specific	a c
η	Hygiene effectiveness relative to social distancing (unitless)	η	$\beta(2,4)$	a
τ_s	Duration of shelter-in-place (d)	τ_s	Validation: $N(45, 30)$ [21, 90] Calibration: state-specific	56
τ_r	Duration of linear increase after shelter-in-place (d)	τ_r	Validation: $N(45, 30)$ [14, 105] Calibration: state-specific	a c
r_{max}	Maximum relative increase in contacts from shelter-in-place (unitless)	r_{max}	Validation: $N(1, 1)$ [0, 2] Calibration: state-specific	a c
τ_{case}	Lag time for observing confirmed case	τ_{case}	$\log_N(7, 2)$ [1, 14]	a
τ_{death}	Lag time for observing confirmed death	τ_{death}	$\log_N(7, 2)$ [1, 14]	a
α_{pos}	Negative binomial shape parameter for case likelihood function	α_{pos}	$\log_U(0.1, 40)$	a
α_{death}	Negative binomial shape parameter for death likelihood function	α_{death}	$\log_U(0.1, 40)$	a

$\log_N$, log-normal distribution with geometric mean and geometric s.d. N, normal distribution with mean and s.d.; U, uniform distribution with minimum and maximum; $\log_U$, log-uniform distribution with minimum and maximum; $\beta(a,b)$, beta distribution with shape parameters a and b; time (t) is measured from $t = 1$, corresponding to 1 January 2020.[a]Assumed, non-informative prior sufficiently wide to have adequate validation coverage. [b]Standard contact-tracing guidance is to self-isolate for 2 weeks. [c]For calibration to 20 June 2020, state-specific priors were derived by fitting to different social-distancing datasets, with each parameter's mean, s.d. and range used to define a normal distribution prior. [d]See Methods for relationship between infected fatality rate (IFR) and δ.

and used the results to inform prior distributions for the transmission model (Fig. 1). We fit our model to state-level daily cases and deaths data using a Bayesian inference approach (Methods). Model performance assessment for several representative states is shown in Fig. 1, with full results in Supplementary Figs. 2 and 3. With respect to validation, the posterior 95% credible interval (CrI) of our model projections, estimated using data to 30 April 2020, covered 84% of the data points from 1 May to 20 June 2020. For seven states (Alaska, Montana, South Dakota, Iowa, Illinois, Michigan and Minnesota), validation had low coverage (<50%) because of insufficient training data to 30 April 2020 to adequately inform sheltering and reopening in those states. This inaccuracy was not unexpected, because the length of sheltering and the degree of reopening could not have been known on 30 April 2020, and thus our model predictions were based on generic prior distributions. However, during model calibration to data to 22 July 2020, these parameters were informed by updated state-specific mobility data. Model performance for fitting all data to 22 July 2020 is shown in Supplementary Figs. 4–6, with posterior parameter distributions shown in Supplementary Fig. 7. Good fits with high coverage (>88% for cases, >92% for deaths) were obtained for all states.

Estimations of effective reproduction number. The effective reproduction number, R_{eff}, is the average number of secondary

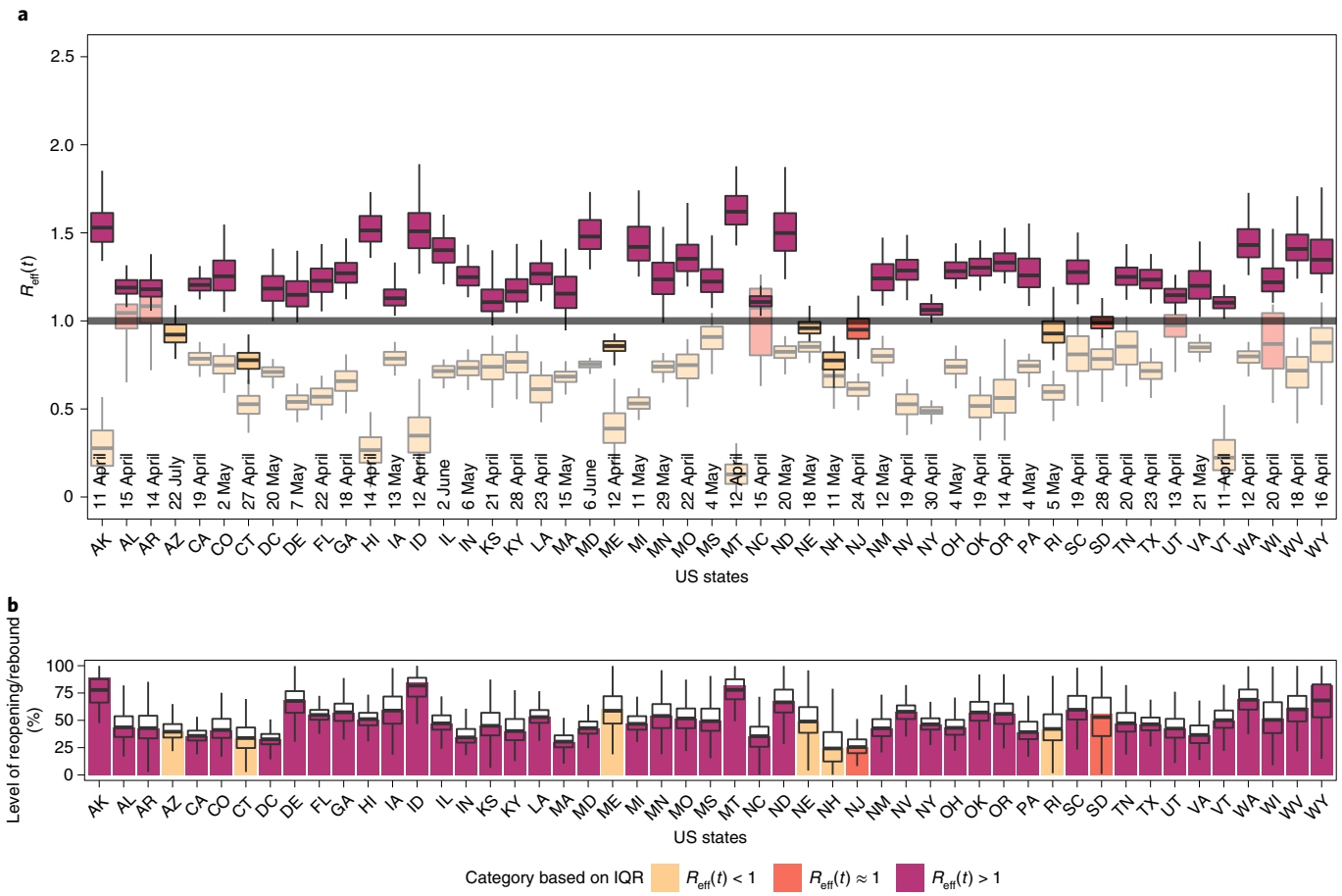


Fig. 2 | Estimated R_{eff} and the level of reopening/rebound in transmission as of 22 July 2020 for all states. a, Estimated R_{eff} (median, IQR and 95% CrI) across states. The figure shows the value of R_{eff} on 22 July 2020, as well as its minimum value between 19 March and 22 July 2020, in lighter shades of each colour. It also includes the date of minimum R_{eff}. **b**, The level of reopening/rebound in disease transmission in each state relative to its minimum value during state shelter-in-place (median, IQR and 95% CrI).

infection cases generated by a single infectious individual during their infectious period[18]. When $R_{eff} > 1$ the epidemic curve is increasing, and when $R_{eff} < 1$ the epidemic curve is decreasing[18]. Using the posterior distribution of our model parameters, we estimated R_{eff} from 19 March to late July 2020 and identified the minimum level of transmission achieved in each state (Fig. 2a). We found that for all except five states (Alabama, Arkansas, North Carolina, Utah and Wisconsin), the interquartile range (IQR) for the minimum R_{eff} value was <1 (varying 0.07–0.98), and these values were mainly achieved during the state shelter-in-place (11 April–29 May 2020) (Fig. 2a). Following states' relaxations of social-distancing measures, disease transmission again started to increase. By 22 July 2020, 42 states and the District of Columbia had at least a 75% probability that $R_{eff} > 1$. The model predicts therefore that, as states are reopening, a majority are at risk of continued increases in the scale of the outbreak and require additional mitigation to contain the spread of the disease.

We conducted an analysis of variance (ANOVA) to evaluate the contribution of each parameter to the variation in R_{eff} (Supplementary Table 1). Across states, we found that the largest drivers of variation in R_{eff} are (1) the power parameter for relating social distancing to hygiene-associated reduction in transmission, η (ANOVA F (one degree of freedom) = 2,989.166, $P < 2.2 \times 10^{-16}$, $\eta^2 > 5\%$, lower 95% confidence interval (CI) of $\eta^2 > 4.5\%$); (2) the degree of mitigation during shelter-in-place, θ_{min} (ANOVA F (one degree of freedom) = 5,177.354, $P < 2.2 \times 10^{-16}$, $\eta^2 > 8.7\%$, lower 95% CI of $\eta^2 > 8.1\%$); (3) the maximum relative increase in contact after shelter-in-place orders, r_{max} (ANOVA F (one degree of freedom) = 8,051.61, $P < 2.2 \times 10^{-16}$, $\eta^2 > 13.5\%$, lower 95% CI of $\eta^2 > 12.8\%$); and (4) the fraction of contacts traced, f_c (ANOVA F (one degree of freedom) = 13,834.053, $P < 2.2 \times 10^{-16}$, $\eta^2 > 23.2\%$, lower 95% CI of $\eta^2 > 22.4\%$), which together contribute >50% of variance (Extended Data Fig. 1 and Supplementary Table 1). This observation is consistent with mobility data alone being insufficient to account for the combined effect of multiple control measures, and suggests that the degree of adoption of non-mobility-related measures, such as enhanced hygiene practices and contact tracing, plays a large role in the extent to which a state may reduce disease transmission.

For each state, we defined Δ as the level of reopening/rebound ($\Delta = 0\%$ at minimum, 100% at full reopening) in disease transmission relative to its lowest transmission rate observed during shelter-in-place, and estimated the current level of reopening/rebound (Fig. 2b). We found that 24 states had an average of 50–80% rebound in COVID-19 transmission by 22 July 2020, while no state had <25% rebound (Fig. 2b).

Impact of testing and contact tracing on easing of social distancing. Bringing and maintaining $R_{eff} < 1$ is necessary to curtail the spread of an outbreak. We evaluated the probability of maintaining $R_{eff} < 1$ for different levels of testing and contact tracing under the level of state reopening as of 22 July 2020. We found that in 42 states

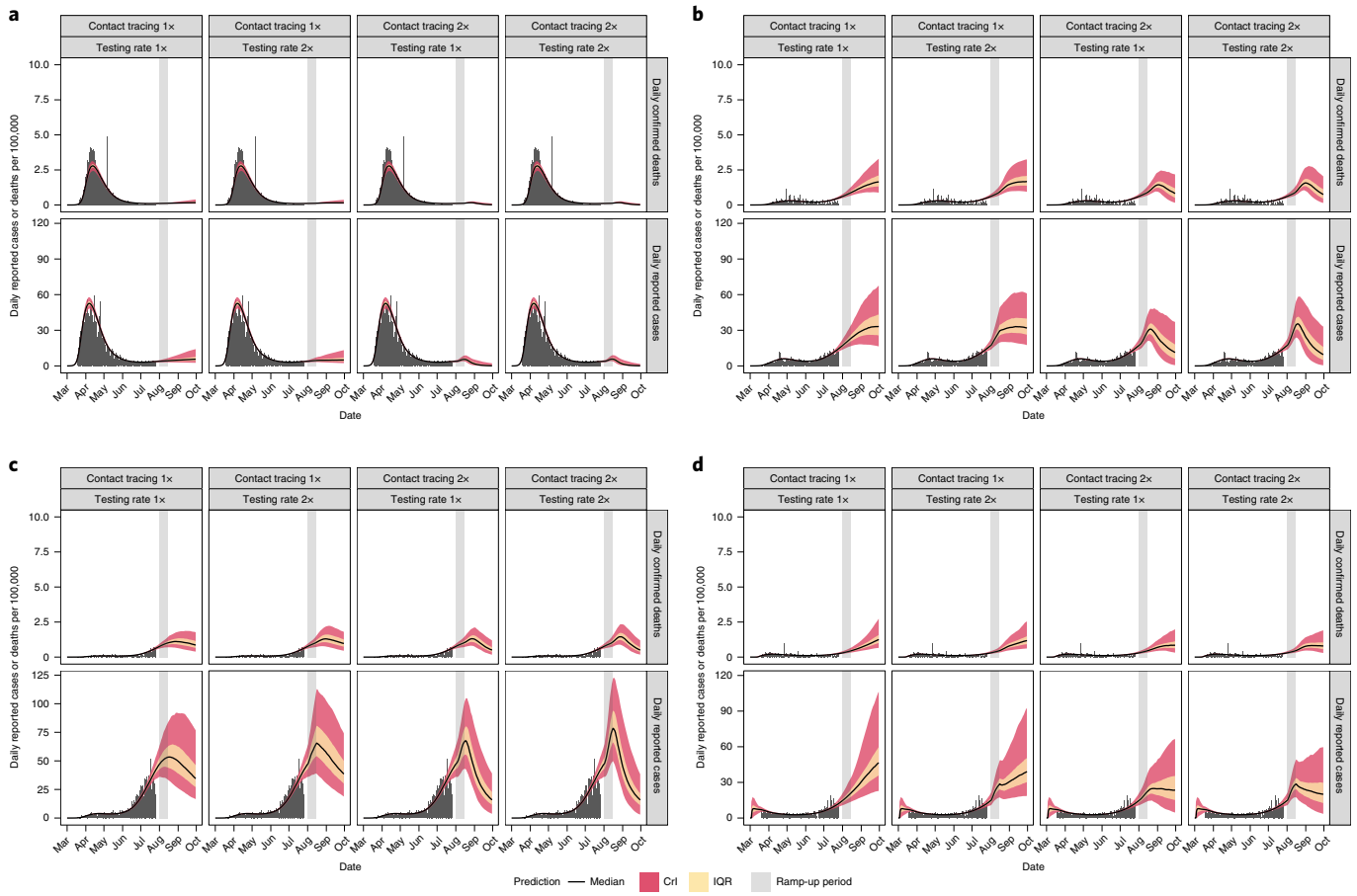


Fig. 3 | a–d, Predicted time course (median, IQR and 95% CrI) of daily reported cases and deaths under different testing and contact-tracing rates (1× and 2×) in New York (**a**), Ohio (**b**), Texas (**c**) and Washington State (**d**).

and the District of Columbia, bringing and maintaining $R_{eff} < 1$ may not be possible without increased contact-tracing efforts because increasing testing and isolation alone would require at least a 3.5-fold increase in coverage to curtail the epidemic curve with 0.975 probability (Extended Data Figs. 2 and 3 and Supplementary Table 2). The challenges are even greater in ensuring continued control of the epidemic with full reopening, because testing and isolation alone would be insufficient to curtail the epidemic in 33 states and, in all states, contact-tracing coverage of 50–75% would be required to curtail the epidemic curve with 0.975 probability (Extended Data Fig. 4 and Supplementary Table 3).

To evaluate the impact of scaling up of testing and contact tracing on epidemic dynamics in each state, we assumed a linear 'ramp-up' of testing and/or contact tracing from 1–14 August 2020, after which both parameters remain constant. We then predicted the daily number of reported cases and deaths (Fig. 3 and Supplementary Fig. 8). We found that, under current levels of reopening and control, 40 states would be unable to curtail the spread of the epidemic within the following 2 months (Supplementary Fig. 8). Even with increased testing and contact tracing, these states will still experience an increase in reported cases and deaths of between 2 weeks and 2 months (Fig. 3 and Supplementary Fig. 8). For example, Ohio, Texas and Washington may experience a 2-week increase in cases and a 1-month increase in deaths even if their current testing and contact-tracing rates were doubled within the following 2 weeks (Fig. 3b–d). Moreover, reported cases increase during the 2-week ramp-up period (Fig. 3). We found that, in 27 states and the district of Columbia, an additional 25% (50%) relaxation of restrictions without simultaneously increasing contact tracing may exacerbate disease dynamics and result, on average, in increases of 25–65% (45–150%) in cases and 22–48% (35–92%) in deaths within the following 2 months (Supplementary Fig. 8).

We next evaluated the maximal degree of rebound in transmission (that is, level of reopening) permitted while maintaining $R_{eff} < 1$ under different testing and contact-tracing scenarios (Fig. 4). We found that, under the current levels of testing and contact-tracing rate, 27 states cannot maintain $R_{eff} < 1$ (at 75% confidence) even with only 25% reopening/rebound in transmission (Fig. 4a). By doubling the current testing rate, eight states could maintain $R_{eff} < 1$ (at 75% confidence) even with a 50% level of reopening (Fig. 4b). By doubling contact tracing, nine states could remove all mobility restrictions while maintaining $R_{eff} < 1$ (at 100% confidence) (Fig. 4c). By doubling both testing rate and contact tracing, ten states could remove all mobility restrictions while maintaining $R_{eff} < 1$ (at 100% confidence) (Fig. 4d).

We categorized states by the additional amount of mitigation efforts needed to maintain $R(t) < 1$ with at least 75% confidence (Fig. 5 and Supplementary Fig. 8). We found that, under current control efforts, no states could reduce and maintain $R(t) < 1$ if their current level of reopening was relaxed by an additional 25% ('Very Low' category), and three states (Connecticut, Maine and New Hampshire) could reduce and maintain $R(t) < 1$ without additional reopening ('Low' category). Eight states could reduce and maintain $R(t) < 1$ by doubling their contact-tracing rate or by implementing additional social-distancing restrictions, a 25% reversal of the current level of reopening ('Moderate' category), while 30 states and

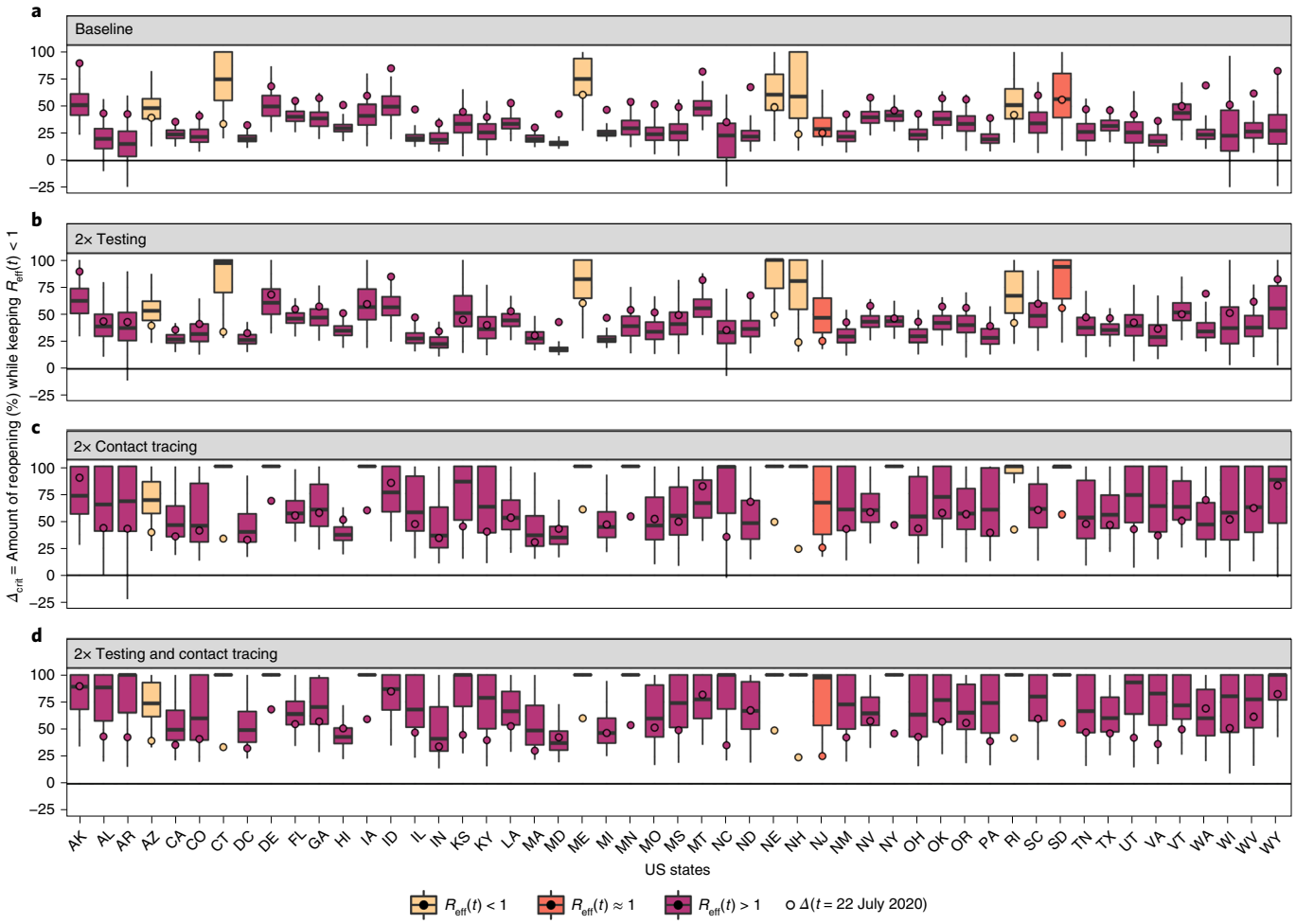


Fig. 4 | Reopening/rebound in transmission Δ_{crit} permitted (0%, minimum shelter-in-place value; 100%, return to no restrictions) to maintain $R_{eff} < 1$. a–d, Levels of reopening/rebound if testing and contact rates are unchanged (**a**), testing rate is doubled (**b**), contact tracing is doubled (**c**) or both testing and contact tracing are doubled (**d**). $\Delta(t)$, the level of reopening/rebound in transmission on 22 July 2020, is denoted by circles. All boxplots show median, IQR and 95% CrI.

the District of Columbia need a combined intervention of doubling both testing and contact tracing and/or 25% reversal of current reopening to reduce and maintain $R(t) < 1$ ('High' category). For the remaining eight states (Arizona, Florida, Idaho, Maryland, North Dakota, Nevada, South Carolina and Washington), a 50% reversal of current reopening, in addition to increased testing and/or contact tracing, are needed to reduce and maintain $R(t) < 1$ ('Very High' category).

Discussion

There is a delicate and continuous balance to strike between the use of social-distancing measures to mitigate the spread of an emerging and deadly disease such as COVID-19 and the need for reopening of various sectors of activities for the social, economic, mental and physical well-being of a community. To address this issue, it is imperative to design measurable, data-driven and flexible milestones to identify when to make specific transitions with regard to easing or re-tightening of specific social-distancing measures. We developed a data-driven SARS-CoV-2 transmission dynamic model, not only to make short-term predictions on COVID-19 incidence and mortality in the United States but, more importantly, to evaluate the impact of relaxation of social-distancing measures and increasing testing and contact tracing on the epidemic in each state.

We showed that, in most states, control strategies implemented during their shelter-in-place period were sufficient to contain the outbreak, defined as reducing and ultimately maintaining $R_{eff} < 1$. However, for the majority of states, our modelling suggests that reopening has proceeded too rapidly and/or without adequate testing and contact tracing to prevent a resurgence of the epidemic. Our model suggests that, for some states, a substantial fraction of the population may have already been infected such that, even without additional intervention, $R_{eff}(t)$ is declining towards (or below) 1 even as $R(t) > 1$. The most extreme example is Arizona, where $R_{eff}(t)$ is estimated to have declined below the previous minimum R_{eff} value achieved during shelter-in-place. However, accurate estimation of the susceptible fraction of the population is difficult due to the uncertain degree of undercounting in the reported case data. Thus, we used $R(t)$ to categorize the mitigation requirements in each state and evaluate the level of control effort needed to curtail the spread of the epidemic in each state.

Moreover, even in states with currently decreasing incidence and mortality, such as Maine and New Jersey, additional relaxation of restrictions is likely to 'bend the epidemic curve upwards' in the absence of increased testing or tracing. However, our model predicts that a combination of increased testing, increased contact tracing and/or scaling back of reopening will be sufficient to curtail the spread of COVID-19 in most states. Specifically,

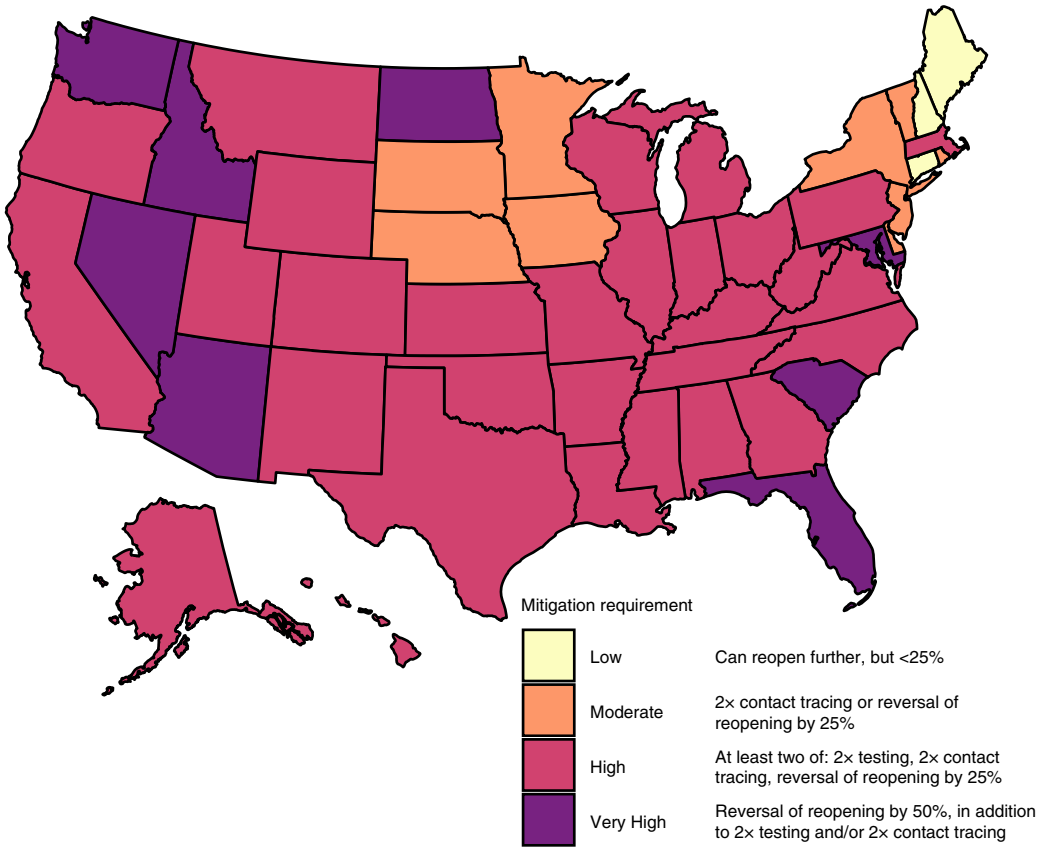


Fig. 5 | State-specific level of mitigation needed, as of 22 July 2020, to curtail the spread of COVID-19. Levels are based on maintaining $R<1$ with at least 75% confidence, equivalent to the upper bound of IQR. Categories are based on evaluation of scenarios with different combinations of baseline/doubling testing, baseline/doubling contact tracing and baseline ±25% in the reopening parameter, Δ. Categories are defined as follows: Very Low (no states): can reopen further by >25% while maintaining $R(t)<1$; Low (three states): can reopen further by <25% with up to 2× increase in testing while maintaining $R(t)<1$; Moderate (nine states): requires 2× contact tracing or reversal of reopening by 25% to maintain $R(t)<1$; High (30 states and DC): requires multiple interventions (2× testing, 2× contract tracing and reversal of reopening by 25%) to maintain $R(t)<1$; Very High (eight states): reversal of reopening by 50% combined with 2× testing and/or 2× contact tracing to maintain $R(t)<1$. Credit: the US map shapefile is derived from the usmap R package, which is open source under GPL-3.

doubling of current testing and contact-tracing rates would enable the majority of states to either maintain or increase the easing of social-distancing restrictions in a 'safe' manner in the short term. Scaling back the current level of reopening by 25%, in combination with doubling of testing and tracing, will be sufficient to control the epidemic in the long term in all but eight Very High risk states. The impact of these interventions on the epidemic curve was evaluated by computing their probability of reducing and maintaining $R<1$. However, in states with high over-dispersion in disease transmission and faced with an epidemic with high super-spreadability characteristics, the reproduction number may be subject to large fluctuation as the number of infection cases decreases. This is more likely to be the case for states with lower dispersion parameters posterior values, such as Arkansas, Connecticut, Idaho, Kansas, Kentucky, Louisiana, Mississippi, New Hampshire, South Carolina and Wyoming (Supplementary Fig. 7).

Increasing testing and contact-tracing rates entails both increasing the number of tests performed per day and requires early identification and effective isolation of COVID-19-infected individuals. This can be accomplished through active case detection via efficient contact-tracing strategies. However, it should also be noted that increased testing and contact tracing will lead to a short-term increase in reported cases because a larger fraction of the infected population is being observed, and that several weeks may pass before these rates begin to show a decline. Therefore, it is imperative that policymakers and the public recognize that such a surge is actually a sign that testing and tracing efforts are succeeding, and exercise the patience to wait several weeks before these successes are reflected as declining rates of reported cases.

Other modelling studies have used SEIR-type compartmental models to assess the impact of social distancing, testing and contact tracing to curb the epidemic curve in Italy and the United Kingdom[14–17]. Consistent with our results, these studies have shown that rapid reopening of the economy without adequate testing and contact tracing could lead to a resurgence of the epidemic[14–17]. Specifically, they show that high testing and contact-tracing rates may enable the maintenance and increase the easing of social-distancing restrictions without an increased rate of COVID-19 transmission[14].

Our study has several limitations, due to modelling assumptions and the quality of available data. Like most COVID-19 transmission models[14–17], we used a compartmental SEIR-type model to model the spread of SARS-CoV-2 because of its simplicity and ability to capture population average dynamics. This modelling approach does not account for heterogeneity in individual-level behaviour, over-dispersion due to super-spreaders, social contact networks and inherent stochasticity, which may play an important role in SARS-CoV-2 transmission dynamics. Although these factors

can be modelled through the use of individual-based models[20–22], individual-based modelling is a more complex modelling framework and may require a substantial amount of individual-level data for model parameterization, calibration and validation.

To characterize the limitations of using cell-phone-based mobility data to infer (prior distributions for) contact rates, we examined the state-to-state variation in mobility data to the corresponding posterior distributions for each mobility-related parameter (Supplementary Fig. 9). Three parameters of particular interest are the minimum relative contact rate (θ_{min}), the duration of the shelter-in-place phase (τ_s) and the maximum amount of reopening (r_{max}). For θ_{min}, none of the r^2 values were consistently <0.2 although the slope and intercept of the regression line for the Unacast Visitation metric were within 15% of 1 and 0, respectively. Similarly, for τ_s, the highest r^2 value was 0.37 for OpenTable Bookings data, which also had a relatively accurate regression line (again within 15%). For r_{max}, the highest r^2 values were for Google retail and recreation (0.49) and Unacast Visitation (0.52) metrics, but the Google data were much more accurate with a slope close to 1 and intercept close to 0. Overall, these results suggest that cell-phone-based mobility data vary substantially in their accuracy (slope and intercept near 1 and 0, respectively) and, overall, have low precision (no r^2 more than about 0.5), and support our use of the range across multiple sources in developing prior distributions, rather than using such data directly for modelling contact rates.

The initiation of social-distancing measures, such as stay-at-home orders in the United States, for mitigating the spread of COVID-19 has occurred concurrently with increased promotion and application of other NPIs, such as hygiene practices (for example, hand hygiene, surface cleaning, cough etiquette and wearing of a face mask). These hygiene practices, coupled with the avoidance of physical contact whenever possible (keeping six feet apart), could impact the spread of COVID-19 by reducing the risk of both exposure and transmission of SARS-CoV-2 from infected patients[23,24]. Though our model explicitly accounts for the differential contribution of social distancing (mobility reduction) versus hygiene practices and physical distancing to reducing COVID-19 transmission, we assume that the impact of hygiene practices and physical distancing was a function of social distancing (mobility reduction). While mobile phone mobility data may continue to be informative in regard to to contact rates, at least in aggregate, the impact of enhanced hygiene practices is more difficult to measure independently. As several states have eased their social-distancing requirements, especially their stay-at-home orders, compliance with hygiene practices would become even more important for reducing individuals' risk of getting or transmitting the pathogen. However, keeping a high population-level adherence to these measures is required to mitigate the spread of the COVID-19 epidemic[25]. As states are reopening various aspects of their economy, data on compliance with enhanced hygiene practices and physical distancing are needed to improve the estimation of these measures' population-level impact on reducing disease transmission.

Additionally, consistent with previous COVID-19 modelling studies[26–28], our model uses a simple functional form to model increases in testing rate from early March to June 2020. This testing rate was estimated through model fitting to daily reported case and mortality data. Particularly in states that have seen a substantial increase in testing capability and efforts during the month of May, our simple time-varying assumption may underestimate the current level of testing and contact tracing. However, it should be noted that increased testing capacity does not necessarily lead to an increased rate of testing if individuals are unaware, unwilling or unable to be tested[29]. Having data on contact tracing and date of symptoms onset would enable us to compute a better estimate of the current testing and contact-tracing rate in each state. Our model also assumes that all individuals who test positive to COVID-19 are effectively isolated for the remainder of their infectious period and no longer contribute to disease transmission. Though voluntary compliance to COVID-19 self-quarantine recommendations may be high across the United States, it is probably not 100%. Therefore, the assumption of effective isolation of all identified cases may cause our model to slightly overestimate the impact of increased testing rate on disease dynamics. However, we anticipate that this assumption would have only a marginal impact on the qualitative nature of our results.

Finally, our model does not explicitly account for age-stratified risk of disease transmission and mortality. This age stratification is important for the design and evaluation of social-distancing and testing strategies that are targeted towards the elderly population, which is at higher risk of COVID-19-induced hospitalization and death[30]. As reopening the economy becomes an imperative for states across the United States, age- or risk-targeted interventions may be a valuable tool to mitigate the burden of the pandemic. Future modelling studies could investigate the effectiveness of age- or risk-targeted non-pharmaceutical and potential pharmaceutical (vaccine or therapeutic) interventions for controlling the spread and burden of COVID-19.

In sum, we use a data-driven mathematical modelling approach to study the impacts of social distancing, testing and contact tracing on the transmission dynamics of SARS-CoV-2. Our findings emphasize the importance for public health authorities not only to monitor the case and mortality dynamics of SARS-CoV-2 in their state, but also to understand the impact of their existing social-distancing measures on SARS-CoV-2 transmission and to evaluate the effectiveness of their testing and contact-tracing programmes for prompt identification and isolation of new cases of COVID-19. As reported, case rates are increasing widely across US states because social-distancing restrictions have been eased to allow the resumption of greater economic activity, and we find that most states need to either substantially scale back reopening or enhance their capacity and scale of testing, case-isolation and contact-tracing programmes to mitigate large-scale increases in COVID-19 cases and deaths.

Methods

Our overall approach is as follows: (1) develop a mathematical model (an SEIR-type compartmental model)[18,19] that incorporates social-distancing data, case identification via testing, isolation of detected cases and contact tracing; (2) assess the model's predictive performance by training (calibrating) it to reported cases and mortality data from 19 March to 30 April 2020 and validating its predictions against data from 1 May to 20 June 2020; and (3) use the model, trained on data to 22 July 2020, to predict future incidence and mortality. The final stage of our approach predicts future events under a set of scenarios that include increased case detection through expansion of testing rate, contact tracing and relaxation or increase of measures to promote social distancing. All model fitting is performed in a Bayesian framework to incorporate available prior information and address multivariate uncertainty in model parameters.

Model formulation. We modified the standard SEIR model to address testing and contact tracing, as well as asymptomatic individuals. A fraction f_A of those exposed (E) to enter the asymptomatic A class (divided into A_U for untested and A_C for contact traced) instead of the infected I class, which in our model formulation also includes infectious presymptomatic individuals. With respect to testing, separate compartments were added for untested, 'freely roaming' infected individuals (I_U), tested/isolated cases (I_T) and fatalities (F_T). Following recovery, untested infected individuals (I_U) and all asymptomatic individuals move to the untested recovered compartment, I_U, and tested infected individuals move to the tested recovered compartment, I_T. In balancing considerations of model fidelity and parameter identifiability, we made the reasonably conservative assumptions that all tested cases are effectively isolated (through self-quarantine or hospitalization) and thus unavailable for transmission, and that all COVID-related deaths are identified/tested.

With respect to contact tracing, the additional compartment S_C represents unexposed contacts who undergo a period of isolation during which they are not susceptible before returning to S, while E_C, A_C and I_C represent contacts who were exposed. Again, the reasonably conservative assumption was made that all exposed

contacts undergo testing, with an accelerated testing rate compared to the general population. We assume a closed population of constant size, N, for each state.

The ordinary differential equations governing our model are as follows:

$$\begin{aligned}
\frac{dS}{dt} &= -S\times c\times[\beta+(1-\beta)\times f_C]\times(I_U+A_U)/N + S_C\times\gamma\\
\frac{dS_C}{dt} &= -S_C\times\gamma + S\times c\times(1-\beta)\times f_C\times(I_U+A_U)/N\\
\frac{dE}{dt} &= -E\times\kappa + S\times c\times\beta\times(1-f_C)\times(I_U+A_U)/N\\
\frac{dE_C}{dt} &= -E_C\times\kappa + S\times c\times\beta\times f_C\times(I_U+A_U)/N\\
\frac{dI_U}{dt} &= -I_U\times(\lambda+\rho) + E\times\kappa\times(1-f_A)\\
\frac{dA_U}{dt} &= -A_U\times\rho + E\times\kappa\times f_A\\
\frac{dI_C}{dt} &= -I_C\times(\lambda_C+\rho_C) + E_C\times\kappa\times(1-f_A)\\
\frac{dA_C}{dt} &= -A_C\times\rho_C + E_C\times\kappa\times f_A\\
\frac{dR_U}{dt} &= (I_U+A_U+A_C)\times\rho + I_C\times\rho_C\\
\frac{dI_T}{dt} &= -I_T\times(\rho+\delta) + I_U\times\lambda + I_C\times\lambda_C\\
\frac{dR_T}{dt} &= I_T\times\rho\\
\frac{dF_T}{dt} &= I_T\times\delta
\end{aligned}$$

where c is the contact rate between individuals, β is the transmission probability per infected contact, f_C is the fraction of contacts identified through contact tracing, $1/\gamma$ is the duration of self-isolation after contact tracing, $1/\kappa$ is the latent period, f_A is the fraction of exposed who are asymptomatic, λ is the testing rate, δ is the fatality rate, ρ is the recovery rate and λ_C and ρ_C are the testing and recovery rates, respectively, of contact-traced individuals. The testing rates λ and λ_C, the fatality rate δ and the recovery rate of traced contacts ρ_C are each composites of several underlying parameters. The testing rate defined as

$$\lambda(t) = F_{\text{test},0}\times\left[1-\frac{1}{1+e^{(t-T50_T)/\tau_T}}\right]\times \text{Sens}_{\text{test}}\times k_{\text{test}},$$

where $F_{\text{test},0}$ is the current testing coverage (fraction of infected individuals tested), $\text{Sens}_{\text{test}}$ is the test sensitivity (true positive rate) and k_{test} is the rate of testing for those tested, with a typical time-to-test equal to $1/k_{\text{test}}$. The time-dependence term models the ramping up of testing using a logistic function with a growth rate of $1/\tau_T$ d^{-1}, where $T50_T$ is the time where 50% of the current testing rate is achieved. Similarly, for testing of traced contacts, the same definition is used with the assumption that all identified contacts are tested, $F_{\text{test},0}=1$ and at a faster assumed testing rate, $k_{C,\text{test}}$:

$$\lambda_C(t) = \left[1-\frac{1}{1+e^{(t-T50_T)/\tau_T}}\right]\times \text{Sens}_{\text{test}}\times k_{C,\text{test}},$$

Because all contacts are assumed to be tested, the rate ρ_C at which they enter the 'recovered' compartment, R_U is simply the rate of false negative test results:

$$\rho_C(t) = \left[1-\frac{1}{1+e^{(t-T50_T)/\tau_T}}\right]\times(1-\text{Sens}_{\text{test}})\times k_{\text{test}}$$

The fatality rate is adjusted to maintain consistency with the assumption that all COVID-19 deaths are identified, assuming constant IFR. Specifically, we first calculated the fraction of infected that is tested and positive:

$$f_{\text{pos}}(t) = f_C\frac{\lambda_C(t)}{\lambda_C(t)+\rho_C(t)} + (1-f_C)\frac{\lambda(t)}{\lambda(t)+\rho}.$$

Then the case fatality rate $\text{CFR}(t)=\text{IFR}/f_{\text{pos}}(t)$. Because $\text{CFR}=\delta/(\delta+\rho)$, this implies

$$\delta(t) = \rho\frac{\text{CFR}(t)}{1-\text{CFR}(t)} = \rho\frac{\text{IFR}}{f_{\text{pos}}(t)-\text{IFR}}.$$

The model is 'seeded' N_{initial} cases on 29 February 2020. Because in the early stages of the outbreak there may be multiple 'imported' cases, we fit to data only from 19 March 2020 onwards, 1 week after the US travel ban was put in place[31].

Our model is fit to daily case y_c and death y_d data (cumulative data are not used for fitting because of autocorrelation). To adequately fit the case and mortality data, we accounted for two lag times. First, a lag is assumed between leaving the I_U compartment and public reporting of a positive test result, accounting for the time it takes to seek a test, obtain testing and have the result reported. No lag is assumed for tests from contact tracing. Second, a lag time is assumed between entering the fatally ill compartment F_T and publicly reported deaths. Additionally, we use a negative binomial likelihood to account for the substantial day-to-day over-dispersion in reporting results. The corresponding equations are as follows:

$$\begin{aligned}
y_{\text{obs},[c,d]}(t) &\approx \text{NegBin}[\alpha_{[c,d]}, p_{[c,d]}(t)]\\
p_{[c,d]}(t) &= \frac{y_{\text{pred},[c,d]}(t)}{\alpha_{[c,d]}+y_{\text{pred},[c,d]}(t)}\\
y_{\text{pred},c}(t) &= I_U(t-\tau_{\text{case}})\times\lambda(t) + I_C(t)\times\lambda_C(t)\\
y_{\text{pred},d}(t) &= I_T(t-\tau_{\text{death}})\times\delta(t)
\end{aligned}$$

In this parameterization, because the dispersion parameter $\alpha\to\infty$, the likelihood becomes a Poisson distribution with expected value $y_{\text{pred},[c,d]}$, whereas for small values of α there is substantial interindividual variability. Case and death data were sourced from The COVID Tracking Project[32].

Finally, we derived the time-dependent reproduction number, $R(t)$ and the effective reproduction number, $R_{\text{eff}}(t)$ of this model, given by

$$R(t) = c\times\beta\times(1-f_C)\left(\frac{1-f_A}{\lambda+\rho}+\frac{f_A}{\rho}\right)$$

and

$$R_{\text{eff}}(t) = R(t)\times\frac{S(t)}{N}$$

$R_{\text{eff}}(t)$ is the average number of secondary infection cases generated by a single infectious individual during their infectious period in partially susceptible population at time t. It is equal to the product of the transmission risk per contact of an infectious individual with their untraced contacts, $c\times\beta\times(1-f_C)$, times their average duration of infection, $\left(\frac{1-f_A}{\lambda+\rho}+\frac{f_A}{\rho}\right)$, and the portion of contacts that are susceptible, $\frac{S(t)}{N}$. This accounts for the relative contribution of asymptomatic, $c\times\beta\times(1-f_C)\left(\frac{f_A}{\rho}\right)\times\frac{S(t)}{N}$ and symptomatic infection, $c\times\beta\times(1-f_C)\left(\frac{1-f_A}{\lambda+\rho}\right)\times\frac{S(t)}{N}$. Using posterior samples for all 50 states and the District of Columbia, we conducted an analysis of variance using a linear model to characterize the contributions to the combined interstate and intrastate variation in R_{eff}. Specifically, we used a linear model for R_{eff} with the model parameters R_0, η, θ_{min}, r_{max}, f_C, f_A, λ and ρ as predictors, and evaluated the percentage of variance in R_{eff} contributed by each parameter.

Incorporating social distancing, enhanced hygiene practices and reopening. The impact of social distancing, hygiene practices and reopening was modelled through a time dependence in the contact rate, c and the transmission probability per infected contact, β:

$$\begin{aligned}
c(t) &= c_0\times[\theta(t)+(1-\theta_{\text{min}})\times r(t)]\\
\beta(t) &= \beta_0\times\theta(t)^{\eta}
\end{aligned}$$

The $\theta(t)$ function parameterized social distancing during the progression to shelter-in-place, and is modelled as a Weibull function:

$$\theta(t) = \theta_{\text{min}} + (1-\theta_{\text{min}})e^{-(t/\tau_\theta)^{n_\theta}},$$

which starts as unity and decreases to θ_{min}, with τ_θ being the Weibull scale parameter and n_θ the Weibull shape parameter (Fig. 1).

The $r(t)$ function parameterized relative increase in contacts due to reopening after shelter-in-place, with $r=1$ corresponding to a return to baseline $c=c_0$.

$$\begin{aligned}
r(t) &= r_{\text{max}}\frac{t-\tau_\theta-\tau_s}{\tau_r}[u(t-t_r)-u(t-t_{\text{rmax}})] + u(t-t_{\text{rmax}})\\
u(t) &= \text{Heaviside}(t) \approx 1-\frac{1}{1+e^{4t}}\\
t_r &= \tau_\theta+\tau_s\\
t_{\text{rmax}} &= \tau_\theta+\tau_s+\tau_r
\end{aligned}$$

The term $r(t)$ is 0 before t_r, linear between t_r and t_{rmax} and constant at a value of r_{max} after that, and made continuous by approximating the Heaviside function by a logistic function. The reopening time is defined as τ_s days after τ_θ, and the maximum relative increase in contacts r_{max} happens τ_r days after that.

We selected the functional form above for $c(t)$ because it was found to be able to represent a wide variety of social-distancing data, including mobile phone mobility data from Unacast[33] and Google[34] as well as restaurant booking data from OpenTable[35]. We used these different mobility sources to derive state-specific prior distributions because different social-distancing datasets had different values for θ_{min}, τ_θ, n_θ, τ_s, r_{max} and τ_r (Supplementary Fig. 1).

With respect to the reduction in transmission probability β, we assumed that during the shelter-in-place phase, hygiene-based mitigation paralleled this decline with an effectiveness power η, and that this mitigation continued through reopening.

Finally, we define an overall reopening parameter Δ that measures the rebound in disease transmission, $c\times\beta$ relative to its minimum, defined to be 0 during shelter-in-place (that is, $R(t)$ is at a minimum) and 1 when all restrictions are removed (when $R(t)=R_0$), which can be derived as:

$$\Delta(t) = \frac{c\times\beta/(c_0\times\beta_0)-\theta_{\text{min}}^{1+\eta}}{1-\theta_{\text{min}}^{1+\eta}}.$$

Our model is illustrated in Fig. 1, with parameters and prior distributions listed in Table 1.

Scenario evaluation. We used the model to make several inferences about the current and future course of the pandemic in each state. First, we consider the effective reproduction number. Two time points of particular interest are the time of minimum R_{eff}, reflecting the degree to which shelter-in-place and other interventions were effective in reducing transmission, and the final time of the simulation, 22 July 2020, reflecting the extent to which reopening has increased R_{eff}.

Additional parameters of interest are the current levels of reopening $\Delta(t)$, testing λ and contact tracing f_C.

We then conducted scenario-based prospective predictions using our model's parameters as estimated to 22 July 2020. We then asked the following questions:

(1) Assuming current levels of reopening, what increases in general testing λ and/or contact tracing f_C would be necessary to bring $R_{eff}<1$?
(2) What level of reopening Δ can maintain $R_{eff}<1$ under four different scenarios: current values of testing and contact tracing, doubling testing, double tracing and doubling both testing and tracing?
(3) What will be the rates of new cases and deaths under different scenarios? Specifically, we evaluate the impact of increases in testing and contact tracing under current levels of reopening, as well as increases or decreases of 25 or 50%.

For (1), we evaluated the posterior probability that $R_{eff}<1$ under scaling transformations $\lambda \rightarrow \lambda \times \mu_\lambda$ and $f_C \rightarrow f_C \times \mu_C$ with scaling factors μ_λ and μ_C:

$$R_{\text{eff}}(t) = S(t) \times c \times \beta \times (1 - \mu_C \times f_C)\left(\frac{1 - f_A}{\mu_\lambda \cdot \lambda + \rho} + \frac{f_A}{\rho}\right)$$

We additionally derived 'critical' values of μ_C and μ_λ where $R_{eff}(t)<1$ under the conditions of increased testing alone ($\mu_C=1$), increased contact tracing alone ($\mu_\lambda=1$) and equal increases in testing and tracing ($\mu_C=\mu_\lambda$). We also performed the same analysis under a full reopening scenario (that is, setting $S(t)=1$, $c=c_0$ and $\beta=\beta_0$).

For (2), we rearranged the equation for R_{eff} in terms of the reopening parameter Δ:

$$R_{\text{eff}}(t) = S(t) \times c_0 \times \beta_0 \times (1 - \mu_C \times f_C)\left(\frac{1 - f_A}{\mu_\lambda \times \lambda + \rho} + \frac{f_A}{\rho}\right)\left[\Delta \times (1 - \theta_{\min}^{1+\eta}) + \theta_{\min}^{1+\eta}\right]$$

We then fixed the scaling factors at 1 or 2, and solved the above equation to determine the percentage of reopening (Δ_{crit}) that can be achieved while keeping $R_{eff}<1$. Values of $\Delta_{crit} \geq \Delta(t)$ indicate the additional degree of reopening possible while maintaining $R_{eff}<1$, while values of $\Delta_{crit}<\Delta(t)$ indicate that reduction of reopening is needed. To convert back to testing and contact-tracing rates, we multiplied the scaling factors μ_C and μ_λ by the original values of f_C and λ, respectively.

Finally, for (3), we additionally evaluated changes in reopening $\Delta \rightarrow \Delta + \Delta_\Delta$ for Δ_Δ values of +25% (+50%) or −25% (−50%), for a total of 20 scenarios (four different levels of testing and tracing and five different levels of reopening). We then ran the SEIR model forward in time to 30 September 2020. For all three intervention parameters, μ_C, μ_λ and Δ_Δ, we assumed a ramp-up period of 2 weeks from 1 to 14 August 2020.

To summarize the relative need for mitigation in each state, we categorized states based on which scenarios resulted in the IQR of $R(t)<1$ on 15 August 2020. The categories were defined as follows:

- Very Low: can reopen further by >25% while maintaining $R(t)<1$
- Low: can reopen further by <25% with up to 2× increase in testing while maintaining $R(t)<1$
- Moderate: requires 2× contact tracing or reversal of reopening by 25% to bring and maintain $R(t)<1$
- High: requires multiple interventions (2× testing, 2× contract tracing and reversal of reopening by 25%) to bring and maintain $R(t)<1$
- Very High: combining 2× testing, 2× contact tracing and reversal of reopening by 50% is needed to bring and maintain $R(t)<1$

We use $R(t)$ instead of $R_{eff}(t)$, to minimize the impact of heterogeneity and uncertainty in the value of $S(t)/N$ on our results. Thus, requiring $R(t)<1$ provides greater assurance of state-wide control of the epidemic.

Software and code. Posterior distributions were sampled with Markov chain Monte Carlo (MCMC) simulation performed using MCSim v.6.1.0 in Metropolis within Gibbs sampling[36]. For each US state, four chains of 200,000 iterations each were run, with the first 20% of runs discarded and 500 posterior samples saved for analysis. For each parameter, comparison of interchain and intrachain variability was assessed to determine convergence, with the potential scale reduction factor $R \leq 1.2$ considered converged[37]. Additional analysis of model outputs was performed in RStudio v.1.2.1335 (ref. [38]) with R v.3.6.1 (ref. [39]).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The following publicly available datasets are used: mobility data from Unacast were sourced from https://covid19-scoreboard-api.unacastapis.com/api/search/covidstateaggregates_v3; mobility data from Google were sourced from https://www.gstatic.com/covid19/mobility/Global_Mobility_Report.csv; restaurant booking data were sourced from OpenTable, https://www.opentable.com/state-of-industry; case and death data were sourced from The COVID Tracking Project, https://covidtracking.com/.

Mobility data are shown in Supplementary Fig. 1. Case and death data are shown in Figs. 1 and 3 and Supplementary Figs. 3–6 and 8. All data used are also available in the software and code repository.

Code availability

The codes used to generate our results will be available on GitHub prior to publication, at https://github.com/wachiuphd/COVID-19-Bayesian-SEIR-US.

Received: 22 July 2020; Accepted: 17 September 2020;
Published online: 6 October 2020

References

1. *Johns Hopkins Coronavirus Resource Center* (Johns Hopkins University of Medicine, accessed 18 May 2020); https://coronavirus.jhu.edu/
2. Holshue, M. L. et al. First case of 2019 novel coronavirus in the united states. *N. Engl. J. Med.* **382**, 929–936 (2020).
3. Weiss, P. & Murdoch, D. R. Clinical course and mortality risk of severe COVID-19. *Lancet* **395**, 1014–1015 (2020).
4. Wu, Z. & McGoogan, J. M. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* **323**, 1239–1242 (2020).
5. Kucharski, A. J. et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *Lancet Infect. Dis.* **20**, 1151–1160 (2020).
6. Chu, D. K. et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *Lancet* **395**, 1973–1987 (2020).
7. Pearce, K. What is social distancing and how can it slow the spread of COVID-19? *John Hopkins University Hub* https://hub.jhu.edu/2020/03/13/what-is-social-distancing/ (2020).
8. Bialek, S. et al. Geographic differences in covid-19 cases, deaths, and Incidence—United States, February 12–April 7, 2020. *Mmwr. Morb. Mortal. Wkly Rep.* **69**, 465–471 (2020).
9. Siedner, M. J. et al. Social distancing to slow the U.S. COVID-19 epidemic: an interrupted time-series analysis. *PLoS Med.* **17**, e1003244 (2020).
10. Wagner, A. B. et al. Social distancing has merely stabilized COVID-19 in the US. *Stat* https://doi.org/10.1002/sta4.302 (2020).
11. Tolbert, J., Kates, J. & Levitt, L. Lifting social distancing measures in America: state actions & metrics. *KFF.org* https://www.kff.org/coronavirus-policy-watch/lifting-social-distancing-measures-in-america-state-actions-metrics/ (2020).
12. *CDC COVID Data Tracker* (Centers for Disease Control and Prevention, accessed 18 May 2020); https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/testing-in-us.html
13. Garnett, G. P., Cousens, S., Hallett, T. B., Steketee, R. & Walker, N. Mathematical models in the evaluation of health programmes. *Lancet* **378**, 515–525 (2011).
14. Giordano, G. et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat. Med.* **26**, 855–860 (2020).
15. Russo, L. et al. Tracing DAY-ZERO and forecasting the fade out of the COVID-19 outbreak in Lombardy, Italy: a compartmental modelling and numerical optimization approach. Preprint at *medRxiv* https://doi.org/10.1101/2020.03.17.20037689 (2020).
16. Gevertz, J., Greene, J., Tapia, C. H. S. & Sontag, E. D. A novel COVID-19 epidemiological model with explicit susceptible and asymptomatic isolation compartments reveals unexpected consequences of timing social distancing. Preprint at *medRxiv* https://doi.org/10.1101/2020.05.11.20098335 (2020).
17. Colbourn, T. et al. Modelling the health and economic impacts of population-wide testing, contact tracing and isolation (PTTI) strategies for COVID-19 in the UK. Preprint at *SSRN* https://doi.org/10.2139/ssrn.3627273 (2020).
18. Anderson, R. M., May, R. M. & Anderson, B. *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ. Press, 1992).
19. Murray, J. D. *Mathematical Biology I. An Introduction* (Springer, 2002).
20. Moghadas, S. M. et al. The implications of silent transmission for the control of COVID-19 outbreaks. *Proc. Natl Acad. Sci.* USA **117**, 17513–17515 (2020).
21. Chinazzi, M. et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
22. Panovska-Griffiths, J. et al. Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and the risk of occurrence of a second COVID-19 epidemic wave in the UK: a modelling study. *Lancet Child Adolesc. Health* https://doi.org/10.1016/S2352-4642(20)30250-9 (2020).
23. Leung, N. H. L. et al. Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat. Med.* **26**, 676–680 (2020).
24. Greenhalgh, T., Schmid, M. B., Czypionka, T., Bassler, D. & Gruer, L. Face masks for the public during the covid-19 crisis. *BMJ* **369**, m1435 (2020).

25. Eikenberry, S. E. et al. To mask or not to mask: modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infect. Dis. Model.* **5**, 293–308 (2020).
26. Pitzer, V. E. et al. The impact of changes in diagnostic testing practices on estimates of COVID-19 transmission in the United States. Preprint at *medRxiv* https://doi.org/10.1101/2020.04.20.20073338 (2020).
27. Yamana, T., Pei, S., Kandula, S. & Shaman, J. Projection of COVID-19 cases and deaths in the US as individual states re-open May 4, 2020. Preprint at *medRxiv* https://doi.org/10.1101/2020.05.04.20090670 (2020).
28. Peirlinck, M., Linka, K., Sahli Costabal, F. & Kuhl, E. Outbreak dynamics of COVID-19 in China and the United States. *Biomech. Model. Mechanobiol.* https://doi.org/10.1007/s10237-020-01332-5 (2020).
29. Thompson, S., Eilperin, J. & Dennis, B. U.S. states see COVID-19 testing supply improvements, but challenges abound. *The Washington Post* (18 May 2020); https://www.washingtonpost.com/health/as-coronavirus-testing-expands-a-new-problem-arises-not-enough-people-to-test/2020/05/17/3f3297de-8bcd-11ea-8ac1-bfb250876b7a_story.html
30. Keeling, M. J. et al. Predictions of COVID-19 dynamics in the UK: short-term forecasting and analysis of potential exit strategies. Preprint at *medRxiv* https://doi.org/10.1101/2020.05.10.20083683 (2020).
31. *Presidential Proclamation—Travel From Europe* (U.S. Department of State, 8 May 2020); https://travel.state.gov/content/travel/en/traveladvisories/presidential-proclamation-travel-from-europe.html
32. *The COVID Tracking Project* (The Atlantic, accessed 18 May 2020); https://covidtracking.com/
33. *Data for Good* (Unacast, accessed 18 May 2020); https://www.unacast.com/data-for-good
34. *COVID-19 Community Mobility Reports* (Google LLC, accessed 15 June 2020); https://www.google.com/covid19/mobility/
35. *The State of the Restaurant Industry* (Open Table, accessed 18 May 2020); https://www.opentable.com/state-of-industry
36. Bois, F. Y. GNU MCSim: Bayesian statistical inference for SBML-coded systems biology models. *Bioinformatics* **25**, 1453–1454 (2009).
37. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472 (1992).
38. RStudio (RStudio Team, 2020); https://rstudio.com/
39. *R: The R Project for Statistical Computing* (R Core Team, 2020); https://www.r-project.org/
40. *Population and Housing Unit Estimates Dataset, 2010–2019* (United States Census Bureau, accessed 19 May 2020); https://www2.census.gov/programs-surveys/popest/datasets/2010-2019/state/detail/
41. Lauer, S. A. et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann. Intern. Med.* https://doi.org/10.7326/M20-0504 (2020).
42. He, X. et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).
43. Mossong, J. et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, e74 (2008).
44. Verity, R. et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* **20**, 669–677 (2020).
45. Li, Q. et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020).
46. Riou, J. & Althaus, C. L. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance* **25**, 2000058 (2020).
47. Park, S. W. et al. Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: a new framework and applications to the novel coronavirus (2019-nCoV) outbreak. *J. R. Soc. Interface* **17**, 20200144 (2020).
48. *America's COVID Warning System* (Covid Act Now, accessed 15 June 2020); https://covidactnow.org/?s=49762
49. Gao, Z. et al. A systematic review of asymptomatic infections with COVID-19. *J. Microbiol. Immunol. Infect.* https://doi.org/10.1016/j.jmii.2020.05.001 (2020).
50. Arons, M. M. et al. Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *N. Engl. J. Med.* **382**, 2081–2090 (2020).
51. Nishiura, H. et al. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *Int. J. Infect. Dis.* **94**, 154–155 (2020).
52. Watson, J., Whiting, P. F. & Brush, J. E. Interpreting a covid-19 test result. *Brit. Med. J.* **369**, m1808 (2020).
53. Cummings, M. J. et al. Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. *Lancet* **395**, 1763–1770 (2020).
54. Sun, K., Chen, J. & Viboud, C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. *Lancet Digit. Health* **2**, e201–e208 (2020).
55. Onder, G., Rezza, G. & Brusaferro, S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA* **323**, 1775–1776 (2020).
56. *Map: How States are Reopening after Coronavirus Shutdown* (Washington Post, accessed 15 June 2020); https://www.washingtonpost.com/graphics/2020/national/states-reopening-coronavirus-map/

Acknowledgements

We acknowledge funding from the National Science Foundation (no. RAPID DEB 2028632) and National Institutes of Health, National Institute of Environmental Health Sciences (no. P30 ES029067). The funders had no role in the conceptualization, design, data collection, analysis, decision to publish or preparation of the manuscript. We thank F. Y. Bois, J. K. Cetina, M. Giannoni, I. Rusyn and W. Więcek for useful input and advice on scenario development, model formulation and MCMC simulation. We also thank Unacast for making their mobility data available to researchers, and The COVID Tracking Project for compiling case and mortality data and providing these to the public. Portions of this research were conducted with the advanced computing resources provided by Texas A&M High Performance Research Computing.

Author contributions

W.A.C. and M.L.N.-M. conceived and designed the study and developed the model. W.A.C. performed the analysis. W.A.C. and M.L.N.-M. drafted the paper, with critical revision by R.F.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41562-020-00969-7.

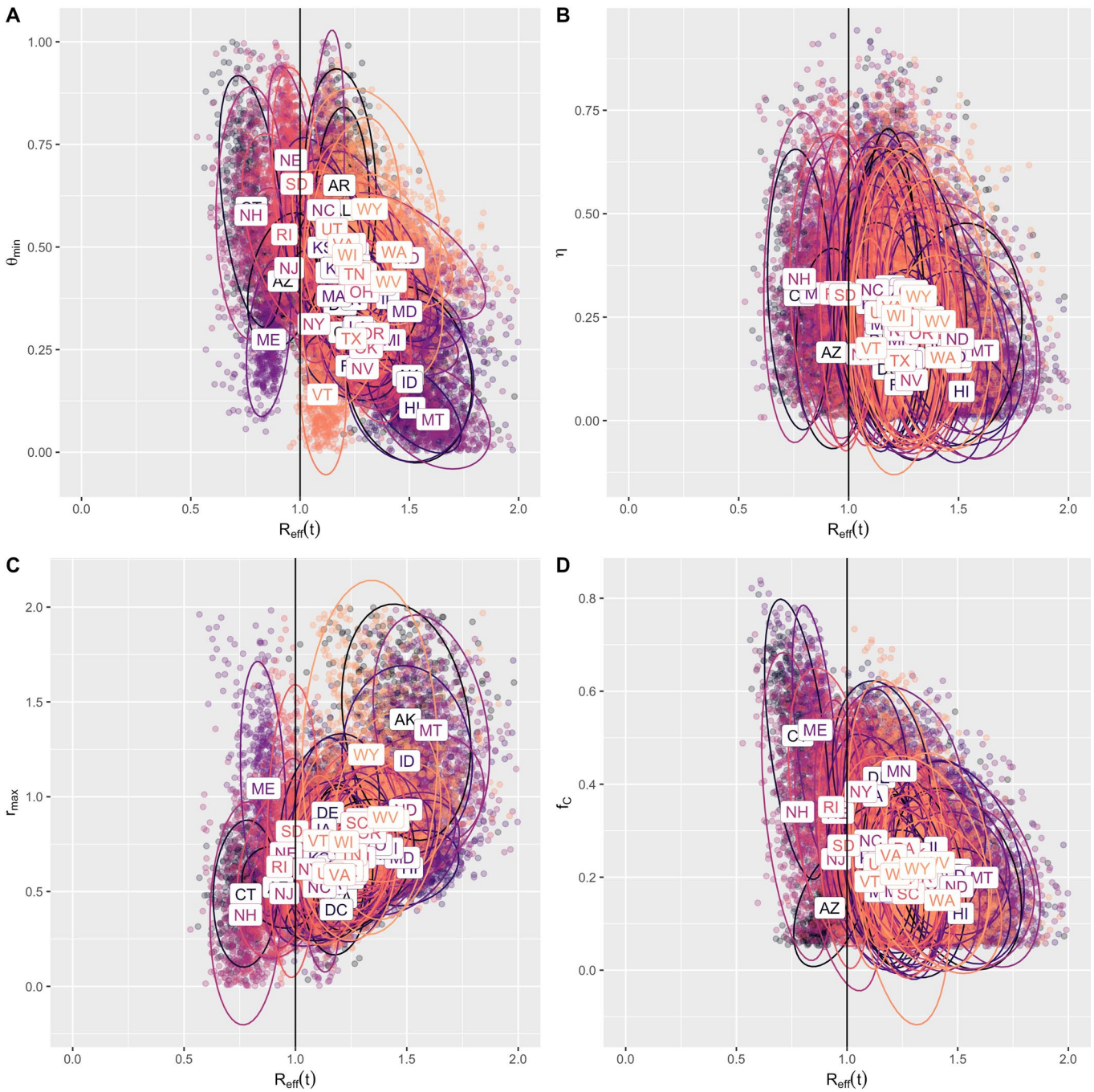
Supplementary information is available for this paper at https://doi.org/10.1038/s41562-020-00969-7.

Correspondence and requests for materials should be addressed to W.A.C. or M.L.N.-M.

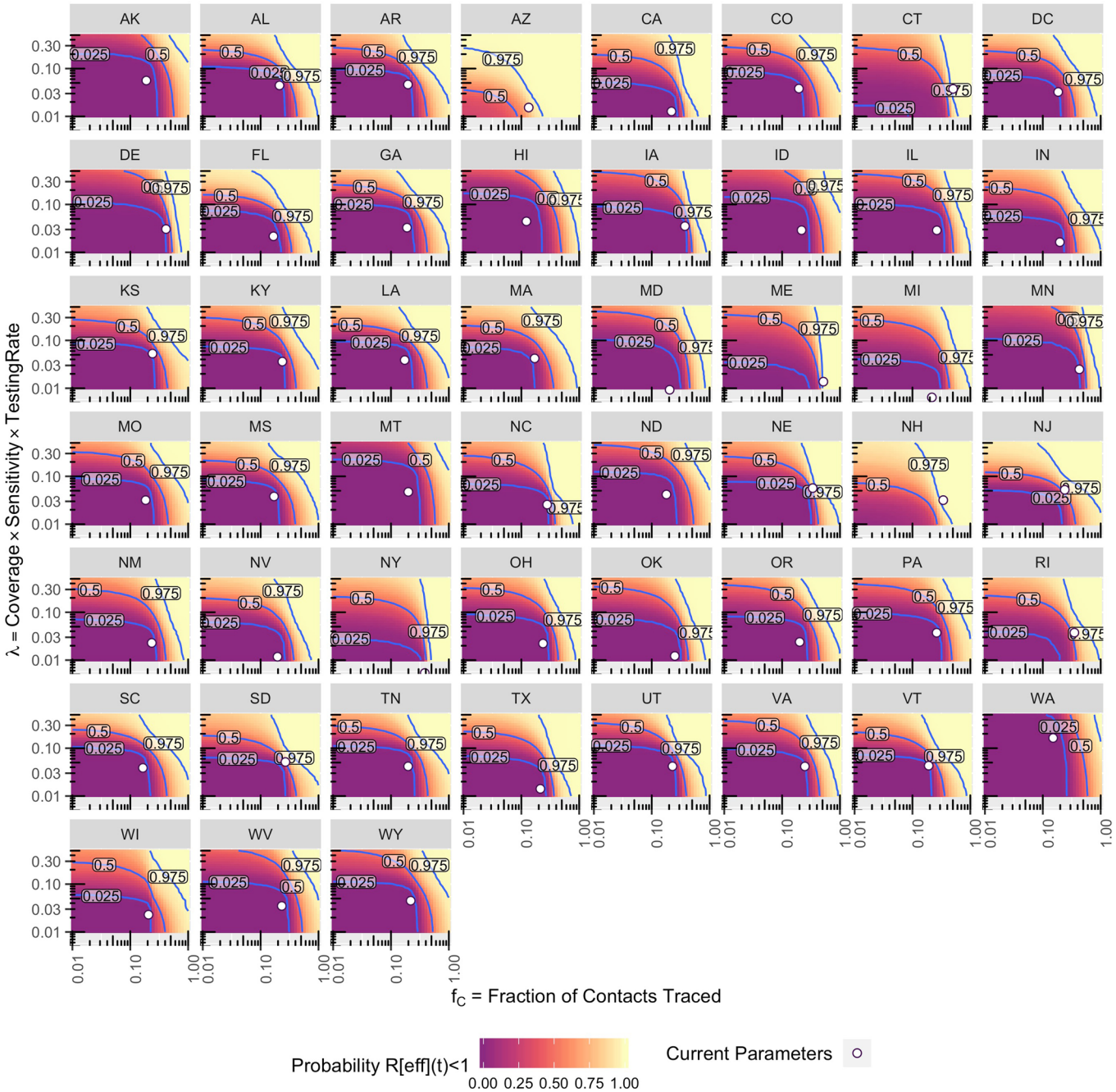
Peer review information Primary handling editor: Charlotte Payne.

Reprints and permissions information is available at www.nature.com/reprints.

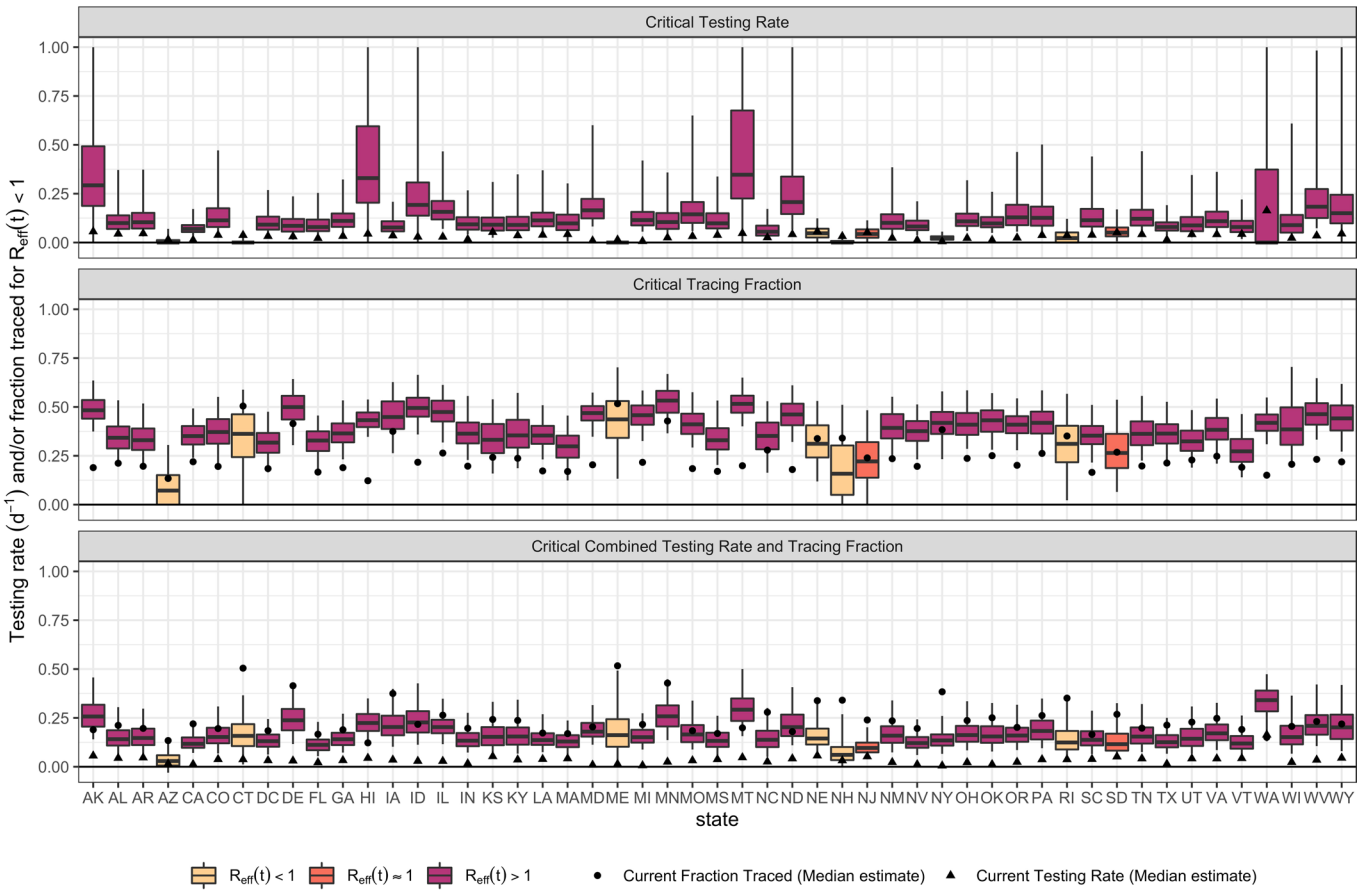
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



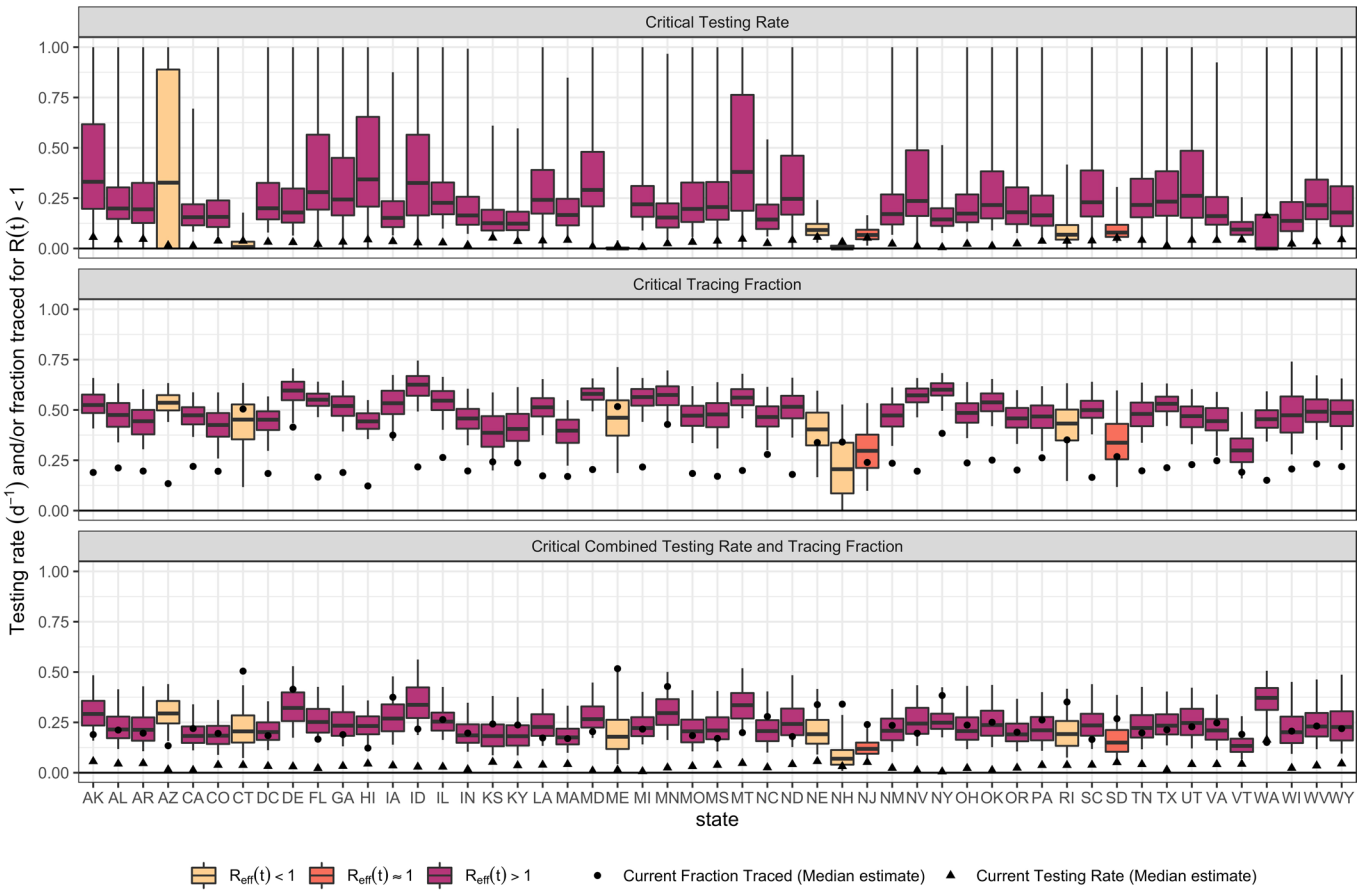
Extended Data Fig. 1 | Correlations across states between $R_{eff}(t)$ and (A) θ_{min}, (B) η, (C) Δ, and (D) f_C. For each state, 500 posterior samples are shown. Substantial state-to-state heterogeneity is evident in all parameters, with η,θ_{min}, r_{max}, and f_C contribute over 50% of the variance in $R_{eff}(t)$ under a linear model (estimated from ANOVA table [Supplementary Table 1] using the sum-of-squares relative to the total sum-of-squares). For θ_{min}, while lower values appear to be associated with greater current values of $R_{eff}(t)$ in a univariate model (linear model coefficient = -0.54, t statistic = -78.14, $p < 2.2e\text{-}16$, 95% CI = [-0.56,-0.53]), the correlation is positive in the multivariate model (coefficient = 0.11, t statistic = 13.65, $p < 2.2e\text{-}16$, 95% CI = [0.09, 0.13]). The other parameters correlate as expected: higher $R_{eff}(t)$ is correlated with lower contribution from hygiene practices (smaller η) (coefficient = -0.39, t statistic = -59.7, $p < 2.2e\text{-}16$, 95% CI = [-0.40,-0.38]), more reopening (larger r_{max}) (coefficient = 0.44, t statistic = 119.0, $p < 2.2e\text{-}16$, 95% CI = [0.43, 0.45]), and lower rates of contact tracing (smaller f_C) (coefficient = -0.90, t statistic = -114.6, $p < 2.2e\text{-}16$, 95% CI = [-0.92, -0.89]).



Extended Data Fig. 2 | Contour maps for each state of the probability that $R_{eff}(t)$ at different levels contact tracing f_c and testing λ. Contours are labelled as by median and 95% Credible interval, and current median estimates of f_c and λ are shown by the circle.



Extended Data Fig. 3 | Estimated testing and contact tracing rates needed for $R_{eff}(t) < 1$ as of July 22, 2020. Boxplots (line = median, box = IQR, whiskers = 95% CrI) are filled based on the estimated $R_{eff}(t)$ on July 22, 2020, as shown in the legend. Top panel is changing testing rate alone, the middle panel is changing contact tracing rate alone, and the bottom panel is changing both to the same value. Also shown are the current median estimates of the testing and contact tracing rates.



Extended Data Fig. 4 | Estimated testing and contact tracing rates needed for R(t) < 1 with complete re-opening (that is, removal of all social distancing and hygiene mitigation). Boxplots (line = median, box = IQR, whiskers = 95% CrI) are filled based on the estimated $R_{eff}(t)$ on July 22, 2020, as shown in the legend. Top panel is changing testing rate alone, the middle panel is changing contact tracing rate alone, and the bottom panel is changing both to the same value. Also shown are the current median estimates of the testing and contact tracing rates.

nature research

Corresponding author(s): Martial L Ndeffo-Mbah

Last updated by author(s): Aug 7, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☒	☐ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	NA
Data analysis	Posterior distributions were sampled using Markov chain Monte Carlo simulation performed using MCSim version 6.1.0 using Metropolis within Gibbs sampling; Additional analysis of model outputs was performed in RStudio version 1.2.1335 with R version 3.6.1. The codes used to generate our results will be available on Github prior to publication at https://github.com/wachiuphd/COVID-19-Bayesian-SEIR-US.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The following publicly available datasets are used:
Mobility data from Unacast were sourced from https://covid19-scoreboard-api.unacastapis.com/api/search/covidstateaggregates_v3.
Mobility data from Google were sourced from https://www.gstatic.com/covid19/mobility/Global_Mobility_Report.csv.
Restaurant booking data were sourced from OpenTable https://www.opentable.com/state-of-industry.
Case and death data were sourced from The COVID Tracking Project (https://covidtracking.com/).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	NA
Data exclusions	NA
Replication	NA
Randomization	NA
Blinding	NA

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging