

RESEARCH ARTICLE

A data-driven semi-parametric model of SARS-CoV-2 transmission in the United States

John M. Drake^{1,2*}, Andreas Handel^{2,3}, Éric Marty^{1,2}, Eamon B. O'Dea^{1,2}, Tierney O'Sullivan^{1,2}, Giovanni Righi^{1,2}, Andrew T. Tredennick^{2,4}

1 Odum School of Ecology, University of Georgia, Athens, Georgia, United States of America, **2** Center for the Ecology of Infectious Diseases, University of Georgia, Athens, Georgia, United States of America, **3** College of Public Health, University of Georgia, Athens, Georgia, United States of America, **4** Western EcoSystems Technology, Inc., Laramie, Wyoming, United States of America

✉ Current address: Metrum Research Group, Tariffville, Connecticut, United States of America

* jdrake@uga.edu



Abstract

To support decision-making and policy for managing epidemics of emerging pathogens, we present a model for inference and scenario analysis of SARS-CoV-2 transmission in the USA. The stochastic SEIR-type model includes compartments for latent, asymptomatic, detected and undetected symptomatic individuals, and hospitalized cases, and features realistic interval distributions for presymptomatic and symptomatic periods, time varying rates of case detection, diagnosis, and mortality. The model accounts for the effects on transmission of human mobility using anonymized mobility data collected from cellular devices, and of difficult to quantify environmental and behavioral factors using a latent process. The baseline transmission rate is the product of a human mobility metric obtained from data and this fitted latent process. We fit the model to incident case and death reports for each state in the USA and Washington D.C., using likelihood Maximization by Iterated particle Filtering (MIF). Observations (daily case and death reports) are modeled as arising from a negative binomial reporting process. We estimate time-varying transmission rate, parameters of a sigmoidal time-varying fraction of hospitalized cases that result in death, extra-demographic process noise, two dispersion parameters of the observation process, and the initial sizes of the latent, asymptomatic, and symptomatic classes. In a retrospective analysis covering March–December 2020, we show how mobility and transmission strength became decoupled across two distinct phases of the pandemic. The decoupling demonstrates the need for flexible, semi-parametric approaches for modeling infectious disease dynamics in real-time.

OPEN ACCESS

Citation: Drake JM, Handel A, Marty É, O'Dea EB, O'Sullivan T, Righi G, et al. (2023) A data-driven semi-parametric model of SARS-CoV-2 transmission in the United States. *PLoS Comput Biol* 19(11): e1011610. <https://doi.org/10.1371/journal.pcbi.1011610>

Editor: Gerardo Chowell, Georgia State University, UNITED STATES

Received: March 1, 2023

Accepted: October 17, 2023

Published: November 8, 2023

Copyright: © 2023 Drake et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All code and data used in this study are publicly available on GitHub (<https://github.com/CEIDatUGA/COVID-stochastic-fitting-manuscript>) and have been archived on Zenodo (<https://zenodo.org/record/8408895>).

Funding: This work was supported by the US National Science Foundation (grant no. NSF DEB-2027786) to JMD. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author summary

Due to the time-varying nature of numerous drivers of disease transmission, flexible, semi-parametric models of disease transmission are necessary to faithfully represent the complex, hidden dynamic processes that drive epidemics of emerging pathogens like SARS-CoV-2. Adequate models are essential to guide policy decisions. We present a data-

Competing interests: The authors have declared that no competing interests exist.

driven semi-parametric model of SARS-CoV-2 transmission that embeds a latent process within a mechanistic compartmental model. The latent process sub-model captures temporal variation in precautionary behaviors that cannot be easily measured. We show that temporal variation in transmission strength is best explained by mobility early in the pandemic and by this latent process later in the pandemic. The model is flexible and incorporates known biological parameters and disease transmission processes.

Introduction

The COVID-19 pandemic in the United States has challenged the capabilities of conventional infectious disease transmission models [1, 2]. Yet, models are critical for guiding policy decisions, updating situational awareness, and retrospectively evaluating the key drivers of transmission and the effectiveness of interventions [3, 4]. The complex interactions of a highly transmissible disease, seasonal forcing, evolving public health messaging and guidance, the rise of new genetic variants, and the long duration of the COVID-19 pandemic have made it difficult to effectively model how the mechanistic drivers of the force of infection changed over time [5]. Lack of adequate data about key behavioral factors, such as the adoption of individual protective behaviors like mask-wearing, adds to the difficulty.

For example, the initial phase of the pandemic was marked by stay-at-home orders across the nation [6]. Human mobility dropped drastically from pre-pandemic levels, slowing the spread of SARS-CoV-2 [7]. Publicly available cell phone mobility data allowed modelers to explicitly estimate the impact of human movement—a proxy for person-to-person contact—in mechanistic models of disease transmission [5, 8, 9]. However, when states began to re-open in April 2020, mobility data became a weaker correlate of transmission strength [4, 10]. Precautionary measures such as masking and social-distancing in public spaces weakened the relationship between mobility and transmission risk [4]. But no consistent and reliable data exist to capture the dynamics of all precautionary behaviors over time. Carefully designed and constrained models can infer these latent dynamics, allowing for more accurate situational awareness and forecasting [4].

Here, we present a data-driven semi-parametric compartmental model of SARS-CoV-2 transmission. The stochastic model includes compartments for latent (exposed but pre-symptomatic), asymptomatic, detected and undetected symptomatic individuals, and hospitalized cases (Fig 1). The model also includes time-varying rates of detection probability, diagnosis time, and mortality, all of which add realism to the model, improve identifiability, and enable stronger inference on remaining dynamics. In our model, transmission rate is allowed to differ among asymptomatic, pre-symptomatic, and symptomatic individuals. Force of infection is a function of the number of individuals in each compartment, the relative mobility of individuals, and a latent process. The latent process is a time-dependent spline function (see [Materials and methods](#)) representing all of the processes influencing the force of infection that cannot be measured. The model was developed in real-time during 2020 and fitted jointly to daily case and deaths data from each U.S. state and the District of Columbia (hereafter collectively referred to as “US states”) from the date of first case notification in each state to December 31, 2020. For this publication, case and death reports were retrieved from the Johns Hopkins University Coronavirus Resource Center [11] on July 18, 2022, meaning any revisions to the data due to time-lags or changes in inclusion criteria for 2020 were applied. Human mobility data were obtained from Unacast [12]. We restricted our analysis to March 1, 2020 to December

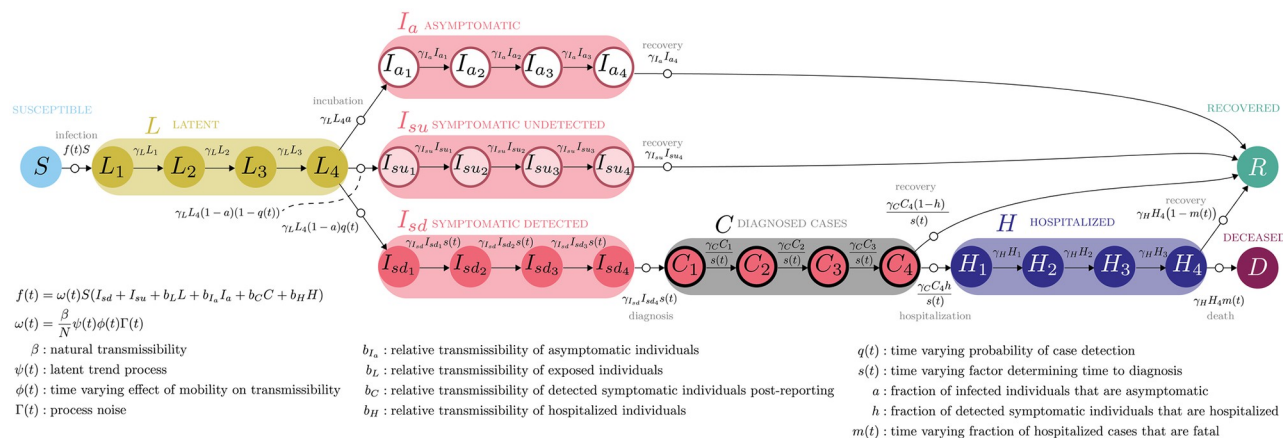


Fig 1. Compartmental model of SARS-CoV-2 transmission.

<https://doi.org/10.1371/journal.pcbi.1011610.g001>

31, 2020, a logical endpoint for retrospective analysis because the model does not include vaccination, and vaccines began to be administered in late December 2020.

To demonstrate the usefulness of our semi-parametric model, we perform a retrospective analysis showing that transmission strength (effective reproduction number) and mobility became decoupled over time. By “decoupled”, we mean two time-varying variables are correlated in one period and not correlated in another. This conclusion could not be arrived at with traditional parametric models. Correlation analysis shows that relative mobility among US states was highly correlated over time, regardless of pandemic phase. To the contrary, however, transmission strength became uncorrelated among states as different states had different levels of adherence to precautionary behaviors [13]. We then show how the relative mobility trend and the latent process trade off in importance over time, so that transmission alternately synchronized with or decoupled from mobility. These dynamics can only be captured with a flexible, data-driven modeling approach when key underlying processes cannot be completely measured.

Results

The model estimated daily cases and daily deaths that closely match the observed data although allowing for reporting errors resulting from aggregation, reporting delays, weekend effects, and other anomalies (Fig 2A and 2B; model fits for all states are shown in Section C in S1 Appendix). Model performance varied by state and by response variable. The model for 45 states (out of 51 total) had mean absolute scaled errors for new cases less than or equal to one (Fig B in S1 Appendix), meaning that the semi-parametric model performed better than a non-mechanistic model comprising a random walk model with weekly periodicity. However, only 30 states had mean absolute scaled errors for new deaths less than or equal to one (Fig B in S1 Appendix), meaning either the process for modeling the transition from hospitalized cases to deaths or the observation model could be improved. The smoothed trajectories of daily deaths capture the general trends of the data well (Fig 2B and Section C in S1 Appendix), which suggests the observation model is the limiting factor in our model. This can be seen in Fig 2B, where the fitted trajectory passes through the cloud of data points but fails to capture cyclical reporting patterns. Cyclical reporting patterns are identified, for example, by large numbers of daily deaths being followed by reports of near 0 daily deaths in the middle of the time series (July–September, Fig 2B).

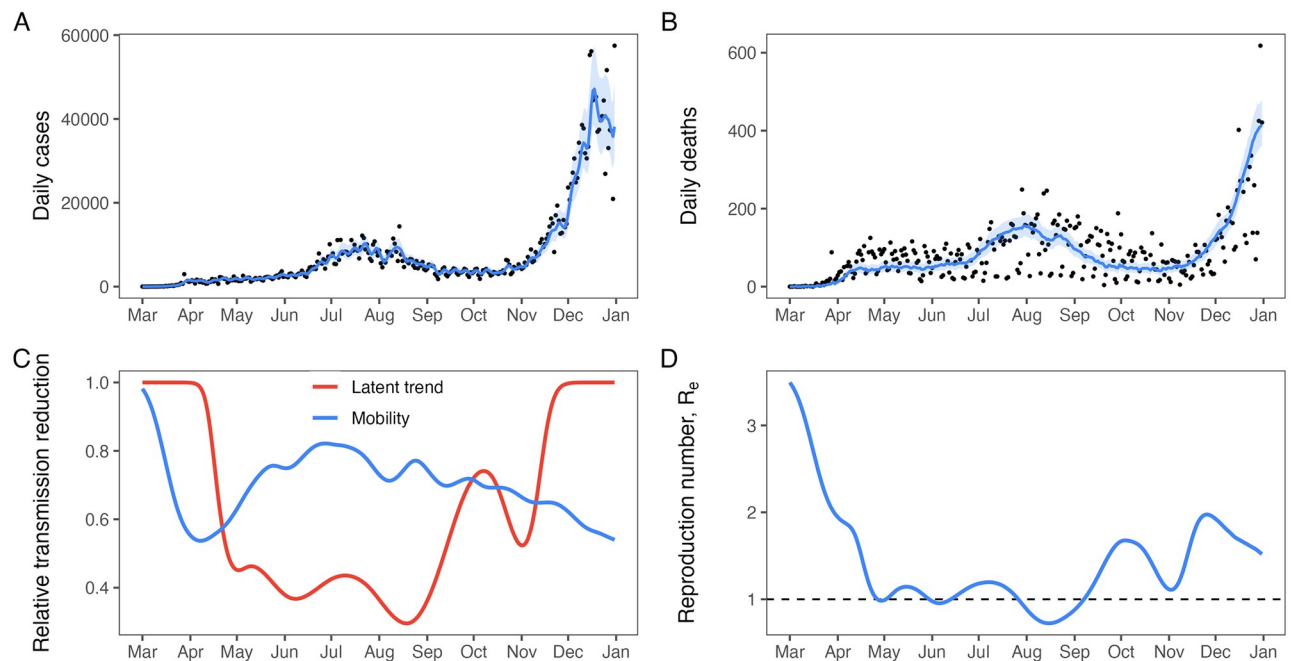


Fig 2. Model fits and estimated quantities for California. A comparison of estimated and observed (A) COVID-19 daily cases and (B) COVID-19 daily deaths show the course of the COVID-19 pandemic in California from March–December 2020. In A and B, lines show the median estimated state variable, ribbons show the 95% prediction interval, and points are noisy, recorded observations. In panel (C), the estimated relative mobility trend (blue line) and the estimated latent trend (red line) are shown to vary considerably over time resulting in dramatic fluctuations in the force of infection, giving rise to the multiple waves infection shown in (A) and (B). These fluctuations in the force of infection are illustrated by (D) the estimated effective reproduction number, which crossed the critical boundary at $R_e = 1$ on numerous occasions during 2020.

<https://doi.org/10.1371/journal.pcbi.1011610.g002>

We also estimated a latent process that, when combined with mobility, uniquely defines transmission strength (Fig 2C). We used the next-generation matrix approach [14] to numerically calculate the time-varying effective reproduction number (Materials and methods), which is a function of mobility, the estimated latent process, and state variables (Figs 2D and 3B).

Analysis of these patterns identified two distinct phases of the early pandemic associated with the time series of relative mobility, the cumulative number of state-issued public health mandates nationwide, and the date at which states began “re-opening” (per CDC definitions: <https://www.cdc.gov/museum/timeline/covid19.html>). On April 24, 2020, states began re-opening and allowing public health mandates to expire, which unsurprisingly coincided with increasing average mobility (Fig 3A). We use this date to notionally partition two phases of the pandemic from March 1, 2020 to April 24, 2020 (Phase 1) and from April 25, 2020 to December 31, 2020 (Phase 2). In Phase 1, we understand transmission to be almost entirely driven by human-to-human contact. The sharp reduction in human mobility compared with the pre-pandemic baseline resulted in dramatically lowering the effective reproduction number, $\mathcal{R}_e(t)$ [4, 8, 15]. In Phase 2, precautionary behaviors like mask wearing and social distancing in public played a much greater role in determining $\mathcal{R}_e(t)$ than mobility [4].

Our estimates of $\mathcal{R}_e(t)$ display the decoupling of transmission strength from mobility between the two phases (Fig 3B). For instance, pairwise correlations of mobility (mean of the absolute value of Pearson’s ρ across all pairs = 0.98, SD = 0.03) and $\mathcal{R}_e(t)$ (mean $|\rho|$ = 0.86, SD = 0.22) are high among states in the first phase of the pandemic (Fig 4A and 4B). In the second phase of the pandemic, mobility remained highly correlated among states (mean

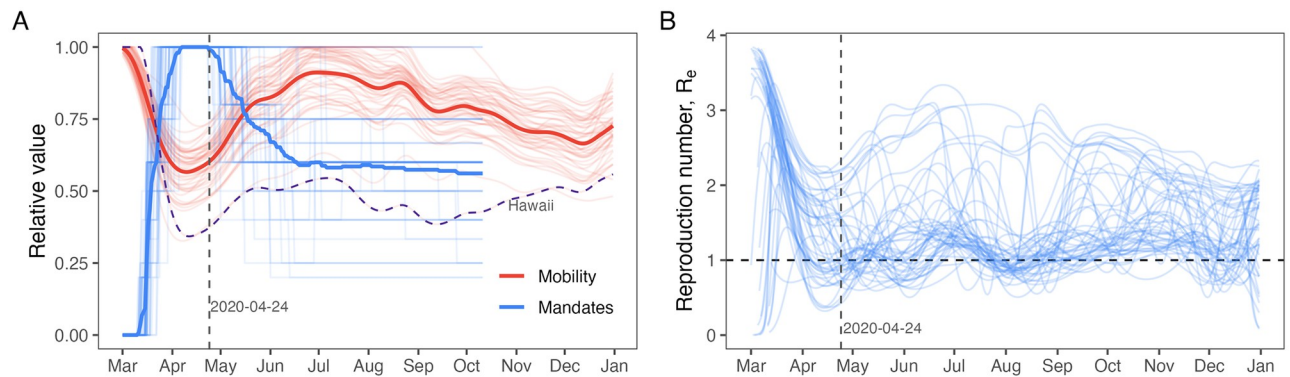


Fig 3. Drivers of force of infection over time. (A) The time series of the relative number of public health mandates (blue lines) and relative mobility (red lines) for each state shows how public health messaging and individual behavior diverge after April 24, 2020 (shown as a vertical dashed line). Relative number of health mandates was calculated as the total number of health mandates issued in each 50 states at each time, divided by the maximum observed over all times in each state. Each light line is a single state. Heavy lines are the average relative mobility and relative number of mandates across all states. The dashed purple line shows the mobility over time for Hawaii, which was the lowest among all states. Before April 24, 2020, public health mandates increased while mobility decreased in all states. After April 24, 2020, public health mandates began to expire and mobility increased, but states began to diverge in their responses, as seen by the larger spread in the light lines after April 24. In panel (B), the estimated effective reproduction number, $\mathcal{R}_e(t)$, for each state over time reflects the changes in public health mandates and mobility. Before April 24, $\mathcal{R}_e(t)$ declined to below the critical value of 1 (dashed horizontal line) for almost all states. After April 24, $\mathcal{R}_e = 1$ is much more dynamic and variable across states, reflecting the differences in mobility and public health messaging. Analysis of these patterns suggest two distinct phases of the pandemic: Phase 1 before April 24 and Phase 2 after April 24. Critically, even as mobility increased in Phase 2 to near pre-pandemic levels, $\mathcal{R}_e(t)$ typically remained near 1.

<https://doi.org/10.1371/journal.pcbi.1011610.g003>

$|\rho| = 0.88$, $SD = 0.16$; Fig 4C) while $\mathcal{R}_e(t)$ became uncorrelated among states (mean $|\rho| = 0.32$, $SD = 0.24$; Fig 4D).

Variance partitioning showed that mobility explained most of the temporal variation of $\mathcal{R}_e(t)$ in the first phase of the pandemic and that the latent trend explained most of the variation in $\mathcal{R}_e(t)$ in the second phase of the pandemic for nearly all states (Fig 5). The latent process was an important model component for some states in Phase 1 of the pandemic (e.g., Nevada (NV), Fig 5). There were no states for which mobility remained more important than the latent process in determining the temporal variation of $\mathcal{R}_e(t)$ in Phase 2 of the pandemic (Fig 5). This is especially noteworthy since relative mobility in Hawaii was the lowest among all states throughout 2020 (Fig 3A). Taken together, these lines of evidence show that $\mathcal{R}_e(t)$ decoupled (i.e., not temporally related) from mobility over the course of the pandemic, a feature of the pandemic that we attribute to “societal learning” as the population adopted increasingly nuanced and localized approaches to reducing transmission [16].

Discussion

The COVID-19 pandemic spurred innovation in epidemiological modeling because models and forecasts of how the pandemic might unfold were in high demand. However, models that cannot reproduce the complex outcomes of nonlinear dynamical systems are not useful for situational awareness, forecasting, or scenario analysis under any but the shortest time horizons [9]. This does not mean that a good model must include mechanistic descriptions of all relevant processes. But, models do need to be able to reproduce observed outcomes (i.e., case, hospitalization, and death reports). We presented a flexible modeling approach that is rooted in the mechanisms of disease transmission and included a latent trend to represent unmeasurable factors that influence transmission. In so doing, we show the decoupling of SARS-CoV-2 transmission from social distancing as other precautionary behaviors became more prevalent.

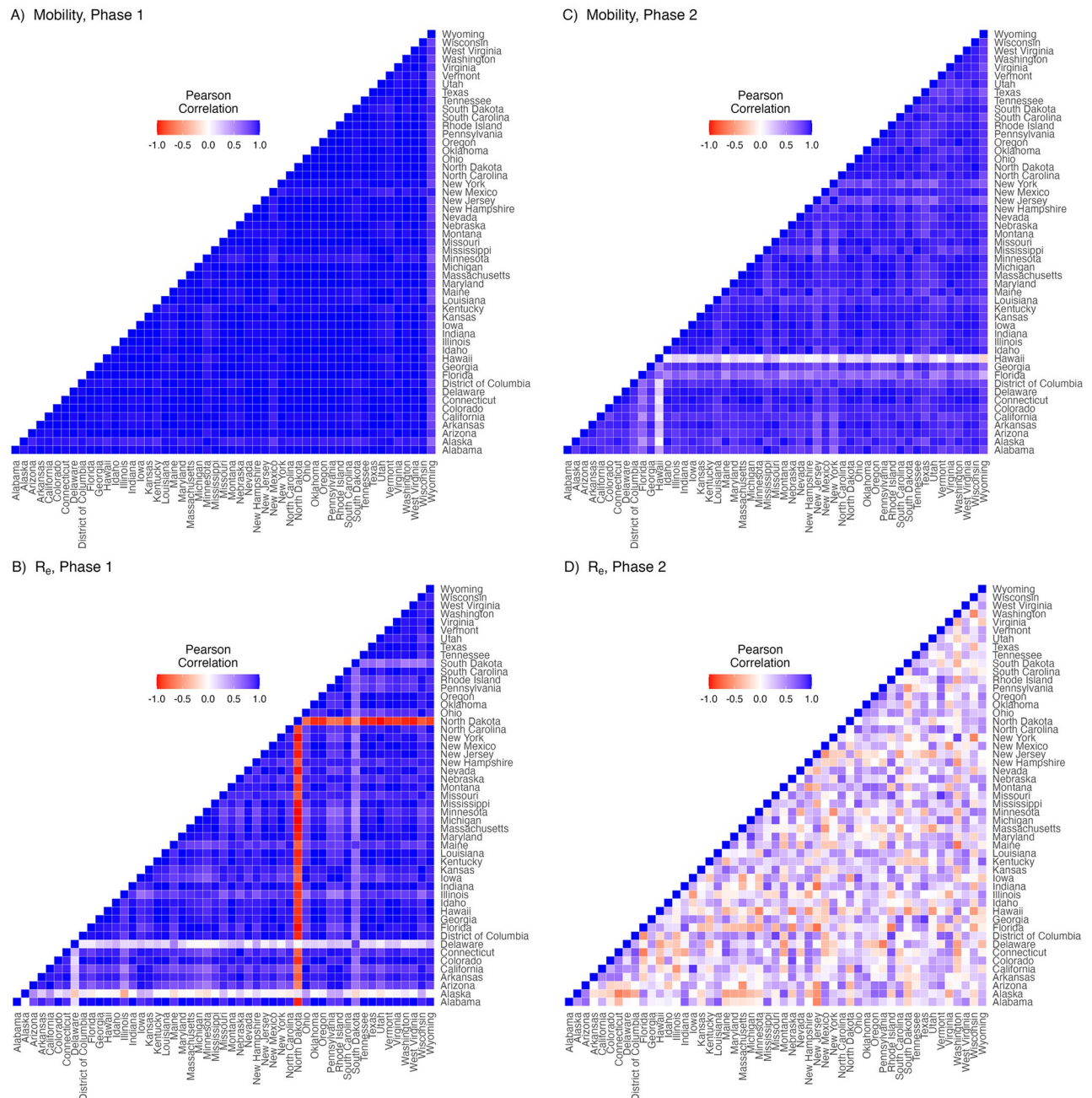


Fig 4. Between-state correlations in mobility and force of infection. Pairwise cross correlations among time series of the (A) relative mobility covariate for each pair of states and (B) among time series of $\mathcal{R}_e(t)$ for each pair of states in Phase 1 of the pandemic (March 1, 2020 to April 24, 2020) show that mobility and force of infection were highly correlated across states in Phase 1 of the pandemic. This indicates a tight coupling between mobility and force of infection in the Phase 1. Panels C and D visualize the same pairwise cross correlations during Phase 2 of the pandemic (April 24, 2020 to December 31, 2020), which shows that mobility remains correlated among states while force of infection is uncorrelated. This suggests that mobility is not the main driver of disease dynamics in Phase 2.

<https://doi.org/10.1371/journal.pcbi.1011610.g004>

Our semi-parametric model mostly performed better than a benchmark random walk model for cases. This benchmark random walk model, which is strictly statistical and lacks mechanistic interpretation, was included in another study that compared twenty-eight COVID-19 forecasting models [17]. The benchmark random walk model was found to have

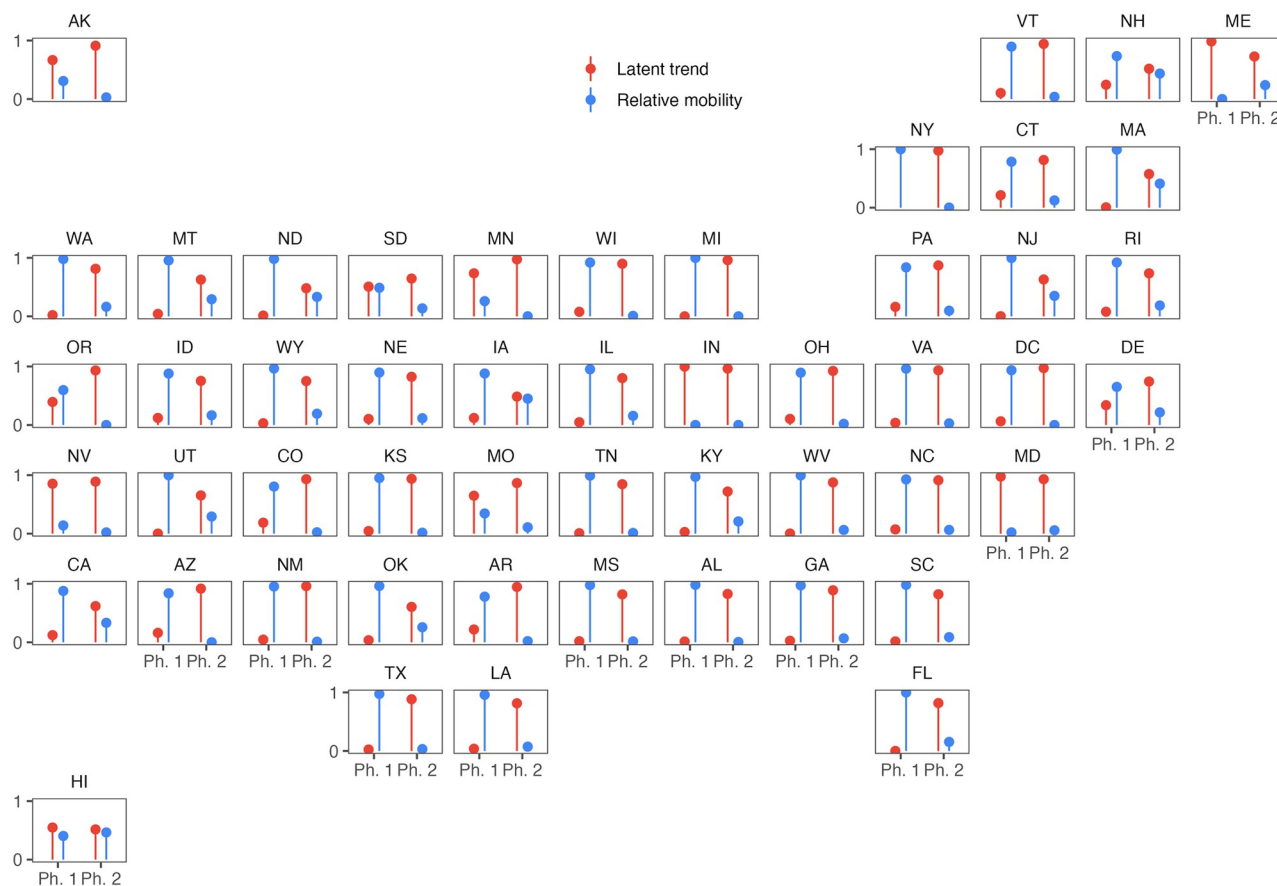


Fig 5. Drivers of temporal variance in force of infection. Comparisons of the amount of temporal variance in $\mathcal{R}_e(t)$ that can be explained by mobility (blue bars) and the latent process (red bars) in the two phases of the pandemic shows that, for most states, the temporal variance of $\mathcal{R}_e(t)$ switches from being explained by mobility in Phase 1 to being explained by the latent process in Phase 2. The temporal variation in $\mathcal{R}_e(t)$ is mostly explained by the latent trend in both phases for several states (e.g., Nevada). Individual panels for each states are arranged by approximate geographic coordinates for each state.

<https://doi.org/10.1371/journal.pcbi.1011610.g005>

intermediate forecasting skill in comparison with the twenty-seven other models [17], making it a good baseline for comparison with our semi-parametric model. The semi-parametric model fits for six states had mean absolute scaled errors (MASE) greater than one (Fig B in S1 Appendix), indicating worse fit than the benchmark random walk model with weekly periodicity. These six states tended to have case reports that oscillated between 0 and 100s in the later weeks of 2020 (Section C in S1 Appendix), likely due to lags in reporting. Because our semi-parametric model did not include an observation process with reporting lags or periodicity, the model performed worse for those states where reporting oscillations were dramatic. The same issue (report lags and periodicity in reporting) resulted in death fits from our model being worse than the benchmark random walk model for 21 states. We suggest that model performance might be improved by including a more complex observation model to account for cyclical variation in reporting [5].

The better fit of the benchmark random walk model in some situations does not invalidate the utility of our model. The benchmark model, a random walk with weekly periodicity, does not include any epidemiology. Therefore, while the random walk model may generate slightly better one-step-ahead predictions in some situations, it does not yield any scientific insight.

Our semi-parametric model did perform better than the random walk model for many states and the filtered trajectories of new cases and deaths matched the observations well, even when $\text{MASE} \geq 1$ (Fig B in [S1 Appendix](#)).

The results of variance partitioning suggest that a purely mechanistic model would have performed just as well as our semi-parametric model in the early stages of the pandemic, when human mobility was the main driver of transmission strength. However, our model performed well across all stages of the pandemic because mobility was allowed to drive transmission early in the pandemic and then trade-off with the latent process in latter stages. The flexibility of the time-varying latent process is critical because the intensity of precautionary behaviors varied over time. Some analyses have sought to address this issue using additional data streams such as the presence of mask mandates over space and time [18]. In contrast, our flexible time-evolving process avoids the need for additional data streams.

Other modelers have used temporal random effects to estimate time-varying transmission strength [4, 19]. For example, Fox et al. [4] modeled time-varying transmission as a function of two time-specific random walks and a normally distributed temporal random effect. Modeling discrete change-points in transmission strength is another approach [20]. But the first- and second-order discontinuities of change-points hardly reflect the way that behaviors change over time and their influence on the ebb and flow of transmission. Our spline approach achieves the desired outcome—letting the relationship between mobility and transmission change over time—while also directly quantifying the contribution of mobility and other forces to moderating transmission. Indeed, our approach is more flexible because the number of basis functions used to estimate the B-spline can also be fitted or, if needed, trimmed to the minimum number possible for computationally efficient model fitting.

We estimated maximum $\mathcal{R}_e(t)$ values in range of 1.52–3.84 across all 50 states. Highest $\mathcal{R}_e(t)$ occurred at the start of the epidemic (Fig 3B and Section D in [S1 Appendix](#)). These estimates are similar to an $\mathcal{R}_0$ estimated from meta-analysis, 3.38 ± 1.40 [21]. Our estimates are lower than those reported for some individual cities and municipalities, probably because our state-level model is too coarse to capture super-spreading events that happen at local scales [22]. For example, Fox et al. [4] reported a 7-day average $\mathcal{R}_e(t)$ of 5.8 at the start of the epidemic in Austin, Texas and Hasan et al. [22] estimated $\mathcal{R}_e(t)$ greater than 5 in the Jakarta–Depok region of Indonesia.

Our analysis has implications for modeling the next pandemic. First, as reported by others, adherence to precautionary behaviors is not static and is not always measurable [4, 23]. This means that modeling approaches need built-in flexibility to estimate latent trends, such as the time-dependent spline function we used. Second, incorporating latent trends makes situational awareness easier (i.e., answering the question “What is the current trend in the force of infection?”), but it makes forecasting harder. Estimating trends in transmission is easier with a data-driven model because the data have a large influence on model dynamics and summary quantities such as $\mathcal{R}_e(t)$. Forecasting is more difficult, however, because the latent trend must be specified for future times at which (by definition) no data exist. This means that latent, smooth trends that were specifically designed to fluctuate over time must be, in practice, extrapolated in some fashion. New methods are arising to handle forecasting with latent processes when the underlying function is periodic [24]. But the transmission dynamics of an emerging pathogen are not periodic, meaning subjective choices must be made to specify the latent process at future times if forecasting is the objective. One possibility is to use the estimated number of cases to predict hospitalizations and/or deaths, but only up to future time

points within the typical duration between case notification, hospitalization, and death. Such forecasts would be relatively short-term by definition.

Flexible, semi-parametric models of disease transmission are an effective way to faithfully represent complex, hidden, and time-evolving transmission in outbreaks of emerging pathogens [4, 5, 25]. Human behavior and disease transmission are inherently linked and each generates feedback to the other. We believe it will never be possible to perfectly and fully represent human behavior. Nonetheless, carefully crafted models, such as the model we presented and others [4, 5, 19], can infer the impact of human behavior when fitted to data. The semi-parametric model we have presented is capable of using more input data (e.g., survey data on mask-wearing) or less (e.g., no mobility data), making it flexible enough to perform statistical inference regardless of how much information about covariates is available. This feature could be of substantial value to models of the ongoing COVID-19 pandemic, where information is accumulating over time, as well as future pandemics, where initial modeling efforts will also be hobbled by the lack of information.

Materials and methods

Data

We fit the model to daily case and death reports for each state in the USA collated by the Center for Systems Science and Engineering at Johns Hopkins University [11]. We did not smooth the reported data but we did redefine all negative case and death reports as NA-values. Time-series of case and death reports for each state used in this analysis are shown in Section C of [S1 Appendix](#). Data were downloaded on 2022-07-18, meaning any revisions to data from 2020 were applied. For example, data sets were often back-filled after reporting lags and some states changed the inclusion criteria for a death to be considered primarily a cause of COVID-19. We do not specifically account for these revisions or changes in data over time in this analysis.

Data on human mobility were downloaded from Unacast (<https://www.unacast.com/>). The data were pre-aggregated by Unacast, and were the difference in daily distance traveled relative to baseline distance (January 2020), averaged over individuals in each state. We converted the data to a relative difference in mobility by adding 1 to each value with a positive sign (greater than baseline) and subtracting the absolute value from 1 for all values with a negative sign (less than baseline). We then fit a smoothing spline through the time series to generate smooth covariate for modeling. The fitted splines for each state are shown in the Section D of [S1 Appendix](#).

Model

The model comprises susceptible, pre-symptomatic, asymptomatic, symptomatic, diagnosed, hospitalized, deceased, and recovered persons ([Table 1](#)). The infectious compartments (pre-symptomatic, asymptomatic, symptomatic, diagnosed, and hospitalized persons) differ in their transmissibility and are thus defined by their impacts on population-level transmission rather than clinical symptoms. To reflect realistic distributions of movement through compartments, the L , I_a , I_{sd} , I_{su} , C , and H compartments are internally split into four stages using the linear chain trick [26] (Fig A in [S1 Appendix](#)). Transitions between compartments were modeled using the Euler multinomial approximation given the size of the “donating” compartment and the specified or fitted rate of transition, as implemented in the R [27] package **pomp** [28] version 2.7.1.0.

Table 1. State variable symbols and definitions in the model.

Symbol	Definition
S	Uninfected and susceptible individuals. Susceptible individuals can become infected by individuals in the L , I_a , I_{su} , I_{sd} , C , and H stages.
L	Individuals with latent infections who do not yet show symptoms. Those individuals can be infectious. At the end of the L stage, a fraction moves into the I_a stage, another fraction moves into the I_{su} stage, and the remainder into the I_{sd} stage.
I_a	Individuals who are infected and asymptomatic.
I_{su}	Individuals who are infected and symptomatic, but are undetected.
I_{sd}	Individuals who are infected and symptomatic, and are detected. Individuals in this compartment will be diagnosed and move to C .
C	Individuals who have been diagnosed as cases. Those individuals are likely isolated and have reduced infectiousness. A fraction of individuals in the C stage will naturally recover, without the need for hospitalization. The remainder move into the H stage.
H	Individuals who have been hospitalized. Those individuals are likely isolated and have reduced infectiousness. A fraction of individuals in the H stage will recover, the remainder will die.
R	Recovered/removed individuals. Those individuals have recovered and are immune.
D	Individuals who died from the infection.

<https://doi.org/10.1371/journal.pcbi.1011610.t001>

Ignoring the internal splits of multi-stage compartments, the stochastic model is defined as the set of difference equations

$$\begin{aligned}
 S(t+1) - S(t) &= -n_1 \\
 L(t+1) - L(t) &= n_1 - n_2 - n_3 - n_4 \\
 I_a(t+1) - I_a(t) &= n_2 - n_5 \\
 I_{su}(t+1) - I_{su}(t) &= n_3 - n_6 \\
 I_{sd}(t+1) - I_{sd}(t) &= n_4 - n_7 \\
 C(t+1) - C(t) &= n_7 - n_8 - n_9 \\
 H(t+1) - H(t) &= n_8 - n_{10} - n_{11} \\
 R(t+1) - R(t) &= n_5 + n_6 + n_9 + n_{10} \\
 D(t+1) - D(t) &= n_{11}
 \end{aligned}$$

with

$$\begin{aligned}
 \begin{pmatrix} n_S \\ n_1 \end{pmatrix} &\sim \text{EM} \left(S(t), \begin{pmatrix} e^{-f(t)\Delta t} \\ 1 - e^{-f(t)\Delta t} \end{pmatrix} \right) \\
 \begin{pmatrix} n_L \\ n_2 \\ n_3 \\ n_4 \end{pmatrix} &\sim \text{EM} \left(L(t), \begin{pmatrix} e^{-\gamma_L \Delta t} \\ a(1 - e^{-\gamma_L \Delta t}) \\ (1 - q(t))(1 - a)(1 - e^{-\gamma_L \Delta t}) \\ q(t)(1 - a)(1 - e^{-\gamma_L \Delta t}) \end{pmatrix} \right) \\
 \begin{pmatrix} n_{I_a} \\ n_5 \end{pmatrix} &\sim \text{EM} \left(I_a(t), \begin{pmatrix} e^{-\gamma_{I_a} \Delta t} \\ 1 - e^{-\gamma_{I_a} \Delta t} \end{pmatrix} \right) \\
 \begin{pmatrix} n_{I_{su}} \\ n_6 \end{pmatrix} &\sim \text{EM} \left(I_{su}(t), \begin{pmatrix} e^{-\gamma_{I_{su}} \Delta t} \\ 1 - e^{-\gamma_{I_{su}} \Delta t} \end{pmatrix} \right) \\
 \begin{pmatrix} n_{I_{sd}} \\ n_7 \end{pmatrix} &\sim \text{EM} \left(I_{sd}(t), \begin{pmatrix} e^{-\gamma_{I_{sd}} \Delta t} \\ 1 - e^{-\gamma_{I_{sd}} \Delta t} \end{pmatrix} \right) \\
 \begin{pmatrix} n_C \\ n_8 \\ n_9 \end{pmatrix} &\sim \text{EM} \left(C(t), \begin{pmatrix} e^{-\gamma_C \Delta t} \\ s(t)^{-1} h(1 - e^{-\gamma_C \Delta t}) \\ s(t)^{-1} (1 - h)(1 - e^{-\gamma_C \Delta t}) \end{pmatrix} \right) \\
 \begin{pmatrix} n_H \\ n_{10} \\ n_{11} \end{pmatrix} &\sim \text{EM} \left(H(t), \begin{pmatrix} e^{-\gamma_H \Delta t} \\ (1 - m(t))(1 - e^{-\gamma_H \Delta t}) \\ m(t)(1 - e^{-\gamma_H \Delta t}) \end{pmatrix} \right),
 \end{aligned}$$

where n_X is the number of individuals that remain in each compartment X ; $f(t)$ is the force of infection at time t ; $\gamma_L, \gamma_{I_a}, \gamma_{I_{su}}, \gamma_{I_{sd}}, \gamma_C$, and γ_H are the transition rates out of each respective compartment; $q(t)$ is the time varying probability of case detection; $s(t)$ is a time varying factor determining time to diagnosis; a is the fraction of infected individuals that are symptomatic; h is the fraction of detected symptomatic individuals that are hospitalized; and $m(t)$ is the time varying fraction of hospitalized cases that are fatal. EM stand for the Euler Multinomial process described by He et al. [29] with the rates (e.g., $e^{-f(t)\Delta t}$) described and stochastic noise descriptions excluded. See He et al. [29] for more details.

The force of infection $f(t)$ was modeled as:

$$f(t) = \omega(t) \times S(t) \times (I_{sd}(t) + I_{su}(t) + b_L L + b_{I_a} I_a(t) + b_C C(t) + b_h H(t)),$$

where b_L, b_{I_a}, b_C , and b_H are relative transmissibility factors for each respective compartment. The time-dependent transmission rate is a function of mobility, a latent trend, and process noise as:

$$\omega(t) = \frac{\beta}{N} \psi(t) \phi(t) \Gamma(t),$$

where N is the constant total population size of the state, $\psi(t)$ is relative human mobility at time t , and $\phi(t)$ is the latent trend of relative transmission strength, specified as the spline function

$$\text{logit}(\phi(t)) = \sum_{i=1}^K g_i \xi_{i_t},$$

where K is the number of basis functions, $\mathbf{g}$ is the vector of spline coefficients, and ξ is the matrix of basis functions. The number of basis functions, K , was defined as the number of days in each time-series divided by 21 (one function every 21 days [3 weeks]). Process noise $\Gamma(t)$ was modeled as gamma-distributed white noise (temporally uncorrelated noise) with mean 1 and variance σ^2 [30].

Time-varying detection probability ($q(t)$) was modeled as a Hill function starting at 10% ($q_{\min}$) on day 0 and increasing to a maximum possible of 40% ($q_{\max}$), reaching that half way point of 25% on the 30th day (q_{half}) since the first case notification:

$$q(t) = q_{\min} + \frac{q_{\max} \times q_r^t}{q_r^{q_{\text{half}}} + q_r^t},$$

where q_r is the rate of increase of the Hill function, which is set to 1.1, and t is the day since first case notification (see Section A of S1 Appendix).

Time-varying decrease in time-to-diagnosis ($1/s(t)$ days) was similarly defined as:

$$s(t) = 1 + \frac{s_{\max} \times s_r^t}{s_r^{s_{\text{half}}} + s_r^t},$$

where $s_{\max}$ is the maximum decrease in days to diagnosis, s_{half} is the day on which the decrease in days to diagnosis is half-way between 0 and $s_{\max}$ (set to day 30), and s_r is rate of increase of Hill function, set to 1.1 (see Section A of S1 Appendix). Note that $s(t)$ is multiplied by $\gamma_{I_{sd}}$ to increase the diagnosis rate, and that γ_C is divided by $s(t)$ to slow the rate from C to H , increasing the amount of time in the C compartment to account for reduced time in the I_{sd} compartment.

Time-varying mortality fraction ($m(t)$) was also specified as a Hill function:

$$m(t) = m_{\text{base}} + \frac{m_{\text{min}} \times m_r^t}{m_r^{m_{\text{half}}} + m_r^t},$$

where m_{base} is the baseline fraction of hospitalizations that result in death, m_{min} is the minimum fraction of hospitalizations that result in death, m_{half} is the days since first case notification at which $m(t)$ is halfway between m_{base} and m_{min} , and m_r is the Hill function rate of change, fixed to 1.

We calculated the effective reproduction number ($\mathcal{R}_e(t)$) using the next-generation matrix approach [14]. Assuming that all of the time dependent functions, S , ω , s , q , are changing slowly over the course of an individual's infection, the expected number of new cases each case generates, $\mathcal{R}_e(t)$, may be approximated as:

$$\begin{aligned} \mathcal{R}_e(t) \approx & S(t) \cdot \omega(t) \cdot 4 \cdot [b_L/\gamma_L + (1-a) \times \\ & (q(t)(b_{I_{sd}}/(s\gamma_{I_{sd}}) + b_C s/\gamma_C + hb_H/\gamma_H) + \\ & (1-q(t))b_{I_{su}}/\gamma_{I_{su}}) + ab_{I_a}/\gamma_{I_a}]. \end{aligned}$$

This equation may be arrived at by assuming that all time-varying parameters are fixed to the value at time t over the course of an infection and multiplying the expected residence time in each infectious stage by the transmission rate, the number of susceptible individuals, and the probability of a newly infected individual experiencing it.

We assumed that new, daily case reports arise from a negative binomial distribution whose central tendency is captured by the flow of individuals from I_{sd4} to C_1 . We denote this quantity by C_{new} , which accumulates over the course of one day in the simulation model and resets to zero at the end of each day (the model is simulated at a time step of 1/20 days). Similarly, we assumed that new, daily death reports arise from a negative binomial distribution whose central tendency is captured by the flow of individuals from C_4 to D . We denote this quantity by D_{new} , which accumulates over the course of one day in the simulation model and resets to zero at the end of each day. Then, for both new cases and deaths, we modeled the observation process as:

$$\begin{aligned} \text{cases}(t) & \sim \text{NB}(C_{\text{new}}(t), \theta_C) \\ \text{deaths}(t) & \sim \text{NB}(D_{\text{new}}(t), \theta_D) \end{aligned}$$

where θ_C and θ_D are the negative binomial (NB) dispersion parameters for cases and deaths, respectively. Note that $\text{cases}(t)$ and $\text{deaths}(t)$ are the observed number of cases or deaths reported on day t .

Model fitting

We fit the model using Maximization by Iterated particle Filtering (MIF). Observations (daily case and death reports) were modeled as arising from a negative binomial reporting process (see above). We estimated 24 to 26 parameters for each state: the baseline fraction of hospitalized cases that result in death (m_{base}), the minimum fraction of hospitalized cases that result in death (m_{min}), the number of days since first case notification at which $m(t)$ is half-way between m_{base} and m_{min} (m_{half}), a parameter accounting for extra-demographic process noise (σ), two negative binomial dispersion parameters (θ_c and θ_d), the initial size of the latent and infectious classes ($I(t=0)$, $I_a(t=0)$, $I_{su}(t=0)$, $I_{sd}(t=0)$), and 14 to 16 coefficients (g_i for $i \in (1, 2, 3, \dots, 14(16))$) specifying the latent trend spline. The number of B-spline coefficients depended on the length of the time series for each state. The estimated latent trend multiplied by the relative

Table 2. Fixed model parameters.

Parameter definition	Parameter symbol	Value	Source
Reproduction number	$\mathcal{R}_0$	7	[31]
Baseline transmission rate of symptomatic individuals	β	$\mathcal{R}_0 \times 0.1 \times \frac{1}{N}$	Assumption
Fraction of infected individuals that are asymptomatic	a	0.18	[32]
Fraction of diagnosed cases that are hospitalized	h	0.12	[33]
Relative transmissibility of L to $I_{su/sd}$	b_L	0.12	[34]
Relative transmissibility of I_a to $I_{su/sd}$	b_a	0.5	[34]
Relative transmissibility of C to $I_{su/sd}$	b_C	0.27	Assumption
Relative transmissibility of H to $I_{su/sd}$	b_H	4.5×10^{-5}	[35]
Duration of time in L stages	$1/\gamma_L$	4 days	[34]
Duration of time in I_a stage	$1/\gamma_{I_a}$	3.5 days	[34]
Duration of time in I_{su} stages	$1/\gamma_{I_{su}}$	6 days	[34]
Duration of time in I_{sd} stage	$1/\gamma_{I_{sd}}$	$0.5 \times 1/\gamma_{I_{su}}$	Assumption
Duration of time in C stages	$1/\gamma_C$	$0.5 \times 1/\gamma_{I_{su}}$	Assumption
Duration of time in H stages	$1/\gamma_H$	6 days	[36]
Minimum detection probability	$q_{\min}$	0.1	Assumption
Maximum detection probability	$q_{\max}$	0.4	Assumption
Day at which detection probability is halfway between $q_{\min}$ and $q_{\max}$	q_{half}	30	Assumption
Rate of increase from $q_{\min}$ to $q_{\max}$	q_r	1.1	Assumption
Maximum factor by which diagnosis speed increases	$s_{\max}$	1.0	[37]
Day at which diagnosis speed-up factor (s) is halfway between 0 and $s_{\max}$	s_{half}	30	Assumption
Rate of increase from 0 to $s_{\max}$	s_r	1.1	Assumption
Initial size of susceptible pool	$S(t = 1)$	N_{state}	Assumption

<https://doi.org/10.1371/journal.pcbi.1011610.t002>

mobility covariate multiplied by $\mathcal{R}_0$ (fixed to 7 for all states) specifies the time-varying transmission rate. All other parameters were fixed at the values reported in Table 2.

Initial size of the susceptible compartment was set as each state's population size minus the number of individuals in other the other compartments that we fix. It is true that the initial size of the susceptible pool will also decrease based on the number of individuals estimated to be in the latent infections compartment at $t = 1$. However, given the small size of the latent and infectious compartments relative to total population size, and the fact that total population size is a point estimate with error, we assume that our simple approach of setting $S(t = 1)$ to each state's population size is valid. For each state, we considered $t = 1$ as the date on which the first case is reported in that state.

We used the IF2 algorithm [38] implemented in the R [27] package **pomp** version 2.7.1.0. [28] to perform MIF. To initialize IF2, we generated 100 parameter sets from a range of parameter values using a Sobol sequence sampling design (Table 3). We then performed two rounds of MIF, each for 100 iterations with 3,500 particles and geometric cooling. For the first round of MIF we set `cooling.factor` = 1.0. For the second round of MIF, which continues from where the first round stopped, we set `cooling.factor` = 0.9. The log likelihoods of the parameter sets following MIF were calculated as the log of the mean likelihoods of 20 replicate particle filters with 5,000 particles each. At this stage, we collected all parameter sets within 10 negative log likelihood points of the maximum and sampled a new collection of 100 parameter sets, with sampling weighted in proportion to the negative log likelihood of the parameter set. All parameters in the new set were perturbed as: $p_{\text{new}} \sim \text{Normal}(p, |p \times 0.25|)$. The perturbed parameter sets were then used to initialize two final rounds of MIF, each run

Table 3. Estimated parameters and starting ranges for MIF estimation procedure. The expit function refers to back-transforming the parameter from the logit scale, which was used for estimation.

Parameter definition	Parameter symbol	Start range
Baseline fraction of hospitalizations that result in death	m_{base}	[expit(-6), expit(6)]
Minimum fraction of hospitalizations that result in death	m_{min}	[expit(-6), expit(6)]
Day at which death fraction is halfway between m_{base} and m_{min}	m_{half}	[exp(-5), exp(5)]
Extra-demographic process noise	σ	[exp(-5), exp(5)]
Case reporting dispersion	θ_c	[exp(-5), exp(5)]
Death reporting dispersion	θ_d	[exp(-5), exp(5)]
Initial size of latent compartment	$L(t=0)$	[exp(0), exp(10)]
Initial size of asymptomatic infectious compartment	$I_a(t=0)$	[exp(0), exp(10)]
Initial size of undetected infectious compartment	$I_{su}(t=0)$	[exp(0), exp(10)]
Initial size of detected infectious compartment	$I_{sd}(t=0)$	[exp(0), exp(10)]
B-spline coefficients	g_i	[-10, 10]

<https://doi.org/10.1371/journal.pcbi.1011610.t003>

for 50 iterations with 3,500 particles. Cooling factors were 1 in the penultimate MIF round and 0.5 in the final MIF round. At this stage, we assume the parameter set with highest log likelihood is the MLE.

Following [4], we calculated smoothed posterior estimates of all time-varying states using replicate particle filtering. Specifically, we ran 500 replicate particle filters with 2,500 particles at the MLE, retaining one randomly sampled complete particle trajectory from each particle filter. The 500 trajectories generated a set of 500 smoothed posterior draws of all time-varying state variables and parameters. $\mathcal{R}_e(t)$ was calculated at each time t for each of the 500 smoothed posteriors using the equation presented above, yielding a smoothed posterior distribution for $\mathcal{R}_e(t)$. We used the median of the posterior distribution of $\mathcal{R}_e(t)$ for all analyses. The estimated, time-varying latent trend was calculated using the MLEs for the B-spline coefficients.

Correlation analysis

We calculated pairwise cross correlations of the time series of relative mobility and estimated latent trends (median of ψ) among all locations (states) using Pearson's correlation coefficient. The correlations were estimated for both phases of the pandemic (see main text) using the `cor()` function in R.

Variance partitioning

We split the time series of $\mathcal{R}_e$ (effective reproduction number), ψ (latent transmission trend), and ϕ (relative human mobility) for each state into two periods: pre-April 24, 2020 and post-April 24, 2020 (see main text). For each period and each state, we partitioned the variance of $\mathcal{R}_e$ over time among ψ and ϕ . We used the median of the smoothed posterior distributions for ψ and $\mathcal{R}_e$. We fit two linear regressions to explain $\mathcal{R}_e$ over time: an intercept only model and a full model with the mobility trend and the latent trend as additive covariates: $\mathcal{R}_e \sim \text{mobility} + \text{latent}$. Models were fitted using the `lm()` function in R. We extracted the sum of squared residuals for each model parameter using the `anova()` function in R. Proportion of variance explained by each covariate was calculated as the sum of squared errors associated with each covariate divided by the sum of squared errors of the null model.

Supporting information

S1 Appendix. Additional model details and results. S1 Appendix comprises five sections: A) graphical displays of the detection probability ($q(t)$) and diagnosis time ($1/s(t)$) functions; B) graphical display of mean absolute scaled errors for each state for new case and new death reports; C) time series of incident case and death reports for each state with model-estimated filtered trajectories overlaid; and D) time series of mobility, estimated latent trend, and effective reproduction number for each state.
(PDF)

Acknowledgments

This study was supported in part by resources and technical expertise from the Georgia Advanced Computing Resource Center, a partnership between the University of Georgia's Office of the Vice President for Research and Office of the Vice President for Information Technology. We thank Shan-Ho Tsai for assistance with high-performance computing.

Author Contributions

Conceptualization: John M. Drake, Andreas Handel, Éric Marty, Eamon B. O'Dea, Andrew T. Tredennick.

Data curation: John M. Drake, Andreas Handel, Tierney O'Sullivan, Giovanni Righi, Andrew T. Tredennick.

Formal analysis: Éric Marty, Eamon B. O'Dea.

Investigation: John M. Drake, Andreas Handel, Eamon B. O'Dea, Tierney O'Sullivan, Giovanni Righi, Andrew T. Tredennick.

Project administration: John M. Drake.

Software: Andreas Handel, Andrew T. Tredennick.

Supervision: John M. Drake.

Visualization: John M. Drake, Éric Marty, Andrew T. Tredennick.

Writing – original draft: John M. Drake, Andrew T. Tredennick.

Writing – review & editing: John M. Drake, Andreas Handel, Éric Marty, Eamon B. O'Dea, Tierney O'Sullivan, Giovanni Righi, Andrew T. Tredennick.

References

1. Bertozzi AL, Franco E, Mohler G, Short MB, Sledge D. The challenges of modeling and forecasting the spread of COVID-19. *Proceedings of the National Academy of Sciences*. 2020; 117(29):16732–16738. <https://doi.org/10.1073/pnas.2006520117> PMID: 32616574
2. Castle JL, Doornik JA, Hendry DF. THE VALUE OF ROBUST STATISTICAL FORECASTS IN THE COVID-19 PANDEMIC. *National Institute Economic Review*. 2021; 256:19–43. <https://doi.org/10.1017/nie.2021.9>
3. Reiner RC, Barber RM, Collins JK, Zheng P, Adolph C, Albright J, et al. Modeling COVID-19 scenarios for the United States. *Nature Medicine*. 2021; 27(1):94–105. <https://doi.org/10.1038/s41591-020-1132-9>
4. Fox SJ, Lachmann M, Tec M, Pasco R, Woody S, Du Z, et al. Real-time pandemic surveillance using hospital admissions and mobility data. *Proceedings of the National Academy of Sciences*. 2022; 119(7): e2111870119. <https://doi.org/10.1073/pnas.2111870119> PMID: 35105729

5. O'Dea EB, Drake JM. A semi-parametric, state-space compartmental model with time-dependent parameters for forecasting COVID-19 cases, hospitalizations and deaths. *Journal of The Royal Society Interface*. 2021; 19(187):20210702.
6. Moreland A, Herlihy C, Tynan MA, Sunshine G, McCord RF, Hilton C, et al. Timing of State and Territorial COVID-19 Stay-at-Home Orders and Changes in Population Movement—United States, March 1–May 31, 2020. *Morbidity and Mortality Weekly Report*. 2020; 69(35):1198–1203. <https://doi.org/10.15585/mmwr.mm6935a2> PMID: 32881851
7. Castillo RC, Staguhn ED, Weston-Farber E. The effect of state-level stay-at-home orders on COVID-19 infection rates. *American Journal of Infection Control*. 2020; 48(8):958–960. <https://doi.org/10.1016/j.ajic.2020.05.017> PMID: 32461066
8. Kraemer MUG, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*. 2020; 368(6490):493–497. <https://doi.org/10.1126/science.abb4218> PMID: 32213647
9. Childs ML, Kain MP, Harris MJ, Kirk D, Couper L, Nova N, et al. The impact of long-term non-pharmaceutical interventions on COVID-19 epidemic dynamics and control: the value and limitations of early models. *Proceedings of the Royal Society B: Biological Sciences*. 2021; 288(1957):20210811. <https://doi.org/10.1098/rspb.2021.0811> PMID: 34428971
10. Nouvellet P, Bhatia S, Cori A, Ainslie KEC, Baguelin M, Bhatt S, et al. Reduction in mobility and COVID-19 transmission. *Nature Communications*. 2021; 12(1):1090. <https://doi.org/10.1038/s41467-021-21358-2> PMID: 33597546
11. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*. 2020; 20(5):533–534. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1) PMID: 32087114
12. Covid-19 Social Distancing Scoreboard—Unacast; 2022. Available from: <https://www.unacast.com/covid19/social-distancing-scoreboard>.
13. Fischer CB, Adrien N, Silguero JJ, Hopper JJ, Chowdhury AI, Werler MM. Mask adherence and rate of COVID-19 across the United States. *PLOS ONE*. 2021; 16(4):e0249891. <https://doi.org/10.1371/journal.pone.0249891> PMID: 33852626
14. van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*. 2002; 180(1):29–48. [https://doi.org/10.1016/S0025-5564\(02\)00108-6](https://doi.org/10.1016/S0025-5564(02)00108-6) PMID: 12387915
15. Badr HS, Du H, Marshall M, Dong E, Squire MM, Gardner LM. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases*. 2020; 20(11):1247–1254. [https://doi.org/10.1016/S1473-3099\(20\)30553-3](https://doi.org/10.1016/S1473-3099(20)30553-3) PMID: 32621869
16. Lodge EK, Schatz AM, Drake JM. Protective population behavior change in outbreaks of emerging infectious disease. *BMC Infect Dis*. 2021; 21(1):577. <https://doi.org/10.1186/s12879-021-06299-x> PMID: 34130652
17. Cramer EY, Ray EL, Lopez VK, Bracher J, Brennen A, Castro Rivadeneira AJ, et al. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. *Proc Natl Acad Sci U S A*. 2022; 119(15):e2113561119. <https://doi.org/10.1073/pnas.2113561119> PMID: 35394862
18. Banerjee S, Lian Y. Data driven covid-19 spread prediction based on mobility and mask mandate information. *Applied Intelligence*. 2022; 52(2):1969–1978. <https://doi.org/10.1007/s10489-021-02381-8> PMID: 34764603
19. Gibson GC, Reich NG, Sheldon D. Real-Time Mechanistic Bayesian Forecasts of Covid-19 Mortality. medRxiv; 2020. Available from: <https://www.medrxiv.org/content/10.1101/2020.12.22.20248736v2>.
20. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science*. 2020; 369(6500):eabb9789. <https://doi.org/10.1126/science.abb9789> PMID: 32414780
21. Alimohamadi Y, Taghdir M, Sepandi M. Estimate of the Basic Reproduction Number for COVID-19: A Systematic Review and Meta-analysis. *Journal of Preventive Medicine and Public Health = Yebang Uihakhoe Chi*. 2020; 53(3):151–157. <https://doi.org/10.3961/jpmph.20.076> PMID: 32498136
22. Hasan A, Susanto H, Kasim MF, Nuraini N, Lestari B, Triany D, et al. Superspreading in early transmissions of COVID-19 in Indonesia. *Scientific Reports*. 2020; 10(1):22386. <https://doi.org/10.1038/s41598-020-79352-5> PMID: 33372191
23. Salomon JA, Reinhart A, Bilinski A, Chua EJ, La Motte-Kerr W, Rönn MM, et al. The US COVID-19 Trends and Impact Survey: Continuous real-time measurement of COVID-19 symptoms, risks, protective behaviors, testing, and vaccination. *Proceedings of the National Academy of Sciences*. 2021; 118(51):e2111454118. <https://doi.org/10.1073/pnas.2111454118> PMID: 34903656

24. Clark NJ, Wells K. Dynamic generalised additive models (DGAMs) for forecasting discrete ecological time series. *Methods in Ecology and Evolution*. 2022; n/a(n/a).
25. Zhou T, Ji Y. Semiparametric Bayesian inference for the transmission dynamics of COVID-19 with a state-space model. *Contemporary Clinical Trials*. 2020; 97:106146. <https://doi.org/10.1016/j.cct.2020.106146> PMID: 32947047
26. Hurtado PJ, Kiro Singh AS. Generalizations of the 'Linear Chain Trick': incorporating more flexible dwell time distributions into mean field ODE models. *Journal of Mathematical Biology*. 2019; 79(5):1831–1883. <https://doi.org/10.1007/s00285-019-01412-w> PMID: 31410551
27. {{R Core Team}}. R: A Language and Environment for Statistical Computing.; 2020. Available from: <https://www.R-project.org/>.
28. King AA, Nguyen D, Ionides EL. Statistical Inference for Partially Observed Markov Processes via the R Package pomp. *Journal of Statistical Software*. 2016; 69:1–43. <https://doi.org/10.18637/jss.v069.i12>
29. He D, Ionides EL, King AA. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *Journal of The Royal Society Interface*. 2010; 7(43):271–283. <https://doi.org/10.1098/rsif.2009.0151> PMID: 19535416
30. Bretó C, Ionides EL. Compound Markov counting processes and their applications to modeling infinitely over-dispersed systems. *Stochastic Processes and their Applications*. 2011; 121(11):2571–2591. <https://doi.org/10.1016/j.spa.2011.07.005>
31. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2—Volume 26, Number 7—July 2020—Emerging Infectious Diseases journal—CDC. *Emerging Infectious Disease Journal—CDC*. 2020;26(7).
32. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance*. 2020; 25(10):2000180. <https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180> PMID: 32183930
33. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*. 2020; 20(6):669–677. [https://doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7) PMID: 32240634
34. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*. 2020; 368(6490):489–493. <https://doi.org/10.1126/science.abb3221> PMID: 32179701
35. Moghadas SM, Shoukat A, Fitzpatrick MC, Wells CR, Sah P, Pandey A, et al. Projecting hospital utilization during the COVID-19 outbreaks in the United States. *Proceedings of the National Academy of Sciences*. 2020; 117(16):9122–9126. <https://doi.org/10.1073/pnas.2004064117> PMID: 32245814
36. Ohsfeldt RL, Choong CKC, Mc Collam PL, Abedtash H, Kelton KA, Burge R. Inpatient Hospital Costs for COVID-19 Patients in the United States. *Advances in Therapy*. 2021; 38(11):5557–5595. <https://doi.org/10.1007/s12325-021-01887-4> PMID: 34609704
37. Zhang J, Litvinova M, Wang W, Wang Y, Deng X, Chen X, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *The Lancet Infectious Diseases*. 2020; 20(7):793–802. [https://doi.org/10.1016/S1473-3099\(20\)30230-9](https://doi.org/10.1016/S1473-3099(20)30230-9) PMID: 32247326
38. Ionides EL, Nguyen D, Atchadé Y, Stoev S, King AA. Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proceedings of the National Academy of Sciences*. 2015; 112(3):719–724. <https://doi.org/10.1073/pnas.1410597112> PMID: 25568084