Journal of the Royal Statistical Society Series A: Statistics in Society, 2024, **00**, 1–26
https://doi.org/10.1093/jrsssa/qnae040

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society Series A Statistics in Society

Original Article

A Bayesian latent variable model for the optimal identification of disease incidence rates given information constraints

Robert Kubinec[1], Luiz Max Carvalho[2], Joan Barceló[1], Cindy Cheng[3], Luca Messerschmidt[3] and Matthew Sean Cottrell[4]

[1]Social Science Division, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates
[2]School of Applied Mathematics, Getulio Vargas Foundation, Rio de Janeiro, Brazil
[3]Hochschule für Politik at the Technical University of Munich (TUM) and the TUM School of Governance, Munich, Germany
[4]University of California Riverside, Riverside, USA

Address for correspondence: Robert Kubinec, Social Science Division, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates. Email: rmk7@nyu.edu

Abstract

We present an original approach for measuring infections as a latent variable and making use of serological and expert surveys to provide ground truth identification during the early pandemic period. Compared to existing approaches, our model relies more on empirical information than strong structural forms, permitting inference with relatively few assumptions of cumulative infections. We also incorporate a range of political, economic, and social covariates to richly parameterize the relationship between epidemic spread and human behaviour. To show the utility of the model, we provide robust estimates of total infections that account for biases in COVID-19 cases and tests counts in the U.S. from March to July of 2020, a period of time when accurate data about the nature of the SARS-CoV-2 virus was of limited availability. In addition, we can show how sociopolitical factors like the Black Lives Matter protests and support for President Donald Trump are associated with the spread of the virus via changes in fear of the virus and cell phone mobility. A reproducible version of this article is available as an Rmarkdown file at https://github.com/CoronaNetDataScience/covid_model.

1 Introduction

The COVID-19 pandemic has led to a significant increase in disease modelling as the demands of a worldwide emergency spurred substantial innovation. Accurately modelling COVID-19 and similar diseases is no simple feat, however, due to multiple forms of selection and measurement bias that are difficult to overcome. The approach we present in this article differs from existing work by conceptualizing the relative level of infections as a time-varying latent variable and applying Bayesian techniques to obtain a posterior distribution over likely estimates. Our model's framework is similar in spirit to election forecasting models that produce latent estimates of candidate support from biased time-varying information, i.e. polls (Linzer, 2013; Lock & Gelman, 2010). Instead of polling data, we directly incorporate serological and expert survey estimates into the model in a way that allows us to identify the scale of the latent variable without having to make extensive assumptions about transmission dynamics.

This approach has two advantages for applied research. First, the minimal set of assumptions makes the model more robust than traditional compartmental models by incorporating as much uncertainty as possible about the infection rate. Existing approaches based on the SIR/

Received: April 1, 2022. **Revised:** March 1, 2024. **Accepted:** March 31, 2024

SEIR framework often require multiple hard-coded values due to the large number of dynamic compartments that must be estimated. When there is a lack of quality empirical information about disease transmission dynamics, such as in the early pandemic period, these assumptions can be difficult to justify. By employing a Bayesian model with informative priors, we can obtain estimates that have realistic uncertainty about the state of the world while also incorporating the best available prior information.

For example, in one of the earliest studies of COVID-19 transmission in the Wuhan area of China by Kucharski et al. (2020), the authors had to use strongly informative prior distributions for the 'incubation period of COVID-19 cases' (5.2 days, SD 3.7), the 'delay from onset to isolation' (6.1 days, SD 2.5), the 'delay from onset to reporting' (6.1 days, SD 2.5), as well as additional assumptions about the contact mixing rate and travel between Chinese regions. We note that these are assumptions required by the model, and even with clever efforts to collect data on all of these variables, such as estimating infection duration from the *Diamond Princess* cruise ship COVID-19 outbreak in February of 2020 (Riou et al., 2020), there are clear limitations to relying on assumptions about disease transmission and reporting dynamics without a large base of empirical evidence.

Second, conceptualizing the infection rate as a latent variable within a generative modelling framework permits us to examine covariate relationships in a manner that is more nuanced than traditional regression modelling. As we show in this article, it is possible for us to employ a wide range of covariates that predict the infection rate and consequently dramatically reduce the uncertainty of our estimates without having to assume any prior relationship between covariates and the infection rate (also known as the attack rate). Second, we are able to examine mediation relationships between covariates and the infection rate, which permits us to be more specific about the pathways through which covariates may be associated with infections. Especially given how difficult it is to achieve causal identification with observational epidemiological data, we believe that employing causal graphs to test for plausible mediation relationships can improve our understanding of potential disease transmission dynamics.

To apply the model, we estimate the infection rate in the U.S. from March to July of 2020, a crucial time in the pandemic as substantial uncertainty existed about how to respond to the disease (Perra, 2021). This uncertainty manifested itself in rapidly changing patterns of human behaviour and also political conflict as leaders diverged over their understanding of the disease's threat. While we have increasing evidence about the relationship between partisan identities and individual beliefs about COVID-19 (Fan et al., 2020; Grossman et al., 2020), data about partisan identity, along with other social and economic covariates, are rarely included in efforts to model and predict the spread of the disease in the population (Haug et al., 2020). As a result, we have difficulty understanding the causal pathways through which political and social variables affect the course of the pandemic (Gadarian et al., 2022), and in turn are affected by it. In this article, we disaggregate the effects of partisanship and demographic factors on disease outcomes by examining the way that these variables are mediated by individual mobility, reported masking and fear of the virus. Even when we cannot make exclusive claims of causal identification, we can still learn in much more detail about time-varying associations between covariates and the disease and the plausible pathways through which beliefs affected actions.

We show with this model that political partisanship in the U.S. had a very strong association with the spread of the pandemic by increasing or reducing people's fear of the virus and also by changing their mobility patterns. A 2% increase in a state's vote share for U.S. President Donald Trump in 2016 is associated with a 0.5% to 0.7% increase in a state's infections mediated by unsafe changes in mobility patterns.

We also find evidence that in-person political activity is positively associated with the spread of the disease, with states that saw a 1-SD increase in social justice protests following the death of George Floyd witnessing an increase in infections as high as 0.7% over time, although only among states that witnessed continuous protest activity. On the other hand, we do not find that the protests reduced people's fears of the disease or noticeably changed mobility patterns, suggesting that the spread of the disease happened solely through increased personal contact at the protests and subsequent chains of transmission rather than by changing behavioural patterns over long term.

2 Background

As more and more data have become available on observed case counts of the SARS-CoV2 coronavirus, there have been increasing attempts to infer how contextual factors like government policies, partisanship, and temperature affect viral spread (Brzezinski et al., 2020; Carleton & Meng, 2020; Dudel et al., 2020; Sajadi et al., 2020; Tasnim et al., 2020). The temptation to make inferences from the observed data, however, can result in misleading conclusions. Modelling approaches that fully account for disease dynamics like the SIR/SEIR specifications are very powerful but also require more information than is known about disease progression in the population—especially in its early stages—requiring researchers to rely on assumptions that are difficult to know with complete confidence (Ferguson et al., 2020). For this reason, in this article we present a retrospective Bayesian model that can adjust for testing bias by estimating the unobservable infection rate up to an unidentified constant. Furthermore, by incorporating informative priors based on serological and expert surveys of infection prevalence, it is possible to put an informative prior on the unobserved infection rate and estimate both recent disease trends and the association of covariates with the historical spread of the disease. By employing a fully Bayesian approach, we are able to allow our uncertainty about this prior information to propagate in the model, ensuring that we are not over-confident in our predictions of the latent infection rate.

An additional advantage of our method is to permit us to make more precise statements about how important theoretical covariates for the spread of disease may or may not be associated with the latent infection rate. A vast and expanding literature documents connections between many political, economic, and social factors with human behaviour related to the COVID-19 pandemic (Abouk & Heydari, 2021; Adolph et al., 2021; Allcott, Boxell, Conway, Gentzkow, et al., 2020; Ashraf, 2020; Barceló et al., 2022; Barceló & Sheen, 2020; Bo et al., 2021; Brauner et al., 2020; Courtemanche et al., 2020; Dave et al., 2020; Fellows et al., 2020; Haug et al., 2020; Islam et al., 2020; Sebhatu et al., 2020; Sharma et al., 2021; Zheng et al., 2020). While existing studies have shown these associations primarily through surveys and other individual-level analyses, it is difficult to test whether these factors jointly have any effect on COVID-19 infections. The reason for this difficulty is due to how these variables affect human behaviour in general equilibrium. For example, non-pharmaceutical interventions (NPIs) like stay-at-home orders have been associated with reduced infections, but stay-at-home orders were also implemented in a rapidly changing environment as public health policies, new suppression practices like masking and the health of the economy varied. People faced myriad influences on their choices during the pandemic, and even if we have a strong reason to believe that a certain factor should influence their behaviour, estimating that effect when many other contravening and contrasting factors were likely at play is challenging.

At the same time, estimating these general equilibrium effects even within the limitations of available data is very important to learn what factors are associated with the spread of COVID-19 in realistic conditions. For example, some argued that masking would lead to increased infections because it would reduce concern over the risk of infection (Abaluck et al., 2020). Evaluating this hypothesis ultimately requires general equilibrium analysis as it involves competing influences on human behaviour. In other words, is the moral hazard of being falsely protected a greater threat than the positive benefits of reducing infections via masking? Being able to sort, rank, and understand socio-economic, political, and healthcare-related factors behind the disease's spread is crucial to better understand why and how COVID-19 overwhelmed countries' disease control systems even when we lack a means of causal identification.

In this article, we seek to address these questions by collecting a set of important covariates, implementing models to adjust for bias in COVID-19 data and employ mediation analysis to understand the pathways that covariates affect the spread of the pandemic. We believe that pathway analysis allows us to uncover meaningful associations that do not obfuscate different time-varying processes. While causal identification in the pandemic is a non-trivial endeavour, employing models that can more realistically evaluate available data is arguably the best path forward.

With this model, we are also able to address important empirical debates about the sociopolitical factors behind the spread of the virus such as political partisanship. Political scientists have investigated to what extent partisanship has inhibited preventive measures against the COVID-19

pandemic as President Trump argued against public health policies like face masks (Gadarian et al., 2022). Research has already shown that Republicans are less likely than Democrats to practice public health behaviours like hand washing (Gadarian et al., 2020), to practice social distancing (Allcott, 2020; Andersen, 2020; Painter & Qiu., 2020), and to comply with policies targeted against COVID-19 (Fan et al., 2020; Grossman et al., 2020).

While partisanship in favour of President Trump and the Republican party has received the most attention, other types of political mobilization have also come under scrutiny. Of particular note were the protest movements against police brutality that spread across the U.S. in the summer and fall of 2020. Existing research suggests the protests have not had an adverse effect on COVID-19 infections (Dave et al., 2020), though the finding is again limited by the measurement bias we describe later. As such, it is clear that political motivations on both the left and the right have been associated with reduced compliance with COVID-19 precautions, though it is not clear through which potential pathways these variables could be affecting disease outcomes.

3 Methods

Fitting models that can realistically model disease trends during the early pandemic period when data are limited can be quite difficult. To address this crucial problem, we present a new Bayesian latent variable model that has a similar aim as epidemiological disease-tracking models in that it is designed explicitly to model disease dynamics. However, our model is a significant simplification of the compartmental models employed by epidemiologists to study viruses, and in particular SARS-CoV2 (Brzezinski et al., 2020; Carleton & Meng, 2020; Dudel et al., 2020; Ferguson et al., 2020; Li et al., 2020; Perkins et al., 2020; Riou et al., 2020; Sajadi et al., 2020; Tasnim et al., 2020). These models suppose different classes (compartments) of individuals in the population, denoted S for susceptible, I for infectious, and R for removed (other compartments may be added, such as E for exposed).

While these models are a powerful expression of the progress of a disease in the population, these models often struggle to provide robust estimates when relying on limited and possibly biased data about disease transmission dynamics. COVID-19 data, especially in the early pandemic period, had serious flaws, including limited testing and under-reporting of hospitalizations and deaths (Bertozzi et al., 2020; Larremore et al., 2021; Moein et al., 2021; Sánchez-Romero et al., 2021). When such data are unavailable, modellers can compensate by simulating plausible random values or using informative prior distributions, but this makes the model estimates tied to the particular set of values used (Grinsztajn et al., 2021). As a result, the challenges in the estimation of compartmental models with empirical data restrict the ability to interpret covariate adjustment. Any estimated associations are tied to the particular values used to identify the models, and it is not always possible to marginalize over all possible (or even a reasonably broad range) of hard-coded values via simulations.

By contrast, this article endeavours to estimate a more limited quantity than each dynamic measure of infected, recovered, and susceptible persons. We believe that many researchers and the general public often only want to learn about what has already happened, or the *empirical* infection rate (also called the attack rate in the epidemiological literature). For a number of time points $t \in T$ since the outbreak's start and states $c \in C$, we aim to identify the following quantity:

$$f_t\left(\frac{I_{ct}}{S_{ct} + R_{ct}}\right),$$

where I_{ct} denotes the number infected with SARS-CoV-2 at time t, and S_{ct} and R_{ct} denote those who remain susceptible to the virus and those who have either died or recovered. In our model, we collapse S_{ct} and R_{ct} to a single quantity—those who are not infected—so we can focus exclusively on identifying I_{ct}.

However, even with this simplification, we do not have estimates of the actual infected rate I_{ct}, only positive COVID-19 cases a_{ct} and numbers of COVID-19 tests q_{ct} due to the aforementioned measurement issues. Given this limitation, the aim of the model is to backwards infer the infection rate I_{ct} as a latent process given observed test and counts. Modelling the latent process is necessary

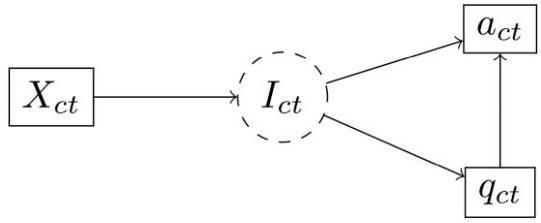


Figure 1. Directed acyclic graph showing confounding of covariate X_{ct} on observed tests q_{ct} and cases a_{ct} due to unobserved infection rate I_{ct}. This figure shows the relationship between a covariate X_{ct} representing a policy or social factor influencing the infection rate I_{ct}. Because the infection rate I_{ct} influences both the number of reported tests q_{ct} and reported cases a_{ct}, any regression of a covariate X_{ct} on the reported data will be biased. Latent variables are shown as circles and observed variables as rectangles.

to avoid bias in using only observed case counts as a proxy for I_{ct}. The reason for this is shown in Figure 1 in which a covariate X_{ct}, such as a stay-at-home order, is hypothesized to affect the infection rate I_{ct}. Unfortunately, increasing infection rates can cause both increasing numbers of observed counts a_{ct} and tests q_{ct}. As more people are infected, more tests are likely to be done, which will increase the number of cases independently of the infection rate. As a result, due to the back-door path from the infection rate I_{ct} to case counts a_{ct} via the number of tests q_{ct}, it is impossible to infer the association of X_{ct} on I_{ct} from the observed data alone without modelling the latent infection rate.

To estimate the process in Figure 1, we assume that the unobserved state-specific cumulative infection rate I_{ct} can be modelled as a time-varying beta-distributed random variable with a mean parameter $\mu \in (0, 1)$ and shape parameter $\phi > 0$. We employ the beta distribution for this parameter because the cumulative infection rate must lie between 0 and 1 (i.e. a proportion) once the pandemic has started.

It is important to understand how the model incorporates time. We estimate one cumulative infection rate for each state for each time point t. In order to take into account time dependence, we include a shared third-order polynomial time trend that is a function of the number of post-outbreak time periods $T_O < T$, where an outbreak begins at the first reported case in a given state. In other words, we expect the residual time process to continue at the same rate across states once the pandemic begins in a given state.

We employ a cubic time function based on theoretical considerations. In our model, the polynomial represents the rate of infection increase in the absence of any other covariates, or equivalently the *counterfactual* rate of infections. We know from the SIR/SEIR simulations that, in the absence of any countervailing measures, epidemics occur in ever-increasing waves until the herd immunity threshold is reached, although the curve is unlikely to be symmetric as a quadratic function would require. As such, we employ this function because it represents a credible baseline for what the epidemic would do if no other factors impeded its spread.

For the same reason, we use a shared polynomial function because, during this early pandemic period, we know that the states are experiencing infections from the same virus. As such, we expect that the counterfactual or residual trajectory of the pandemic to be the same across states. If we were to allow this time trend to vary across states in a multilevel structure, we would mistakenly absorb the effect of covariates as heterogeneity in the viral strain.

We define the conditional distribution of the unobserved infection rate I_{ct} as:

$$\Pr(I_{ct} \mid t = T) \sim \text{Beta}(\mu\phi, (1 - \mu)\phi) \tag{1}$$

$$\mu = g^{-1}(\alpha_1 + \beta_{I1} t_o + \beta_{I2} t_o^2 + \beta_{I3} t_o^3 + \beta_C X_{ct}). \tag{2}$$

This parameterization of the beta distribution in terms of μ and ϕ follows from the beta regression literature (Ferrari & Cribari-Neto, 2004) so that we can model the expected value $E[I_{ct}]$ directly via μ. As such, we use $g^{-1}(\cdot)$, the inverse logit function, to scale the linear model in μ to the (0, 1) interval. The three β_{Ii} are polynomial coefficients of the number of post-outbreak time periods t_o.

The parameter vector β_C represents the effect of independent covariate matrix X_{ct} on the latent infection rate. These are our main variables of interest and have effects in addition to the polynomial time trends. Finally, the parameter ϕ becomes a dispersion parameter which, intuitively, equals the effective sample size of the beta-binomial model. For each day t, ϕ is equal to the amount of information in the linear model about cases and tests expressed as the size of a random sample from the population.

Because we do not have measures of I_{ct}, we need to use the observed data, tests q_{ct} and cases a_{ct}, to infer I_{ct}. First, we propose that the number of infections will almost certainly increase the number of tests as states try to stop the disease's spread via surveillance. Second, we can assume that a rising infection rate is associated with a higher ratio of positive results (reported cases) conditional on the number of tests, that is, COVID-19 is causing positive test results. We model both of these observed indicators, tests and cases, jointly to simultaneously adjust for the infection rate's influence on both factors. It is this joint modelling that permits us to directly incorporate testing bias. In fact, our model learns about the infection rate from the absolute number of tests.

To model the number of tests, we assume that each state has an unobserved level of testing capacity, which increases at a non-linear rate during the course of the epidemic. We employ a quadratic function of testing capacity to express the concept of diminishing marginal returns. States were able to ramp up testing once PCR tests were approved by the FDA but faced constraints due to shortages of supplies, personnel, and labs. The cumulative number of observed tests q_{ct} for a given time point t and state c and as a fraction of the states' population, c_p, then has a binomial distribution:

$$q_{ct} \sim \text{Binomial}(c_p, g^{-1}(\alpha_2 + \beta_b I_{ct} + \beta_{cq1} L_t + \beta_{cq2} L_t^2)). \tag{3}$$

The parameters β_{cq1} and β_{cq2} represent the quadratic increase in testing capacity that varies by state c for each post-outbreak time point L_t. We similarly allow for partial pooling of these coefficients as testing capacity will show a limited level of variability across states. The parameter β_b then represents the independent contribution of the level of infections I_{ct} on the total number of requests demanded marginal of testing capacity. The intercept α_2 indicates how many tests would be performed in a state with an infection rate of zero and at time $t = 0$, and as such is likely to be very low.

The binomial model for the number of observed tests q_{ct} provides some information about I_{ct}, but not enough for useful estimates. We can learn much more about I_{ct} by also modelling the number of observed cases a_{ct} as another binomial random variable expressed as a proportion of the state population, c_p:

$$a_{ct} \sim \text{Binomial}(c_p, g^{-1}(\alpha_3 + \beta_a I_{ct})), \tag{4}$$

where $g^{-1}(\cdot)$ is again the inverse logit function, α_3 is an intercept that indicates how many cases would test positive with a cumulative infection rate of 50% (i.e. zero on the logit scale), and β_a is a scaling parameter that reflects how much information about case counts comes from the latent infection process. The multiplication of this parameter and the infection rate determines the cumulative number of cases, a_{ct}, as a proportion of the state population, c_p.

To summarize the model, infection rates determine how many tests a state is likely to undertake and also the number of positive tests (confirmed cases). This simultaneous adjustment helps takes care of mis-interpreting the observed data without taking into account varying testing levels, which has made it hard to generalize findings concerning the disease across health jurisdictions. It also allows us to learn the likely location of the infection rate conditional on what we observe in terms of tests and cases.

Because sampling from a model with a hierarchical Beta parameter can be difficult, we simplify the likelihood by combining the beta distribution and the binomial counts into a beta-binomial model for tests:

$$q_{ct} \sim \text{Beta-Binomial}(c_p, \mu_q \phi_q, (1 - \mu_q)\phi_q) \tag{5}$$

$$\mu_q = g^{-1}(\alpha_2 + \beta_b I_{ct} + \beta_{cq1} L_t + \beta_{cq2} L_t^2) \tag{6}$$

and cases:

$$a_{ct} \sim \text{Beta-Binomial}(q_{ct}, \mu_a\phi_a, (1-\mu_a)\phi_a) \tag{7}$$

$$\mu_a = g^{-1}(\alpha_3 + \beta_a I_{ct}), \tag{8}$$

where I_{ct} is now equal to the linear model shown in (2) and implicitly mapped to (0, 1) as a component of μ_a.

3.1 Identifiability

This model contains an unobserved latent process I_{ct}, and as such the model as presented is not identified from the data alone without further information. For example, the parameters that control the influence of the infection rate on tests and cases could increase and the latent infection rate could decrease without the probability of the observed data changing.

We take two further steps to identify this model that we believe represent very limited additional assumptions, especially compared to existing modelling approaches. First, we impose the assumption that I_{ct} is a non-decreasing quantity. The number of infected people cannot decrease in an epidemic without a significant virus mutation, but the model as expressed does not require that to be true. We can eliminate that possibility from the model by imposing an ordered constraint on I_{ct}:

$$I_{ct} = \begin{cases} I_{ct} & \text{if } t = 1 \\ I_{ct-1} + e^{I_{ct}} & \text{if } 1 < t < T. \end{cases} \tag{9}$$

This transformation forces I_{ct} to be no less than I_{ct-1}. At the same time, we do not need to impose any constraints on the covariates themselves, allowing us to sample those in an unconstrained space before we transform I_{ct}.

However, we also need some information about the empirical scale of testing bias to produce identified estimates of I_{ct}. We could do so by adding a prior to the model about the plausible range of total infections to reported cases, though we prefer to use information that is more precise. Our information about the possible level of infections comes from two sources. First, the Centers for Disease Control's serology surveys conducted during the pandemic represent an empirical way of relating I_{ct} to plausible estimates of infections at varying time points. We include a list of these surveys for the time period under study in the online supplementary material. Second, we incorporate expert survey data from McAndrew and Reich (2022) who surveyed epidemiologists in the early weeks of the pandemic to obtain their best estimates of the total level of infections. Helpfully, this survey provided a robust estimate of uncertainty by eliciting distributions over infections. Furthermore, this type of empirical data can be collected relatively rapidly, which increases this model's utility for future pandemics.

By including this information as informative priors, we also implicitly account for many of the variables explicitly parameterized in compartmental models such as reporting delays. Because we have an estimate of the number infected at time t that is independent of reported cases and tests, the model will find the parameter estimates that are most likely given the observed differences between the surveys and the reported data.

Because we model the infection rate as a cumulative count, we can directly parameterize this information in the model. For the serological surveys, given a state c and time point t for which we have survey information, we model the count of infected S^P_{ct} as a proportion of the total subjects in each serology survey S^N_{ct} with the Binomial distribution:

$$S_{ct} \sim \text{Binomial}(S^N_{ct}, g^{-1}(I_{ct})). \tag{10}$$

For the expert survey data, which is expressed as a distribution of proportions E_{ct}, we use the beta distribution to represent our uncertainty:

$$E_{ct} \sim \text{Beta}(S^N_{ct}, g^{-1}(I_{ct})\phi_E, (1 - g^{-1}(I_{ct}))\phi_E)), \tag{11}$$

where E_{ct} is the average expert estimate, and ϕ_E is an estimated shape parameter from the empirical distribution reported in McAndrew and Reich (2022). This empirical distribution is further stratified by state as the original estimates are for national totals; to do so we divide E_t by the proportion of cases and tests that a given state c had reported by that time t.

It is important that the serology and expert surveys enter the model in this fashion so that we can model the survey count stochastically and propagate our uncertainty from sample size and expert judgment through to our estimates of I_{ct}. This uncertainty matters as well because the serology surveys exhibit random noise and do not always increase over time, as can be seen in the online supplementary material. By modelling the relationship as a probabilistic one, we are making the weaker assumption that the infected rate is probably close to the serology estimate, but the two do not need to be the identical. The combined posterior estimates for I_{ct} will then be weighted with the case and test likelihoods to produce the most credible estimate of I_{ct}.

As we show in the online supplementary material with simulations, no other identification restrictions are necessary to estimate the model beyond weakly informative priors assigned to parameters.

These are:

$$\beta_a \sim \text{Normal}(0, 5), \tag{12}$$

$$\beta_{qci} \sim \text{Normal}(\mu_{qi}, \sigma_{qi}), \tag{13}$$

$$\sigma_{qi} \sim \text{Exponential}(100), \tag{14}$$

$$\mu_{qi} \sim \text{Normal}(0, 50), \tag{15}$$

$$\beta_C \sim \text{Normal}(0, 5), \tag{16}$$

$$\beta_{Ii} \sim \text{Normal}(0, 50), \tag{17}$$

$$\alpha_1 \sim \text{Normal}(0, 10), \tag{18}$$

$$\alpha_2 \sim \text{Normal}(0, 10), \tag{19}$$

$$\alpha_3 \sim \text{Normal}(0, 10), \tag{20}$$

where the Normal distribution is parameterized in terms of mean and standard deviation. The testing effect variance parameter σ_{qi} receives a relatively tight prior to reflect that while we expect there to be some variability in inter-state testing production, this variability should not explain all of the variation in tests as infection rates also play an important role.

We also extend this model in order to analyse the mediation of a subset of covariates X'_{ct} by adding mediators M_{ct} for mobility and F_{ct} for fear of the disease to the causal diagram as shown in Figure 2. Figure 2 has several paths due to the fact that the influence of covariates X_{ct} affects the two mediators differently. Given that beliefs and preferences precede actions, the covariates X'_{ct} first influence I_{ct} along the *ae* and *abd* path through perceptions of how dangerous the disease is. These beliefs both affect the chance of an individual getting infected and thus I_{ct} directly on the path *ae*, such as by causing an individual to adopt social distancing behaviours, and also on an indirect path *abd* by which an increase in a people's fear of the disease reduces mobility as people prefer to stay home.

In addition to pathways through the fear mediator F_{ct}, a causal factor like NPIs could influence infections along the pathway through mobility *ed* without increasing or decreasing fear. This

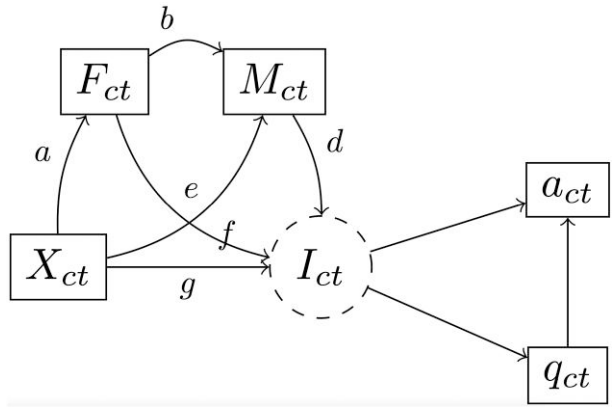


Figure 2. Directed acyclic graph for latent infection rate with mediators. This figure adds mediators M_{ct} (mobility data) and F_{ct} (fear of COVID-19) that mediate the relationship between state-level covariates X'_{ct} and the latent infection rate I_{ct}. Because beliefs precede actions, F_{ct} is causally prior to M_{ct} and can affect infections both via reducing mobility (path *abd*) and directly apart from mobility (path *ae*), such as by encouraging individuals to remain socially distant. Latent variables are shown as circles and observed variables as rectangles.

situation could arise if government policies forced people to stay at home against their will and despite their unconcern about the disease. Finally, a covariate could have an unmediated direct effect g on the infection rate. The total effect of a covariate X_{ct} on the spread of the disease is then the sum of all the paths, $abd + af + ed + g$. To calculate the indirect effects and direct effects given the use of the inverse logit function $g^{-1}(\cdot)$, we employ the chain rule as in Winship and Mare (1983) to calculate the marginal effect of covariates with respect to different pathways to I_{ct}.

Adding the mediators to the model does not require significant additional parametric assumptions as they can be included as Normal distributions (i.e. OLS regression) as in Yuan and MacKinnon (2009). It should be noted that there are in fact five mobility covariates as explained in the following section, and so we explicitly model the covariance in mobility via a multivariate Normal distribution with a covariance matrix parameter Σ_m.

To add our mediation covariates M_{ct} and F_{ct}, which we describe in more detail in the next section, we multiply the following likelihoods with the joint posterior:

$$M_{ct} \sim MVN(\alpha_m + \beta_m X'_{ct}, \Sigma_m) \tag{21}$$

$$F_{ct} \sim N(\alpha_f + \beta_f X'_{ct}, \sigma_f) \tag{22}$$

We also include all of M_{ct} and F_{ct} as linear predictors in (2).

We fit this model using Markov Chain Monte Carlo in the Stan software package (Carpenter et al., 2017). We run the sampler for 4,000 iterations (2,000 iterations used for warmup) and four independent chains to test for convergence.

4 Data

The only data required to fit the model, in addition to the covariates of interest and serology surveys, are observed cases and tests for COVID-19 by day. In this section, we fit the model to numbers of COVID-19 case counts on US states and territories provided by The New York Times. By doing so, we can use the differences in trajectories across states to help identify the effect of state-level covariates on the infection rate. We supplement these observed case counts with testing data by day from the COVID-19 Tracking Project. We then take the 7-day rolling average of both series to account for reporting fluctuations and weekly reporting effects.

We note that COVID-19 cases and deaths are available at the county level in the U.S. We do not use this reduced level of aggregation for two reasons. First, and most importantly, our aim is to better understand the mechanisms of COVID transmission, which requires us to have access to

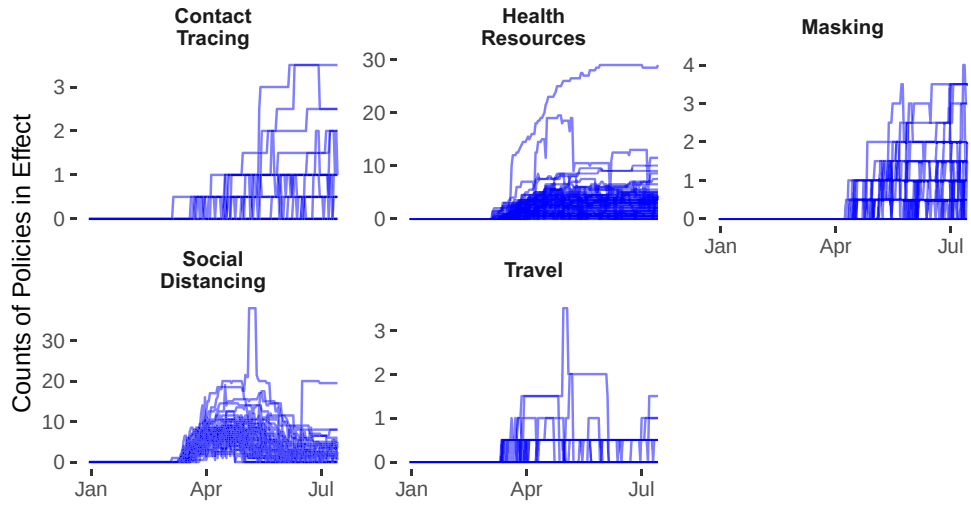


Figure 3. Aggregation of state-level policies in effect by day and by state from the COVID-AMP dataset.

daily polling data which is not available at the country level. Second, we note that what data are available is much more prone to measurement error due to issues with reporting that vary by county (Stoto et al., 2022). Aggregating to the state level can reduce this idiosyncratic measurement error and permit more stable inferences, especially when looking at day-to-day changes in these covariates.

To analyse the effect of suppression policies, we use data on counts of social distancing policies, restrictions on mass gatherings, restrictions on businesses, mandatory mask orders, restrictions on government services, and stay-at-home orders from the COVID AMP dataset (Katz et al., 2023). For each type of policy, we include a variable representing the count of policies in that category effective for a particular day. For each update to an existing policy, we code it as +1 if the update increases the scope of the policy or −1 if it decreases the scope of the policy (down to a minimum of 0). While this is a simplification of the underlying data, we are still able to capture relative complexity over time without having to make judgments about stringency or other qualitative criteria. We then interact these policy counts with a linear trend to examine time-varying policy effects. We separately include policies designed to increase health resources like personal protective equipment (PPE) and also policies requiring mask use as we do not examine time-varying effects of these covariates.

The policy data are plotted by state in Figure 3. As can be seen, there is a rise in policies after the pandemic begins in the middle of March, though the number of policies varies across categories. The count of policies is an admittedly imperfect measure though it communicates more information about policy activity than a binary coding. Generally speaking, states imposed many more policies designed to increase their access to PPE for health staff than they were willing to take on lockdowns, social distancing, and restrictions on businesses and government services. This difference likely has to do with the increased cost and salience of these policies vis-a-vis relatively less politically difficult options like gathering more masks and face shields for health care workers (Cheng et al., 2020).

To better understand over-time factors that may also affect COVID-19, we include polling data from Civiqs and YouGov at the state level. From Civiqs, we include state-level polling averages by day for the percentage of respondents favouring Trump, percentage reporting the economy is 'very good', and the percentage reporting that they are 'extremely concerned' about the coronavirus. From YouGov, we use a poll from May 8th reporting average number of respondents who said

they used masks by U.S. state. As this poll does not vary over time, we set the mask prevalence at one-half the minimum value of the poll prior to the WHO's revision of guidance concerning wearing masks on April 3rd, and equal to the poll's values thereafter. As described in the previous section, the poll asking respondents whether they are 'extremely concerned' about COVID-19 represents our fear mediator, and is also included as a separate outcome with other covariates as predictors.

To better understand the mediating effects of suppression policies, we include Google mobility data[1] for retail, residential, parks, workplaces, transit, and retail establishments. These estimates are by day and aggregated to the state level. They are measured in terms of an index that is initialized with a value of 100 at the index start on 15 February 2020. To test for mediation, we include these as predictors of the infection rate, and separately fit a likelihood with each mobility covariate as an outcome and the other covariates as predictors.

We note that it is important to measure mediation for mobility because mobility is hypothesized to affect the spread of COVID-19 (Gao et al., 2020). As such, measuring the simultaneous effect on mobility for covariates in our model is important as the covariates could be affecting mobility, which subsequently affects COVID-19 spread. Ignoring this association would result in post-treatment bias that deflates the effect of predictors in the model, though our main interest in including these variables is because this mediation is substantively interesting to decompose.

To measure protest activity, we include a covariate reflecting the proportion of a state's population engaged in social justice protests following the death of George Floyd on 25 May 2020. These data are drawn from publicly available information about the number and size of protests from three online sources: Wikipedia protest data, the Count Love protest web-crawling web site,[2] and list of protests compiled by Ipsos.[3] For protests present in only one of the three sources, we used information on both size and location. If a protest was present in three sources, we averaged reported protest size. If the sources had contradictory information about the type of protest, we had research assistants re-code the protest using secondary sources. For protests for which size was not available, we imputed missing data using random forest algorithms (Stekhoven & Bühlmann, 2012).

All time-varying covariates—polling, protests, policies, and mobility data—are lagged by 14 days to account for the likely delay in events showing up in reported cases. This 14-day lag comes from the epidemiology literature and is meant to take into the account the amount of time required for people to be infected, be tested and then have the test results reflected in case counts.

We further add in non-varying state-level data on Donald Trump's vote share for the 2016 election from the MIT Election Lab, a 2019 estimate of state GDP from the Bureau of Economic Analysis, the 2018 percentage of foreign born residents, population under 18 years of age and population density from the U.S. Census Bureau, 2019 state-level average data on air pollution,[4] cardiovascular deaths per capita, percentage of residents under age 18, number of dedicated health care providers, public health funding, and smoking rates provided by the United Health Foundation ('America's Health Rankings 2019 Report,' 2019). We include these variables to address several known factors that relate to vulnerability to infectious disease (i.e. population density, public health funding) and respiratory diseases in particular (air quality, smoking). All variables are standardized to permit comparability.

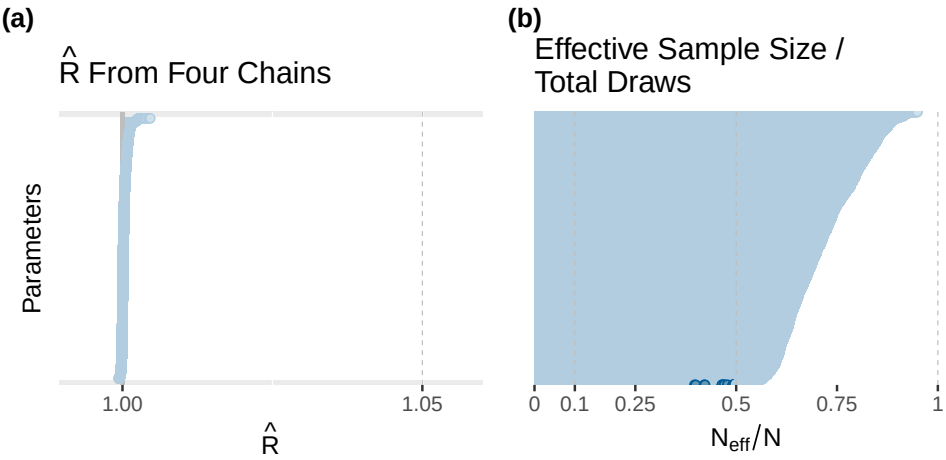
We note that in general we cannot make claims of causal identification as we can with our claims of statistical identification of the latent infection rate. COVID-19 is not a very likely candidate for meeting any kind of assumption about ignorable selection into treatment; it is a disease that is indirectly caused by human behaviour. What we are able to do is decompose variance into causally relevant pathways, but our results remain in the realm of association and are only identified insofar as our causal model is correct and our data are unbiased.

[1] See https://www.google.com/covid19/mobility/

[2] https://countlove.org/

[3] See https://www.ipsos.com/en-us/knowledge/society/Protests-in-the-wake-of-George-Floyd-killing-touch-all-50-states

[4] Defined as average exposure of the general public to particulate matter of 2.5 microns or less ($PM_{2.5}$) measured in micrograms per cubic metre (3-year estimate).



Panel A shows split−Rhat calculated from four independent Hamiltonian Markov Chains estimated with Stan for each model parameter. Values below 1.05 are considered to be converged. Panel B shows the ratio of effective samples (N_{eff}) to Markov Chain draws (N), which is an indicator of sampling efficiency. Values above 0.5 indicate that Markov samples are reasonably efficient.

Figure 4. Convergence diagnostics for MCMC chains with Stan.

4.1 Model convergence

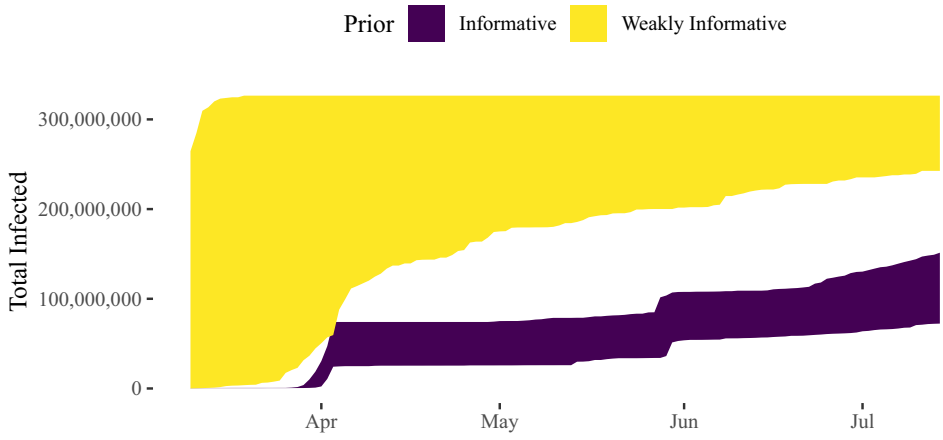
While this model has a fairly high level of complexity, our Stan Hamiltonian Monte Carlo sampler is able to reach a stationary distribution as evidenced by convergence metrics. Figure 4 shows strong convergence diagnostics based on four independent chains with 2,000 post-warmup iterations each. Figure 4a shows that all parameters had very low values of $\hat{R}$ calculated from the four chains, which suggests that the chains all converged to the same stationary distribution. Figure 4b suggests that sampling efficiency was quite high as the number of effective samples (N_{eff}) was 50% or higher as a proportion of the total samples N. Because we had 8,000 total post-warmup samples, this means that each parameter had an effective sample size (equivalent to Monte Carlo independent draws) of at least 4,000 draws. This sampling efficiency again suggests that the Stan sampler reached convergence and that we have a large enough sample size to obtain quality estimates even in the tails of the distribution.

4.2 Prior predictive distribution

We next show what the joint posterior distribution looks like when we sample only from the priors—that is, when we ignore any information in either the cases or tests data. The prior predictive distribution helps us understand the robustness of the priors—do they cover a wide enough range of possible infection rates—as well as the scope for our model to learn from the data. For our model, we can fit two kinds of prior distributions: a weakly informative prior distribution that samples solely from the independent parameter distributions shown in equations (12)–(20) and a strongly informative prior distribution that includes information from serology surveys and expert surveys about the total number of infections. We plot estimated cumulative national infections for both types of prior distributions in Figure 5.

As can be seen with the weakly informative priors, the possible infection rate at $t = 1$ is equal to anywhere from 0% of the U.S. infected up to 100% infected. The prior predictive distribution moves upward due to the ordered transformation imposed on the latent infection rates, but that is simply an artefact of the transformation. A priori, the weakly informative priors permit virtually any infection level at baseline and consequently any possible infection trajectory.

The informative prior, by contrast, shows very little infections until April, when infections rise considerably and then rise at a slower rate through July. Uncertainty with the informative priors is still quite high, with the final estimate in July allowing for between 50 million and 100 million



Plot shows the predicted infection rate for weakly informative priors (yellow) and informative priors incorporating serology surveys and expert survey data (purple). The width of the intervals indicate that weakly informative priors allow for all possible infection rates at baseline while informative priors restrict plausible estimates to between 0 and 100,000 infections depending on the day of the pandemic.

Figure 5. Prior predictive distributions.

infections in the U.S. at the time. We believe that this estimate is reasonable and admits a substantial amount of residual uncertainty. Importantly, this informative prior is expressed with uncertainty in terms of sample size for serology surveys and infection rate distributions for expert surveys, so the uncertainty in the infection rates reflects this propagated uncertainty. It is useful as well to know that we can estimate this model without any data from cases or tests and still obtain reasonable estimates of infection trajectories, albeit with substantial uncertainty.

Our presentation of these different priors shows the flexibility of Bayesian inference in being able to incorporate diverse sources of information while propagating uncertainty to final results. Theoretically, we want the priors to incorporate all available information during the early pandemic period. If we are able to do so, we know that Bayesian inference with these priors will be optimal in the sense that we cannot make inferences that are more precise or less biased (Zellner, 1988). We do have to have at least some confidence in the information we provide. We can incorporate measures of uncertainty, but if such information about disease trends is patently false, we would not want to include it in the model. However, absent such abjectly wrong priors, Bayesian inference shows how utilizing as much available prior information as possible can provide precise estimates even when the full picture of the disease outbreak is not known.

For these reasons, it is important to note that the use of informative priors is not only to identify the model—although this is required as the weakly informative priors cannot identify the scale of the latent variable—but also to permit the best inferences within an information-constrained environment. During an early pandemic period, very limited information may be available, though some of the prior information we employ, such as expert surveys, can be collected very quickly. So long as the prior information represents the best information available and adequate efforts have been made to capture uncertainty in this information, then including this information will necessarily result in better and less biased posterior inferences about disease trends. For these reasons, we believe the flexibility of this method allows it to be applied to a variety of fast-moving disease outbreaks, not just that of the COVID-19 pandemic.

5 Results

The results of drawing from posterior values for the beta-binomial distribution of cases and tests are shown relative to the original observed values in Figure 6 for five states. The plots show that although there is noise in the predictions (represented by the black shaded region), the model is

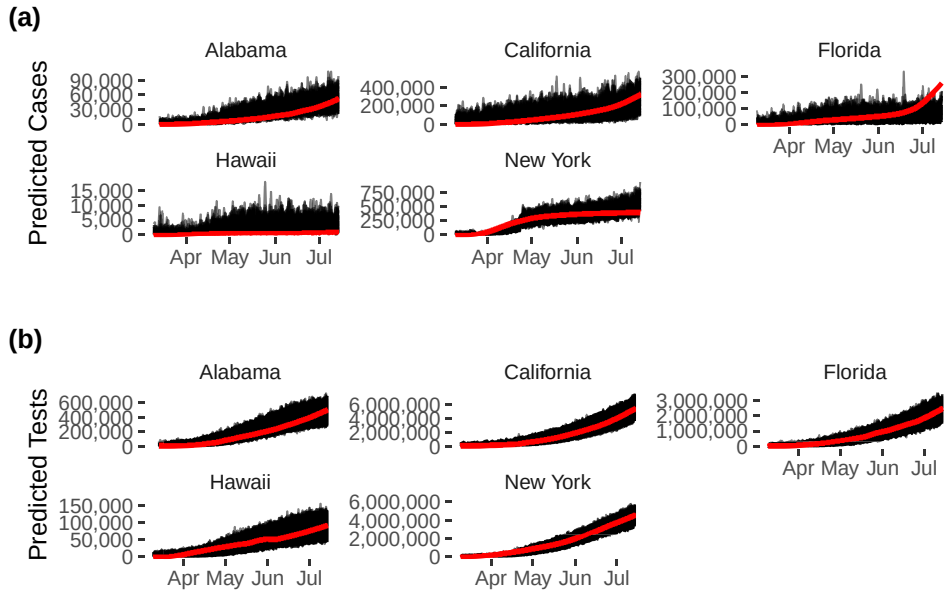


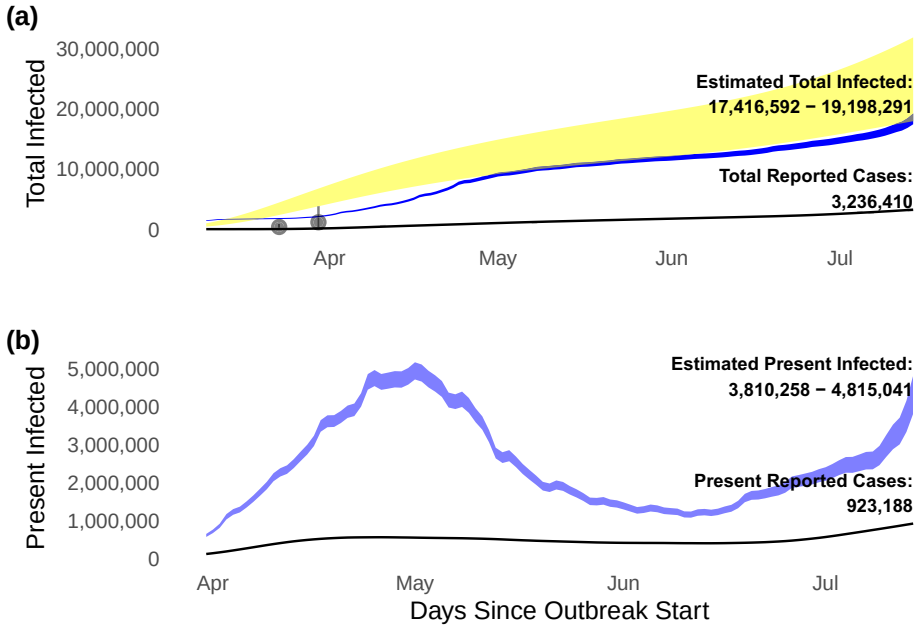
Figure 6. Predictive validity of model vis-a-vis observed cases and tests.

generally able to capture the empirical values (represented by a red line) with reasonable uncertainty. There is more uncertainty with the case rates than the test rates, but that is generally due to the fact that the case rates are much smaller and hence the relative uncertainty appears greater. Our posterior estimates of the dispersion parameter ϕ for cases is 1,175.9 UI (1,125.7, 1,227.5) and for tests is 584.9 UI (564.9, 605.9), which implies that our implicit sample size for estimating cases is roughly double that for tests. This difference is unsurprising as we would expect confirmed cases to have more information about infection rates than the total number of tests.

We next report the model's estimates of infected counts for the US population as a whole in Figure 7. Panel A in this plot shows the cumulative total both for reported cases (thin black line) and for the model's estimate of total infected (blue line). The interval in this plot, as with all figures presented, are the 5% and 95% quantiles of the empirical posterior distribution. As can be seen, the model estimates that there are approximately 5–6 times as many infected people in the U.S. as reported cases, with the total cumulative number of infected persons reaching approximately 18 million total infected with around 4 million presently infected as of mid-July. The expert survey estimates we incorporate in the model are shown as black points in panel A, showing that our estimates hew fairly closely to this expert survey information, although our estimates are somewhat higher at the very early stage of the pandemic relative to the expert estimates.

Comparing the trajectory of the blue ribbon posterior estimate in Figure 7 with the weakly informative (yellow) and informative prior infection (black) estimates in Figure 5 shows that adding in the cases and tests data substantially reduces uncertainty and also reveals much more of the latent trend in infections. In particular, the addition of the cases and tests data reveals a spike in the latent infection rate towards the end of the time series that our serology survey understandably did not identify. As such, comparing these figures reveals how adding in additional data and appropriately adjusting for bias can result in relatively precise estimates of a latent quantity.

We compare our estimates with a popular COVID-19 forecaster employing SEIR models from Gu (2020) by plotting their estimates as a yellow ribbon in Figure 7a relative to our posterior model estimates that are shown as a blue ribbon. As can be seen, the trajectories are similar although we



Blue 5% – 95% HPD intervals show estimated infected and the black line shows observed cases from the New York Times. These estimates are based on CDC seroprevalence data and a Bayesian model of how cases and tests are influenced by infection rates. Black dots in Panel A show early expert estimates of COVID–19 prevalence in the United States. Yellow ribbon shows 5% – 95% predicted cumulative infections from covid19–projections.com empirical model.

Figure 7. Total cumulative and present COVID-19 infections in the U.S.

show a somewhat slower and more non-linear rate of growth during the early pandemic period. On the whole it would seem that our estimate of infected individuals is on the conservative end compared to other approaches. Furthermore, our intervals are far more precise than other approaches, which is likely because we employ extensive covariate adjustment to better infer human behaviour during the course of the pandemic.

Panel B in the plot shows our estimates of infected individuals, except that it adjusts the cumulative number with a 19-day lag to account for the approximate time that recovery from COVID-19 requires (deaths are first subtracted). This plot displays an imperfect but useful formulation of the likely number of people infected at any given time point. As of 14 July, it would appear that there were approximately 4 million infected individuals in the U.S., while the number peaked at about 5 million in early May.

By comparison, Figure 8 shows the cumulative totals of estimated infections by state. Figure 8a in this figure has the count of infections by state, while Figure 8b shows the percentage of the population infected by state. Both the overall S-shape of the epidemic can be seen along with the substantial heterogeneity in infections, with early infected states like New York and New Jersey still in the top quartile of states with infections even though they successfully reduced the rate of disease spread. Figure 8b in the figure includes the serology survey estimates as blue dots, showing that these estimates map fairly close to our posterior estimates except for very high serology numbers. Our posterior estimates shrink these outlier serology surveys, such as an estimate of almost 15% infected for New York in mid-April, back towards the overall mean infection rate.

We examine the prior information to posterior estimates directly in Figure 9. On this plot, we show our posterior infection estimate on the y axis and the prior estimate (either expert survey or serology survey) on the x axis. The horizontal dotted line shows what an exact 1:1 relationship between these two infection estimates would look like. Figure 9a shows the relationship for the complete posterior with cases and tests data while Figure 9b shows this relationship for the prior

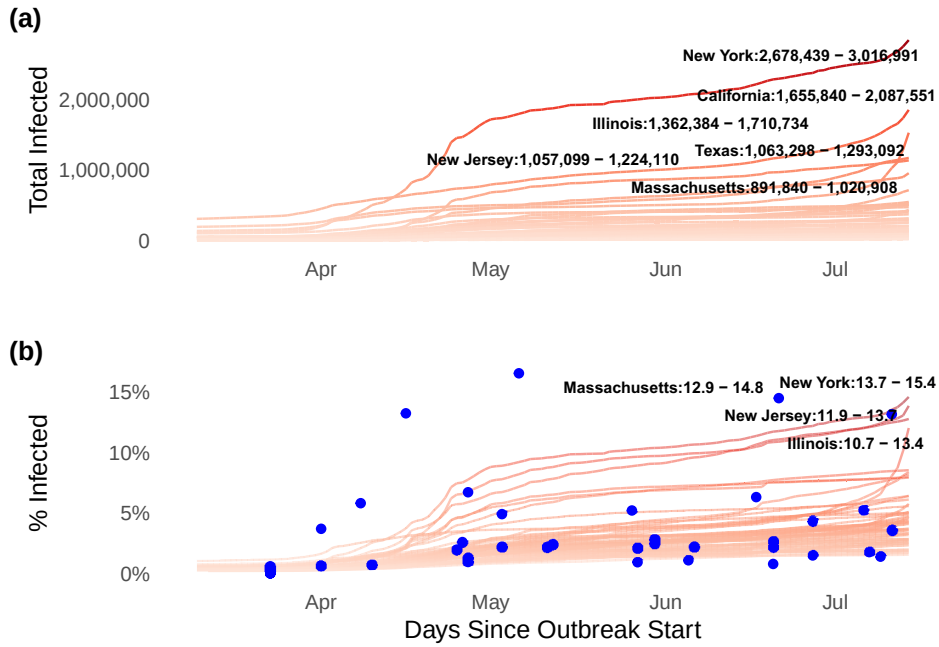


Figure 8. Average cumulative count of infected people by US state as of July 14th.

predictive distribution with the serology and expert survey information included. As can be seen, in Figure 9a, the posterior distribution diverges from the informative priors, with higher estimates of infections at lower levels of prior infection information and lower estimates of infections at higher levels of prior infection information. By contrast, panel B reveals a nearly exact 1:1 relationship between the informative priors and prior predictive distribution. These two panels show that the addition of information about cases and tests leads the model to shrink the estimates away from expert surveys and serology surveys that appear unlikely given the combined likelihood.

In addition to the estimation of the cumulative count of infected individuals, the model provides further useful information by parameterizing the relationship between the unobserved infection rate and the number of tests conducted in a given state. These individual parameters are shown in Figure 10. The scale of the *y* axis shows the number of people that a state was able to test relative to each person infected. The plot shows that those states that were worst-hit early in the pandemic were also those that under-counted infections because of testing (New York, New Jersey).

We would note that this information is also helpful to policy makers and others trying to make sense of observed case counts given the limitation in testing thus far. Our estimates help take into account these known biases and adjust them based on differences between states and within states in terms of disease trajectories. We believe this model can be used to help understand disease trends and factors associated with it even in the relatively data-poor environment that characterizes the initial period of a pandemic. Unlike SEIR/SIR approaches, we do not employ information about hospitalization and death reporting delays, the infection-age distribution, or initial seeds. While these other empirical observations can provide additional information about the progress and severity of the disease, they can also make inferences fragile when empirical data is of poor quality.

5.1 Covariates

To calculate the effect of covariates on the infection rate, we report here average cumulative marginal effects by state, i.e. by how much a given covariate increased the proportion infected for a

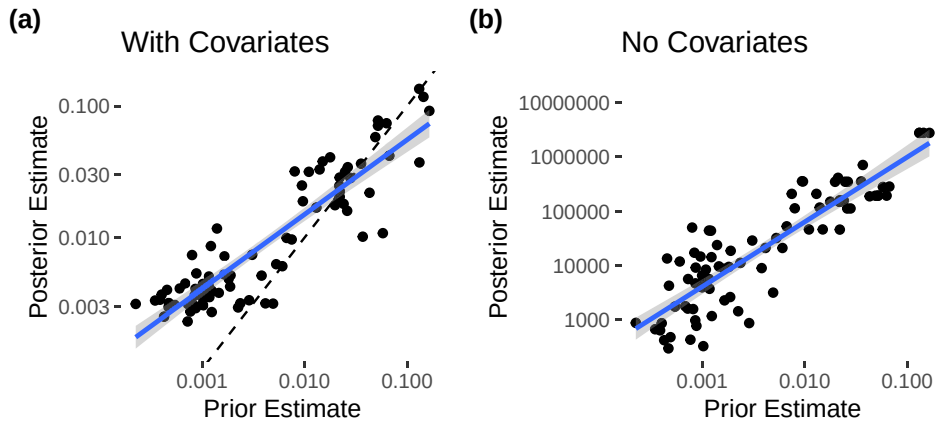


Figure 9. Comparison of posterior cumulative infection rates compared to priors on infection rates.

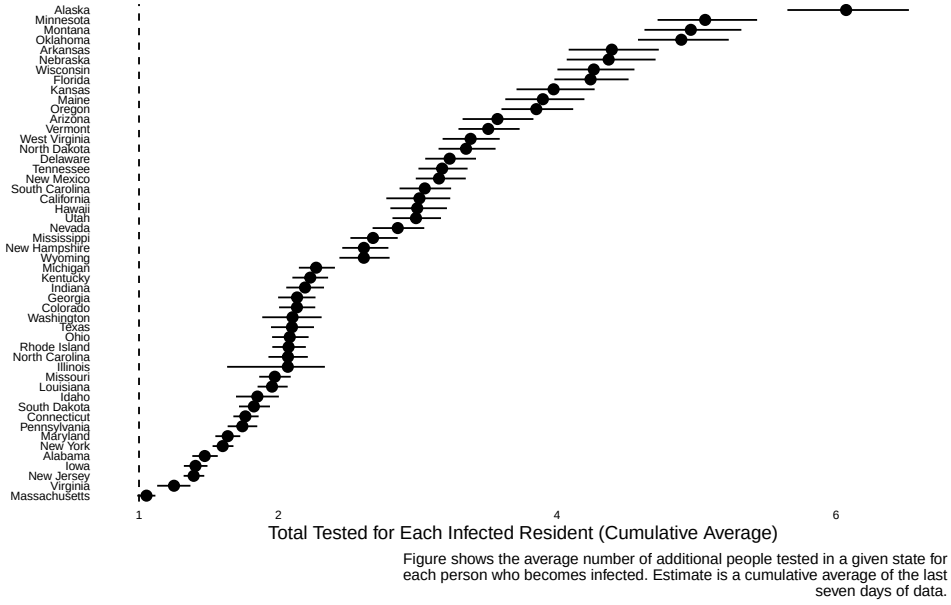


Figure 10. Measuring states' testing rates relative to infection rates.

one-unit increase in the covariate over time. We report cumulative marginal effects rather than the sample average marginal effect because the outcome monotonically increases, and so the marginal effect at any one point in time is not as meaningful a statistic. The way to interpret the coefficients presented is how a 1-unit (1-SD) change would affect the infection rate if that increase were sustained for an average state's entire time series (March to July). As such, these estimates represent ceilings of effects in terms of what a given state could experience.

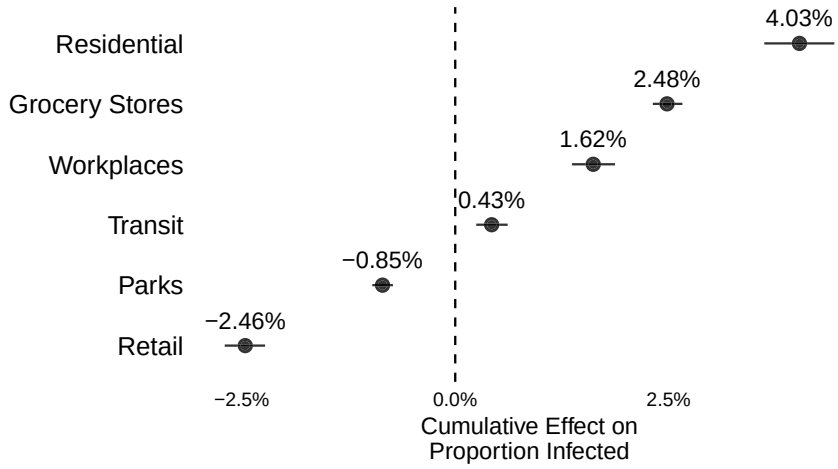


Figure 11. Effect of Google mobility data on COVID-19 spread.

We first show the association of mobility types with the infection rate. In Figure 11, we show the marginal effect of a 1-SD increase in different types of Google mobility on the infection rate expressed as a fraction of a state's population. In line with the growing research on cellphone mobility and the epidemic, there are strong positive effects of some types of mobility on the spread of the disease, especially residential, workplace, and grocery store mobility. Movement in parks and retail stores, on the other hand, is negatively associated with COVID-19 occurrence. While these results are somewhat surprising given that both residential and workplace mobility are very large, other results confirm with prior suspicions that outdoor activities like attending parks are relatively low-risk for COVID exposure. In fact, increased mobility in parks is associated with reduced infections, probably because it substitutes for more high-risk types of mobility.

To interpret these coefficients correctly, it is important to take into account the multivariate normal distribution that was used to model each of these mobility measures as one joint distribution. The residual correlations for the mobility measures model are shown in Figure 12. These correlations are intuitive, with transit positively correlated with other mobility measures except residential (people tend to be at home if they are not in transit). What is quite important is that workplace and residential mobility are strongly inversely correlated at −0.88; in other words, people tend to be at home if they are not working and vice versa. As a result, the association of residential and workplace mobility with COVID-19 is complicated due to this displacement effect. The fact that residential mobility is positively associated with infections once this displacement effect is taken into account accords with the modelling literature that warned that stay-at-home orders would paradoxically increase infections in the home as people were kept in close quarters with each other (Ferguson et al., 2020). We believe these strong correlations provide compelling evidence for employing the multivariate normal distribution in our model so that we do not assume these measures are conditionally independent. At the same time, it does render the interpretation of mediation effects somewhat more complicated as the model is explicitly taking into account that changes in one type of mobility are likely to displace or effect other types of mobility.

We next turn to an analysis of the rest of the covariates used to predict the latent infection rate. Figure 13 shows the marginal effect of all other covariates in the model on the latent infection rate expressed as average cumulative marginal effects. The estimates are further broken out in terms of mediation. The mobility effect is equivalent to the *ed* path in Figure 2, i.e. it is the path from the covariates to mobility that does not go through increased fear of COVID-19 measured by daily polls. The fear of COVID-19 pathway, on the other hand, is equivalent to the $abd + ae$ paths, or the sum of the path from fear through mobility and the path from fear to infections apart

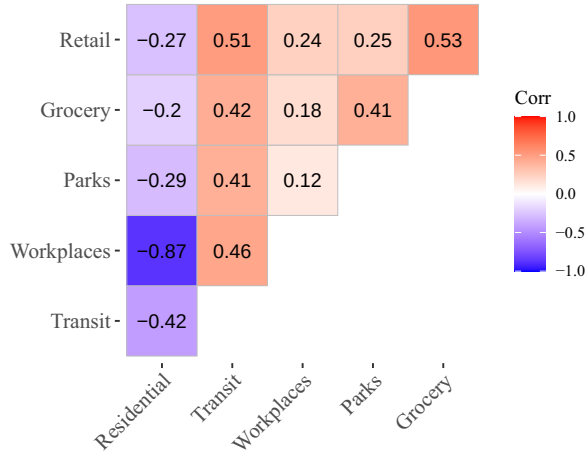


Figure 12. Estimated correlation of mobility measures.

from mobility. In other words, a covariate's effect mediated by increased fear of COVID-19 can both immediately impact the outcome by heightening sensitivity to the severity of the pandemic and affect the outcome by reducing an individual's willingness to engage in dangerous types of mobility. The direct effects, which represent the unexplained effect of covariates independent of either concern over COVID-19 or changes in mobility, are then equivalent to the *g* path in Figure 2, and the total effects are the sum of all paths. The direct and indirect effects are disaggregated in panel A while the total effects are shown in panel B.

The use of mediation analysis shows substantial heterogeneity in the types of associations and whether direct and indirect effects tend to complement or substitute each other. First, it is important to note that the single strongest associations in panel B come from the YouGov mask-wearing poll, the Civiqs concern over coronavirus poll, count of protests, and the relative level of a state's cardiovascular disease. As these are cumulative average marginal effects, that number reflects what an *average* state might experience; the effect could well be larger for states with higher infection rates than average. On the other hand, as these effects are cumulative, they reflect a state that experienced a sustained increase in the covariates and so it might overstate the effects somewhat.

On the whole, panel B would suggest that masking and concern over the coronavirus are leading indicators of infections given that these covariates were lagged by 14 days. In other words, people only changed their behaviour once they believed that the pandemic was about to worsen, and indeed their presuppositions were correct. However, this pattern suggests that increased masking and fear of the virus were unable to prevent waves of infection from occurring. Sadly, these empirical associations correspond with what we knew of the progression of the pandemic in the United States up until the arrival of effective vaccines.

Mediation analysis is helpful at understanding what may be driving these associations. We can learn more about the meaning of the results when we can identify effects through pathways which we have a theoretical reason to believe matter for fighting the epidemic: individual concern over COVID-19 and individual mobility. The COVID poll's positive association with the spread of infections appears via mobility: as people become more concerned about the pandemic, their mobility patterns change, which is a leading indicator of the spread of the disease. However, again the positive association shows that these precautionary changes were not sufficient to prevent outbreaks. Likewise, masking has a positive association with disease, but only via an association with increased fear of the virus. Reported masking has a clear negative *direct* association, which we can reasonably infer to be the beneficial effect of wearing masks that can prevent infection from occurring.

In contrast to other research, we find that social justice protests are positively associated with COVID-19 spread via a direct pathway, which could be evidence of the interpersonal contact brought on by the social movement. However, as we report cumulative marginal effects, it is unlikely that states experienced protests every day in the sample, suggesting that the reported

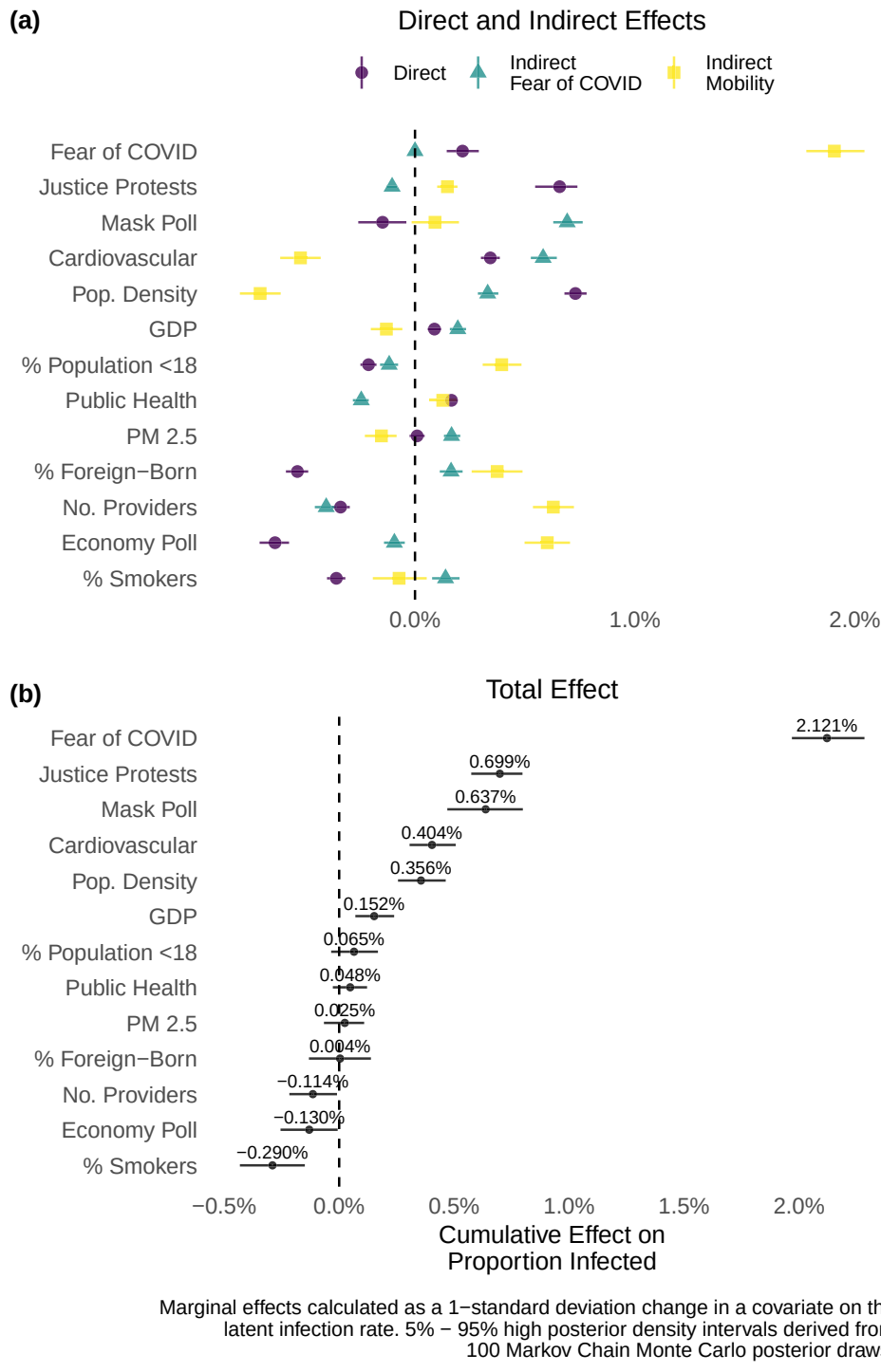


Figure 13. Marginal effects of covariates on latent infection rates for U.S. states.

association is more of an upper bound for what most states experienced. On the other hand, we do not find much evidence, as Dave et al. (2020) suggest, that the positive effect of the protests was offset by reduced mobility by non-protesters as the indirect effects are almost zero.

There are other interesting associations in Figure 11. States with more people with cardiovascular issues tended to see more infections due to a direct pathway, possibly vulnerability to the harmful effects of the coronavirus, but also have reduced infections via mobility patterns, suggesting that these states exhibited more caution with respect to infection vectors. Similarly, population density has a positive direct association with infections—possibly due to the increased crowding and risk of interpersonal contact—but likewise offsetting associations via safer mobility patterns. While we cannot infer causality, these patterns suggest that decomposing total effects can provide new hypotheses and exploratory findings that help us understand the complex and offsetting patterns in human behaviour.

The economy poll is another interesting case as it is weakly negatively associated with infections in terms of total effects, but it is strongly negatively correlated through a direct pathway. At the same time, improved economic perceptions are associated with increased infections via unsafe changes in mobility patterns. This result is theoretically interesting as trade-offs over the economy were often framed as a willingness to combat the epidemic versus the economic consequences of social distancing (Bonaccorsi et al., 2020). The empirical analysis shows that this trade-off may exist and that reduced infections is associated with improved perceptions of economic growth but also encouraging risky behaviours via mobility and interpersonal contact.

What is clear is that the strongest time-varying factors present in the model concern individual behaviour more than policies or state preparedness. Considering that the percentage of foreign residents (i.e. exposure to international travel) and risk of cardiovascular disease were fixed before the arrival of the pandemic, the most important manipulable factors are those involving beliefs, such as in the strength of the economy and the relative threat of COVID-19, along with personal behaviours like mask-wearing and attendance to protests.

We next turn to the prominence of partisanship variables in explaining the spread of the disease, which we did not include in the previous figures as we interacted Trump vote share and within-state changes in approval polls in our model. Instead, we explore this interaction graphically in Figure 14. In this figure, the effect of Trump 2016 vote share is plotted conditional on the relative level of daily Trump approval polling on the x axis. The effects are shown aggregated in panels A and B and disaggregated across mobility types in panel C. Panel A shows that in general, the effect of partisanship for Trump has both direct and indirect effects, with the direct effect highly conditional on the above/below polling average of approval for Trump in a given state (which has a maximum swing of about $+/-4$ pp). When Trump approval rose, states with high Trump vote share witnessed fewer infections later on. These high conditional associations are likely due to the rally-around-the-flag effect in which Trump's approval rating spiked when the epidemic first appeared in March and April, leading to an association between high approval levels and low infection counts in conservative areas of the country.

However, it is important to note opposite effects through the mediated pathways. Figure 14b and c shows that Trump vote share mediated through mobility is strongly positive in terms of infection counts. While the effects are not as large as the direct effects, they are still substantial. Trump vote share's effect on COVID-19 mediated through these important channels shows that pro-Trump states tend to implement social-distancing behaviours at lower rates, as previous research has shown, with consequent relative increases in infections. For the fear of COVID-19 path, rising Trump approval is associated with reduced infections, though this association does not vary with Trump vote share, which suggests that this simply reflects people's improved opinion of Trump when infections are relatively low (and vice versa when infections are high).

On the whole, this finding points to very strong associations between partisanship and the spread of the COVID-19, comparable or greater than the demographic and socio-economic factors in the model. States with higher Trump vote shares have seen significantly fewer infections via unexplained pathways, but very importantly, this decrease did not come through reduced mobility nor increased concern over COVID-19. The direct relationship is likely an artefact of the pattern of the early spread of the virus. After all, it is well-known that the early states that were infected with COVID tended to vote against Trump, although partisanship is not why they were more vulnerable to COVID initially. We believe that pro-Trump states received fortuitous outcomes by happening to not be on major travel routes from early COVID-19 hot spots; rising Trump approval in these states occurred as pro-Trump residents believed their president's dismissal of the virus' threat. In other words, the unexplained direct effect justified the relative

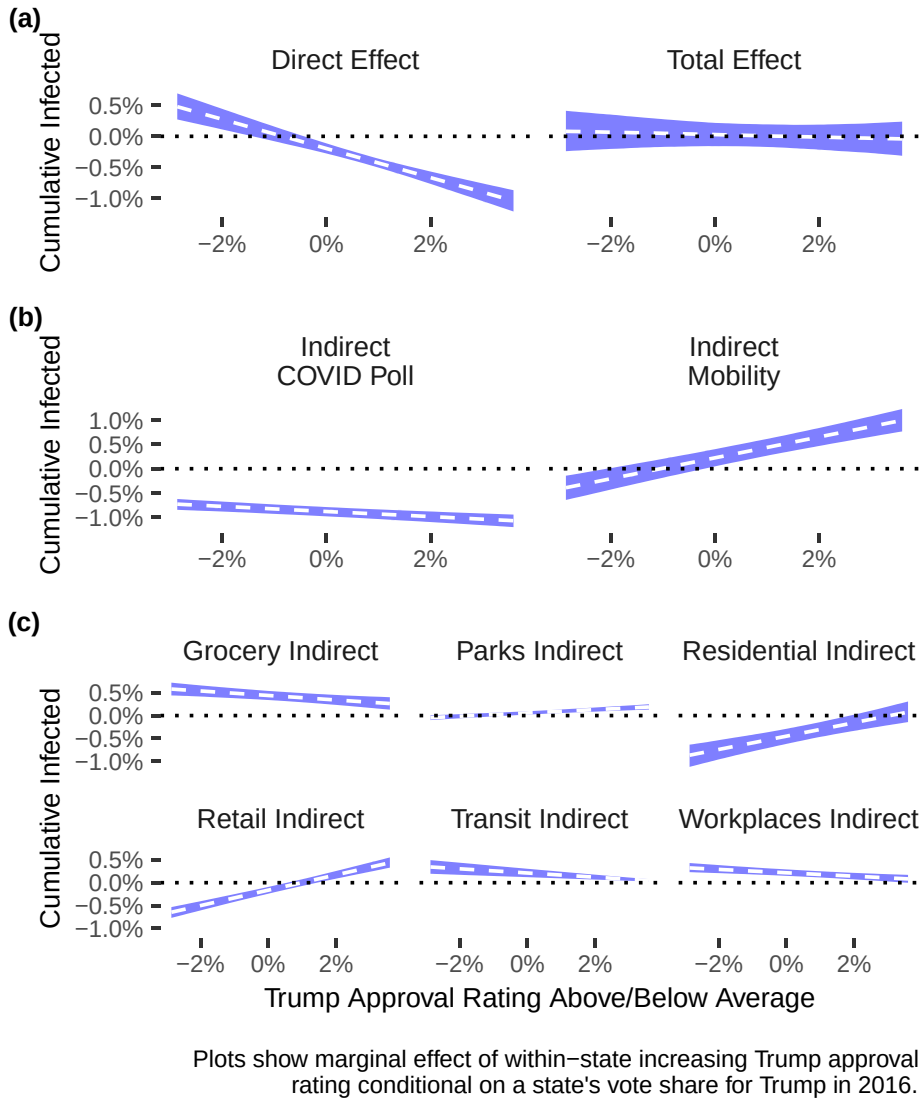


Figure 14. Marginal effects of Trump vote share in 2016 conditional on state approval polls.

inattention to important behaviours that could prevent infection. Given the increase in COVID-19 infections in the last 2 months in heavily Republican states, it would seem that this tendency would lead pro-Trump states to suffer in the long run as behaviour caught up with initial conditions.

6 Conclusion

Identifying the spread of a disease in the early period of a pandemic can be quite difficult, and the model presented in this article is a way to maximize the use of available information while also propagating uncertainty in the available evidence. Furthermore, the model permits sophisticated inference in terms of relating covariates to the spread of the disease, which allows us to disentangle possible theoretical pathways through which relevant risk factors might have played a role in the spread of the pandemic. At the same time, all of the data in this article are observational in nature and causality cannot be either inferred or assumed. Nonetheless, we believe that there is still a considerable amount of information that can be gleaned from

studying this period of the COVID-19 pandemic, especially as this early period coincides with very high risk as the virus is relatively poorly understood and treatment options are limited. We hope that this modelling framework will prove useful for quickly and accurately modelling the outbreak of future acute infectious diseases.

Conflict of interest: We have no conflicts of interest to disclose for this work.

Funding

We have no specific funding to acknowledge for this work.

Data availability

All data and code to reproduce this work are available in a Github repository at https://github.com/CoronaNetDataScience/covid_model.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series A*.

References

Abaluck J., Chevalier J. A., Christakis N. A., Forman H. P., Kaplan E. H., Ko A. I., & Vermund S. H. (2020). *The case for universal cloth mask adoption and policies to increase supply of medical masks for health workers* (SSRN 191). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3567438.

Abouk R., & Heydari B. (2021). The immediate effect of COVID-19 policies on social-distancing behavior in the United States. *Public Health Reports*, *136*(2), 245–252. https://doi.org/10.1177/0033354920976575

Adolph C., Amano K., Bang-Jensen B., Fullman N., & Wilkerson J. (2021). Pandemic politics: Timing state-level social distancing responses to COVID-19. *Journal of Health Politics, Policy and Law*, *46*(2), 211–233. https://doi.org/10.1215/03616878-8802162

Allcott H., Boxell L., Conway J., Gentzkow M., Thaler M., & Yang D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of Public Economics*, *191*, https://doi.org/10.1016/j.jpubeco.2020.104254

Allcott H., Boxell L., Conway J. C., Ferguson B. A., Gentzkow M., & Goldman B. (2020). What explains temporal and geographic variation in the early US coronavirus pandemic? https://doi.org/10.3386/w27965

"America's Health Rankings 2019 Report." (2019). United Health Foundation. https://www.americashealthrankings.org/learn/reports/2019-annual-report.

Andersen M. (2020). *Early evidence on social distancing in response to COVID-19 in the United States* (SSRN). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3569368.

Ashraf B. N. (2020). Economic impact of government interventions during the COVID-19 pandemic: International evidence from financial markets. *Journal of Behavioral and Experimental Finance*, *27*, 100371. https://doi.org/10.1016/j.jbef.2020.100371

Barceló J., Kubinec R., Cheng C., Rahn T. H., & Messerschmidt L. (2022). Windows of repression: Using COVID-19 policies against political dissidents? *Journal of Peace Research*, *59*(1), 73–89. https://doi.org/10.1177/00223433211062389

Barceló J., & Sheen G. C.-H. (2020). Voluntary adoption of social welfare-enhancing behavior: Mask-wearing in Spain during the COVID-19 outbreak. *PLoS One*, *15*(12), e0242764. https://doi.org/10.1371/journal.pone.0242764

Bertozzi A. L., Franco E., Mohler G., Short M. B., & Sledge D. (2020). The challenges of modeling and forecasting the spread of COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(29), 16732–16738. https://doi.org/10.1073/pnas.2006520117

Bo Y., Guo C., Lin C., Zeng Y., Li H. B., Zhang Y., Hossain M. S., Chan J. W. M., Yeung D. W., Kwok K. O., Wong S. Y. S., Lau A. K. H., & Lao X. Q. (2021). Effectiveness of non-pharmaceutical interventions on COVID-19 transmission in 190 countries from 23 January to 13 April 2020. *International Journal of Infectious Diseases*, *102*, 247–253. https://doi.org/10.1016/j.ijid.2020.10.066

Bonaccorsi G., Pierri F., Cinelli M., Flori A., Galeazzi A., Porcelli F., Schmidt A. L., Valensise C. M., Scala A., Quattrociocchi W., & Pammolli F. (2020). Economic and social consequences of human mobility restrictions under COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(27), 15530–15535. https://doi.org/10.1073/pnas.2007658117

Brauner J. M., Mindermann S., Sharma M., Johnston D., Salvatier J., Gavenčiak T., Stephenson A. B., Leech G., Altman G., Mikulik V., Norman A. J., Monrad J. T., Besiroglu T., Ge H., Hartwick M. A., Teh Y. W., Chindelevitch L., Gal Y., & Kulveit J. (2020). Inferring the effectiveness of government interventions against COVID-19. *Science*, *371*(6531), eabd9338. https://doi.org/10.1126/science.abd9338

Brzezinski A., Deiana G., Kecht V., & Van Dijcke D. (2020). *The COVID-19 pandemic: Government versus community action across the United States* (CEPR Press, no. 7: 115–47).

Carleton T., & Meng K. C. (2020). *Causal empirical estimates suggest COVID-19 transmission rates are highly seasonal* [Working Paper]. https://t.co/69vR0LUGsT?amp=1.

Carpenter B., Gelman A., Hoffman M. D., Lee D., Goodrich B., Betancourt M., Brubaker M., Guo J., Li P., & Riddell A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Cheng C., Barceló J., Hartnett A. S., Kubinec R., & Messerschmidt L. (2020). COVID-19 government response event dataset (CoronaNet v.1.0). *Nature Human Behaviour*, *4*(7), 756–768. https://doi.org/10.1038/s41562-020-0909-7

Courtemanche C., Garuccio J., Le A., Pinkston J., & Yelowitz A. (2020). Strong social distancing measures in the United States reduced the COVID-19 growth rate. *Health Affairs*, *39*(7), 1237–1246. https://doi.org/10.1377/hlthaff.2020.00608

Dave D. M., Friedson A. I., Matsuzawa K., & Sabia J. J. (2020). When do shelter-in-place orders fight COVID-19 best? Policy heterogeneity across states and adoption time. https://doi.org//10.3386/w27091

Dave D. M., Friedson A. I., Matsuzawa K., Sabia J. J., & Safford S. (2020). *Black lives matter protests, social distancing, and COVID-19* (NBER).

Dudel C., Riffe T., Acosta E., van Raalte A. A., & Myrskyla M. (2020). *Monitoring trends and differences in COVID-19 case fatality rates using decomposition methods: Contributions of age structure and age-specific fatality* [Working Paper]. https://doi.org/10.31235/osf.io/j4a3d.

Fan Y., Orhun A. Y., & Turjeman D. (2020). *Heterogeneous actions, beliefs, constraints and risk tolerance during the COVID-19 pandemic* (NBER).

Fellows I. E., Slayton R. B., & Hakim A. J. (2020). 'The COVID-19 pandemic, community mobility and the effectiveness of non-pharmaceutical interventions: The United States of America, February to May 2020', arXiv, arXiv:2007.12644 [q-Bio, Stat], preprint: not peer reviewed. https://doi.org//10.48550/arXiv.2007.12644

Ferguson N. M., Nedjati-Gilani G., & Laydon D. (2020). *Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand* [Working Paper]. Imperial College of London https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf.

Ferrari S., & Cribari-Neto F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, *31*(7), 799–815. https://doi.org/10.1080/0266476042000214501

Flaxman S., Mishra S., Gandy A., Unwin H. J. T., Mellan T. A., Coupland H., Whittaker C., Zhu H., Berah T., Eaton J. W., Monod M., Ghani A. C., Donnelly C. A., Riley S., Vollmer M. A. C., Ferguson N. M., Okell L. C., Bhatt S., & Imperial College COVID-19 Response Team (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7

Gadarian S. K., Goodman S. W., & Pepinsky T. B. (2020). *Partisanship, health behavior and policy attitudes in the early stages of the COVID-19 pandemic* (SSRN). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3562796.

Gadarian S. K., Goodman S. W., & Pepinsky T. B. (2022). *Pandemic politics: The deadly toll of partisanship in the age of COVID*. Princeton University Press.

Gao S., Rao J., Kang Y., Liang Y., & Kruse J. (2020). Mapping county-level mobility pattern changes in the United States in response to COVID-19. *SIGSpatial Special*, *12*(1), 16–26. https://doi.org/10.1145/3404820.3404824

Grinsztajn L., Semenova E., Margossian C. C., & Riou J. (2021). Bayesian workflow for disease transmission modeling in Stan. *Statistics in Medicine*, *40*(27), 6209–6234. https://doi.org/10.1002/sim.v40.27

Grossman G., Kim S., Rexer J. M., & Thirumurthy H. (2020). Political partisanship influences behavioral responses to governors' recommendations for COVID-19 prevention in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(39), 24144–24153. https://doi.org/10.1073/pnas.2007835117

Gu Y. (2020). "Covid19-Projections.com." https://covid19-projections.com/about/#about-the-model.

Haug N., Geyrhofer L., Londei A., Dervic E., Desvars-Larrive A., Loreto V., Pinior B., Thurner S., & Klimek P. (2020). Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature Human Behaviour*, *4*(12), 1303–1312. https://doi.org/10.1038/s41562-020-01009-0

Islam N., Sharp S. J., Chowell G., Shabnam S., Kawachi I., Lacey B., Massaro J. M., D'Agostino R. B., & White M. (2020). Physical distancing interventions and incidence of coronavirus disease 2019: Natural experiment in 149 countries. *BMJ*, *370*, m2743. https://doi.org/10.1136/bmj.m2743

Katz R., Toole K., Robertson H., Case A., Kerr J., Robinson-Marshall S., Schermerhorn J., Orsborn S., Van Maele M., Zimmerman R., Stevens T., Phelan A., Carlson C., Graeden E., & COVID AMP Coding Team (2023). 'COVID AMP: An open access database of COVID-19 response policies', medRxiv, preprint: not peer reviewed. https://doi.org//10.1101/2023.05.01.23289163

Kucharski A. J., Russell T. W., Diamond C., Liu Y., Edmunds J., Funk S., Eggo R. M., & Centre for Mathematical Modelling of Infectious Diseases COVID-19 working group (2020). Early dynamics of transmission and control of COVID-19: A mathematical modelling study. *The Lancet Infectious Diseases*, *20*(5), 553–558. https://doi.org/10.1016/S1473-3099(20)30144-4

Larremore D. B., Fosdick B. K., Bubar K. M., Zhang S., Kissler S. M., Metcalf C. J. E., Buckee C. O., & Grad Y. H. (2021). Estimating SARS-CoV-2 seroprevalence and epidemiological parameters with uncertainty from serological surveys. *Elife*, *10*, e64206. https://doi.org/10.7554/eLife.64206

Li R., Chen B., & Pei S. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*, *368*(6490). https://doi.org/10.1126/science.abb3221

Linzer D. A. (2013). Dynamic Bayesian forecasting of presidential elections in the states. *Journal of the American Statistical Association*, *108*(501), 124–134. https://doi.org/10.1080/01621459.2012.737735

Lock K., & Gelman A. (2010). Bayesian combination of state polls and election forecasts. *Political Analysis*, *18*(3), 337–348. https://doi.org/10.1093/pan/mpq002

McAndrew T., & Reich N. G. (2022). An expert judgment model to predict early stages of the COVID-19 pandemic in the United States. *PLoS Computational Biology*, *18*(9), e1010485. https://doi.org/10.1371/journal.pcbi.1010485

Moein S., Nickaeen N., Roointan A., Borhani N., Heidary Z., Javanmard S. H., Ghaisari J., & Gheisari Y. (2021). Inefficiency of SIR models in forecasting COVID-19 epidemic: A case study of Isfahan. *Scientific Reports*, *11*(1), 4725. https://doi.org/10.1038/s41598-021-84055-6

Painter M., & Qiu T. (2020). *Political beliefs affect compliance with COVID-19 social distancing orders* (SSRN). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3569098.

Perkins T. A., Cavany S. M., Moore S. M., Oidtman R. J., Lerch A., & Poterek M. (2020). *Estimating unobserved SARS-CoV-2 infections in the United States* [Working Paper]. http://perkinslab.weebly.com/uploads/2/5/6/2/25629832/perkins_etal_sarscov2.pdf.

Perra N. (2021). Non-pharmaceutical interventions during the COVID-19 pandemic: A review. *Physics Reports*, *913*, 1–52. https://doi.org/10.1016/j.physrep.2021.02.001

Riou J., Hauser A., Counotte M. J., & Althaus C. L. (2020). Adjusted age-specific case fatality ratio during the COVID-19 epidemic in Hubei, China, January and February 2020. *PLoS Medicine*, *17*(7). https://doi.org//10.1371/journal.pmed.1003189.

Sajadi M. M., Habibzadeh P., Vintzileos A., Shokouhi S., Miralles-Wilhelm F., & Amoroso A. (2020). *Temperature, humidity and latitude analysis to predict potential spread and seasonality for COVID-19* (SSRN). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3550308.

Sánchez-Romero M., di Lego V., Prskawetz A., & Queiroz B. L. (2021). An indirect method to monitor the fraction of people ever infected with COVID-19: An application to the United States. *PLoS One*, *16*(1), e0245845. https://doi.org/10.1371/journal.pone.0245845

Sebhatu A., Wennberg K., Arora-Jonsson S., & Lindberg S. I. (2020). Explaining the homogeneous diffusion of COVID-19 nonpharmaceutical interventions across heterogeneous countries. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(35), 21201–21208. https://doi.org/10.1073/pnas.2010625117

Sharma M., Mindermann S., Rogers-Smith C., Leech G., Snodin B., Ahuja J., Sandbrink J. B., Monrad J. T., Altman G., Dhaliwal G., Finnveden L., Norman A. J., Oehm S. B., Sandkühler J. F., Aitchison L., Gavenčiak T., Mellan T., Kulveit J., Chindelevitch L., …Brauner J. M. (2021). Understanding the effectiveness of government interventions in Europe's second wave of COVID-19. *Nature Communications*, *12*(1), 5820. https://doi.org/10.1038/s41467-021-26013-4

Stekhoven D. J., & Bühlmann P. (2012). MissForest-non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118. https://doi.org/10.1093/bioinformatics/btr597

Stoto M. A., Woolverton A., Kraemer J., Barlow P., & Clarke M. (2022). COVID-19 data are messy: Analytic methods for rigorous impact analyses with imperfect data. *Globalization and Health*, *18*(1), 2. https://doi.org/10.1186/s12992-021-00795-0

Tasnim S., Hossain M. M., & Mazumder H. (2020). Impact of rumors or misinformation on coronavirus disease (COVID-19) in social media. *Journal of Preventive Medicine and Public Health*, *53*(3), 171–174. https://doi.org/10.3961/jpmph.20.094

Winship C., & Mare R. D. (1983). Structural equations and path analysis for discrete data. *The American Journal of Sociology*, *89*(1), 54–110. https://doi.org/10.1086/227834

Yuan Y., & MacKinnon D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14*(4), 301–322. https://doi.org/10.1037/a0016972

Zellner A. (1988). Optimal information processing and Bayes's theorem. *The American Statistician*, *42*(4), 278–280. https://doi.org/10.1080/00031305.1988.10475585

Zheng Q., Jones F. K., Leavitt S. V., Ung L., Labrique A. B., Peters D. H., Lee E. C., Azman A. S., & HIT-COVID Collaboration (2020). HIT-COVID, a global database tracking public health interventions to COVID-19. *Scientific Data*, *7*(1), 286. https://doi.org/10.1038/s41597-020-00610-2