**Harvard Data Science Review • Special Issue 1: COVID-19:
Unprecedented Challenges and Chances**

Estimating the Number of SARS-CoV-2 Infections and the Impact of Mitigation Policies in the United States

James Johndrow¹ Patrick Ball² Maria Gargiulo^{2,3} Kristian Lum^{2,4}

¹Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America,

²Human Rights Data Analysis Group, San Francisco, California, United States of America,

³Department of Sociology, University of Oxford, Oxford, England, United Kingdom,

⁴Department of Computer and Information Science, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America

Published on: Nov 23, 2020

DOI: <https://doi.org/10.1162/99608f92.7679a1ed>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

ABSTRACT

Knowledge of the number of individuals who have been infected with the novel coronavirus SARS-CoV-2 and the extent to which attempts for mitigation by executive order have been effective at limiting its spread are critical for effective policy going forward. Directly assessing prevalence and policy effects is complicated by the fact that case counts are unreliable. In this article, we present a model for using death-only data—in our opinion, the most stable and reliable source of COVID-19 information—to estimate the underlying epidemic curves. Our model links observed deaths to a susceptible-infected-removed (SIR) model of disease spread via a likelihood that accounts for the lag in time from infection to death and the infection fatality rate. We present estimates of the extent to which confirmed cases in the United States undercount the true number of infections, and analyze how effective social distancing orders have been at mitigating or suppressing the virus. We provide analysis for four states with significant epidemics: California, Florida, New York, and Washington.

Keywords: COVID-19, SARS-CoV-2, compartmental model, SIR model, Bayesian

Media Summary

Our analysis offers two findings. First, we estimate that the true number of COVID-19 infections is likely 6 to 10 times larger than the number of confirmed, officially reported cases as of late March. Second, we find that the first round of executive orders for social distancing had mixed effects across states. While in New York the effective reproductive number (a measure of the spread of the virus) dropped low enough to suppress the spread of the virus, this was less clear in states like Florida, where our estimates suggest that the executive orders caused the infection rate only to plateau rather than decrease.

Our model offers two technical improvements, in addition to these findings. First, we use what we believe is the most reliable, or least unreliable, source of data about the pandemic: reported deaths. Next, we build a Bayesian model that accounts for the time lag between infection and death.

We model the transmission of the virus (creating new infections) by a susceptible-infected-removed (SIR) model, a mechanistic description of the spread of infectious disease. Based on prior research, we assume an infection fatality rate of 1%. The true infection fatality rate is still contested, even among experts, so we provide results for infection fatality rates covering a range of plausible scenarios.

Our work is motivated by the problem that official data about confirmed COVID-19 infections drastically understates the true extent of the disease. Many people who have been infected have not sought testing or have not succeeded in being tested due to restrictive policies delineating who is eligible for testing. As testing capabilities in the United States have ramped up, larger proportions of infected individuals have been recorded

as infected in the data. This has even made comparative statements across time based on case counts unreliable. Subsequent declines in testing further complicate the use of the reported cases.

These findings offer insights into the true scale of the pandemic. Furthermore, with estimated infection rates that correct for underreporting bias, these findings enable comparisons among regions and over time. A measure of the magnitude and pattern of the infections is an essential input to health care planning, public policy about distancing, and public communication about the progress of the pandemic.

1. Introduction

The coronavirus SARS-CoV-2, which causes the disease COVID-19, has already changed the lives of billions of people globally. Many people have been ordered to stay in their homes, and economies worldwide have largely come to a halt. Thousands have died and at least several million have been infected. In the United States, social distancing policies began being implemented in mid-March, as states such as Washington, California, and New York saw a sharp rise in the number of hospitalizations attributable to the virus. Government response has been hindered by an insufficient supply of materials needed for testing, and by the large proportion of infected individuals who are asymptomatic and therefore are unlikely to seek testing even if it is available. These factors make it difficult to know the true size of the infected population. Because effective surveillance has not been possible, policymakers have instead turned to social distancing policies as the best available tool to slow the spread of the virus.

Here, we seek to address two key questions: (1) How many people are actually infected or have ever been infected with SARS-CoV-2?; and (2) Are the social distancing policies currently in place effective at suppressing the virus? That is, can they be expected to lead to a decrease in the number of infections? Effectively addressing these questions requires innovative modeling due to severe limitations in commonly used sources of data for tracking the spread of the virus ([Angelopoulos et al., 2020](#)).

1.1. (Limitations of) Data on COVID-19

Ideally, to address these questions we would use data on the number of confirmed cases to understand the prevalence of the disease and assess policy measures. However, we view confirmed case counts for COVID-19 to be unreliable and ill-suited to this type of analysis for a number of reasons. Media reports have made clear that testing is more available in some regions than others, and so case counts are primarily an indication of where testing is most comprehensive.¹ In fact, there may be as much as a 20-fold difference in case detection between countries, leading to incomparable numbers ([Golding et al., 2020](#)).

Evidence for the insufficiency of case counts within the United States can also be seen in the large observed difference in test-positive rates—the proportion of positives among all tests performed—across states. For example, as of May 17, 2020, California has conducted 1.24 million tests and enumerated about 80,000

confirmed cases, for a test positive rate of about 6.5%. By contrast, New York has conducted 1.41 million tests and enumerated about 350,000 confirmed cases, for a test positive rate of close to 25%. For case counts to represent an accurate count of the number of infections in both locations, one would need to believe that testing policies and test-seeking behavior in New York are much more likely to identify individuals who have the disease while simultaneously not missing infected people at a higher rate than in California, a highly dubious prospect. One way such large discrepancies in test positive rates could arise is through random testing. If people were tested at random in both states, we would expect to see higher test positive rates in locations with higher prevalence, such as New York. While the prevalence likely does vary between the states, outside of limited designed experiments accounting for a tiny fraction of the total number of tests, testing is not administered to a random sample of people. Instead, tests are administered mainly to symptomatic individuals and people who seek testing. Rather, the most likely explanation for this disparity in test positive rates is that in locations like New York—where the epidemic is large—the number of people infected has persistently outstripped the capacity to test, while in California—where the epidemic is much smaller—this is less true. Thus, the case counts likely do not accurately reflect even the *relative* size of the infected population across states. It is even more questionable that they could accurately reflect the *absolute* size of the infected population across states.

Furthermore, even within a single administrative boundary, case counts from earlier in the epidemic cannot meaningfully be compared with more recent case counts. Early in the epidemic, tests were administered only to people meeting strict criteria that likely excluded many sick individuals ([Johnson & McGinley, 2020, March 7](#)). In some cases, these criteria were chosen due to shortages of testing resources, such as reagents ([Cavitt, 2020, April 12](#)). Over time, shortages of key components of testing have slowly resolved, and the administrative rules and guidelines have themselves changed. For example, according to the Centers for Disease Control’s (CDC) website, several revisions to the official guidance on which patients should be tested have been made, including a March 4, 2020, revision that modified the criteria for testing to expand the pool of eligible people ([Centers for Disease Control and Prevention, 2020](#)). Thus, some of the trends in cases we observe over time within a given location are likely attributable to changes in test availability and criteria for test eligibility rather than purely to changes in prevalence of the disease. This makes case data of questionable value for modeling, even if model results are not compared across different locations.

Another, potentially more reliable, data source is hospitalizations. Unfortunately, these data have become universally available only fairly recently. For example, the Johns Hopkins COVID-19 database provides hospitalization data by state beginning only on April 14, 2020—more than two months after the epidemic began in New York, Washington, and California. High-quality data on the size of the epidemic in the early days are critical to fitting epidemiological models, which tend to be sensitive to initial conditions, and thus a partial time series lacks information that is critical for obtaining the best estimates. Moreover, even if the data source as it currently exists had been available since the beginning of the outbreak, it is a measure of *total* hospitalizations by day. Without supplementary information on the length of stay for each patient, it is not

possible to calculate the number of *new* hospital admissions over time, which is much more useful from the perspective of modeling the total number of infected individuals. Thus, while this could be useful as an additional data source, particularly in the future as the length of the available time series grows, at the moment it is not adequate for our purposes.

This brings us to data on deaths attributable to COVID-19. Death data have been recorded since the earliest days of the epidemic and are universally available across states. Because gravely ill patients who die from severe disease are more likely than the average infected person to be hospitalized and tested, we believe death data are the most complete and representative data source available on a timely basis that could be used to estimate the number of SARS-CoV-2 infections. Even so, death data are imperfect. Excess mortality calculations suggest that the death-only data will miss some people who died of COVID-19 but whose cause of death was not listed as such on their death certificates [Weinberger et al. \(2020\)](#). Why, then, not use excess mortality data? Excess mortality estimates offer an invaluable approach to evaluating the disease's overall impact. However, excess mortality almost certainly overcounts actual deaths directly attributable to COVID-19, since some of the excess mortality is a result of people with other health conditions avoiding hospitals and clinics, cancellation of procedures to reduce hospital census, and so forth. Such second-order effects of the disease don't fit into traditional models of the spread of infectious disease that are built around the assumption that people who die from the disease were at some point themselves *infectious*. Thus, while approaches based on excess mortality are incredibly useful for revealing the total effect of the epidemic on mortality, we find them less suitable for fitting models of the spread of the virus. As a result, we fit our model to the data that we believe are best suited for modeling the spread of the virus and have a reasonable chance of being accurate, or—more precisely—are likely the *least inaccurate* of available measures of the extent of COVID-19 infections: death data. Notably, several other influential COVID-19 modeling efforts have come to the same conclusion and have turned to death data to fit or calibrate their models ([Altieri et al., 2020](#); [Ferguson et al. \(2020, March 16\)](#); [Flaxman et al. \(2020\)](#); [Golding et al., 2020](#)).

1.2. Our Modeling Approach

A common approach in the statistical epidemiological literature focuses on fitting or calibrating ordinary differential equation (ODE) models to observed data, and this underpins our approach as well ([Hethcote, 2000](#)). However, because we build our model using death-only data, we propose modifications to the standard approaches that rely on case data. In order to do this, the observed deaths need to be linked to the underlying state variables of the ODE model via a sensible, scientifically motivated likelihood. We do this by means of a distribution for the time from infection to death, and an assumption about the infection fatality rate (IFR), that is, the proportion of infected individuals who will eventually die of COVID-19. Using death-only data in the early stage of the epidemic, these parameters and the parameters of the underlying ODE model are not separately identifiable. For this reason, the time to death distribution and IFR are *assumptions* of our model.

There exist external estimates of the time to death distribution based on high-quality data. For the IFR, precisely because of the difficulty outlined above in establishing the true number of infections, there remains considerable uncertainty about its true value. Because the IFR is such an important assumption of our model, we perform analysis for five scenarios with different values of the IFR that span the (relatively small) range of plausible values. Our model assumes that the IFR is constant over the time period considered. While we believe this is a useful approximation to the truth, in reality there may be drift over time. For example, as physicians and researchers gain more insight into the most efficacious treatments for the disease, outcomes for seriously ill patients may improve over time, leading to a lower rate of death given infection later in the epidemic. It has also been speculated that the demographics of who becomes infected may change the IFR.² However, because our focus is on the first months of the pandemic during which we believe the IFR was likely close to constant, we do not incorporate a time-varying IFR in our model.

Conditional on these assumptions and a sampling model for deaths given infections, we fit the parameters of an underlying ODE model of epidemic dynamics to the observed death data. We emphasize that our analysis should be read as ‘if the infection fatality rate is X , then the following would be true,’ rather than primarily as advocating for a particular value of the IFR over others.

As a side-effect of fitting epidemic curves and evaluating the impact of interventions, our model allows us to make projections for the likely trajectory of infections and deaths. However, exact projections is not the primary goal of this work. A model focused more on precise predictions would likely include location-specific measures of containment and treatment efficacy, as well as age- and comorbidity-specific infection fatality rates. In order to minimize the prediction error, it would also account for reporting artifacts, such as delays in recording of deaths in quickly updated records over weekends or holidays. It would also account for increased mortality if the sick overwhelm hospital systems. It would also emphasize finding external data sources that provide strong leading indicators of future deaths. This kind of information could be included in a more prediction-optimized version of the model. Our focus here is to outline a modeling approach using minimal but relatively reliable data that are described by a likelihood and priors that incorporate our understanding of the data-generating process, is fitted to data, and is underpinned by a widely used epidemiological model that is designed to approximate the real dynamics of disease spread. It also turns out that this model predicts reasonably well over a two-week time horizon—the only prediction that we assess here.

The remainder of this article is organized as follows. In section 2 we review related work. In section 3 we introduce our model and explain how the model can be used to infer the number of infections and the effect of executive orders for social distancing. In section 4 we describe an Markov chain Monte Carlo (MCMC) algorithm for fitting our model. In section 5 we give results. In section 6 we conduct a sensitivity analysis. Section 7 concludes.

2. Related Work

Several previous studies have attempted to model the dynamics of the pandemic in various geographic locations and with varying goals. Several of these have provided some analysis or estimate of the amount by which confirmed cases undercount the true number of infections. For example, [R. Li et al. \(2020\)](#) propose that in the first month of the epidemic in China, 82–90% of infections were undocumented. [Riou et al. \(2020\)](#) use a susceptible-exposed-infected-removed (SEIR) model, an epidemiological ordinary differential equation (ODE) model, and calibrate their model to the time series of reported deaths and reported infections. By modeling the underreporting of symptomatic cases, and by assuming that approximately half of infections lead to symptomatic cases, they estimate the infected population in Hubei, finding that approximately 30% of infections were documented. [Perkins et al. \(2020\)](#) estimate directly that in the United States, more than 90% of infections have been undocumented by tests using Chinese data and initial reports in the United States. The first wave of a random sampling design in Indiana conducted by researchers at the University of Indiana and the Indiana State Department of Health concluded in a preliminary report that the number of positive tests undercounted the number of people ever infected by a factor of 9.6 ([Menachemi et al., 2020](#)).

[Ferguson et al. \(2020, March 16\)](#) model the effect of transmission between susceptible and infectious individuals using a microsimulation model built on synthetic populations designed to mimic the populations of the United Kingdom and United States. They assume a fixed time-to-onset and a range of R_0 values from 2.0–2.6, and they assume symptomatic cases to be 50% more infectious than asymptomatic cases. Having calibrated their model to the cumulative number of deaths, they estimate deaths and hospital loads under different nonpharmaceutical interventions involving social distancing and isolation. This analysis was arguably the most influential in triggering the adoption of social distancing policies in the United States, as it indicated that without social distancing measures in place, there would likely be around 2.2 million deaths, and hospitals nationwide would be completely overwhelmed. They also predicted that social distancing measures could reduce the number of deaths substantially and prevent exceedance of hospital resources, but only if they were dynamically turned ‘on and off’ by triggering mechanisms based on the current number of COVID-19 patients. The report envisioned social distancing remaining in place for roughly 18 months, the amount of time they think it will take for a vaccine to become available.

The [CHIME app \(Weissman et al., 2020\)](#) is an online tool created by researchers at the University of Pennsylvania to help hospitals anticipate the number of incoming COVID-19 patients and their needs. The CHIME model uses the current number of COVID-19 hospitalizations to ‘back out’ the total number of cases based on a user-provided hospitalization rate conditional on infection. Similar to our work, their model does not rely on the case counts. They make forward projections for the number of hospital admissions, ICU admissions, and ventilators needed over the coming weeks. They allow the user to specify the parameters of their underlying epidemiological model as inputs in terms of the doubling time for the infected population.

The model given in [Murray \(2020\)](#) has a goal similar to CHIME (hospital use planning). This model takes a different approach by fitting parametric curves to observed cumulative death rates. They use a hierarchical model on the parameters of the parametric curve. The model essentially projects that the future course of the cumulative death rate curve in the United States will follow a path similar to that observed in other locations that are farther along in the course of their epidemics. There is no underlying model of epidemic dynamics, but there is a sampling model for death rates, which allows them to give confidence intervals. The research team developing this model has updated the details of their estimation procedure several times since their model was made public. Instead of an arbitrary curve-fitting procedure, the current version as of writing fits an underlying SEIR model ([Institute for Health Metrics & Evaluation, 2020](#)). The *New York Times* [online tool](#) allows the user to specify inputs to understand how those inputs affect likely infections, hospital loads, and deaths; infections are a side-effect of the rest of the model.

[Flaxman et al. \(2020\)](#) take a similar approach to ours to evaluate the effect of social distancing orders across countries in Europe. They define a likelihood for death-only data and incorporate mitigation efforts in the model. They build a hierarchical model to estimate the number of infections and effects of mitigation across countries in Europe. The overall approach is similar, but their analysis is done for Europe, whereas here we consider the United States. [Lewnard et al. \(2020\)](#) similarly evaluate the effect of mitigation efforts. They observe reductions in estimates of the effective reproduction number for patients in three hospital systems in Northern California, Southern California, and Washington State as a consequence of the implementation of nonpharmaceutical interventions, like social distancing. [Song et al., 2020](#) describe a statistical software package that can be used to estimate the impact of quarantine protocols on the spread of COVID-19, and they apply their method to data from China.

3. Model

Let ν_t be the number of *new* infections on day t of the epidemic, and let p denote the infection fatality rate, that is, the probability of death given infection. We denote the day of the first infection by T_0 . Let $\theta = \{\theta_s : s = 0, 1, \dots, m\}$ be the set of probabilities defining the discrete time-to-death distribution, where θ_s denotes the probability that, for those who die, death from COVID-19 occurs s days after the initial infection. Let $X(t, t')$ denote the number of individuals newly infected on day t who die on day t' . Our death model is

$$(3.1) \quad X(t, t') \mid p, \theta, \nu_t \sim \text{Poisson}(p\nu_t\theta_{(t'-t)}).$$

The observed deaths on day r are thus given by

$$(3.2) \quad D(r) = \sum_{t=1}^r X(t, r),$$

the sum over all previous days of the number of individuals infected on that day who went on to die on day r . The distribution of $D(r)$ marginal of X is given by

$$(3.3) \quad D(r) \mid p, \theta, \nu \sim \text{Poisson} \left(p \sum_{t=1}^r \nu_t \theta_{(r-t)} \right),$$

and for $r \neq s$, $D(r) \perp\!\!\!\perp D(s) \mid p, \theta, \nu$. This defines our likelihood. The use of a Poisson distribution in specifying our model may seem unnatural compared to the specification $X(t, t') \mid p, \theta, \nu_t \sim \text{Binomial}(\nu_t, p\theta_{(t-t')})$. The Poisson specification allows ν_t to take real values as opposed to integer values, as would be required for a Binomial distribution. This allows us to use simpler, deterministic models for the underlying epidemiological curves defining the ν_t s, simplifying computation. Fortunately, in cases where $p\theta_s$ is small and ν is large—precisely the situation in which we find ourselves after the very early days of the epidemic— $\text{Poisson}(p\theta_s \nu)$ is a good approximation to $\text{Binomial}(\nu, p\theta_s)$.

The observed number of deaths D are linked to a compartmental epidemiological model via the total number of newly infected individuals on day t , ν_t . To generate realistic ν_t curves, we use a SIR model of epidemic dynamics, with state evolution given by the following ODE:

$$(3.4) \quad \begin{aligned} \frac{ds_t}{dt} &= \begin{cases} -\beta s_t i_t & t < T_1 \\ -\phi \beta s_t i_t & t \geq T_1 \end{cases} \\ \frac{di_t}{dt} &= \begin{cases} \beta s_t i_t - \gamma i_t & t < T_1 \\ \phi \beta s_t i_t - \eta \gamma i_t & t \geq T_1 \end{cases} \\ \frac{dr_t}{dt} &= \begin{cases} \gamma i_t & t < T_1 \\ \eta \gamma i_t & t \geq T_1 \end{cases}, \end{aligned}$$

where s_t is the proportion of susceptible individuals at time t , i_t is the number of infected individuals at time t , r_t is the number of removed individuals at time t , and T_1 is the time at which social distancing orders were issued. The ν_t variables on which our likelihood is conditioned can be extracted directly from this model. The number of new infections that occurred during day t is simply the difference in the size of the S compartment at time t and $t + 1$ multiplied by N . This is because, under this model, individuals may leave the susceptible compartment only by becoming infected, so the number of new infections that occurred in any time period is precisely the reduction in the number susceptible in that same period.

Up to time T_1 , these equations represent a standard SIR model with parameters $\{\beta, \gamma\}$; postmitigation, it is a SIR model with parameters $\{\phi\beta, \eta\gamma\}$. In a SIR model with parameters $\{\beta, \gamma\}$, infected individuals are all

considered contagious, and the mean time between infection and removal (the end of the contagious period) is given by γ^{-1} . R_0 is given by $\beta\gamma^{-1}$ —this is the number of people who are infected by a single infected individual in a population in which everyone is susceptible. The quantity $(1 - \phi) \in (0, 1)$ represents the percent reduction in the rate of infection that occurred following implementation of mitigation policies; the quantity $\eta > 1$ represents the increase in the rate at which infected individuals cease to be contagious—whether by recovery, death, self-isolation, or medical quarantine—following the implementation of mitigation policies.

The SIR model is a stylized mathematical model of the spread of infectious disease, and—like any model of a complex process—is an approximation of a more complicated reality. In particular, there are several ways in which its compartments only approximately represent distinct subpopulations in the real world. For example, although the R compartment is typically conceptualized as containing people who have been ‘removed’ from the population of infectious people due to recovery or death, reality is not so simple. Quarantine or hospitalization may largely ‘remove’ an individual from the population of infectious people because the rate of transmission for isolated people is substantially reduced; indeed, that is the point of quarantine. Although some models explicitly include a quarantine compartment (e.g., [Song et al., 2020](#)), we allow the postinfection quarantines to be subsumed by the removed compartment. This more expansive definition of the R compartment justifies the increase in the rate of removal from γ to $\eta\gamma$ following the promulgation of mitigation policies. As awareness and fear of the virus grows, individuals become more likely to self-isolate or seek care when they begin to show symptoms. One further consequence of this simplification is that the number of individuals who die on the t th day is not directly comparable with the number of people who enter the R compartment during that period because many of the individuals who died at that time could have been effectively removed at an earlier time.

These limitations notwithstanding, we elect to use the SIR model to underpin our likelihood because it is among the simplest options that generate epidemic simulations incorporating the key facets of an epidemic. Specifically, the SIR model manifests herd immunity, rapid growth in the early phase, and is constrained to produce a finite number of infections ([Hethcote, 2000](#)). The main purpose of incorporating a compartmental model into our method is to have a realistic, low-dimensional generative model of the new infection curves, ν_t , on which our likelihood is based. Because of its simplicity, ease of interpretability, and well-understood properties, we find the SIR model to be an appealing choice. However, the framework we have defined for creating a sensible likelihood-based model of death-only data could easily be extended to embed more elaborate compartmental models or any other generative model of new infections.

Figure 1 shows one realization of our SIR model. The top panel shows one draw of the daily number of deaths. Notice that this is not a smooth curve. When fitting our model, these D s will be our data and will be the only thing we observe directly. The other panels are what we will infer from the D s. The middle panel of the figure shows the ν s, the number of daily new infections. The series of D s has roughly the same shape as the ν s,

though it lags it by about 25 days. This lag is due to the shape of the time from infection to death distribution θ , which we describe following. The bottom panel gives the standard view of the SIR model, showing the total number of susceptible, infected, and removed at any given time point.

Figure 1. One realization of our modified SIR models. The top two panels are displayed on the scale of counts. The bottom panel, which shows our underlying SIR model, is given on the scale of proportion of the population.

3.1. Fixed Model Parameters

Due to identifiability constraints, we do not estimate all of the parameters of the model. Rather, we fix those parameters for which there exists relatively higher quality information on reasonable values applicable in the context studied here. By and large, this includes those parameters pertaining more to the biology of the disease than to the social dynamics of its spread, which we expect to vary more widely across cultural and geographic contexts.³ These include θ and p .

For θ , we draw from two studies. (Zhou et al., 2020) reports that in Wuhan, China, the time from symptom onset to death had a median of 18.5 days with an interquartile range of 15 to 22 days. A Gamma mixture of Poisson distributions (henceforth referred to as the ‘Poisson-Gamma’ distribution) with parameters $\{a, b\} = \{27.75, 1.5\}$ matches the reported quantiles well. Lauer et al. (2020) estimate the incubation period of the disease, also using data from China. They report a median incubation time of 5.1 days with an

estimated 97.5th percentile of 11.5 days and a 99th percentile of 14 days. The quantiles of a Poisson-Gamma distribution with parameters $\{a, b\} = \{5.5, 1.1\}$ match these reported quantiles well. We compute numerically the mass function of the distribution of the total time from infection to death by generating 100,000 samples from the distribution of the incubation period and the distribution of the time from symptom onset to death that were just described. The time to death is the sum of these two numbers. We truncate the maximum time to death from infection to be the 99th percentile of the generated samples; we call this maximum time to death m . The parameter θ is a length $m + 1$ vector giving the probability that death occurs θ_s days after infection given that death occurs. The probability of death s days after infection (marginal of death occurring) is given by $p\theta_s$. This time to death distribution is shown in Figure 2.

Figure 2. Distribution of time (in days) from infection to death.

Because of its importance in the undercount analysis and the considerable uncertainty surrounding the value of the infection fatality rate p , most of our results are given for five different values of p . To set this range of values for p , we rely on several external data points. [Russell et al. \(2020\)](#) use data from individuals on the Diamond Princess cruise ship to estimate an infection fatality rate of 1.2% (95% CI: 0.38%-2.7%) after adjusting for delays between infection confirmation and death. The ship was a closed population: We know who was on the ship and therefore who to test, so we have confidence in the denominator (all individuals infected with COVID-19 were identified because all people on the ship could be tested). We base the upper end of the range of values we consider for p on the upper end of the confidence interval given by [Russell et al. \(2020\)](#).

We base the lower bound of values considered on data from the United States. As of May 18, 2020, the [New York City Health website](#) reported 20,806 deaths for which the deceased tested positive for COVID-19 or COVID-19 was listed as the cause of death on their death certificate. Compared with an estimated population of about 8.4 million people, this gives an *overall* mortality rate of about 0.25%. Given that each day there are more reported COVID-19 deaths in New York City, and that it is unlikely that every person in New York City

has already been infected, this offers a compelling lower bound on the infection fatality rate in the United States.

This range of values is consistent with other external sources of information. One high-quality source is reported case fatality rates in countries that have done aggressive testing and contact tracing, including testing asymptomatic individuals. In these places, the case numbers may approach the true number of infections, so case fatality information in these locations provides a reasonably tight upper bound on infection fatality rates. Here, we look to South Korea, which has done the most testing per capita. As of a May 17, 2020 press release from the Korea Centers for Disease Control and Prevention ([Korea Centers for Disease Control and Prevention, 2020, May 17](#)), there were 11,050 confirmed positive cases and 262 deaths in South Korea. This gives a naive estimate of the case fatality rate of 2.4%.

For each of these data sources, differences in underlying age and income structure, comorbidities, and other risk factors could limit the generalizability of these estimates to the United States at large. For example, the age distribution on the Diamond Princess cruise ship skews considerably older than the U.S. population in general with about 58% of its passengers 60 years of age or older, while only about 16% of the U.S. population is 62 years of age or older.⁴ The Diamond Princess IFR might be higher than the IFR for the U.S. population in general, as older age is a known risk factor for COVID-19 ([Zhou et al., 2020](#)). However, the fact that the people on the Diamond Princess were able to travel may indicate that these people suffer fewer other comorbidities on average than their similarly aged U.S. cohort. IFR estimates from other geographic regions, such as South Korea, may be limited by similar demographic and risk-factor differences relative to the United States. Additionally, the IFR is influenced by the quality and availability of medical care. Data from locations where medical care is unavailable or inadequate, or where the health care system has been overwhelmed by a surge of COVID-19 patients, may overestimate the IFR relative to situations where quality care is available to anyone who needs it. Such factors are difficult to adjust for, since data on risk factors is preliminary and limited and health care availability can change dramatically if the number of infections in an area explodes.

The differences between the context from which we draw our estimates of p and the population to which they are applied notwithstanding, we believe that the comparison is still useful in providing a range of rough estimates. Indeed, there are many similarities as well. The Diamond Princess cruise ship had a large American population—about 400 of the 3,700 total passengers and crew on board were U.S. citizens. South Korea's healthcare system has not been overwhelmed by COVID-19 cases. This is similar to the United States where, although hospitals in New York City were very full, health care providers have not reached a situation where they have had to ration ventilators.

With all of these qualifications, we cannot be certain that any given value of p is correct when applied in our context. We also do not think that it is possible to precisely estimate the IFR for this virus with the currently available data. However, we believe these external estimates are useful in providing a range of possible values. Thus, because the infection fatality rate is arguably the most important parameter in our model that cannot be

learned from the data, we consider five different scenarios that cover the spectrum of plausible values put forth above: 0.2%, 0.5%, 1.0%, 1.5%, and 2.5%. We focus on the 1.0% case in some figures because we consider values of p in this range to be most likely.

3.2. Estimated Model Parameters

The parameters we estimate in our model are γ , β , ϕ , η , and T_0 . We base our prior for the infectious period on previously published studies. [Ferguson et al. \(2020, March 16\)](#) give a mean generation time (γ^{-1} in the SIR model) of 6.4 days; [Prem et al. \(2020\)](#) assume an infectious period of 3 or 7 days in their simulations with an incubation period of about 6 days. While these aren't directly comparable due to differences in the model—as described above, people can actually move to the R compartment while still infected with the virus—we use these as rough guidance. We choose the prior $\gamma^{-1} \sim \text{Uniform}(1, 15)$, corresponding to an infectious period that ranges from about 1 to 15 days. The value of 1 day would be consistent with a situation in which most symptomatic people rapidly self-isolate, and asymptomatic people are uncommon. Rather than specifying a prior on β , we choose a prior on $R_0 = \beta\gamma^{-1}$, which induces a prior on β . Using the reported confidence interval in [Q. Li et al. \(2020\)](#) to establish rough upper and lower bounds on R_0 , we specify the prior $R_0 \sim \text{Uniform}(1, 4)$. We place a uniform prior on T_0 between January 1, 2020, and February 20, 2020.

We place a $\text{Uniform}(0.2, 0.9)$ prior on ϕ , allowing it to encompass the possibility that mitigation policies had significant effect (ϕ small) and that the policy had a very limited effect (ϕ large). This prior rules out the possibility that the mitigation policies, in fact, increased the transmissibility. This prior is consistent with information on the [University of Maryland COVID-19 Impact Analysis Platform](#), which calculates several social distancing-relevant metrics as a function of mobility data. In each of the states considered, the percent of people not staying home according to this data was roughly 60-80% post-executive orders as it was prior. We do not expect that decreases in mobility perfectly map to decreases in transmissions, as other behaviors that likely changed in tandem with social distancing policies, such as mask wearing or not shaking hands, also have an effect. However, from this data it is clear that we should expect that there was *some* effect, as there was a significant increase in the proportion of the population staying home. Furthermore, based on this, we also believe that transmissions were not completely eliminated, as there still remain a significant proportion of the population interacting outside of their homes.

Finally, we use a $\text{Uniform}(1, 4)$ prior on η , allowing the rate at which people are removed from the infectious population to increase after the implementation of social distancing policies. A value of η substantially larger than 1 would indicate that after social distancing measures and other policies were promulgated, symptomatic people more rapidly self-isolated or otherwise ‘removed’ themselves from human contact after becoming symptomatic. This would be consistent with increasing awareness of the virus due to persistent messaging from media and the authorities resulting in greater individual efforts to self-isolate. On the other hand, values near 1 would indicate that symptomatic people did not change their behavior much with regard to self-isolation after social distancing policies went into effect.

4. Computation

We carry out computation by MCMC using the adaptive Metropolis algorithm [Haario et al. \(2001\)](#). The algorithm produces samples from the Bayesian posterior distribution of the parameters of our model, $\xi = \{\beta, \gamma, T_0, \phi, \eta\}$. Specifically, we update ξ by proposing a new set of parameters, ξ^* from $N(\xi, \Sigma)$, with the time-inhomogeneous covariance Σ computed using the method of [Haario et al. \(2001\)](#). In brief, during a pre-adaptation phase, proposals are made from the usual multivariate Gaussian random-walk, that is,

$$(4.1) \quad \xi^* \sim N(\xi, c_0 \Sigma_0),$$

where Σ_0 is an initial estimate of the posterior covariance, and c_0 is a scalar tuning parameter. After adaptation begins, the proposal becomes

$$(4.2) \quad \xi_t^* \sim N(\xi, c_1 \hat{\Sigma}_t),$$

where $\hat{\Sigma}_t$ is the usual time-averaging estimate of the posterior covariance based on the Markov path up until step $t - 1$. For details, see [Haario et al. \(2001\)](#). Once a new state ξ^* is proposed, we solve the SIR model with these parameters using the Runge-Kutta (4,5) method. We then calculate the sequence ν_t^* from the s_t sequence. We accept or reject the proposed η^* using the usual Metropolis acceptance probability, with target density proportional to

$$(4.3) \quad \ell(\nu(\xi), \xi) = \sum_{t=1}^T p(D(t) \mid \nu) + \log(\pi(\xi)),$$

where $p(D(t) \mid \nu)$ is the Poisson probability mass function implied by (3.3), and $\pi(\cdot)$ represents the prior density. Notice that, because ν is a deterministic function of ξ , both terms in (4.3) depend on the values of these parameters. The tuning parameters of the adaptive Metropolis algorithm are chosen to give acceptance rates between about .2 and .4.

To obtain the initial value of the proposal covariance Σ_0 , we compute an estimate of the sampling covariance of the maximum likelihood estimator using a parametric bootstrap. We tune the constants c_0 and c_1 to obtain acceptance probabilities of approximately .2 to .4, consistent with the guidance of [Roberts and Rosenthal \(2001\)](#). We run for 100,000 iterations, begin adaptation after 20,000 iterations, and use 50,000 iterations of burn-in. Representative trace plots are shown in Appendix A. We also compute Gelman-Rubin statistics based on the last 50,000 iterations and five replicate chains initiated by sampling from a Gaussian centered at the maximum likelihood estimate with the bootstrap estimate of the covariance. This is done for every state for $p = .01$. All of the statistics are less than 1.03, which does not suggest problems with nonstationarity.⁵

5. Results

We fit our model to the daily number of deaths for several states in the United States: California, Florida, New York, and Washington. We use the [state-level data compiled by the New York Times](#). The last date in our training data is April 30, 2020. We base the state populations on [2018 US state population estimates from World Population Review](#), which pulls data from the U.S. Census. We take the date of implementation of social distancing to be the first day on which restaurants and schools were both closed statewide by executive order, as recorded in a [GitHub repository](#) maintained by researchers at the University of Washington. This definition is of course somewhat arbitrary, but given limitations of the available data, it is difficult to conceive of a richer model that would allow the effects of social distancing to phase in gradually without making similarly strong assumptions about how much each type of measure is ‘worth’ compared to the eventual statewide lockdowns that were implemented everywhere.

5.1. Model Fit and Projections

Figure 3 shows model fit for each of the states for $p = .01$. The gray bands show point-wise 95% posterior predictive intervals; the black line shows the posterior mean. The red lines show a 7-day moving average of the daily deaths, while the gray lines show the raw daily reported deaths. We include the moving average for reference because we believe some of the variability in the daily raw data is likely due to delays in reporting rather than inherent variability in the timing of deaths. Despite the low-dimensionality of the free parameters in our model, it fits the data reasonably well. On several occasions, the true values exhibit large spikes or valleys that stray outside of the intervals. This suggests that perhaps a negative binomial or other overdispersed count distribution be considered for future models, though of course one expects to occasionally see values outside of a pointwise 95% interval. We also believe that some of the extremes of day-to-day variation in the raw data are recording artifacts. For example, there is a persistent trend of reported deaths temporarily falling on weekends and spiking early in the week. There is no obvious way to reallocate the data to weekend days, and since our model is not very sensitive to variation on the scale of a few days, we do not attempt to do so. The comparison of the posterior mean to the 7-day moving average provides another way to visualize the performance of the model that is not as sensitive to recording artifacts.

Figure 3 includes data through May 13, 2020, the last 13 days of which were not used for training. The values shown after April 30, 2020 (marked with a vertical line in Figure 3) are projections from the model and associated 95% pointwise posterior credible intervals. For Florida, New York, and Washington, the projections are fairly accurate over this 2-week time span and the moving average largely falls within the intervals. For California, the model predicts a slowly moving increase in deaths, whereas the data in that time period appear to have reached a plateau. This discrepancy could occur if the true time to death distribution were slightly longer than that indicated by our chosen θ . If that were true, the model would ‘expect’ changes to the trajectory of deaths to show up in the data before they actually do. Because the time elapsed between the promulgation of social distancing orders and the end of our training data is only about 45 days, whereas the time from infection

to death can be as long as 40 days, this might cause the model to erroneously infer that the trajectory had not slowed. Another possibility is that behavioral changes did not coincide with the executive orders but slowly increased since then, causing additional slowing in the death rate. Because the parameters are constant during the full time period after the executive orders have been issued, this would cause the estimated parameters to overestimate the average effect immediately after the orders were issued and underestimate it later.

Table 1 shows posterior means and 95% posterior credible intervals of preintervention R_0 for each state for all five values of p . With the exception of Washington, where a larger value of p leads to considerably smaller estimated values of R_0 , these estimates are not especially sensitive to p (good news since this is the parameter about which we are most uncertain), and all fall within the range between about 2.5 and 4. That the estimates are not very sensitive to p may seem surprising, since for smaller values of p , the number of infected individuals at any time must be larger to produce the observed deaths. However, because our model also estimates T_0 (the time of the initial infection) and γ , it is possible to produce quite different time series of new infections with similar R_0 values. The data are thus indicating that changing the assumption about p would imply changes in the parameters that keep R_0 nearly constant.

Table 1. Estimated posterior mean of R_0 , with 95% posterior credible interval shown in parentheses.

	CA	FL	NY	WA
.002	2.99 (1.91,3.83)	3.25 (2.59,3.77)	3.96 (3.87,4.00)	3.06 (2.37,3.39)
.005	2.85 (1.45,3.79)	3.19 (2.45,3.68)	3.91 (3.70,4.00)	2.86 (2.13,3.20)

.01	3.13 (1.82,3.88)	3.19 (2.32,3.68)	3.89 (3.63,4.00)	2.72 (2.05,3.06)
.015	3.12 (2.06,3.83)	3.21 (2.56,3.70)	3.88 (3.65,4.00)	2.66 (1.99,2.99)
.025	3.11 (1.99,3.82)	3.21 (2.49,3.69)	3.81 (3.52,3.99)	2.56 (1.94,2.88)

Figure 4. Estimates of the cumulative undercount factor for each state by day. Plotted on log-10 scale.

5.2. Undercount Estimates

One quantity estimated by our model is the cumulative number of SARS-CoV-2 infections over time. Comparing these estimates with the reported number of confirmed cases by state allows us to estimate the extent by which confirmed cases of COVID-19 undercount the number of infections. We define the undercount at each time point to be our estimated number of cumulative infections as of that time point divided by the cumulative number of confirmed cases at that time. That is, the undercount is the multiplicative factor by which the recorded number of confirmed cases underestimates the true number. Figure 4 shows the undercount for each state across time for each value of p , based on the posterior mean of the total number of people ever infected by day. Naturally, when p is smaller, and therefore the number of deaths is a smaller proportion of the true number of infections, we estimate the undercount to be greater. Under the range of values for p

considered, we find that the undercount was somewhere between 5-fold (for Washington with $p = .025$) and 175-fold (for New York with $p = .002$) across states as of April 1. As testing availability has increased, the undercount factor has fallen to around 2.5-fold (for Washington and Florida with $p = .025$) to about 45-fold (for both New York and California with $p = .002$) as of April 30. Table 2 shows the undercount and associated 95% posterior credible intervals as of the last day that appears in our data, April 30, 2020.

Table 2. Estimated posterior mean of undercount, with 95% posterior credible interval shown in parentheses.

	CA	FL	NY	WA
.002	43.30 (39.29,47.72)	31.92 (28.04,35.88)	44.90 (44.30,45.46)	33.77 (30.87,36.92)
.005	17.60 (15.95,19.45)	12.89 (11.16,14.49)	19.25 (18.89,19.59)	13.59 (12.29,14.88)
.01	8.96 (8.05,10.02)	6.47 (5.68,7.30)	9.78 (9.58,9.96)	6.86 (6.23,7.52)
.015	5.99 (5.36,6.67)	4.30 (3.77,4.84)	6.55 (6.41,6.68)	4.59 (4.16,5.04)
.025	3.58 (3.23,3.97)	2.59 (2.26,2.93)	3.93 (3.84,4.01)	2.77 (2.51,3.05)

These estimated undercounts—particularly those for $p = .01$, $p = .015$, and $p = .025$ —are consistent with other estimates of the extent of undercounting. The first wave of random sampling design in Indiana conducted by researchers at the University of Indiana and the Indiana State Department of Health concluded in a preliminary report that the number of positive tests undercounted the number of people ever infected by a factor of 9.6 ([Menachemi et al., 2020](#)). They also estimate the infection fatality rate at .006. We arrive at undercount factors of between 6.5 and 9.8 for the four states analyzed with $p = .01$ and between 12.9 and 19.25 with $p = .005$. This places our estimates within a similar range of values as those found using a combination of serological and viral RNA testing in Indiana.

New York State also recently conducted a seroprevalence study. In this study, they found that about 21.2% of New York City residents tested positive for having had the disease. Applying this to New York City’s population of 8.4 million people, this results in an estimate of about 1.8 million infections in the New York City compared to the number of reported cases at the time of about 153,000. This implies an undercount factor of a little over 11. Compared with around 11,460 reported deaths at the time of the release of this study, this also implies a raw IFR of about $p = .006$.⁶ Given that some people who are currently infected will, unfortunately, go on to die, this likely provides a slight underestimate of p .

These kinds of estimates can also help inform discussions about herd immunity. For example, in New York as of April 30, there had been about 311,000 cases. In our $p = .01$ scenario, we estimate the undercount factor

at 10 times, as of this date in New York. This would mean that about 3.1 million had been infected in New York between the start of the epidemic and April 30, which is between one sixth and one seventh of the population. This is far too small a proportion for herd immunity to be a major factor at that point. However, an analyst who assigns high probability to the $p = .002$ case would conclude that close to 14 million people (the vast majority of the population) had been infected by April 30, and therefore that herd immunity is sufficient that almost all mitigation efforts could now be lifted and the number of infections would continue to decline. We find this latter scenario to be highly unlikely.

5.3. Effects of Executive Orders for Social Distancing

Executive orders were issued and social distancing policies began taking effect in parts of the United States around March 15, and thus as of the writing of this article, enough time has now passed since the orders in at least some states that the effects will be visible in the deaths data. Recalling the time from infection to death distribution from Figure 2, changes in the dynamics of new infections should begin to be visible in death data around two weeks following the onset of the policy change, and should be mostly visible by around four weeks. This makes it now appropriate to attempt to estimate the effect of such social distancing orders on infection dynamics. In the SIR model, the number of infected individuals will grow when the rate of new infections is higher than the rate of removal and will decrease otherwise. In other words, the rates are equal when the time derivative of i_t is zero, that is,

$$(5.1) \quad \begin{aligned} \phi\beta s_t i_t &= \eta\gamma i_t \\ \frac{\phi\beta s_t}{\eta\gamma} &= 1 \equiv \rho_t. \end{aligned}$$

We thus focus on estimating the quantity ρ_T (where T is the last day that appears in our training data). When $\rho_T > 1$, the number of infected individuals is still growing, while if $\rho_T < 1$, it is declining and the virus is being suppressed. This quantity is sometimes called R_t , but to avoid confusion with compartments of the SIR model, we have used unconventional notation. It is important to note that, because this quantity depends on the current state of s_t , which is decreasing in time, it is possible for the current measures to fail to suppress the virus today yet be sufficient to suppress the virus at some later time because s_t will have declined.

Table 3 shows estimates of the posterior mean of ρ_T along with 95% posterior credible intervals. Results are reported for four states and for all five values of p . It is clear that this quantity is not very sensitive to the assumption about p , which is encouraging since this is the major assumption of our model. Based on these results, the current measures appear sufficient to suppress the virus in New York and Washington, and just at the boundary of being sufficient in California and Florida. It will be interesting to see how these conclusions change in the coming weeks as social distancing and other mitigation policies are relaxed.

Table 3. Estimated posterior mean of ρ_T , with 95% posterior credible interval shown in parentheses.

	CA	FL	NY	WA
.002	1.06 (1.01,1.13)	1.00 (0.86,1.14)	0.44 (0.43,0.44)	0.71 (0.63,0.79)
.005	1.07 (1.02,1.16)	1.01 (0.87,1.16)	0.63 (0.62,0.64)	0.73 (0.64,0.82)
.01	1.10 (1.04,1.20)	1.02 (0.88,1.17)	0.68 (0.67,0.69)	0.75 (0.66,0.83)
.015	1.10 (1.04,1.19)	1.02 (0.88,1.17)	0.69 (0.68,0.71)	0.75 (0.67,0.84)
.025	1.10 (1.04,1.18)	1.03 (0.88,1.18)	0.70 (0.69,0.72)	0.77 (0.68,0.85)

Our model could be sensitive to various types of misspecification, including incorrect specification of θ or p , incorrect specification of the time point T_1 at which the parameters of the model change, and incorrect specification of the SIR model itself. Because we give results for multiple values of p , this type of misspecification is less of a concern. In Appendix B, we conduct a simulation study to assess sensitivity of estimates of the quantities of interest, such as the undercount factor and R_0 , to these other types of misspecification. The results suggest that the model is fairly robust to these sources of misspecification, except in the case of estimation of R_0 when T_1 is misspecified. However, we conclude that badly misspecifying T_1 is often detectable in poor model fit. Details are provided in Appendix B.

5.4. A Prior on p

We have so far presented results for different values of p . The parameter p is not identifiable when the parameters of the SIR model are estimated from the data. When performing subjective Bayesian analysis for an application in which there is widespread disagreement about a non-identifiable parameter—such as p in this case—it makes sense to present multiple cases and allow readers to weight the different cases according to their own priors. ‘Group’ Bayesian decision making is an especially fraught problem with no real solution, and thus when a parameter cannot be updated with data and there clearly exists widespread disagreement about the value of this parameter, it is especially important to allow analysts to impose their own prior on these parameters. In this section, we present results using our own prior on p to demonstrate how one could obtain a single posterior estimate for parameters of the SIR model, averaging over their prior on p .

We assign a prior on p based on our own reading of the literature, media accounts, statements of public health officials, and interactions with epidemiologists and other experts. Despite this, our prior is certainly subjective, and the results we present here should not be taken to supplant the results earlier in the article. Our prior is a $\text{Gamma}(30, 3000)$ distribution, truncated to the interval $[0.002, 0.025]$, and discretized to support

points 0.002, 0.003, . . . , 0.024, 0.025. This prior has mean of approximately 0.01 and places most of its mass between 0.005 and 0.015.

To obtain results, we run the MCMC algorithm described in section 4 for p taking each value in the support, then average the samples with weights given by the prior. Results are shown in Table 4. As expected, the posterior mean for the undercount is similar to what we obtained with p fixed at .01, but credible intervals are considerably wider. This is indeed the main difference between putting a prior on p and fixing p : because the undercount is sensitive to p , putting a prior on p results in much wider credible intervals for the undercount than with any fixed value of p . In most cases, the results for R_0 and ρ_T do not change considerably compared to any of the results with fixed p , since these quantities were not particularly sensitive to the assumed value of p .

Table 4. R_0 , ρ_T , and the estimated undercount (95% posterior credible intervals) calculated by averaging samples over a prior on p .

	R_0	ρ_T	Undercount
CA	3.09 (1.89,3.87)	1.09 (1.03,1.19)	9.00 (6.33,13.29)
FL	3.20 (2.46,3.68)	1.02 (0.88,1.17)	7.00 (4.54,9.75)
NY	3.89 (3.64,4.00)	0.68 (0.65,0.70)	10.00 (7.02,14.00)
WA	2.74 (2.08,3.09)	0.75 (0.66,0.83)	7.00 (4.89,10.20)

6. Sensitivity Analysis

In this section we address sensitivity of our findings to the value we have chosen for θ and to the assumption that the death count data is an accurate representation of the number of COVID-19 attributable deaths.

6.1. Alternative θ

One of the first major interventions outside of China to stem the spread of COVID-19 was the lockdown in Italy on March 9, 2020. Around 20 days later, the number of daily deaths in Italy showed a sustained downward trend. We adjust the θ used in the main text to an average time to death of 20 days consistent with Italy. We then see what effects using this θ might have on our results.

To create a θ based on an average time to death of about 20 days, we induce a correlation between the time to symptom onset and the time to death for those who died. Specifically, we simulate the time from infection to symptom onset as before from a Poisson-Gamma distribution with parameters $\{a, b\} = \{5.5, 1.1\}$ to be consistent with [Lauer et al. \(2020\)](#). We then simulate the time from symptom onset to death from a Poisson-

Gamma distribution with parameters $\{a, b\} = \{t_1 - \bar{t}_1 + 5.5, 1.1\}$, where t_1 is the simulated time from infection to symptom onset. This induces a positive correlation such that those who had faster symptom onset also had faster time to death from symptom onset on average. This results in an alternative θ with average time to death of about 20.5 days and a correlation between incubation time and time from symptom onset to death of about .35. This alternative θ is shown in Figure 5.

Figure 5. Alternative θ .

6.2. Excess Mortality-Inflated Death Counts

Using the raw death count data assumes that deaths due to COVID-19 have been accurately recorded. Recent studies on excess mortality suggest that, in some states, there has been underreporting of COVID-19 deaths. For example, [Weinberger et al. \(2020\)](#) calculate excess mortality due to pneumonia and influenza (P&I) from February 9, 2020, to March 28, 2020. They note that in California during this time period, while only 101 COVID-19 deaths were reported, there were 399 excess deaths due to P&I during that same time period. If we interpret all of those deaths to be COVID-related, this implies that that only about one quarter of COVID-19 deaths were reported as such. Similar analysis for Florida and Washington showed reporting rates of about 30% and 100%, respectively. Though New York did not show underreporting when comparing only to pneumonia and influenza deaths, due to the large volume of cases there, the authors noted a reporting rate of 60% in New York City and 40% in New York State (excluding New York City) based on all-cause excess mortality. We assume a rough estimate of underreporting for New York State in its entirety is around 50%.

Based on these findings, we inflate the number of observed deaths on each day by a factor of four, three, two, and one for California, Florida, New York, and Washington, respectively. By inflating the data in this way, this implicitly assumes that excess mortality during this time period is all due to COVID-19 infections. While many of the excess deaths during this time period may be the result of the environment caused by COVID-19 (e.g., not going to the hospital for other illnesses out of fear of contracting COVID-19, reduction in medical services

available to non COVID-19 patients, and saturation of emergency medical services), it is unlikely that all are directly attributable to infection itself. This calculation also assumes that the rate at which COVID-related deaths were not attributed to COVID-19 in official records is constant across time. As testing capacity has expanded over time and the estimates we use are based on data only through March 28, this is unlikely to be true. Because not all excess mortality is due to COVID-19 disease, these two cases (raw death counts presented as our main results and the excess mortality-inflated death counts presented here) likely bracket the true number of deaths. However, more updated estimates of P&I excess mortality do not seem to be available. Even all-cause excess mortality estimates from the CDC are lagged by a few weeks.

6.3. Alternative Results

Results from applying our model with the alternative θ and for excess mortality (labeled as ‘multiplier’) are given below. Each of these were run for $p = .01$. Table 5 gives the estimated R_0 under each of these alternative scenarios. In neither case do the estimates of R_0 substantively change under these alternative scenarios. Point estimates in the alternative scenarios fall within the posterior credible intervals of the model fit in the primary scenarios given in the main text.

Table 5. Estimated posterior mean of R_0 , with 95% posterior credible interval shown in parentheses.

	CA	FL	NY	WA
multiplier	2.98 (1.73,3.74)	3.38 (2.93,3.66)	3.94 (3.81,4.00)	2.72 (2.00,3.06)
alternative theta	3.26 (2.00,3.97)	3.43 (2.40,3.98)	3.58 (2.87,3.99)	2.58 (1.72,3.04)

Note: Rows correspond to two alternative scenarios considered: death counts adjusted using an excess mortality multiplier and alternative value of θ used. In both cases, $p = .01$.

Table 6 gives estimates of ρ_T under each of the alternative scenarios. Once again, there are no substantive changes to our estimates. It is still the case that the estimated ρ_T s imply a slow rise in deaths in California, a plateau or very slow increase in Florida, and continued downward trends in New York and Washington.

Table 6. Estimated posterior mean of ρ_T , with 95% posterior credible interval shown in parentheses.

	CA	FL	NY	WA
multiplier	1.06 (1.02,1.11)	1.07 (0.97,1.16)	0.63 (0.63,0.64)	0.75 (0.66,0.84)
alternative theta	1.09 (1.05,1.14)	1.02 (0.91,1.19)	0.78 (0.77,0.80)	0.83 (0.75,0.91)

Note: Rows correspond to two alternative scenarios considered: death counts adjusted using an excess mortality multiplier and alternative value of θ used. In both cases, $p = .01$.

Table 7 shows the estimated factor by which current testing undercounts the actual number of infections as estimated by our model. This is analogous to Table 2 in the main text. Unsurprisingly, we found that inflating the number of COVID-19 deaths results in higher estimates of the total infected and, consequently, a larger undercount factor (relative to results for $p = .01$ in the main text). Using the alternative value of θ has minimal impact on estimates of the undercount, with three of the four point estimates calculated under the alternative θ falling within the posterior credible intervals obtained in the main analysis. Notably, under the alternative θ , New York sees a statistically significant though not substantively meaningful decrease in the undercount factor.

Table 7. Estimated posterior mean of undercount, with 95% posterior credible interval shown in parentheses for alternative scenarios: Deaths adjusted according to excess mortality multipliers or using the alternative θ .

	CA	FL	NY	WA
multiplier	34.69 (32.93,36.79)	19.97 (18.58,21.31)	19.31 (19.07,19.56)	6.85 (6.22,7.50)
alternative theta	8.51 (7.84,9.19)	6.20 (5.46,7.01)	8.87 (8.72,9.03)	6.93 (6.28,7.64)

Note: For both alternative scenarios, $p = .01$.

In all but one of the calculations under alternative scenarios, we find that the alternative parameters do not substantively change the results. The only exception to this is the undercount factors, which—unsurprisingly—exhibit an increase when we add a multiplier to the time series to account for possible undercounting. However, early evidence suggests that accounting for under-reporting would change the death counts by at most a factor of 4, and this is likely a significant overestimate for the reasons outlined above. Since increasing the death counts with p fixed has a similar effect on our analysis to decreasing p with the death counts fixed, and the values of p we consider in the main analysis range from 0.2% up to 2.5% (a difference of a factor of 10), the variation in results that is possible due to undercounting of deaths is already well represented by variation in p in the scenarios we consider.

7. Conclusion

We have built a model for the transmission of SARS-CoV-2 using only information that has a reasonable chance of being measured correctly: the observed number of daily deaths, timing of containment measures, and information on the clinical progression of the disease. In contrast to models that use a wider range of information that may be less precisely measured, we believe the main strengths of our approach are that it

prioritizes simplicity, interpretability, and identifiability. Because not all parameters one might want to estimate in this setting are separately identifiable, we have carefully specified our model so that all estimated parameters are identifiable conditional on quantities for which we have high-quality auxiliary information. Our model also has a proper likelihood and is fit to data in the Bayesian paradigm, rather than relying on ad hoc calibration of model parameters to produce trajectories resembling the observed data. This allows us to formally account for uncertainty in all of our estimates by posterior credible intervals. Our model is underpinned by an SIR model of infection dynamics to link observed deaths to the underlying unobserved infections. It would not be difficult with our approach to substitute some other model in place of the SIR. In particular, any other compartmental model, such as an SEIR model could be used, or the model could be elaborated to incorporate more change points in the parameters of the compartmental model to account for fine grained policy analysis.²

We estimate that official case counts substantially undercount the number of infections. This is not a surprise. Despite recent increases in testing capabilities, our analysis suggests that testing continues to undercount the number of infections by a factor between 6 and 20 for $p = .01$ and $p = .005$, the values of p we consider most likely. Future work could incorporate IFRs that vary by time or other covariates.

Seroprevalence studies done in two states in the United States result in similar conclusions with respect to undercount. Such studies have only recently become available and do not give snapshots of undercount over time. We believe that the correspondence between our estimates and serological studies that have since been done points to the utility of such a model for estimating key quantities of interest (such as the extent of undercounting), especially early in the epidemic when testing capabilities are yet built up.

While the seroprevalence studies using random design are just now reporting results, our initial estimates were first posted in a preprint in late March. This suggests that approaches similar to what we've described here could be useful in providing indications of the undercount and the total number of infections of emerging diseases in the early days of its spread. During this time, when the number of cases is relatively low and the rate of spread is nearly exponential, accounting for the time lag between infection and death as we do here can lead to substantial differences in estimates of the total number of cases relative to naive methods, such as simply dividing the total number of deaths by the IFR.

Our model also suggests wide variability in the initial transmissibility of the disease prior to intervention policies as well as wide variability in the effects of those policies. For example, we estimate that New York had the highest R_0 s at the outset of the outbreak. This is unsurprising considering New York City has so far been the epicenter of the epidemic in the United States. However, estimates of ρ_T suggest that the spread of the virus in New York has slowed to the lowest rate among the states considered in this analysis. Whereas in New York, estimates of ρ_T indicate a continued, fairly rapid decrease in deaths, estimates of ρ_T for California and Florida are very near to one, indicating that sharp declines in deaths are unlikely.

All of these projections are conditioned on the current policies staying in place and their effects on transmission remaining constant. Tightening or relaxing mitigation policies, or ‘quarantine fatigue’ leading to a decrease in compliance by the public, would impact the trajectory of the spread. Estimates of ρ_T near one indicate that there is little room for relaxation of mitigation policies without causing an increase in the number of infections and deaths. As mitigation measures are tightened and relaxed over time, an elaboration of our model incorporating multiple change points in the parameters may be useful for future projections and evaluation of the continued efficacy of the measures.

Acknowledgments

We thank Alexander D’Amour for finding a misstated equation in an earlier version. We thank Amy Herring and David Dunson for comments on an early draft.

Disclosure Statement

This work was supported by grants to the Human Rights Data Analysis Group by the John D. and Catherine T. MacArthur Foundation and the Oak Foundation. Kristian Lum and James Johndrow acknowledge funding from National Institute of Environmental Health Sciences (NIEHS) grant 3R01ES028804-03S.

References

- Altieri, N., Barter, R. L., Duncan, J., Dwivedi, R., Kumbier, K., Li, X., Netzorg, R., Park, B., Singh, C., Tan, Y. S., Tang, T., Wang, Y. W., Zhang, C., & Yu, B. (2020). Curating a COVID-19 data repository and forecasting county-level death counts in the United States. *Harvard Data Science Review*, (Special Issue 1). <https://doi.org/10.1162/99608f92.1d4e0dae>
- Angelopoulos, A. N., Pathak, R., Varma, R., & Jordan, M. I. (2020). On identifying and mitigating bias in the estimation of the COVID-19 case fatality rate. *Harvard Data Science Review*, (Special Issue 1). <https://doi.org/10.1162/99608f92.f01ee285>
- Bonislawski, A. (2020, April 27). New York, California serology studies give early estimates of COVID-19 prevalence. *360Dx*. <https://www.360dx.com/infectious-disease/new-york-california-serology-studies-give-early-estimates-covid-19-prevalence>
- Cavitt, M. (2020, April 12). Shortage of key chemical forces hospitals to prioritize and limit coronavirus testing. *The Oakland Press*. https://www.theoaklandpress.com/news/coronavirus/shortage-of-key-chemical-forces-hospitals-to-prioritize-and-limit-coronavirus-testing/article_1861bcba-7b52-11ea-8749-8f80d7d167fe.html

Centers for Disease Control and Prevention. (2020). *Overview of testing for SARS-CoV-2*.

https://www.cdc.gov/coronavirus/2019-ncov/hcp/testing-overview.html?CDC_AA_refVal

Ferguson, N. M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G., Dighe, A., Dorigatti, I., Fu, H., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Okell, L. C., van-Elsland, S., & Thompson, H. (2020, March 16). *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand*.

<https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf>

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., Méléodie, M., Team, I. C. C.-19 R., Chani, A. C., Donnelly, C. A., Riley, S., Vollmer, M. A. C., Ferguson, N. M., Okell, L. C., & Bhatt, S. (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>

Fleshler, D., & Goodman, C. K. (2020, June 20). The changing face of coronavirus in Florida: Younger victims, and what that means. *Sun Sentinel*. <https://www.sun-sentinel.com/coronavirus/fl-ne-coronavirus-florida-younger-20200620-zdbyrk6h25cwxak5h74cchirkestory.html>

Golding, N., Russell, T. W., Abbott, S., Hellewell, J., Pearson, C. A. B., Zandvoort, van K., Jarvis, C. I., Gibbs, H., Liu, Y., Eggo, R. M., Edmunds, J. W., & Kucharski, A. J. (2020). Reconstructing the global dynamics of under-ascertained COVID-19 cases and infections. *medRxiv*. <https://doi.org/10.1101/2020.07.07.20148460>

Haario, H., Saksman, E., & Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 7(2), 223–242. <https://doi.org/10.2307/3318737>

Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM Review*, 42(4), 599–653. <https://doi.org/10.1137/S0036144500371907>

Howden, L. M., & Meyer, J. A. (2011). *Age and sex composition, 2010*. U.S. Department of Commerce, Economics; Statistics Administration. <https://www.census.gov/content/dam/Census/library/publications/2011/dec/c2010br-03.pdf>

Institute for Health Metrics and Evaluation. (2020). *Main updates on IHME COVID-19 predictions since April 29, 2020*. http://www.healthdata.org/sites/default/files/files/Projects/COVID/Estimation_update_050420.pdf

Johnson, C. Y., & McGinley, L. (2020, March 7). What went wrong with the coronavirus tests in the U.S. *Washington Post*. https://www.washingtonpost.com/health/what-went-wrong-with-the-coronavirus-tests/2020/03/07/915f5dea-5d82-11ea-b29b-9db42f7803a7_story.html

Korea Centers for Disease Control and Prevention. (2020, May 17). Updates on COVID-19 in Republic of Korea [Press release]. https://is.cdc.go.kr/upload_comm/syview/doc.html?fn=158968959671100.pdf&rs=/upload_comm/docu/0030/

Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A., Reich, N. G., & Lessler, J. (2020). The incubation period of Coronavirus Disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, 172(9), 577–582. <https://doi.org/10.7326/M20-0504>

Lecher, C., Varner, M., & Martin, E. (2020, April 16). Can you get a Coronavirus test? It may depend on where you live. *The Markup*. <https://themarkup.org/coronavirus/2020/04/16/can-you-get-a-coronavirus-test>

Lewnard, J. A., Liu, V. X., Jackson, M. L., Schmidt, M. A., Jewell, B. L., Flores, J. P., Jentz, C., Northrup, G. R., Mahmud, A., Reingold, A. L., Peterson, M., Jewell, N. P., Young, S., & Bellows, J. (2020). Incidence, clinical outcomes, and transmission dynamics of hospitalized 2019 coronavirus disease among 9,596,321 individuals residing in California and Washington, United States: A prospective cohort study. *medRxiv*. <https://doi.org/10.1101/2020.04.12.20062943>

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., ... Feng, Z. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 382(13), 1199–1207. <https://doi.org/10.1056/NEJMoa2001316>

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., & Shaman, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490), 489–493. <https://doi.org/10.1126/science.abb3221>

Menachemi, N., Yiannoutsos, C. T., Dixon, B. E., Duszynski, T. J., Fadel, W. F., Wools-Kaloustian, K. K., Needleman, N. U., Box, K., Caine, V., Norwood, C., Weaver, L., & Halverson, P. K. (2020). Population point prevalence of SARS-CoV-2 infection based on a statewide random sample—Indiana, April 25–29, 2020. *Morbidity and Mortality Weekly Report*, 69(29), 960–964. <https://doi.org/10.15585/mmwr.mm6929e1>

Murray, C. (2020). Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator days and deaths by US state in the next 4 months. *medRxiv*. <https://doi.org/10.1101/2020.03.27.20043752>

Patino, M. (2020, May 1). Coronavirus data in the U.S. is terrible, and here's why. *Bloomberg CityLab*. <https://www.bloomberg.com/news/articles/2020-05-01/why-coronavirus-reporting-data-is-so-bad>

Perkins, A., Cavany, S. M., Moore, S. M., Oidtmann, R. J., Lerch, A., & Poterek, M. (2020). Estimating unobserved SARS-CoV-2 infections in the United States. *medRxiv*. <https://doi.org/10.1101/2020.03.15.20036582>

Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. M., Davies, N., for Mathematical Modelling of Infectious Diseases COVID-19 Working Group, C., Jit, M., & Klepac, P. (2020). The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. *The Lancet Public Health*, 5(5), e261–e270. [https://doi.org/10.1016/S2468-2667\(20\)30073-6](https://doi.org/10.1016/S2468-2667(20)30073-6)

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. *medRxiv*. <https://doi.org/10.1101/2020.03.04.20031104>

Roberts, G. O., & Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4), 351–367. <https://doi.org/10.1214/ss/1015346320>

Russell, T. W., Hellewell, J., Jarvis, C. I., van-Zandvoort, K., Abbott, S., Ratnayake, R., nCov working group, C., Eggo, R. M., & Kucharski, A. J. (2020). Estimating the infection and case fatality ratio for COVID-19 using age-adjusted data from the outbreak on the Diamond Princess cruise ship. *medRxiv*. <https://doi.org/10.1101/2020.03.05.20031773>

Silver, N. (2020, April 4). Coronavirus case counts are meaningless. *FiveThirtyEight*. <https://fivethirtyeight.com/features/coronavirus-case-counts-are-meaningless>

Song, P. X., Wang, L., Zhou, Y., He, J., Zhu, B., Wang, F., Tang, L., & Eisenberg, M. (2020). An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China. *medRxiv*. <https://doi.org/10.1101/2020.02.29.20029421>

Weinberger, D., Cohen, T., Crawford, F., Mostashari, F., Olson, D., Pitzer, V., Reich, N. G., Russi, M., Simonsen, L., Watkins, A., & Vibound, C. (2020). Estimating the early death toll of COVID-19 in the United States. *medRxiv*. <https://doi.org/10.1101/2020.04.15.20066431>

Weissman, G. E., Crane-Droesch, A., Chivers, C., Luong, T., Hanish, A., Levy, M. Z., Lubken, J., Becker, M., Draugelis, M. E., Anesi, G. L., Brennan, P. J., Christie, J. D., Hanson, C. W., Mikkelsen, M. E., & Halpern, S. D. (2020). Locally informed simulation to predict hospital capacity needs during the COVID-19 pandemic. *Annals of Internal Medicine*, 173(1), 21–28. <https://doi.org/10.7326/M20-1260>

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., & Cao, B. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *The Lancet*, 395(10229), 1054–1062. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)

Appendices

Appendix A. Markov Chain Monte Carlo Diagnostics

Figure A1 shows representative trace plots of model parameters for $p = .01$ for New York. Once adaptation begins, the algorithm evidently mixes quite well. A more formal assessment of convergence is given in Table A1, which shows the estimate (and upper 95% confidence interval) for the Gelman-Rubin potential scale reduction factor for each parameter, calculated using five independent chains with different starting values. Table A2 shows the multivariate potential scale reduction factor. Values of the statistics ‘near one’ indicate that nonconvergence has not been detected.

Figure A1. Markov chain Monte Carlo samples of γ^{-1} , η , ϕ , R_0 , and T_0 for New York in the case where $p = .01$.

Table A1. Gelman-Rubin convergence diagnostic point estimate (upper 95% confidence interval) for parameters of model calculated with five independent runs of Markov Chain Monte Carlo.

	California	Florida	New York	Washington

T_0	1.013 (1.035)	1.002 (1.003)	1.001 (1.001)	1.001 (1.002)
R_0	1.011 (1.028)	1.002 (1.004)	1.001 (1.001)	1.001 (1.002)
$1/\text{Gamma}$	1.012 (1.03)	1.003 (1.004)	1 (1.001)	1.001 (1.002)
Φ	1.004 (1.008)	1.004 (1.006)	1.001 (1.002)	1.001 (1.002)
Eta	1.006 (1.012)	1.012 (1.023)	1.001 (1.002)	1.005 (1.01)

Table A2. Multivariate Gelman-Rubin potential scale reduction factor.

	MPSRF
California	1.03
Florida	1.01
New York	1.00
Washington	1.00

Appendix B. Simulation Study

We conduct a simulation study to assess sensitivity of the method to various assumptions. We analyze five simulation cases. In the first four cases, the daily new infections are obtained from the two-period susceptible-infected-recovered (SIR) model in (3.4), but we explore several ways in which other assumptions can be violated. The parameters of the SIR model are: $T_0 = 35$ (corresponding to February 4, since we index January 1 as time 1), $\gamma = 1/9$, $R_0 = 3.8$, $\phi = .3$, $\eta = 1.4$, and $T_1 = 78$ (corresponding to March 18, 2020). These parameters were selected to give a death series looking roughly like that observed in New York with $p = .01$. In every case, the Poisson model in (3.3) is used to generate deaths given new infections with $p = .01$ and θ shown in Figure 2. We then statistically infer using our method. Since we already provide results for a range of values of p , we do not include variation in p here.

1. The model is properly specified.
2. We use the alternative (incorrect) value of θ displayed in Figure 5 when statistically inferring.
3. We assume that T_1 is one week too early (corresponding to March 11 instead of the true value of March 18) when statistically inferring.
4. We assume that T_1 is one week too late (corresponding to March 25 instead of the true value of March 18) when statistically inferring.

5. The true new infection series ν is generated by a susceptible-exposed-infected-removed (SEIR) model with two time periods rather than an SIR model with two time periods, but otherwise the model is correctly specified. The SEIR model is a popular alternative to the SIR model for modeling SARS-CoV-2. We choose the parameters of the SEIR model to give a death series similar to that for the other four cases.

Figure A2 shows the simulated data used in Cases 1–4 (left column) and Case 5 (right column). The top panels show deaths, the middle row of panels shows new infections, and the bottom row shows the state variables of the ODE model (SIR in the left column and SEIR in the right column) through a period of 121 days (corresponding to the dates January 1, 2020, through April 30, 2020).

Figure A2. Simulated data for examples. Case 1 (shown in the left column) is a SIR model. Case 2 (shown in the right column) is a SEIR model. Values for the bottom panel displaying the SEIR model are displayed as a proportion of the total population and the S compartment has been omitted so that detail for the E, I, and R compartments would be visible.

Table A3. Estimated posterior mean and 95% posterior credible intervals for each simulation scenario.

	R_0	ρ_T	Total Infections
1	3.54 (2.82,3.98)	0.72 (0.70,0.74)	11.57 (11.23,11.91)

2	3.40 (2.30,3.98)	0.83 (0.82,0.86)	10.70 (10.51,10.90)
3	1.77 (1.61,2.14)	0.92 (0.92,0.92)	10.63 (10.44,10.81)
4	2.87 (2.63,3.33)	0.48 (0.45,0.51)	10.79 (10.54,11.06)
5	3.38 (2.45,3.98)	0.74 (0.72,0.77)	12.19 (11.85,12.54)

We report estimates for R_0 , ρ_T , and the proportion of the total population that has ever been infected as of the last day of the simulation in Table A3. In Cases 1-4, the true value of ρ_T is 0.72 and the true value of the total proportion of the population that has ever been infected is 11.7%, while in case 5 they are 0.72 and 12.1%, respectively. Case 1 demonstrates that we recover the true parameters using our method (all values lie within the posterior credible intervals). Case 2 shows that the only result that has noticeable sensitivity to incorrect specification of θ is the total proportion infected. Since the true θ has a longer expected time to death than the θ used for inference in this case, it makes sense that we slightly underestimate the true proportion of the population that has been infected. Cases 3 and 4 show that estimates of R_0 , and, to a lesser extent, ρ_T are sensitive to incorrect specification of the time point at which the model parameters change T_1 , but the total proportion of the population that has ever been infected is actually not very sensitive to misspecification of this parameter. Note, however, that being wrong by one week about T_1 typically has a noticeable effect on the model fit, to the extent that one would suspect a problem simply by looking at posterior summaries. This can be seen in Figure A3, which shows the observed death data and the posterior predictive point estimates and intervals for deaths for Case 4. The model fits poorly and is unable to find any set of parameters that reproduces the observed deaths. Thus while there is considerable sensitivity to misspecification of T_1 , in the cases we considered, this misspecification can be easily detected by a posterior predictive check. Finally, Case 5 shows that, even when the SIR model itself is misspecified and the true generating process includes an exposed period in which people are not infectious, we are able to recover the truth with reasonable accuracy.

Overall, the results suggest that, although some types of misspecification can lead to poor estimates of R_0 , these types of misspecification may be detectable by posterior predictive checks. Furthermore, the total proportion of the population that has ever been infected, while highly sensitive to the p parameter, is reasonably robust to the types of misspecification we explore in this section. This suggests that estimates of the undercount—which will be good if and only if the estimate of the total population ever infected is good—are fairly robust to misspecification when p is known.

Figure A3. Observed death series (red), posterior mean of deaths (dashed black line), and posterior 95% credible interval for deaths (gray region) for Case 4.

©2020 James Johndrow, Patrick Ball, Maria Gargiulo, and Kristian Lum. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](#), except where otherwise indicated with respect to particular material included in the article.

Footnotes

1. See reports on variability in the United States in Lecher et al. (2020, April 16), Silver (2020, April 4), and Patino (2020, May 1). [↗](#)
2. See, e.g., Fleshler and Goodman (2020, June 20) which notes the median age of COVID-19 cases has shifted younger. [↗](#)
3. It is impossible to make a clean distinction between these two types of parameters. For example, if the rate of spread allows for the number of people requiring care to overwhelm existing health care

infrastructures, the case fatality rate will increase, as those who require medical support to survive but cannot receive it will die. [↵](#)

4. See 2010 U.S. Census data (Howden & Meyer, 2011). [↵](#)

5. Code for our analysis is available at <https://github.com/jamesjohndrow/Covid19-modeling>. [↵](#)

6. The study results have not yet been released, but Bonislowski (2020) provides a news media summary. [↵](#)

7. A GitHub repository with extensions to the current model and applications to other countries' data can be found at <https://github.com/paulo-o/covid19>. This repository is still in development and contains ongoing work by two of this article's authors, James Johndrow and Kristian Lum, as well as Paulo Orenstein [Instituto de Matemática Pura e Aplicada (IMPA)]. [↵](#)

References

- Altieri, N., Barter, R. L., Duncan, J., Dwivedi, R., Kumbier, K., Li, X., Netzorg, R., Park, B., Singh, C., Tan, Y. S., Tang, T., Wang, Y. W., Zhang, C., & Yu, B. (2020). Curating a covid-19 data repository and forecasting county-level death counts in the united states. *aRxiv*. <https://arxiv.org/abs/2005.07882> [↵](#)
- Angelopoulos, A. N., Pathak, R., Varma, R., & Jordan, M. I. (2020). On identifying and mitigating bias in the estimation of the COVID-19 case fatality rate. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.f01ee285> [↵](#)
- Cavitt, M. (n.d.). Shortage of key chemical forces hospitals to prioritize and limit coronavirus testing. *The Oakland Press*. https://www.theoaklandpress.com/news/coronavirus/shortage-of-key-chemical-forces-hospitals-to-prioritize-and-limit-coronavirus-testing/article_1861bcba-7b52-11ea-8749-8f80d7d167fe.html [↵](#)
- Ferguson, N. M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G., Dighe, A., Dorigatti, I., Fu, H., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Okell, L. C., van Elsland, S., & Thompson, H. (n.d.). *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand*. <https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf> [↵](#)
- Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., Méléodie, M., Team, I. C. C.-19 R., Chani, A. C., Donnelly, C. A., Riley, S., Vollmer, M. A. C., Ferguson, N. M., Okell, L. C., & Bhatt, S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584, 257–261. <https://doi.org/10.1038/s41586-020-2405-7> [↵](#)
- for Disease Control, C., & Prevention. (2020). Overview of testing for sars-cov-2. https://www.cdc.gov/coronavirus/2019-ncov/hcp/testing-overview.html?CDC_AA_refVal=https%253A%252F%252Fwww.cdc.gov%252Fcoronavirus%252F2019-ncov%252Fhcp%252Fclinical-criteria.html [↵](#)

- for Disease Control, K. C., & Prevention. (n.d.). *Updates on covid-19 in republic of korea*.
https://is.cdc.go.kr/upload_comm/syview/doc.html?fn=158968959671100.pdf&rs=/upload_comm/docu/0030/ ↵
- for Health Metrics, I., & Evaluation. (2020). Main updates on ihme covid-19 predictions since april 29, 2020. http://www.healthdata.org/sites/default/files/files/Projects/COVID/Estimation_update_050420.pdf ↵
- Golding, N., Russell, T. W., Abbott, S., Hellewell, J., Pearson, C. A. B., Zandvoort, van K., Jarvis, C. I., Gibbs, H., Liu, Y., Eggo, R. M., Edmunds, J. W., & Kucharski, A. J. (2020). Reconstructing the global dynamics of under-ascertained covid-19 cases and infections. medRxiv.
<https://doi.org/10.1101/2020.07.07.20148460> ↵
- Haario, H., Saksman, E., & Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 7(2), 223–242. <https://doi.org/10.2307/3318737> ↵
- Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM Review*, 42(4), 599–653.
<https://doi.org/10.1137/S0036144500371907> ↵
- Johnson, C. Y., & McGinley, L. (n.d.). What went wrong with the coronavirus tests in the U.S. *The Washington Post*. https://www.washingtonpost.com/health/what-went-wrong-with-the-coronavirus-tests/2020/03/07/915f5dea-5d82-11ea-b29b-9db42f7803a7_story.html ↵
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A., Reich, N. G., & Lessler, J. (2020). The incubation period of Coronavirus Disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine*, 172(9), 577–582.
<https://doi.org/doi.org/10.7326/M20-0504> ↵
- Lewnard, J. A., Liu, V. X., Jackson, M. L., Schmidt, M. A., Jewell, B. L., Flores, J. P., Jentz, C., Northrup, G. R., Mahmud, A., Reingold, A. L., Peterson, M., Jewell, N. P., Young, S., & Bellows, J. (2020). Incidence, clinical outcomes, and transmission dynamics of hospitalized 2019 coronavirus disease among 9,596,321 individuals residing in california and washington, united states: A prospective cohort study. medRxiv.
<https://doi.org/10.1101/2020.04.12.20062943> ↵
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., ... Feng, Z. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 382(13), 1199–1207. <https://doi.org/10.1056/NEJMoa2001316> ↵
- Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., & Shaman, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*, 368(6490), 489–493. <https://doi.org/10.1126/science.abb3221> ↵
- Menachemi, N., Yiannoutsos, C. T., Dixon, B. E., Duszynski, T. J., Fadel, W. F., Wools-Kaloustian, K. K., Needleman, N. U., Box, K., Caine, V., Norwood, C., Weaver, L., & Halverson, P. K. (2020). Population point prevalence of SARS-CoV-2 infection based on a statewide random sample—Indiana, April 25–29, 2020. *Morbidity and Mortality Weekly Report*, 69(29), 960–964. <https://doi.org/10.15585/mmwr.mm6929e1> ↵

- Murray, C. J. L. (2020). Forecasting covid-19 impact on hospital bed-days, icu-days, ventilator days and deaths by us state in the next 4 months. medRxiv.
<https://doi.org/https://doi.org/10.1101/2020.03.27.20043752> ↵
- Perkins, A., Cavany, S. M., Moore, S. M., Oidtmann, R. J., Lerch, A., & Poterek, M. (2020). Estimating unobserved sars-cov-2 infections in the united states. medRxiv.
<https://doi.org/10.1101/2020.03.15.20036582> ↵
- Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. M., Davies, N., Mathematical Modelling of Infectious Diseases COVID-19 Working Group, for C., Jit, M., & Klepac, P. (2020). The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. *The Lancet Public Health*. [https://doi.org/10.1016/S2468-2667\(20\)30073-6](https://doi.org/10.1016/S2468-2667(20)30073-6) ↵
- 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. *medRxiv*.
<https://doi.org/10.1101/2020.03.04.20031104> ↵
- Roberts, G. O., & Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4), 351–367. <https://doi.org/10.1214/ss/1015346320> ↵
- Russell, T. W., Hellewell, J., Jarvis, C. I., van Zandvoort, K., Abbott, S., Ratnayake, R., group, nCov working C., Eggo, R. M., & Kucharski, A. J. (2020). Estimating the infection and case fatality ratio for covid-19 using age-adjusted data from the outbreak on the diamond princess cruise ship. medRxiv.
<https://doi.org/10.1101/2020.03.05.20031773> ↵
- Song, P. X., Wang, L., Zhou, Y., He, J., Zhu, B., Wang, F., Tang, L., & Eisenberg, M. (2020). An epidemiological forecast model and software assessing interventions on covid-19 epidemic in china. medRxiv. <https://doi.org/10.1101/2020.02.29.20029421> ↵
- Weinberger, D., Cohen, T., Crawford, F., Mostashari, F., Olson, D., Pitzer, V., Reich, N. G., Russi, M., Simonsen, L., Watkins, A., & Vibound, C. (2020). Estimating the early death toll of covid-19 in the united states. medRxiv. <https://doi.org/10.1101/2020.04.15.20066431> ↵
- Weissman, G. E., Crane-Droesch, A., Chivers, C., Luong, T., Hanish, A., Levy, M. Z., Lubken, J., Becker, M., Draugelis, M. E., Anesi, G. L., Brennan, P. J., Christie, J. D., Hanson, C. W., Mikkelsen, M. E., & Halpern, S. D. (2020). Locally informed simulation to predict hospital capacity needs during the COVID-19 pandemic. *Annals of Internal Medicine*, 173(1), 21–28. <https://doi.org/10.7326/M20-1260> ↵
- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., & Cao, B. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 395, 1054–1062. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3) ↵