

# **Are we there yet? Adaptive SIR model for continuous estimation of COVID-19 infection rate and reproduction number in the United States**

Mark Shapiro, Fazle Karim, Guido Muscioni, Abel Saju Augustine

Submitted to: Journal of Medical Internet Research  
on: September 17, 2020

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 4

Supplementary Files..... 19

    Figures ..... 20

        Figure 1..... 21

        Figure 2..... 22

        Figure 3..... 23

# Are we there yet? Adaptive SIR model for continuous estimation of COVID-19 infection rate and reproduction number in the United States

Mark Shapiro<sup>1</sup> PhD; Fazle Karim<sup>1</sup> PhD; Guido Muscioni<sup>1</sup> MSc; Abel Saju Augustine<sup>1</sup> MSc

<sup>1</sup>Anthem, Inc. Indianapolis US

## Corresponding Author:

Mark Shapiro PhD  
Anthem, Inc.  
220 Virginia Avenue  
Indianapolis  
US

## Abstract

**Background:** The dynamics of the COVID-19 epidemic vary due to local population density and policy measures. When making decisions, policy makers consider an estimate of the effective reproduction number  $R_t$  which is the expected number of secondary infections by a single infected individual.

**Objective:** We propose a simple method for estimating the time-varying infection rate and reproduction number  $R_t$ .

**Methods:** We use a sliding window approach applied to a Susceptible-Infectious-Removed model. The infection rate is estimated using the reported cases for a seven-day window to obtain continuous estimation of  $R_t$ . The proposed adaptive SIR (aSIR) model was applied to data at the state and county levels.

**Results:** The aSIR model showed an excellent fit for the number of reported COVID-19 positive cases, a one-day forecast MAPE was less than 2.6% across all states. However, a seven-day forecast MAPE reached 16.2% and strongly overestimated the number of cases when the reproduction number was high and changing fast. The maximal  $R_t$  showed a wide range of 2.0 to 4.5 across all states, with the highest values for New York (4.4) and Michigan (4.5). We demonstrate that the aSIR model can quickly adapt to an increase in the number of tests and associated increase in the reported cases of infections. Our results also suggest that intensive testing may be one of the effective methods of reducing  $R_t$ .

**Conclusions:** The aSIR model provides a simple and accurate computational tool to obtain continuous estimation of the reproduction number and evaluate the impact of mitigation measures.

(JMIR Preprints 17/09/2020:24389)

DOI: <https://doi.org/10.2196/preprints.24389>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.  
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.  
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [http](#)

## Original Manuscript

## Are we there yet? Adaptive SIR model for continuous estimation of COVID-19 infection rate and reproduction number in the United States

Mark B Shapiro, Fazle Karim, Guido Muscioni, Abel Saju Augustine  
Anthem, Inc., USA

### Abstract

**Background:** The dynamics of the COVID-19 epidemic vary due to local population density and policy measures. When making decisions, policy makers consider an estimate of the effective reproduction number  $R_t$  which is the expected number of secondary infections by a single infected individual.

**Objective:** We propose a simple method for estimating the time-varying infection rate and reproduction number  $R_t$ .

**Methods:** We use a sliding window approach applied to a Susceptible-Infectious-Removed model. The infection rate is estimated using the reported cases for a seven-day window to obtain continuous estimation of  $R_t$ . The proposed adaptive SIR (*aSIR*) model was applied to data at the state and county levels.

**Results:** The *aSIR* model showed an excellent fit for the number of reported COVID-19 positive cases, a one-day forecast MAPE was less than 2.6% across all states. However, a seven-day forecast MAPE reached 16.2% and strongly overestimated the number of cases when the reproduction number was high and changing fast. The maximal  $R_t$  showed a wide range of 2.0 to 4.5 across all states, with the highest values for New York (4.4) and Michigan (4.5). We demonstrate that the *aSIR* model can quickly adapt to an increase in the number of tests and associated increase in the reported cases of infections. Our results also suggest that intensive testing may be one of the effective methods of reducing  $R_t$ .

**Conclusion:** The *aSIR* model provides a simple and accurate computational tool to obtain continuous estimation of the reproduction number and evaluate the efficacy of mitigation measures.

### Abbreviations

SIR: Susceptible-Infectious-Removed

aSIR: Adaptive Susceptible-Infectious-Removed

JHU: Johns Hopkins University

MAPE: mean absolute prediction error

SEIR: Susceptible-Exposed-Infected-Removed

### Introduction

We are in the middle of a global COVID-19 pandemic caused by the SARS-CoV-2 virus. As of September 2, 2020, over 6 million individuals in the United States have been reported positive for SARS-CoV-2. Modeling studies are key for understanding factors that drive the spread of the

disease and for developing mitigation strategies. Early modeling efforts forecasted very large numbers of infected individuals which would overwhelm healthcare systems in many countries [1–3]. These forecasts served as a call to action for policy makers to introduce policy measures including social distancing, travel restrictions, and eventually lockdowns to avoid the predicted catastrophe [4–6]. The mitigating policy measures have been successful in changing the dynamics of the epidemic and “flattening the curve” so that fewer people needed to seek treatment at any given time and as such not overwhelm the healthcare system.

One of the most fundamental metrics that describes the epidemic’s dynamics is the reproduction number  $R_t$  which is the expected number of secondary infections by a single infectious individual [7]. The idea that the course of an epidemic is determined by the rate of contact between susceptible and infectious individuals was proposed by William Hamer in 1906 [8]. Later, Kermack and McKendrick [9] showed that epidemics stop not when there are no susceptible individuals left, but rather when each infected individual can infect on average fewer than one more person. The reproduction number  $R_t$  depends on three factors: 1) the likelihood of infection per contact, 2) the period during which infectious individuals freely interact with those susceptible to contract the disease, and 3) rate of contact. The likelihood of infection per contact (factor 1) is determined by pathogen virulence and also by protective measures such as social distancing or wearing masks. Free interactions between infectious and susceptible individuals (factor 2) occur until the infectious individual is self-quarantined or hospitalized, either when the individual tests positive or symptoms become severe. Finally, the rate of contact (factor 3) is strongly affected by public health measures to mitigate risk [10], such as lockdowns during the COVID-19 epidemic. Thus, the reproduction number is determined by the biological properties of the pathogen and multiple aspects of social behavior. When  $R_t > 1$  the number of cases is expected to grow exponentially. The epidemic is contained when  $R_t$  decreases and remains below 1. Real-time estimation of  $R_t$  is critical for determining the effect of implemented mitigation measures and planning for the future.

We propose a method for the continuous estimation of infection rate and reproduction number  $R_t$  that reflects the effects of mitigation measures as well as immunity acquired by those who recover from the disease. We estimate  $R_t$  using a Susceptible-Infectious-Removed (SIR) model [9] that describes the dynamics of population compartments as follows: individuals start as Susceptible, are infected with the virus and become Infectious, and then move to the Removed compartment once they are quarantined or hospitalized, recover, or die. The SIR model is one of the simplest epidemiological models that still captures the main properties of an epidemic [11,12] and it has been widely used in epidemic modeling studies. In the majority of SIR modeling studies, the model parameters were constant. An SIR model with constant parameters, however, cannot be applied to the COVID-19 epidemic because various mitigating measures were introduced as the epidemic progressed. The effect of policy changes on COVID-19 dynamics has been modeled using a combination of an SIR model and Bayesian inference [13,14]. In these modeling studies, the infection spreading rate was assumed to be piece-wise linear between the three dates of policy changes. In another approach, continuous estimation

of the reproduction number and the effect of mitigation measures were obtained based on estimates of the distribution of the serial interval between the symptom onset in the primary and secondary cases [15–17]. The Bayesian inference methods as well as methods based on estimations of the serial interval include multiple parameters whose values are not estimated from the data. In contrast, we propose an adaptive SIR model (*aSIR*) in which only one parameter, the removal rate, is taken from the literature, while the second parameter, the infection rate, is continuously estimated from the data using a sliding window. A continuous estimate of the reproduction number  $R_t$  is then calculated using the infection rate estimate. The SIR model is described as a system of differential equations, and the key idea in our proposed method is that the initial conditions for each window are taken as values estimated for the previous window. The only additional hyperparameter is the length of the sliding window. The proposed method retains the conceptual and computational simplicity of SIR-type models and can be easily extended through the introduction of additional compartments supported by data.

## Methods

### Data

The data on daily and cumulative confirmed cases between February 29, 2020 and September 2, 2020 were obtained from John Hopkins University (JHU) and the dates of interventions by states (e.g. state of emergency, stay-at-home order) were obtained from Wikipedia [CITATION Wik20 \l 1033 ]. The JHU data were available at two levels of aggregation: county and state. JHU considers many sources for reporting these data; county level information is extracted from the website of the state's departments of health [CITATION Joh20 \l 1033 ] and state level data are extracted directly from the Centers for Disease Control website.

### Model

The SIR model is a system of ordinary differential equations:

$$\begin{aligned}\frac{dS}{dt} &= -\beta \frac{I}{N} S \\ \frac{dI}{dt} &= \beta \frac{I}{N} S - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

Here,  $S$  is the number of susceptible individuals;  $I$  is the number of infectious individuals who freely interact with others and can transmit the infection;  $R$  is the number of individuals removed from the other two compartments because they are quarantined or hospitalized, recover and acquire immunity, or die. Several COVID-19 government data sources provide the daily number of newly confirmed cases as well as a cumulative number of confirmed cases. Careful consideration is required to determine if these numbers should be attributed to the  $I$  or

*R* compartment. In the US, once an individual has been confirmed COVID-19 positive that person is expected to be either self-isolated or hospitalized. Therefore, we assigned the data on confirmed cases to the *R* compartment, and we fit the model on the cumulative number of confirmed cases.

The infection rate  $\beta$  is  $\beta = p * c$ , where  $p$  is the probability of infection during contact with an infectious individual, and  $c$  is the average number of contacts per day. We have no data that would allow us to estimate  $p$  and  $c$  separately, so we directly estimate  $\beta$  as is usually done when using SIR models.

The removal rate  $\gamma$  determines the rate with which the infected are removed from *I* to the *R* compartment. In the context of the COVID-19 epidemic,  $\gamma$  is determined by the time it takes for severe symptoms to appear so the person gets tested and is self-quarantined or hospitalized. Therefore, we will assume the duration of the infectious period as the average time it takes for the infected person to become isolated, not the overall time to recover. We assume that the person is infectious from the day they get infected before the symptoms appear [18–20]. The average time to develop symptoms has been reported as 5 to 6 days [21–23]. We assume the infectious period before developing severe symptoms is 6 days, so  $\gamma = 1/6$ .

## Time-variant Parameter Estimation

The *aSIR* model contains two parameters,  $\beta$  and  $\gamma$ , with  $\gamma = 1/6$  taken from the literature and  $\beta$  estimated using the reported data for each region of interest. The time-variant  $\beta(t)$  was estimated using a sliding window of  $\tau = 7$  days and step of  $s = 1$  day, with the estimated values for *S* and *I* from the previous window used as the initial conditions for the next window.

The reproduction number was calculated as  $R_t(t) = \beta(t) / \gamma$ .

1. For the first window, we determined the date when the number of confirmed cases began to increase exponentially. This is important because for many states or counties, very few confirmed cases were initially reported for a number of days or even weeks, which suggests that either the epidemic had not started or the true number of infected people was not known. It is not reasonable to apply a SIR model for this initial period. We took the onset of the epidemic as the first of the four consecutive days in which the number of reported confirmed cases rose in at least three days. The initial conditions for system (1) for window 0 were  $S_0(0) = N$ , where *N* is the population in the region of interest,  $I_0(0) = 1$ , and  $R_0(0) = 0$ . Infection rate  $\beta_i$  and  $S(t)$ ,  $I(t)$  for  $t \in [0, \tau - 1]$  were estimated given the initial conditions and actual *R*.
2. Slide the window by  $s = 1$  point. For the new  $i + 1$  window, take the initial conditions as the estimated values from the previous window  $S_{i+1}(0) = S_i(s)$ ,  $I_{i+1}(0) = I_i(s)$ , and actual  $R_{i+1}(0) = R(s)$ . Use actual values of  $R(t)$ , estimate infection rate  $\beta_{i+1}$  and  $S_{i+1}(t)$ ,  $I_{i+1}(t)$ .
3. For each window, calculate  $R_{t,i} = \beta_i / \gamma$ , assign the  $R_{t,i}$  to the last time point of the window. To get a smooth estimate of  $R_t$  we used a rolling average of 5 points.



## Results

We fit the model for each state and county in the United States. The model performance was evaluated by calculating the quality of fit as the root mean squared error between the actual and fitted  $P$  data for all windows concatenated ( $wRMSE$ ). The fit was excellent with  $wRMSE < 6$  across all states. We also calculated a 1-day forecast, 3-day forecast, and 7-day forecast of  $R$  after each window (Fig. 1A). The mean absolute prediction error (MAPE) for the forecasts is given in Table 1. The 1-day forecast error did not exceed 2.6% across all states while the 7-day forecast error was large and reached 16.2% for the state of New York. In particular, the 7-day forecast strongly overestimated the number of cases when the reproduction number was high and changing fast (Fig. 1).

The estimated time course of  $R_t$  for the state of New York and for Nassau county, one of the most affected in the beginning of COVID-19 epidemic, are shown in Figure 1. The estimated daily number of infectious individuals rapidly increased and then gradually declined after the lockdown was introduced on March 22, 2020 (Fig. 1A). The estimated reproduction number also declined after the lockdown began (Fig. 1B). The time course of  $R_t$  shows weekly seasonality which likely reflects the effect of social interactions and possibly the effect of fluctuations in case reporting on weekdays vs. weekends. For New York state and Nassau county,  $R_t$  exhibited an initial increase which may reflect the fact that the epidemic in the New York region was continuously seeded by travelers arriving to JFK airport until a ban on international travel was introduced on March 12, 2020. It may also reflect the fact that not all severe cases were initially recognized and reported as COVID-19. In Florida,  $R_t$  decreased close to 1 by mid-April but then began increasing at the end of May (Fig. 1B). In June 2020, Florida authorities introduced more stringent measures to control the epidemic which is reflected in decreasing  $R_t$  in the second half of July 2020. The opening of multiple states since June 2020 has been accompanied by  $R_t$  rising above 1 (not shown here), and close monitoring of reproductive number is needed to contain another wave of the epidemic.

We also compared *aSIR* with the model by Cori and colleagues [15] implemented as R package *EpiEstim* and a model implemented by Kevin Systrom and Thomas Vladeck available from a popular and influential COVID-19 data tracking website [rt.live](https://www.rtt.live) (Fig. 2). In *EpiEstim*, we assumed equal probability of infection within the infectious period of 6 days, the  $R_t$  estimate was smoothed with a 7-point rolling average window, same as in *aSIR*. While all three models show similar estimates when  $R_t$  is close to 1, their estimates differ considerably in the beginning of the epidemic. In particular, the *rt.live* model returned lower max  $R_t$  than the other two models, and estimated that  $R_t$  already decreased to 1 by the time the lockdown was announced in NY state on March 22, 2020 (Fig. 2, shaded region). The *EpiEstim* and the *aSIR* models estimated similar peak values of  $R_t$ , and both models estimated that  $R_t$  decreased close to 1 in the first week of April, 2020. Although both models show a rapidly decreasing  $R_t$  in March, the *aSIR* model shows a lagged change. We are not aware of ground truth data however to determine which model produces a more accurate estimate.

Next, we investigated the effect of an abrupt increase in testing on the  $R_t$  estimate (Fig. 3). We assumed a step-wise 50% increase in testing that persisted after April 12 (Fig. 3, left panel). Both *aSIR* and *EpiEstim* models exhibited a spike in  $R_t$ . It can also be argued that if testing increases then infectious individuals may be detected and quarantined sooner, resulting in a shorter infectious period and larger removal rate  $\gamma$ , in turn lowering  $R_t$ . We did not model a possible increase in  $\gamma$ . Instead, we assumed that the underlying dynamics of the epidemic did not change, and within 2 weeks both models returned to the  $R_t$  time course estimated without the testing increase.

## Discussion

We have developed a simple approach to adaptively estimate time-varying parameters of the SIR model using reported data on the number of confirmed cases. This approach adds to the already large literature on COVID-19 modeling in two ways. First, we estimate the parameters of the SIR model using a sliding window of a limited duration, 7 days, to account for fast changes in transmissibility and contact patterns in response to changes in social behavior and government mitigation measures. The window duration is a hyperparameter that can be changed as needed, the trade-off being the accuracy of the parameter estimates versus rapid reaction to changes in the underlying epidemic. Because the proposed model is so simple, a number of scenarios can be explored as needed.

Second, we attribute the data on reported cases to the Removed compartment rather than Infectious. This modeling decision is based on the realities of the COVID-19 epidemic in the US where confirmed positive individuals are supposed to self-isolate or are hospitalized. Although these individuals remain infectious and can infect other family members or caretakers even when self-isolated or hospitalized, they are not freely interacting with the susceptible population as would be required to attribute them to the  $I$  compartment. It has been proposed to add a new  $X$  compartment in the SIR model to model symptomatic quarantined infectious individuals [24]. We have no data to independently estimate this additional parameter of quarantine rate, however. For the same reason, we did not use the Susceptible-Exposed-Infected-Removed (SEIR) model because we are not aware of reliable data about the duration of the exposed period during which an infected person is not yet infectious. Moreover, it has been found that the SIR model performed better than an SEIR model in representing the information contained in the confirmed-case data on COVID-19 [25].

The reported numbers of positive cases represent a fraction of infected individuals because of the limited availability of testing in March and April of 2020 with the result that only those who developed severe symptoms were tested. Up to 80% of the infected individuals may be asymptomatic or develop mild symptoms [26] and were not tested, so for that period our model applies only to the small sub-population who developed severe symptoms. This sub-population, however, is of particular interest because it represents those most at risk, and the

reproduction number estimated from this limited data can be used to guide policy decisions aimed at protecting the most vulnerable population [27]. At the same time, as the numbers of the tested individuals increase, the short sliding window approach makes our model adaptable to an ever-larger proportion of the population (Fig. 3).

Across all US states, the maximal  $R_t$  values were estimated for New York (4.4) and Michigan (4.5) (Table 1) which is close to the mean value of 4.34 estimated for Italy [28] but higher than that obtained by a stochastic transmission model [29,30]. The wide range of maximal values of  $R_t$  from 2.0 to 4.5 (Table 1) likely reflects the differences in contact rates due to population density [31,32]. Increased social distancing is required to contain the spread of the epidemic [33,34], with more stringent mitigation measures, including lockdown, considered necessary to decrease the contact rate in high-density states and counties. Another measure to lower  $R_t$  is to increase the removal rate  $\gamma$  by intensive testing and quarantine of individuals tested positive. This targeted intervention would strongly decrease the interaction between the infectious and susceptible individuals and keep  $R_t < 1$  until a vaccine becomes available and while vaccination efforts are ramping up. Intensive testing combined with social distancing and mask wearing followed by isolation of individuals tested positive are key features in reopening strategies for schools and universities [35–37]. Our model allows researchers as well as policy makers to monitor the reproduction number in different geographical regions of the US, better understand the effect of government policies on the dynamics of the epidemic, and develop further mitigation strategies as we continue to battle COVID-19 [38,39].

## Limitations

The SIR model is perhaps the simplest model that captures the dynamics of an epidemic. It is based on several assumptions that are valid only to some degree as we consider real-life scenarios. The two main limitations of the original SIR model are that it has constant parameters and it is deterministic. Our proposed aSIR model allows to estimate time-varying parameters and thus removes the first limitation. The other limitation remains, however. It is assumed that infectious individuals freely interact with the susceptible population. The infection rate  $\beta$  encompasses both the probability of transmission and the average number of contacts per day. The SIR model does not reflect interaction dynamics that are stochastic in nature and are described by stochastic epidemiologic models [15–17,40]. **In its simplest form, the SIR model** does not reflect the heterogeneity of viral transmission reflected in overdispersion or superspreading where few numbers of infected individuals infect large number of others [41–43]. The removal rate  $\gamma$  is an average number of days until an infectious individual is removed and does not reflect the variability of this interval, nor does it allow one to model a possibility of a subsequent, albeit reduced transmission to caregivers or other susceptible individuals as it may happen in real life. As a result, the reproduction number  $R_t$ , which is calculated using constants  $\beta$  and  $\gamma$  rather than their distributions, does not reflect the stochastic nature of the epidemic dynamics. Parameter distributions can be obtained by applying Bayesian methods to SIR modeling [44]. Moreover, a single value of  $R_t$  estimated for a large population does not reflect differences in sub-populations such as age groups which are especially relevant for COVID-19 [45–47]. Building a bank of aSIR models for each sub-population or region can provide a more realistic insight into the epidemic dynamics across a

larger population [48,49]. Another critical assumption is that once an individual is infected and recovered, they are no longer susceptible to repeated infection. This assumption does not appear to be strongly violated, however. Although some cases of repeated infection with SARS-CoV-2 have been reported [50–52], the risk of re-infection is considered small [53,54]. Overall, the SIR model is a trade-off between the computational simplicity and veracity of describing the real-life complexities of an epidemic.

**We argue that SIR type models, in particular the proposed time-variant aSIR model, have an advantage over more complicated models in the beginning stages of an epidemic when critical public policy decisions need to be made while the empirical data on interaction dynamics, transmission rates, and the disease progression and contagiousness from the moment of infection is not yet readily available. It provides a simple and efficient way to assess the efficacy of interventions as the epidemic progresses.**

**Table 1.** Reproduction numbers  $R_t$  and forecast accuracy for 50 states.

| State                | $R_t$<br>Max | MAPE<br>1-Day<br>Forecast | MAPE<br>3-Day<br>Forecast | MAPE<br>7-Day<br>Forecast |
|----------------------|--------------|---------------------------|---------------------------|---------------------------|
| Alabama              | 2.9          | 1.5%                      | 4.2%                      | 10.0%                     |
| Alaska               | 2.8          | 1.6%                      | 3.7%                      | 11.3%                     |
| Arizona              | 3.3          | 1.3%                      | 2.9%                      | 10.1%                     |
| Arkansas             | 2.8          | 1.5%                      | 4.0%                      | 12.6%                     |
| California           | 2.5          | 1.7%                      | 2.9%                      | 6.3%                      |
| Colorado             | 2.6          | 1.1%                      | 2.9%                      | 7.3%                      |
| Connecticut          | 4.1          | 2.0%                      | 3.1%                      | 9.3%                      |
| Delaware             | 2.4          | 1.7%                      | 2.9%                      | 7.9%                      |
| District of Columbia | 2.1          | 0.8%                      | 1.8%                      | 4.4%                      |
| Florida              | 3.6          | 2.0%                      | 4.4%                      | 9.3%                      |
| Georgia              | 3.0          | 1.8%                      | 3.7%                      | 7.4%                      |
| Hawaii               | 2.7          | 2.0%                      | 3.6%                      | 9.7%                      |
| Idaho                | 3.4          | 2.4%                      | 4.8%                      | 13.6%                     |
| Illinois             | 4.0          | 1.3%                      | 2.5%                      | 8.5%                      |
| Indiana              | 3.8          | 1.4%                      | 4.0%                      | 10.4%                     |
| Iowa                 | 2.8          | 1.8%                      | 3.6%                      | 8.0%                      |
| Kansas               | 3.0          | 1.6%                      | 3.5%                      | 8.6%                      |
| Kentucky             | 3.0          | 2.6%                      | 4.9%                      | 11.2%                     |
| Louisiana            | 3.7          | 1.8%                      | 4.0%                      | 12.1%                     |
| Maine                | 2.0          | 1.2%                      | 2.8%                      | 6.7%                      |
| Maryland             | 3.3          | 1.2%                      | 2.8%                      | 6.2%                      |
| Massachusetts        | 3.4          | 1.3%                      | 3.6%                      | 9.7%                      |

|                |     |      |      |       |
|----------------|-----|------|------|-------|
| Michigan       | 4.5 | 1.6% | 3.5% | 12.8% |
| Minnesota      | 2.7 | 1.3% | 2.9% | 8.0%  |
| Mississippi    | 2.9 | 1.2% | 3.0% | 9.3%  |
| Missouri       | 3.6 | 1.8% | 3.3% | 11.4% |
| Montana        | 3.3 | 1.6% | 3.7% | 11.9% |
| Nebraska       | 2.5 | 2.0% | 4.1% | 9.7%  |
| Nevada         | 2.9 | 2.3% | 3.8% | 10.0% |
| New Hampshire  | 2.3 | 1.7% | 3.3% | 8.5%  |
| New Jersey     | 4.1 | 1.5% | 2.3% | 7.8%  |
| New Mexico     | 2.3 | 2.2% | 3.4% | 7.3%  |
| New York       | 4.4 | 1.5% | 4.2% | 16.2% |
| North Carolina | 3.2 | 1.3% | 2.4% | 7.2%  |
| North Dakota   | 2.4 | 1.8% | 4.8% | 12.6% |
| Ohio           | 3.3 | 1.2% | 3.4% | 9.8%  |
| Oklahoma       | 3.1 | 1.4% | 3.5% | 10.2% |
| Oregon         | 2.5 | 1.3% | 2.8% | 6.6%  |
| Pennsylvania   | 3.2 | 1.7% | 2.9% | 6.3%  |
| Rhode Island   | 2.4 | 1.5% | 3.1% | 6.8%  |
| South Carolina | 3.5 | 2.1% | 4.3% | 10.6% |
| South Dakota   | 2.1 | 1.3% | 3.2% | 8.7%  |
| Tennessee      | 3.5 | 2.2% | 4.8% | 12.5% |
| Texas          | 3.6 | 2.0% | 3.9% | 9.3%  |
| Utah           | 3.2 | 1.4% | 3.1% | 8.3%  |
| Vermont        | 2.9 | 0.8% | 2.4% | 7.7%  |
| Virginia       | 2.5 | 1.1% | 2.1% | 5.1%  |
| Washington     | 3.0 | 2.0% | 4.8% | 8.6%  |
| West Virginia  | 3.5 | 1.6% | 3.9% | 14.0% |
| Wisconsin      | 3.6 | 1.5% | 3.2% | 10.0% |
| Wyoming        | 2.9 | 1.9% | 4.7% | 14.2% |

## References

1. Arenas A, Cota W, Gomez-Gardenes J, Gómez S, Granell C, Matamalas JT, Soriano-Panos D, Steinegger B. A mathematical model for the spatiotemporal epidemic spreading of COVID19. medRxiv Cold Spring Harbor Laboratory Press; 2020 Mar 23;2020.03.21.20040022. [doi: 10.1101/2020.03.21.20040022]
2. Ferguson N, Laydon D, Nedjati Gilani G, Imai N, Ainslie K, Baguelin M, Bhatia S, Boonyasiri A, Cucunuba Perez Z, Cuomo-Dannenburg G, Dighe A, Dorigatti I, Fu H, Gaythorpe K, Green W, Hamlet A, Hinsley W, Okell L, Van Elsland S, Thompson H, Verity R, Volz E, Wang H, Wang Y, Walker P, Winskill P, Whittaker C, Donnelly C, Riley S, Ghani A. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand [Internet]. Imperial College London; 2020 Mar. [doi: 10.25561/77482]
3. Mitjà O, Arenas À, Rodó X, Tobias A, Brew J, Benlloch JM. Experts' request to the Spanish Government: move Spain towards complete lockdown. *The Lancet* 2020 Apr 11;395(10231):1193–1194. [doi: 10.1016/S0140-6736(20)30753-4]
4. Adam D. Special report: The simulations driving the world's response to COVID-19. *Nature* Nature Publishing Group; 2020 Apr 2;580(7803):316–318. [doi: 10.1038/d41586-020-01003-6]
5. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet* 2020 Mar 21;395(10228):931–934. [doi: 10.1016/S0140-6736(20)30567-5]
6. Enserink M, Kupferschmidt K. With COVID-19, modeling takes on life and death importance. *Science American Association for the Advancement of Science*; 2020 Mar 27;367(6485):1414–1415. PMID:32217707
7. Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the Basic Reproduction Number ( $R_0$ ). *Emerg Infect Dis* 2019 Jan;25(1):1–4. PMID:30560777
8. Hamer W. The Milroy Lectures ON EPIDEMIC DISEASE IN ENGLAND—THE EVIDENCE OF VARIABILITY AND OF PERSISTENCY OF TYPE. *The Lancet* 1906 Mar 3;167(4305):569–574. [doi: 10.1016/S0140-6736(01)80187-2]
9. Kermack WO, McKendrick AG. A Contribution to the Mathematical Theory of Epidemics. *Proc R Soc Lond Ser Contain Pap Math Phys Character The Royal Society*; 1927;115(772):700–721.
10. Non-pharmaceutical public health measures for mitigating the risk and impact of epidemic and pandemic influenza [Internet]. World Health Organ. World Health Organization; 2019 [cited 2020 Jul 7]. Available from: [http://www.who.int/influenza/publications/public\\_health\\_measures/publication/en/](http://www.who.int/influenza/publications/public_health_measures/publication/en/)
11. Anderson RM. Discussion: The Kermack-McKendrick epidemic threshold theorem. *Bull Math Biol* 1991 Mar 1;53(1):1. [doi: 10.1007/BF02464422]
12. Anderson RM, May RM. *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press; 1991. ISBN:0-19-854040-X
13. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, Priesemann V. Inferring

change points in the spread of COVID-19 reveals the effectiveness of interventions. Science [Internet] American Association for the Advancement of Science; 2020 May 15 [cited 2020 Jun 21]; PMID:32414780

14. Karnakov P, Arampatzis G, Kičić I, Wermelinger F, Wälchli D, Papadimitriou C, Koumoutsakos P. Data driven inference of the reproduction number ( $R_0$ ) for COVID-19 before and after interventions for 51 European countries. medRxiv Cold Spring Harbor Laboratory Press; 2020 May 23;2020.05.21.20109314. [doi: 10.1101/2020.05.21.20109314]
15. Cori A, Ferguson NM, Fraser C, Cauchemez S. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *Am J Epidemiol* Oxford Academic; 2013 Nov 1;178(9):1505–1512. [doi: 10.1093/aje/kwt133]
16. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc Biol Sci* 2007 Feb 22;274(1609):599–604. PMID:17476782
17. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol* 2004 Sep 15;160(6):509–516. PMID:15353409
18. Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, Jacobs JR, Taylor J, Spicer K, Bardossy AC, Oakley LP, Tanwar S, Dyal JW, Harney J, Chisty Z, Bell JM, Methner M, Paul P, Carlson CM, McLaughlin HP, Thornburg N, Tong S, Tamin A, Tao Y, Uehara A, Harcourt J, Clark S, Brostrom-Smith C, Page LC, Kay M, Lewis J, Montgomery P, Stone ND, Clark TA, Honein MA, Duchin JS, Jernigan JA. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. *N Engl J Med* Massachusetts Medical Society; 2020 May 28;382(22):2081–2090. [doi: 10.1056/NEJMoa2008457]
19. Hu Z, Song C, Xu C, Jin G, Chen Y, Xu X, Ma H, Chen W, Lin Y, Zheng Y, Wang J, Hu Z, Yi Y, Shen H. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Sci China Life Sci* 2020 Mar 4;1–6. PMID:32146694
20. Tindale LC, Stockdale JE, Coombe M, Garlock ES, Lau WYV, Saraswat M, Zhang L, Chen D, Wallinga J, Colijn C. Evidence for transmission of COVID-19 prior to symptom onset. *eLife* 2020 Jun 22;9. PMID:32568070
21. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, Xing X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Liu M, Tu W, Chen C, Jin L, Yang R, Wang Q, Zhou S, Wang R, Liu H, Luo Y, Liu Y, Shao G, Li H, Tao Z, Yang Y, Deng Z, Liu B, Ma Z, Zhang Y, Shi G, Lam TTY, Wu JT, Gao GF, Cowling BJ, Yang B, Leung GM, Feng Z. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N Engl J Med* [Internet] Massachusetts Medical Society; 2020 Jan 29 [cited 2020 Jun 27]; [doi: 10.1056/NEJMoa2001316]
22. Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung S, Yuan B, Kinoshita R, Nishiura H. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. *J Clin Med* [Internet] 2020 Feb 17 [cited 2020 Jun 27];9(2). PMID:32079150
23. Q&A on coronaviruses (COVID-19) [Internet]. World Health Organ. 2020. Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/>

q-a-detail/q-a-coronaviruses

24. Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science American Association for the Advancement of Science*; 2020 May 15;368(6492):742–746. PMID:32269067
25. Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic? *Infect Dis Model* 2020;5:271–281. PMID:32289100
26. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA American Medical Association*; 2020 Apr 7;323(13):1239–1242. [doi: 10.1001/jama.2020.2648]
27. Lourenco J, Paton R, Ghafari M, Kraemer M, Thompson C, Simmonds P, Klenerman P, Gupta S. Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic. *medRxiv Cold Spring Harbor Laboratory Press*; 2020 Mar 26;2020.03.24.20042291. [doi: 10.1101/2020.03.24.20042291]
28. Wangping J, Ke H, Yang S, Wenzhe C, Shengshu W, Shanshan Y, Jianwei W, Fuyin K, Penggang T, Jing L, Miao L, Yao H. Extended SIR Prediction of the Epidemics Trend of COVID-19 in Italy and Compared With Hunan, China. *Front Med [Internet] Frontiers*; 2020 [cited 2020 Jun 23];7. [doi: 10.3389/fmed.2020.00169]
29. Abbott S, Hellewell J, Thompson RN, Sherratt K, Gibbs HP, Bosse NI, Munday JD, Meakin S, Doughty EL, Chun JY, Chan Y-WD, Finger F, Campbell P, Endo A, Pearson CAB, Gimma A, Russell T, CMMID COVID modelling group, Flasche S, Kucharski AJ, Eggo RM, Funk S. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Res* 2020 Jun 1;5:112. [doi: 10.12688/wellcomeopenres.16006.1]
30. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, Eggo RM, Sun F, Jit M, Munday JD, Davies N, Gimma A, van Zandvoort K, Gibbs H, Hellewell J, Jarvis CI, Clifford S, Quilty BJ, Bosse NI, Abbott S, Klepac P, Flasche S. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis* 2020 May 1;20(5):553–558. [doi: 10.1016/S1473-3099(20)30144-4]
31. Hu H, Nigmatulina K, Eckhoff P. The scaling of contact rates with population density for the infectious disease models. *Math Biosci* 2013 Aug 1;244(2):125–134. [doi: 10.1016/j.mbs.2013.04.013]
32. Sy KTL, White LF, Nichols BE. Population density and basic reproductive number of COVID-19 across United States counties. *medRxiv Cold Spring Harbor Laboratory Press*; 2020 Jun 13;2020.06.12.20130021. [doi: 10.1101/2020.06.12.20130021]
33. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Flasche S, Clifford S, Pearson CAB, Munday JD, Abbott S, Gibbs H, Rosello A, Quilty BJ, Jombart T, Sun F, Diamond C, Gimma A, Zandvoort K van, Funk S, Jarvis CI, Edmunds WJ, Bosse NI, Hellewell J, Jit M, Klepac P. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health Elsevier*; 2020 May



1;5(5):e261–e270. PMID:32220655

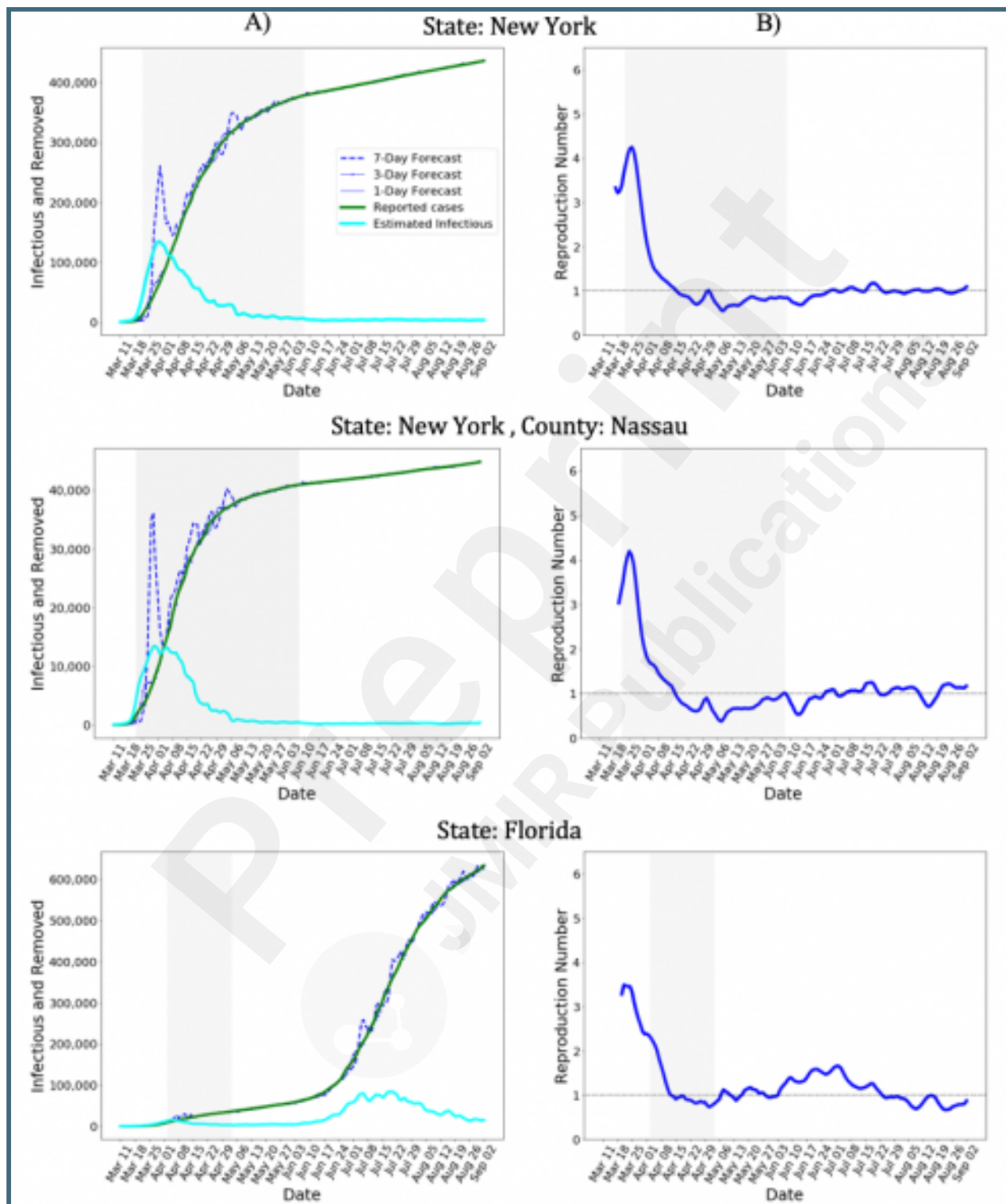
34. Delen D, Eryarsoy E, Davazdahemami B. No Place Like Home: Cross-National Data Analysis of the Efficacy of Social Distancing During the COVID-19 Pandemic. *JMIR Public Health Surveill* 2020 May 28;6(2):e19862. [doi: 10.2196/19862]
35. Mukherjee UK, Bose S, Ivanov A, Souyris S, Seshadri S, Sridhar P, Watkins R, Xu Y. Evaluation of reopening strategies for educational institutions during COVID-19 through agent based simulation. *Sci Rep* 2021 Mar 17;11(1):6264. PMID:33731722
36. Paltiel AD, Zheng A, Walensky RP. Assessment of SARS-CoV-2 Screening Strategies to Permit the Safe Reopening of College Campuses in the United States. *JAMA Netw Open* 2020 Jul 31;3(7):e2016818. [doi: 10.1001/jamanetworkopen.2020.16818]
37. Panovska-Griffiths J, Kerr CC, Stuart RM, Mistry D, Klein DJ, Viner RM, Bonell C. Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and the risk of occurrence of a second COVID-19 epidemic wave in the UK: a modelling study. *Lancet Child Adolesc Health Elsevier*; 2020 Nov 1;4(11):817–827. PMID:32758453
38. Friedman E, Friedman J, Johnson S, Landsberg A. Transitioning Out of the Coronavirus Lockdown: A Framework for Evaluating Zone-Based Social Distancing. *Front Public Health* [Internet] *Frontiers*; 2020 [cited 2020 Jun 27];8. [doi: 10.3389/fpubh.2020.00266]
39. Inglesby TV. Public Health Measures and the Reproduction Number of SARS-CoV-2. *JAMA American Medical Association*; 2020 Jun 2;323(21):2186–2187. [doi: 10.1001/jama.2020.7878]
40. Allen LJS. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infect Dis Model* 2017 Mar 11;2(2):128–142. PMID:29928733
41. Zhang Y, Li Y, Wang L, Li M, Zhou X. Evaluating Transmission Heterogeneity and Super-Spreading Event of COVID-19 in a Metropolis of China. *Int J Environ Res Public Health* [Internet] 2020 May [cited 2021 Mar 6];17(10). PMID:32456346
42. Lau MSY, Grenfell B, Thomas M, Bryan M, Nelson K, Lopman B. Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. *Proc Natl Acad Sci National Academy of Sciences*; 2020 Sep 8;117(36):22430–22435. PMID:32820074
43. Adam DC, Wu P, Wong JY, Lau EHY, Tsang TK, Cauchemez S, Leung GM, Cowling BJ. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat Med Nature Publishing Group*; 2020 Nov;26(11):1714–1719. [doi: 10.1038/s41591-020-1092-0]
44. Lawson AB, Kim J. Space-time covid-19 Bayesian SIR modeling in South Carolina. *PLOS ONE Public Library of Science*; 2021 Mar 17;16(3):e0242777. [doi: 10.1371/journal.pone.0242777]
45. Goldstein E, Lipsitch M, Cevik M. On the effect of age on the transmission of SARS-CoV-2 in households, schools and the community. *medRxiv* [Internet] 2020 Jul 28 [cited 2021 Mar 6]; PMID:32743609

46. Davies NG, Klepac P, Liu Y, Prem K, Jit M, Eggo RM. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med Nature Publishing Group*; 2020 Aug;26(8):1205–1211. [doi: 10.1038/s41591-020-0962-9]
47. Liu Y, Gu Z, Xia S, Shi B, Zhou X-N, Shi Y, Liu J. What are the underlying transmission patterns of COVID-19 outbreak? An age-specific social contact characterization. *EClinicalMedicine* [Internet] Elsevier; 2020 May 1 [cited 2021 Mar 6];22. PMID:32313879
48. Turk PJ, Chou S-H, Kowalkowski MA, Palmer PP, Priem JS, Spencer MD, Taylor YJ, McWilliams AD. Modeling COVID-19 Latent Prevalence to Assess a Public Health Intervention at a State and Regional Scale: Retrospective Cohort Study. *JMIR Public Health Surveill* 2020 Jun 19;6(2):e19353. [doi: 10.2196/19353]
49. Mehta M, Julaiti J, Griffin P, Kumara S. Early Stage Machine Learning–Based Prediction of US County Vulnerability to the COVID-19 Pandemic: Machine Learning Approach. *JMIR Public Health Surveill* 2020 Sep 11;6(3):e19446. [doi: 10.2196/19446]
50. Murillo-Zamora E, Mendoza-Cano O, Delgado-Enciso I, Hernandez-Suarez CM. Predictors of severe symptomatic laboratory-confirmed SARS-COV-2 reinfection. *medRxiv Cold Spring Harbor Laboratory Press*; 2020 Oct 18;2020.10.14.20212720. [doi: 10.1101/2020.10.14.20212720]
51. Gousseff M, Penot P, Gallay L, Batisse D, Benech N, Bouiller K, Collarino R, Conrad A, Slama D, Joseph C, Lemaignen A, Lescure F-X, Levy B, Mahevas M, Pozzetto B, Vignier N, Wyplosz B, Salmon D, Goehring F, Botelho-Nevers E. Clinical recurrences of COVID-19 symptoms after recovery: Viral relapse, reinfection or inflammatory rebound? *J Infect Elsevier*; 2020 Nov 1;81(5):816–846. PMID:32619697
52. To KK-W, Hung IF-N, Ip JD, Chu AW-H, Chan W-M, Tam AR, Fong CH-Y, Yuan S, Tsoi H-W, Ng AC-K, Lee LL-Y, Wan P, Tso EY-K, To W-K, Tsang DN-C, Chan K-H, Huang J-D, Kok K-H, Cheng VC-C, Yuen K-Y. Coronavirus Disease 2019 (COVID-19) Re-infection by a Phylogenetically Distinct Severe Acute Respiratory Syndrome Coronavirus 2 Strain Confirmed by Whole Genome Sequencing. *Clin Infect Dis* [Internet] 2020 Aug 25 [cited 2021 Mar 6]; (ciaa1275). [doi: 10.1093/cid/ciaa1275]
53. Raddad LJA, Chemaitelly H, Malek JA, Ahmed AA, Mohamoud YA, Younuskunju S, Ayoub HH, Kanaani ZA, Khal AA, Kuwari EA, Butt AA, Coyle P, Jeremijenko A, Kaleeckal AH, Latif AN, Shaik RM, Rahim HFA, Yassine HM, Kuwari MGA, Romaihi HEA, Thani SMA, Bertollini R. Assessment of the risk of SARS-CoV-2 reinfection in an intense re-exposure setting. *medRxiv Cold Spring Harbor Laboratory Press*; 2020 Sep 28;2020.08.24.20179457. [doi: 10.1101/2020.08.24.20179457]
54. McMahon A, Robb NC. Reinfection with SARS-CoV-2: Discrete SIR (Susceptible, Infected, Recovered) Modeling Using Empirical Infection Data. *JMIR Public Health Surveill* 2020 Nov 16;6(4):e21168. [doi: 10.2196/21168]

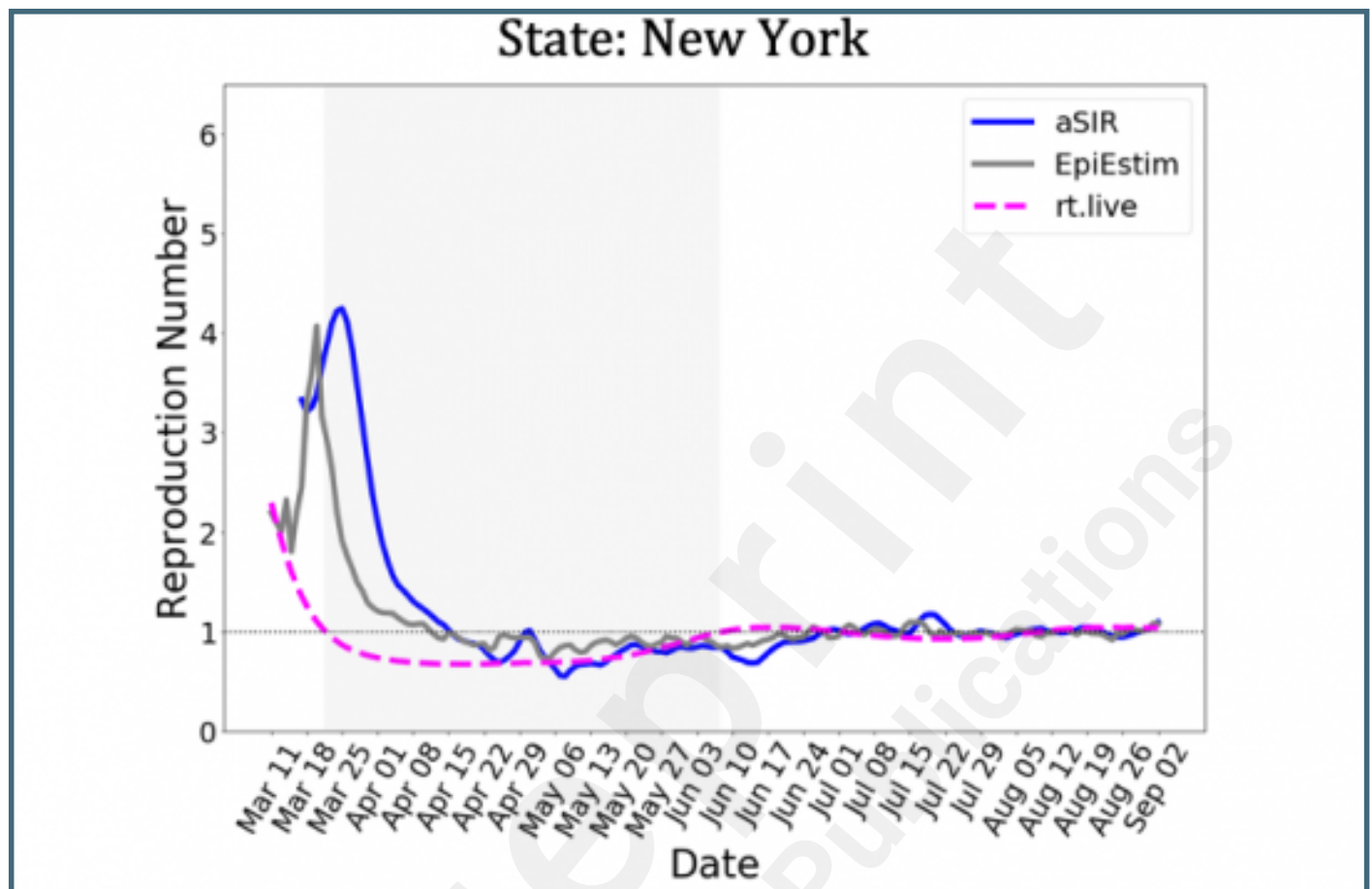
## Supplementary Files

## Figures

A) Estimated Infectious and forecasted Removed. B) Estimated reproduction number  $R_t$ . The shaded region indicates the dates of the lockdown. While the 1-day and 3-day forecasts are accurate, the 7-day forecast exhibits large errors when  $R_t > 1$  and is changing fast.



Comparison of models that generate continuous  $R_t$  estimates. The three  $R_t$  estimates differ widely in the beginning of the epidemic. In particular, the  $R_t$  estimated by rt.live model decreased to 1 by the lockdown onset on March 22 (shaded region).



The effect of a step-wise 50% increase in testing (left panel, dashed line). The 1-day forecast by the aSIR model adapts within a week. For  $R_t$  estimate, both EpiEstim and our aSIR models produced a spike followed by a decrease (right panel, dashed lines) before returning to the unperturbed  $R_t$  time course (solid lines).

