# PLOS ONE

# Estimation of time-varying reproduction numbers underlying epidemiological processes: A new statistical tool for the COVID-19 pandemic

Hyokyoung G. Hong[1], Yi Li[2]*

**1** Department of Statistics and Probability, Michigan State University, East Lansing, MI, United States of America, **2** Department of Biostatistics, University of Michigan, Ann Arbor, MI, United States of America

* yili@umich.edu
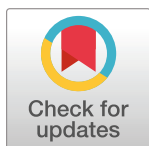
## Abstract

The coronavirus pandemic has rapidly evolved into an unprecedented crisis. The susceptible-infectious-removed (SIR) model and its variants have been used for modeling the pandemic. However, time-independent parameters in the classical models may not capture the dynamic transmission and removal processes, governed by virus containment strategies taken at various phases of the epidemic. Moreover, few models account for possible inaccuracies of the reported cases. We propose a Poisson model with time-dependent transmission and removal rates to account for possible random errors in reporting and estimate a time-dependent disease reproduction number, which may reflect the effectiveness of virus control strategies. We apply our method to study the pandemic in several severely impacted countries, and analyze and forecast the evolving spread of the coronavirus. We have developed an interactive web application to facilitate readers' use of our method.

## 1 Introduction

Coronaviruses are enveloped single-stranded positive-sense RNA viruses belonging to a broad family of coronaviridae and are widely harbored in animals [1–3]. Most of the coronaviruses only cause mild respiratory infections, but SARS-CoV-2, a newly identified member of the coronavirus family, initiated the contagious and lethal coronavirus disease 2019 (COVID-19) in December 2019 [4, 5]. Since the detection of the first case in Wuhan, the COVID-19 pandemic has evolved into a global crisis within only four months. As of June 30, 2020, the virus has infected more than 10 million individuals, caused about 518,000 deaths [6], and altered the life of billions of people.

The pandemic has been closely monitored by the international society. For example, the World Health Organization (WHO) and Johns Hopkins University's Coronavirus Resource Center [6] have, since the outbreak, reported the daily numbers of infectious and recovered cases, and deaths for nearly every country. The governmental websites of many counties, such as Australia, the US, Singapore, also have been tracking these numbers starting from various

time points. These websites have become valuable resources to help advance the understanding of spread of the virus. We have access to a time-series data repository on GitHub (https://github.com/ulklc/covid19-timeseries), which consolidates and updates information obtained from these data sources. Our data analysis is based on the data obtained from this GitHub data repository.

Much effort has been devoted by the affected countries to battling the disease. However, the crisis has not been over, with new infections detected every day. To forecast when the pandemic gets controlled and evaluate the effects of virus control measures, it is imperative to develop appropriate models to describe and understand the change trend of the pandemic [7–10].

The susceptible-infectious-removed (SIR) model was utilized to explain the rapid rise and fall of the infected individuals from the epidemics of severe acute respiratory syndrome (SARS), influenza A virus subtype (H1N1) and middle east respiratory syndrome (MERS) [11–15]. The key idea is to divide a total population into three compartments: the susceptible, *S*, who are healthy individuals capable of contracting the disease; the infectious, *I*, who have the disease and are infectious; and the removed, *R*, who have recovered from the disease and gained immunity or who have died from the disease [16]. The model assumes a one-way flow from susceptible to infectious to removed, and is reasonable for infectious diseases, which are transmitted from human to human, and where recovery confers lasting resistance [17]. SIR models originated from the Kermack-McKendrick model [18], consisting of three coupled differential equations to describe the dynamics of the numbers in the *S*, *I*, and *R* compartments, which tend to fluctuate over time. For example, the number of infectious individuals increases drastically at the start of the epidemic, with a surge in susceptible individuals becoming infectious. As the epidemic develops, the number of infectious individuals decreases when more infectious individuals die or recover than susceptible individuals become infectious. The epidemic ends when the infectious compartment ceases to exist [16, 18].

SIR models and the modified versions, such as susceptible-exposed-infectious-recovered model (SEIR), were applied to analyze the COVID-19 outbreak [19–23]. Many of these models assume constant transmission and removal rates, which may not hold in reality. For example, as a result of various virus containment strategies, such as self-quarantine and social distancing mandates, the transmission and removal rates may vary over time [24].

Recently, a number of researchers [25–27] considered time-dependent SIR models adapted to the dynamical epidemiological processes evolving over time. However, few considered random errors in reporting, such as under-reporting (e.g. asymptomatic cases or virus mutation) or over-reporting (e.g. false positives of testing), or characterized the uncertainty of predictions.

Poisson models naturally fit count data [28]. Several works [29–31] used Poisson distributions to model *I* and *R* from frequentist or Bayesian perspectives; however, most of the works only considered constant transmission and removal rates. How to extend these works to accommodate time-dependent rates remains elusive.

We propose to adopt a Poisson model to estimate the time-varying transmission and removal rates, and understand the trends of the pandemic across countries. For example, we can predict the number of the infectious persons and the number of removed persons at a certain time for each country, and forecast when the curves of cases become flattened.

An important epidemiological index that characterizes the transmission potential is the basic reproduction number, $\mathcal{R}_0$, defined as the expected number of secondary cases produced by an infectious case [32–34]. Our model leads to a temporally dynamical $\mathcal{R}_0$, which measures at a given time how many people one infectious person, during the infectious period, will infect [35]. This may help evaluate the quarantine policies implemented by various authorities.

A recent work [35] demonstrated that $\mathcal{R}_0$ is likely to vary "due to the impact of the performed intervention strategies and behavioral changes in the population".

The merits of our work are summarized as follows. First, unlike the deterministic ODE-based SIR models, our method does not require transmission and removal rates to be known, but estimates them using the data. Second, we allow these rates to be time-varying. Some time-varying SIR approaches [27] directly integrate into the model the information on when governments enforced, for example, quarantine, social-distancing, compulsory mask-wearing and city lockdowns. Our method differs by computing a time-varying $\mathcal{R}_0$, which gauges the status of coronavirus containment and assesses the effectiveness of virus control strategies. Third, our Poisson model accounts for possible random errors in reporting, and quantifies the uncertainty of the predicted numbers of susceptible, infectious and removed. Finally, we apply our method to analyze the data collected from the aforementioned GitHub time-series data repository. We have created an interactive web application (https://younghhk.shinyapps.io/tvSIRforCOVID19/) to facilitate users' application of the proposed method.

## 2 A Poisson model with time-dependent transmission and removal rates

We introduce a Poisson model with time-varying transmission and removal rates, denoted by $\beta(t)$ and $\gamma(t)$. Consider a population with $N$ individuals, and denote by $S(t)$, $I(t)$, $R(t)$ the true but unknown numbers of susceptible, infectious and removed, respectively, at time $t$, and by $s(t) = S(t)/N$, $i(t) = I(t)/N$, $r(t) = R(t)/N$ the fractions of these compartments.

### 2.1 Time-varying transmission, removal rates and reproduction number

The following ordinary differential equations (ODE) describe the change rates of $s(t)$, $i(t)$ and $r(t)$:

$$\frac{ds(t)}{dt} = -\beta(t)s(t)i(t), \tag{1}$$

$$\frac{di(t)}{dt} = \beta(t)s(t)i(t) - \gamma(t)i(t), \tag{2}$$

$$\frac{dr(t)}{dt} = \gamma(t)i(t), \tag{3}$$

with an initial condition: $i(0) = i_0$ and $r(0) = r_0$, where $i_0 > 0$ in order to let the epidemic develop [36]. Here, $\beta(t) > 0$ is the time-varying transmission rate of an infection at time $t$, which is the number of infectious contacts that result in infections per unit time, and $\gamma(t) > 0$ is the time-varying removal rate at $t$, at which infectious subjects are removed from being infectious due to death or recovery [33]. Moreover, $\gamma^{-1}(t)$ can be interpreted as the infectious duration of an infection caught at time $t$[37].

From (1)–(3), we derive an important quantity, which is the time-dependent reproduction number

$$\mathcal{R}_0(t) = \frac{\beta(t)}{\gamma(t)}.$$

To see this, dividing (2) by (3) leads to

$$\mathcal{R}_0(t) = \frac{1}{s(t)} \left\{ \frac{di}{dr}(t) + 1 \right\}, \tag{4}$$

where $(di/dr)(t)$ is the ratio of the change rate of $i(t)$ to that of $r(t)$. Therefore, compared to its time-independent counterpart, $\mathcal{R}_0(t)$ is an instantaneous reproduction number and provides a real-time picture of an outbreak. For example, at the onset of the outbreak and in the absence of any containment actions, we may see a rapid ramp-up of cases compared to those removed, leading to a large $(di/dr)(t)$ in (4), and hence a large $\mathcal{R}_0(t)$. With the implemented policies for disease mitigation, we will see a drastically decreasing $(di/dr)(t)$ and, therefore, declining of $\mathcal{R}_0(t)$ over time. The turning point is $t_0$ such that $\mathcal{R}_0(t_0) = 1$, when the outbreak is controlled with $(di/dr)(t_0) < 0$.

Under the fixed population size assumption, i.e., $s(t) + i(t) + r(t) = 1$, we only need to study $i(t)$ and $r(t)$, and re-express (1)–(3) as

$$\begin{aligned} \frac{di(t)}{dt} &= \beta(t)i(t)\{1 - i(t) - r(t)\} - \gamma(t)i(t), \\ \frac{dr(t)}{dt} &= \gamma(t)i(t), \end{aligned} \tag{5}$$

with the same initial condition.

## 2.2 A Poisson model based on discrete time-varying SIR

As the numbers of cases and removed are reported on a daily basis, $t$ is measured in days, e.g. $t = 1, \ldots, T$. Replacing derivatives in (5) with finite differences, we can consider a discrete version of (5):

$$\begin{aligned} i(t+1) - i(t) &= \beta(t)i(t)\{1 - i(t) - r(t)\} - \gamma(t)i(t), \\ r(t+1) - r(t) &= \gamma(t)i(t), \end{aligned} \tag{6}$$

where $\beta(t)$ and $\gamma(t)$ are positive functions of $t$. We set $i(0) = i_0$ and $r(0) = r_0$ with $t = 0$ being the starting date.

Model (6) admits a recursive way to compute $i(t)$ and $r(t)$:

$$\begin{aligned} i(t+1) &= \{1 + \beta(t) - \gamma(t)\}i(t) - \beta(t)i(t)\{i(t) + r(t)\}, \\ r(t+1) &= r(t) + \gamma(t)i(t) \end{aligned} \tag{7}$$

for $t = 0, \ldots, T - 1$. The first equation of (7) implies that $\beta(t) < \gamma(t)$ or $\mathcal{R}_0(t) = \beta(t)\gamma^{-1}(t) < 1$ leads to that $i(t+1) < i(t)$ or the number of infectious cases drops, meaning the spread of virus is controlled; otherwise, the number of infectious cases will keep increasing.

## 2.3 Estimation and inference

To fit the model and estimate the time-dependent parameters, we can use nonparametric techniques, such as splines [38–43], local polynomial regression [44] and reproducible kernel Hilbert space method [45]. In particular, we consider a cubic B-spline approximation [46].

Denote by $\mathbf{B}(t) = \{B_1(t), \ldots, B_q(t)\}_T$ the $q$ cubic B-spline basis functions over $[0, T]$ associated with the knots $0 = w_0 < w_1 < \ldots < w_{q-2} < w_{q-1} = T$. For added flexibility, we allow the

number of knots to differ between $\beta(t)$ and $\gamma(t)$ and specify

$$
\begin{aligned}
\log \beta(t) &= \sum_{j=1}^{q_1} \beta_j B_j(t), \\
\log \gamma(t) &= \sum_{j=1}^{q_2} \gamma_j B_j(t).
\end{aligned}
\tag{8}
$$

When $\beta_1 = \cdots = \beta_{q_1}$ and $\gamma_1 = \cdots = \gamma_{q_2}$, the model reduces to a constant SIR model [46]. We use cross-validation to choose $q_1$ and $q_2$ in our numerical experiments.

Denote by $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{q_1})$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{q_2})$ the unknown parameters, by $Z_I(t)$ and $Z_R(t)$ the reported numbers of infectious and removed, respectively, and by $z_I(t) = Z_I(t)/N$ and $z_R(t) = Z_R(t)/N$, the reported proportions. Also, denote by $I(t)$ and $R(t)$ the true numbers of infectious and removed, respectively at time $t$. We propose a Poisson model to link $Z_I(t)$ and $Z_R(t)$ to $I(t)$ and $R(t)$ as follows:

$$
\begin{aligned}
Z_R(t) &\sim \text{Pois}\{R(t)\}, \\
Z_I(t) &\sim \text{Pois}\{I(t)\}.
\end{aligned}
\tag{9}
$$

We also assume that, given $I(t)$ and $R(t)$, the observed daily number $\{Z_I(t), Z_R(t)\}$ are independent across $t = 1, \ldots, T$, meaning the random reporting errors are "white" noise. We note that (9) is directly based on "true" numbers of infectious cases and removed cases derived from the discrete SIR model (6). This differs from the Markov process approach, which is based on the past observations.

With (6), (7) and (8), $R(t)$ and $I(t)$ are the functions of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, since $R(t) = N \times r(t)$ and $I(t) = N \times i(t)$. Given the data $(Z_I(t), Z_R(t))$, $t = 1, \ldots, T$, we obtain $(\hat{\boldsymbol{\beta}}, \hat{\gamma})$, the estimates of $(\boldsymbol{\beta}, \boldsymbol{\gamma})$, by maximizing the following likelihood

$$
L(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{t=1}^{T} \frac{e^{-R(t)} R(t)^{Z_R(t)}}{Z_R(t)!} \times \prod_{t=1}^{T} \frac{e^{-I(t)} I(t)^{Z_I(t)}}{Z_I(t)!},
$$

or, equivalently, maximizing the log likelihood function

$$
\ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) = N \sum_{t=1}^{T} \{-r(t) + z_R(t) \log r(t) - i(t) + z_I(t) \log i(t)\} + C,
\tag{10}
$$

where $C$ is a constant free of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. See the S1 Appendix for additional details of optimization.

We then estimate the variance-covariance matrix of $(\hat{\boldsymbol{\beta}}, \hat{\gamma})$ by inverting the second derivative of $-\ell(\boldsymbol{\beta}, \boldsymbol{\gamma})$ evaluated at $(\hat{\boldsymbol{\beta}}, \hat{\gamma})$. Finally, for $t = 1, \ldots, T$, we estimate $I(t)$ and $R(t)$ by $\hat{I}(t) = N\hat{i}(t)$ and $\hat{R}(t) = N\hat{r}(t)$, where $\hat{i}(t)$ and $\hat{r}(t)$ are obtained from (7) with all unknown quantities replaced by their estimates; estimate $\beta(t)$ and $\gamma(t)$ by $\hat{\beta}(t)$ and $\hat{\gamma}(t)$, obtained by using (8) with $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ replaced by $(\hat{\boldsymbol{\beta}}, \hat{\gamma})$; and estimate $\mathcal{R}_0(t)$ by $\hat{\mathcal{R}}_0(t) = \hat{\beta}(t)/\hat{\gamma}(t)$.

**Summary of estimation and inference for $\beta(t)$, $\gamma(t)$, $\mathcal{R}_0(t)$, $I(t)$, $R(t)$**

**Estimation:** Let $N$ be the size of population of a given country. The date when the first case was reported is set to be the starting date with $t = 1$, $i_0 = Z_I(1)/N$ and $r_0 = Z_R(1)/N$. The observed data are $\{Z_I(t), Z_R(t), t = 1, \ldots, T\}$, obtained from the GitHub data repository website mentioned in the introduction. We maximize (10) to obtain $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_{q_1})$ and $\hat{\gamma} = (\hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_{q_2})$. The optimal $q_1$ and $q_2$ are obtained via cross-validation. We denote by

$\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_{q_1})$ and $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_{q_2})$, based on which we calculate $\hat{\beta}(t), \hat{\gamma}(t), \hat{\mathcal{R}}_0(t), \hat{R}(t), \hat{I}(t)$.

**Inference:** The estimated variance-covariance matrix of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$, denoted by $\hat{V}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$, can be obtained by inverting the second derivative of $-\ell(\boldsymbol{\beta}, \boldsymbol{\gamma})$ evaluated at $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$. For each $t$, as $\hat{\beta}(t), \hat{\gamma}(t), \hat{\mathcal{R}}_0(t), \hat{R}(t)$ and $\hat{I}(t)$ are smooth functions of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$, we apply the delta method [47] to estimate their variances and obtain the confidence intervals. As an illustration, we compute $\widehat{\text{var}}(\hat{R}(t)) = \dot{\hat{R}}(t)^{\text{T}} \hat{V}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) \dot{\hat{R}}(t)$ and $\widehat{\text{var}}(\hat{I}(t)) = \dot{\hat{I}}(t)^{\text{T}} \hat{V}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) \dot{\hat{I}}(t)$, where $\dot{\hat{R}}(t)$ and $\dot{\hat{I}}(t)$ are the partial derivative vectors of $\hat{R}(t)$ and $\hat{I}(t)$ with respect to $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$.

# 3 Analysis of the COVID-19 pandemic among severely affected countries

Since the first case of COVID-19 was detected in China, it quickly spread to nearly every part of the world [6]. COVID-19, conjectured to be more contagious than the previous SARS and H1N1 [48], has put great strain on healthcare systems worldwide, especially among the severely affected countries [49]. We apply our method to assess the epidemiological processes of COVID-19 in some severely impacted countries.

## 3.1 Data descriptions and robustness of the method towards specifications of the initial conditions

The country-specific time-series data of confirmed, recovered, and death cases were obtained from a GitHub data repository website (https://github.com/ulklc/covid19-timeseries). This site collects information from various sources listed below on a daily basis at GMT 0:00, converts the data to the CSV format, and conducts data normalization and harmonization if inconsistencies are found. The data sources include

- World Health Organization (WHO): https://www.who.int/

- DXY.cn. Pneumonia 2020: http://3g.dxy.cn/newh5/view/pneumonia.

- BNO News: https://bnonews.com/index.php/2020/02/the-latest-coronavirus-cases/

- National Health Commission of China (NHC): http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml

- China CDC (CCDC): http://weekly.chinacdc.cn/news/TrackingtheEpidemic.htm

- Hong Kong Department of Health: https://www.chp.gov.hk/en/features/102465.html

- Macau Government: https://www.ssm.gov.mo/portal/

- Taiwan CDC: https://sites.google.com/cdc.gov.tw/2019ncov/taiwan?authuser=0

- US CDC: https://www.cdc.gov/coronavirus/2019-ncov/index.html

- Government of Canada: https://www.canada.ca/en/public-health/services/diseases/coronavirus.html

- Australia Government Department of Health: https://www.health.gov.au/news/coronavirus-update-at-a-glance

- European Centre for Disease Prevention and Control (ECDC): https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases

- Ministry of Health Singapore (MOH): https://www.moh.gov.sg/covid-19

- Italy Ministry of Health: http://www.salute.gov.it/nuovocoronavirus

- Johns Hopkins CSSE: https://github.com/CSSEGISandData/COVID-19

- WorldoMeter: https://www.worldometers.info/coronavirus/

In particular, the current population size of each country, $N$, came from the website of WorldoMeter. Our analyses covered the periods between the date of the first reported coronavirus case in each nation and June 30, 2020. In the beginning of the outbreak, assessment of $i_0$ and $r_0$ was problematic as infectious but asymptomatic cases tended to be undetected due to lack of awareness and testing. To investigate how our method depends on the correct specification of the initial values $r_0$ and $i_0$, we conducted Monte Carlo simulations. As a comparison, we also studied the performance of the deterministic SIR model in the same settings. Fig 1 shows that, when the initial value $i_0$ was mis-specified to be 5 times of the truth, the curves of $i(t)$ and $r(t)$ obtained by the deterministic SIR model (6) were considerably biased. On the other hand, our proposed model (9), by accounting for the randomness of the observed data, was robust toward the mis-specification of $i_0$ and $r_0$: the estimates of $r(t)$ and $i(t)$ had negligible biases even with mis-specified initial values. In an omitted analysis, we mis-specified $i_0$ and $r_0$ to be only twice of the truth, and obtain the similar results.

Our numerical experiments also suggested that using the time series, starting from the date when both cases and removed were reported, may generate more reasonable estimates.

## 3.2 Estimation of country-specific transmission, removal rates and reproduction numbers

Using the cubic B-splines (8), we estimated the time-dependent transmission rate $\beta(t)$ and removal rate $\gamma(t)$, based on which we further estimated $\mathcal{R}_0(t)$, $I(t)$ and $R(t)$. To choose the optimal number of knots for each country when implementing the spline approach, we used 5-fold cross-validation by minimizing the combined mean squared error for the estimated infectious and removed cases.

Fig 2 shows sharp variations in transmission rates and removal rates across different time periods, indicating the time-varying nature of these rates. The estimated $I(t)$ and $R(t)$ overlapped well with the observed number of infectious and removed cases, indicating the reasonableness of the method. The pointwise 95% confidence intervals (in yellow) represent the uncertainty of the estimates, which may be due to error in reporting. Fig 3 presents the estimated time-varying reproduction number, $\hat{\beta}(t)\hat{\gamma}(t)^{-1}$, for several countries. The curves capture the evolving trends of the epidemic for each country.

In the US, though the first confirmed case was reported on January 20, 2020, lack of immediate actions in the early stage let the epidemic spread widely. As a result, the US had seen soaring infectious cases, and $\mathcal{R}_0(t)$ reached its peak around mid-March. From mid-March to early April, the US tightened the virus control policy by suspending foreign travels and closing borders, and the federal government and most states issued mandatory or advisory stay-home orders, which seemed to have substantially contained the virus.

The high reproduction numbers with China, Italy, and Sweden at the onset of the pandemic imply that the spread of the infectious disease was not well controlled in its early phases. With the extremely stringent mitigation policies such as city lockdown and mandatory mask-wearing implemented in the end of January, China was reported to bring its epidemic under
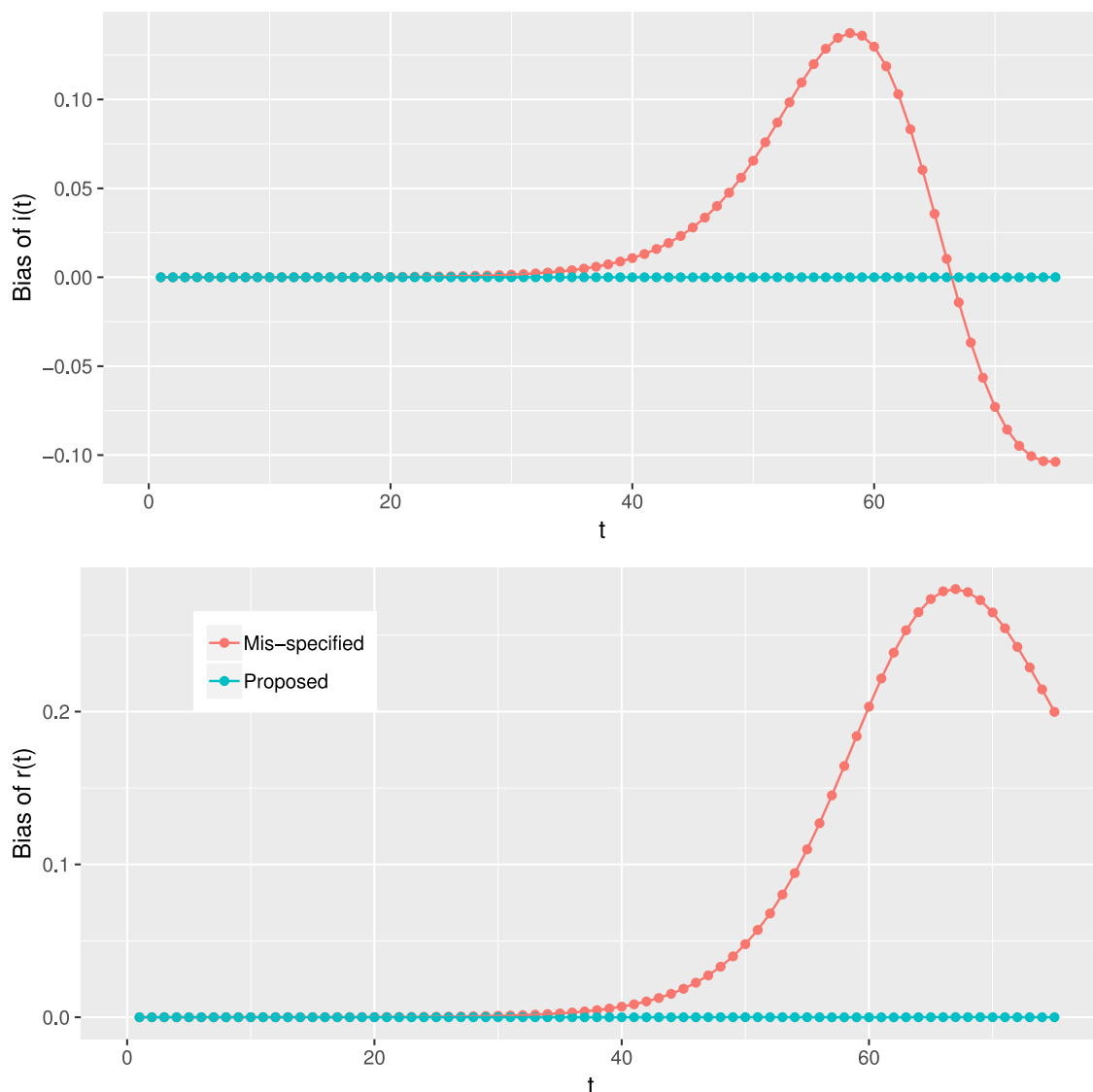
**Fig 1. The impact of mis-specification of $i_0$ and $r_0$.** Plots of the relative biases of $\hat{i}(t)$ (upper) and $\hat{r}(t)$ (lower) when $\hat{i}(t)$ and $\hat{r}(t)$ are derived 1) by using the ODE framework with the mis-specified initials ("Mis-specified") and 2) by using proposed model with mis-specified initials ("Proposed"). In the model, the true $(\beta, \gamma) = (e^{-1}, e^{-1.95})$ and $(i_0, r_0) = (10^{-6}, 10^{-6})$. These values are roughly equal to the constant estimates of the real situation. The mis-specified initials are set as $(5 \times 10^{-6}, 5 \times 10^{-6})$.

https://doi.org/10.1371/journal.pone.0236464.g001

control with a quickly dropping $\mathcal{R}_0(t)$ in February. This indicates that China might have contained the epidemic, with more people removed from infectious status than those who became infectious.

Sweden is among the few countries that imposed more relaxed measures to control coronavirus and advocated herd immunity. The Swedish approach has initiated much debate. While some criticized that this may endanger the general population in a reckless way, some felt this might terminate the pandemic more effectively in the absence of vaccines [50]. Fig 3 demonstrates that Sweden has a large reproduction number, which however keeps decreasing. The "big V" shape of the reproduction number around May 1 might be due to the reporting errors or lags. Our investigation found that the reported number of infectious cases in that period suddenly dropped and then quickly rose back, which was unusual.
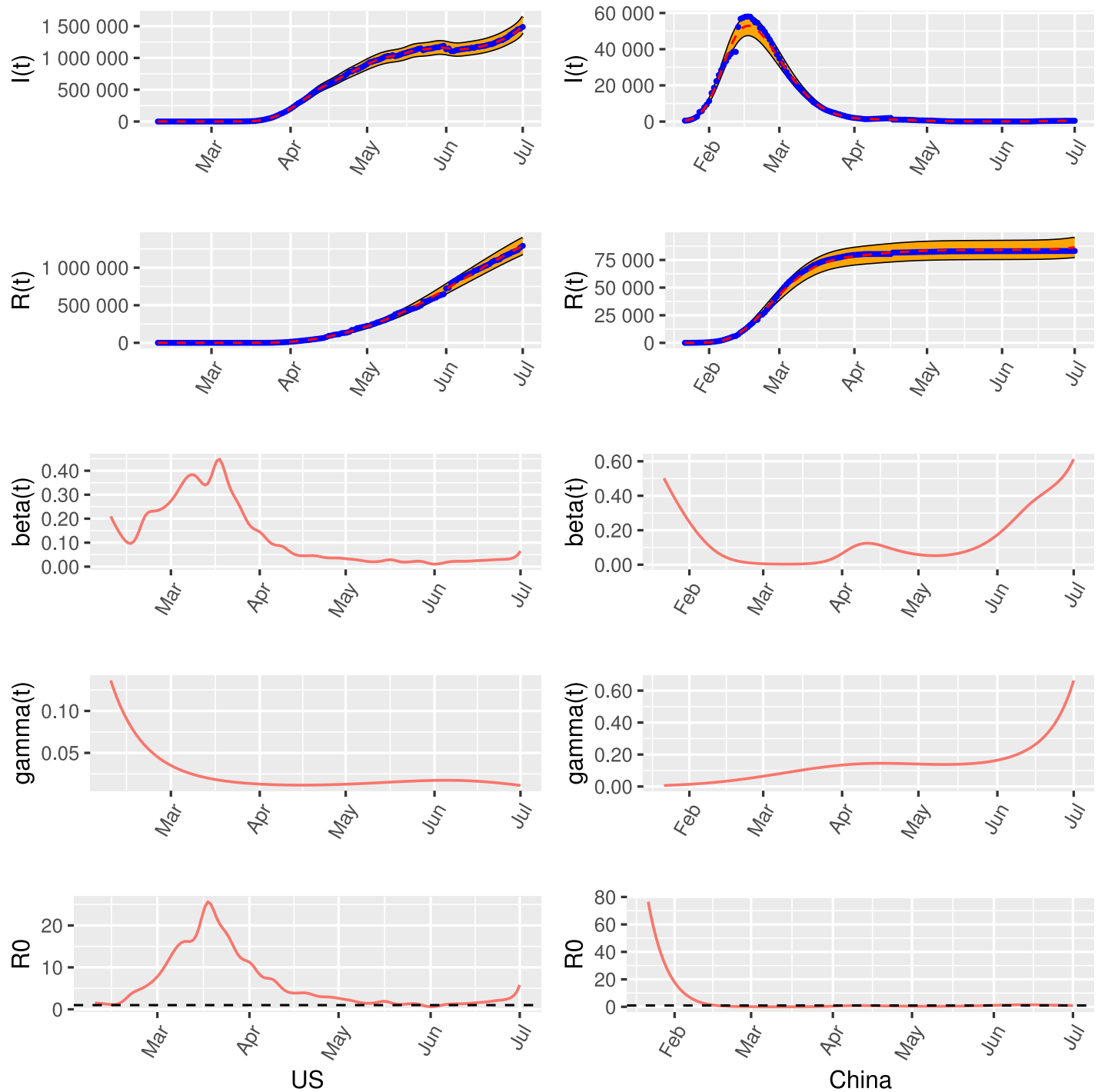
**Fig 2. Estimated reproduction number $\mathcal{R}_0(t)$ for selected countries based on the data up to June 30, 2020.** The US (left) and China (right) are shown based on the data up to June 30, 2020. The blue dots and the red dashed curves represent the observed data and the model-based predictions, respectively, with 95% confidence interval.

https://doi.org/10.1371/journal.pone.0236464.g002

**Fig 3. Estimated $I(t)$, $R(t)$, $\beta(t)$, $\gamma(t)$, and $\mathcal{R}_0(t)$.** The US (left) and China (right) are shown based on the data up to June 30, 2020. The blue dots and the red dashed curves represent the observed data and the model-based predictions, respectively, with 95% confidence interval.
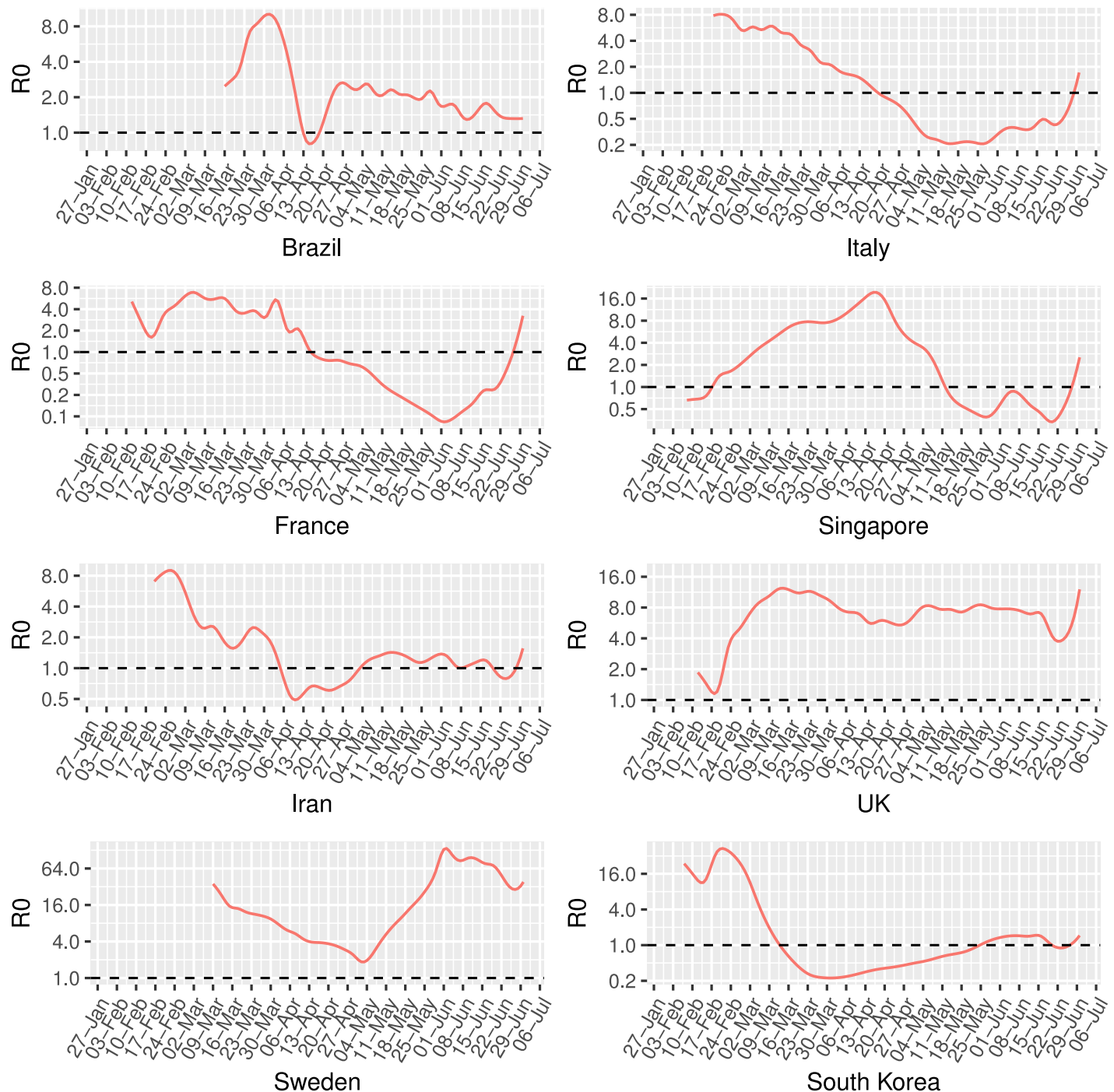
Around February 18, a surge in South Korea was linked to a massive cluster of more than 5,000 cases [51]. The outbreak was clearly depicted in the time-varying $\mathcal{R}_0(t)$ curve. Since then, South Korea appeared to have slowed its epidemic, likely due to expansive testing programs and extensive efforts to trace and isolate patients and their contacts [52].

More broadly, Fig 3 categorizes countries into two groups. One group features the countries which have contained coronavirus. Countries, such as China and South Korea, took aggressive actions after the outbreak and presented sharper downward slopes. Some European countries such as Italy and Spain and Mideastern countries such as Iran, which were hit later than the East Asian countries, share a similar pattern, though with much flatter slopes. On the other hand, the US, Brazil, and Sweden are still struggling to contain the virus, with the $\mathcal{R}_0(t)$ curves hovering over 1. We also caution that, among the countries whose $\mathcal{R}_0(t)$ dropped below 1, the curves of the reproduction numbers are beginning to uptick, possibly due to the resumed economy activities.

## 3.3 An interactive web application and R code

We have developed a web application (https://younghhk.shinyapps.io/tvSIRforCOVID19/) to facilitate users' application of the proposed method to compute the time-varying reproduction number, and estimated and predict the daily numbers of active cases and removed cases for the presented countries and other countries; see Fig 4 for an illustration.

Our code was written in R [53], using the `bs` function in the `splines` package for cubic B-spline approximation, the `nlm` function in the `stats` package for nonlinear minimization, and the `jacobian` function in the `numDeriv` package for computation of gradients and hessian matrices. Graphs were made by using the `ggplot2` package. Our code can be found on the aforementioned shiny website.
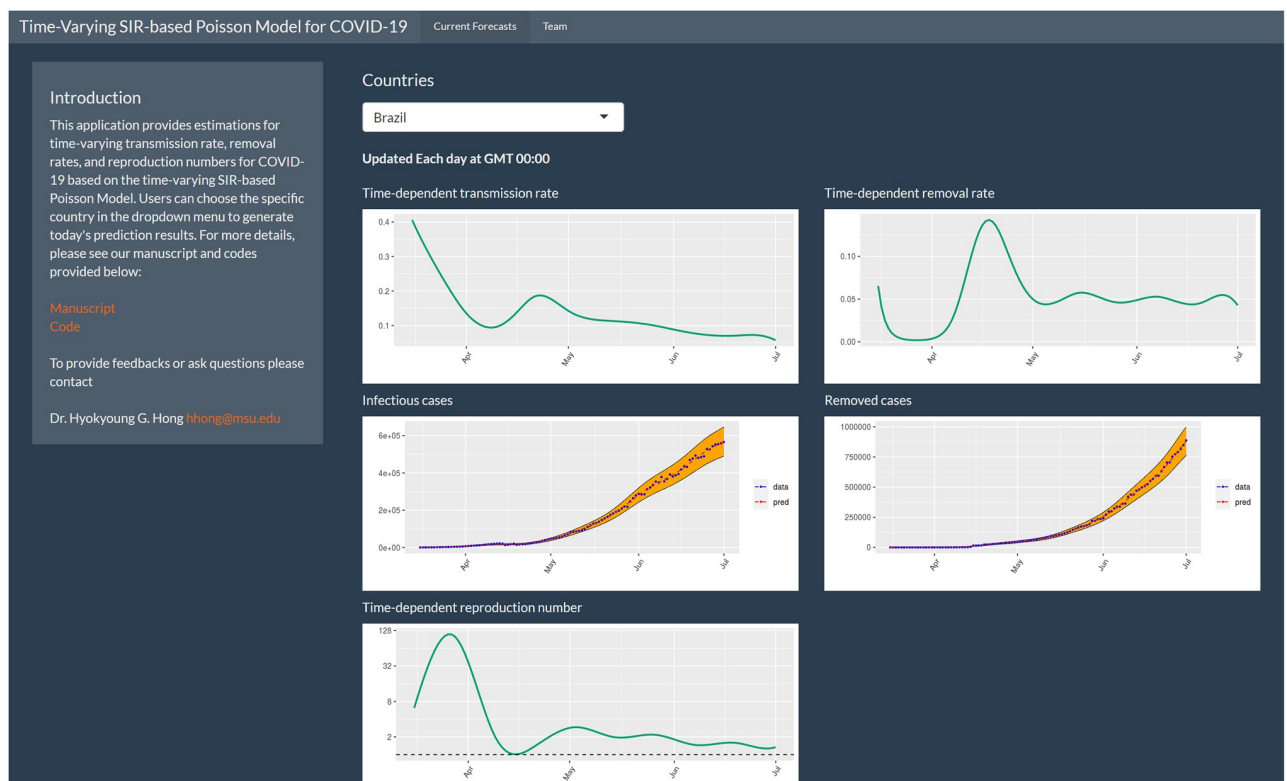


**Fig 4. An illustration of the developed interactive web application.**

https://doi.org/10.1371/journal.pone.0236464.g004

## 4 Discussion

The rampaging pandemic of COVID-19 has called for developing proper computational and statistical tools to understand the trend of the spread of the disease and evaluate the efficacy of mitigation measures [54–57]. We propose a Poisson model with time-dependent transmission and removal rates. Our model accommodates possible random errors and estimates a time-dependent disease reproduction number, $\mathcal{R}_0(t)$, which can serve as a metric for timely evaluating the effects of health policies.

There have been substantial issues, such as biases and lags, in reporting infectious cases, recovery, and deaths, especially at the early stage of the outbreak. As opposed to the deterministic SIR models that heavily rely on accurate reporting of initial infectious and removed cases, our model is more robust towards mis-specifications of such initial conditions. Applications of our method to study the epidemics in selected countries illustrate the results of the virus containment policies implemented in these countries, and may serve as the epidemiological benchmarks for the future preventive measures.

Several methodological questions need to be addressed. First, we analyzed each country separately, without considering the traffic flows among these countries. We will develop a joint model for the global epidemic, which accounts for the geographic locations of and the connectivity among the countries.

Second, incorporating timing of public health interventions such as the shelter-in-place order into the model might be interesting. However, we opted not to follow this approach as no such information exists for the majority countries. On the other hand, the impact of the interventions or the change point can be embedded into our nonparametric time-dependent estimates.

Third, the validity of the results of statistical models eventually hinges on the data transparency and accuracy. For example, the results of Chinazzi et al. [58] suggested that in China only one of four cases were detected and confirmed. Also, asymptomatic cases might have been undetected in many countries. All of these might have led to underestimation of the actual number of cases. Moreover, the collected data could be biased toward patients with severe infection and with insurance, as these patients were more likely to seek care or get tested. More in-depth research is warranted to address the issue selection bias.

Finally, our present work is within the SIR framework, where removed individuals include recovery and deaths, who hypothetically are unlikely to infect others. Although this makes the model simpler and widely adopted, the interpretation of the $\gamma$ parameter is not straightforward. Our subsequent work is to develop a susceptible-infectious-recovered-deceased (SIRD) model, in which the number of deaths and the number of recovered are separately considered. We will report this elsewhere.

## 5 Conclusion

Containment of COVID-19 requires the concerted effort of health care workers, health policy makers as well as citizens. Measures, e.g. self-quarantine, social distancing, and shelter in place, have been executed at various phases by each country to prevent the community transmission. Timely and effective assessment of these actions constitutes a critical component of the effort. SIR models have been widely used to model this pandemic. However, constant transmission and removal rates may not capture the timely influences of these policies.

We propose a time-varying SIR Poisson model to assess the dynamic transmission patterns of COVID-19. With the virus containment measures taken at various time points, $\mathcal{R}_0$ may vary substantially over time. Our model provides a systematic and daily updatable tool to evaluate the immediate outcomes of these actions. It is likely that the pandemic is ending and

many countries are now shifting gear to reopen the economy, while preparing to battle the second wave of virus attack [59, 60]. Our tool may shed light on and aid the implementation of future containment strategies.

## Supporting information

**S1 Appendix. Details of optimization.**
(PDF)

## Author Contributions

**Conceptualization:** Hyokyoung G. Hong, Yi Li.

**Data curation:** Hyokyoung G. Hong.

**Formal analysis:** Hyokyoung G. Hong, Yi Li.

**Investigation:** Hyokyoung G. Hong.

**Methodology:** Hyokyoung G. Hong, Yi Li.

**Software:** Hyokyoung G. Hong, Yi Li.

**Writing – review & editing:** Hyokyoung G. Hong, Yi Li.

## References

1. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. Methods Mol Biol. 2015; 1282:1–23. https://doi.org/10.1007/978-1-4939-2438-7_1

2. Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, et al. Bats are natural reservoirs of SARS-like coronaviruses. Science. 2005; 310(5748):676–679. https://doi.org/10.1126/science.1118391 PMID: 16195424

3. Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, Lau JH, et al. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. Journal of Virology. 2012; 86(7):3995–4008. https://doi.org/10.1128/JVI. 06540-11 PMID: 22278237

4. Trombetta H, Faggion HZ, Leotte J, Nogueira MB, Vidal LR, Raboni SM. Human coronavirus and severe acute respiratory infection in Southern Brazil. Pathogens and Global Health. 2016; 110(3):113–118. https://doi.org/10.1080/20477724.2016.1181294

5. Wang C, Liu L, Hao X, Guo H, Wang Q, Huang J, et al. Evolving epidemiology and impact of non-pharmaceutical interventions on the outbreak of coronavirus disease 2019 in Wuhan, China. 2020;

6. Johns Hopkins Cornonavirus Resource Center; 2020. Available from: https://coronavirus.jhu.edu/.

7. Hall I, Gani R, Hughes H, Leach S. Real-time epidemic forecasting for pandemic influenza. Epidemiology & Infection. 2007; 135(3):372–385. https://doi.org/10.1017/S0950268806007084

8. Grassly NC, Fraser C. Mathematical models of infectious disease transmission. Nature Reviews Microbiology. 2008; 6(6):477–487. https://doi.org/10.1038/nrmicro1845

9. Chang SL, Harding N, Zachreson C, Cliff OM, Prokopenko M. Modelling transmission and control of the COVID-19 pandemic in Australia. arXiv:2003.10218; 2020.

10. Pellis L, Scarabel F, Stage HB, Overton CE, Chappell LH, Lythgoe KA, et al. Challenges in control of Covid-19: short doubling time and long delay to effect of interventions. arXiv:2004.00117; 2020.

11. Laguzet L, Turinici G. Individual vaccination as Nash equilibrium in a SIR model with application to the 2009–2010 influenza A (H1N1) epidemic in France. Bulletin of Mathematical Biology. 2015; 77 (10):1955–1984. https://doi.org/10.1007/s11538-015-0111-7

12. Schwartz EJ, Choi B, Rempala GA. Estimating epidemic parameters: Application to H1N1 pandemic data. Mathematical Biosciences. 2015; 270:198–203. https://doi.org/10.1016/j.mbs.2015.03.007

13. Huang X, Clements AC, Williams G, Mengersen K, Tong S, Hu W. Bayesian estimation of the dynamics of pandemic (H1N1) 2009 influenza transmission in Queensland: A space–time SIR-based model. Environmental Research. 2016; 146:308–314. https://doi.org/10.1016/j.envres.2016.01.013

14. Mkhatshwa T, Mummert A. Modeling super-spreading events for infectious diseases: case study SARS. arXiv:1007.0908; 2010.

15. Giraldo JO, Palacio DH. Deterministic SIR (Susceptible–Infected–Removed) models applied to varicella outbreaks. Epidemiology & Infection. 2008; 136(5):679–687. https://doi.org/10.1017/S0950268807009260

16. Blackwood JC, Childs LM. An introduction to compartmental modeling for the budding infectious disease modeler. Letters in Biomathematics. 2018; 5(1):195–221. https://doi.org/10.30707/LiB5.1Blackwood

17. Cox LA, Jr. Risk Analysis Foundations, Models, and Methods. Springer Science & Business Media; 2012.

18. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London Series A. 1927; 115(772):700–721.

19. Nesteruk I. Statistics based predictions of coronavirus 2019-nCoV spreading in mainland China. medRxiv. 2020;

20. Chen Y, Cheng J, Jiang Y, Liu K. A time delay dynamical model for outbreak of 2019-nCoV and the parameter identification. arXiv:2002.00418; 2020.

21. Peng L, Yang W, Zhang D, Zhuge C, Hong L. Epidemic analysis of COVID-19 in China by dynamical modeling. arXiv:2002.06563; 2020.

22. Zhou T, Liu Q, Yang Z, Liao J, Yang K, Bai W, et al. Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. Journal of Evidence-Based Medicine; 2020. https://doi.org/10.1111/jebm.12376 PMID: 32048815

23. Maier BF, Brockmann D. Effective containment explains sub-exponential growth in confirmed cases of recent COVID-19 outbreak in mainland China. arXiv:2002.07572; 2020.

24. Tognotti E. Lessons from the history of quarantine, from plague to influenza A. Emerging Infectious Diseases. 2013; 19(2):254. https://doi.org/10.3201/eid1902.120312

25. Chen YC, Lu PE, Chang CS, Liu TH. A Time-dependent SIR model for COVID-19 with undetectable infected persons. arXiv:2003.00122 [q-bio.PE]; 2020.

26. Boatto S, Bonnet C, Cazelles B, Mazenc F. SIR model with time dependent infectivity parameter: approximating the epidemic attractor and the importance of the initial phase; 2018. Available from: https://hal.inria.fr/hal-01677886.

27. Song PX, Wang L, Zhou Y, He J, Zhu B, Wang F, et al. An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China. medRxiv. 2020;

28. Hilbe JM. Modeling Count Data. Cambridge University Press; 2014.

29. Kirkeby CT, Halasa T, Gussmann MK, Toft N, Græsbøll K. Methods for estimating disease transmission rates: Evaluating the precision of Poisson regression and two novel methods. Scientific Reports. 2018; 8(1). https://doi.org/10.1038/s41598-018-26491-5

30. O'Dea EB, Pepin KM, Lopman BA, Wilke CO. Fitting outbreak models to data from many small norovirus outbreaks. Epidemics. 2014; 6:18–29. https://doi.org/10.1016/j.epidem.2013.12.002

31. Zhuang L, Cressie N, Pomeroy L, Janies D. Multi-species SIR models from a dynamical Bayesian perspective. Theoretical ecology. 2013; 6(4):457–473. https://doi.org/10.1007/s12080-013-0180-x

32. Dietz K. The estimation of the basic reproduction number for infectious diseases. Statistical Methods in Medical Research. 1993; 2(1):23–41. https://doi.org/10.1177/096228029300200103

33. Diekmann O, Heesterbeek JAP. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. vol. 5. John Wiley & Sons; 2000.

34. Eichner M, Dietz K. Transmission potential of smallpox: estimates based on detailed data from an outbreak. American Journal of Epidemiology. 2003; 158(2):110–117. https://doi.org/10.1093/aje/kwg103

35. Liu QH, Ajelli M, Aleta A, Merler S, Moreno Y, Vespignani A. Measurability of the epidemic reproduction number in data-driven contact networks. Proceedings of the National Academy of Sciences. 2018; 115(50):12680–12685. https://doi.org/10.1073/pnas.1811115115

36. Chen YC, Lu PE, Chang CS, Liu T. A Time-dependent SIR model for COVID-19 with undetectable infected persons. arXiv preprint arXiv:200300122. 2020;.

37. Jones H. Notes on R0, Stanford University, Stanford; 2007.

38. De Boor C. A Practical Guide to Splines. vol. 27. Springer-Verlag New York; 1978.

39. Ramsay JO, Hooker G, Campbell D, Cao J. Parameter estimation for differential equations: a generalized smoothing approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2007; 69(5):741–796. https://doi.org/10.1111/j.1467-9868.2007.00610.x

**40.** Lawrence ND, Sanguinetti G, Rattray M. Modelling transcriptional regulation using Gaussian processes. In: Advances in Neural Information Processing Systems; 2007. p. 785–792.

**41.** Alvarez MA, Luengo D, Lawrence ND. Linear latent force models using Gaussian processes. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013; 35(11):2693–2705. https://doi.org/10.1109/TPAMI.2013.86

**42.** Alvarez M, Luengo D, Lawrence ND. Latent force models. In: Artificial Intelligence and Statistics; 2009. p. 9–16.

**43.** Wheeler MW, Dunson DB, Pandalai SP, Baker BA, Herring AH. Mechanistic hierarchical Gaussian processes. Journal of the American Statistical Association. 2014; 109(507):894–904. https://doi.org/10.1080/01621459.2014.899234

**44.** Ruppert D. Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. Journal of the American Statistical Association. 1997; 92(439):1049–1062. https://doi.org/10.1080/01621459.1997.10474061

**45.** Akgül A. New reproducing kernel functions. Mathematical Problems in Engineering; 2015.

**46.** Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M. A review of spline function procedures in R. BMC medical research methodology. 2019; 19(1):46. https://doi.org/10.1186/s12874-019-0666-3

**47.** Parr WC. A note on the jackknife, the bootstrap and the delta method estimators of bias and variance. Biometrika. 1983; 70(3):719–722. https://doi.org/10.1093/biomet/70.3.719

**48.** Sansonetti P. COVID-19, chronicle of an expected pandemic. EMBO Molecular Medicine. 2020; https://doi.org/10.15252/emmm.202012463.

**49.** Tanne JH, Hayasaki E, Zastrow M, Pulla P, Smith P, Rada AG. Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide. BMJ. 2020; 368.

**50.** Paterlini M. 'Closing borders is ridiculous': the epidemiologist behind Sweden's controversial coronavirus strategy; 2020 (accessed 04/21/20). https://www.nature.com/articles/d41586-020-01098-x.

**51.** Yoon D, Martin TW. Why a South Korean church was the perfect petri dish for coronavirus; 2020 (accessed 03/02/20). Available from: https://www.wsj.com/articles/why-a-south-korean-church-was-the-perfect-petri-dish-for-coronavirus-11583082110.

**52.** Normile D. Coronavirus cases have dropped sharply in South Korea. What's the secret to its success?; 2020 (accessed 03/02/20). Available from: https://www.sciencemag.org/news/2020/03/coronavirus-cases-have-dropped-sharply-south-korea-whats-secret-its-success.

**53.** R Core Team. R: A Language and Environment for Statistical Computing; 2013. Available from: http://www.R-project.org/.

**54.** Hossain MM. Current status of global research on novel coronavirus disease (COVID-19): A bibliometric analysis and knowledge mapping. Available at SSRN 3547824. 2020;.

**55.** Sarkodie SA, Owusu PA. Investigating the cases of novel coronavirus disease (COVID-19) in China using dynamic statistical techniques. Heliyon. 2020; 6:e03747. https://doi.org/10.1016/j.heliyon.2020.e03747

**56.** Zhang Y, Jiang B, Yuan J, Tao Y. The impact of social distancing and epicenter lockdown on the COVID-19 epidemic in mainland China: A data-driven SEIQR model study. medRxiv. 2020;

**57.** Picchiotti N, Salvioli M, Zanardini E, Missale F. COVID-19 Italian and Europe epidemic evolution: A SEIR model with lockdown-dependent transmission rate based on Chinese data; 2020.

**58.** Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science. 2020; 368(6489):395–400. https://doi.org/10.1126/science.aba9757 PMID: 32144116

**59.** As China's virus cases reach zero, experts warn of second wave; 2020 (accessed 03/18/20). Available from: https://www.bloomberg.com/news/articles/2020-03-18/as-china-virus-cases-near-zero-experts-warn-of-second-wave.

**60.** Coronavirus: Asian nations face second wave of imported cases; 2020 (accessed 03/19/20). Available from: https://www.bbc.com/news/world-asia-51955931.