*Mathematical Biosciences and Engineering*

http://www.aimspress.com/journal/MBE

MBE, 18(6): 9727–9742.
DOI: 10.3934/mbe.2021476
Received: 19 July 2021
Accepted: 28 October 2021
Published: 04 November 2021

*Research article*

# Impact of insufficient detection in COVID-19 outbreaks

**Yue Deng[1], Siming Xing[2], Meixia Zhu[3] and Jinzhi Lei[2,4,*]**

[1] School of Computer Science and Technology, Tiangong University, Tianjin, 300387, China
[2] School of Mathematical Sciences, Tiangong University, Tianjin, 300387, China
[3] School of Software, Tiangong University, Tianjin, 300387, China
[4] Center for Applied Mathematics, Tiangong University, Tianjin, 300387, China

* **Correspondence:** Email: jzlei@tiangong.edu.cn.

**Abstract:** The COVID-19 (novel coronavirus disease 2019) pandemic has tremendously impacted global health and economics. Early detection of COVID-19 infections is important for patient treatment and for controlling the epidemic. However, many countries/regions suffer from a shortage of nucleic acid testing (NAT) due to either resource limitations or epidemic control measures. The exact number of infective cases is mostly unknown in counties/regions with insufficient NAT, which has been a major issue in predicting and controlling the epidemic. In this paper, we propose a mathematical model to quantitatively identify the influences of insufficient detection on the COVID-19 epidemic. We extend the classical SEIR (susceptible-exposed-infections-recovered) model to include random detections which are described by Poisson processes. We apply the model to the epidemic in Guam, Texas, the Virgin Islands, and Wyoming in the United States and determine the detection probabilities by fitting model simulations with the reported number of infected, recovered, and dead cases. We further study the effects of varying the detection probabilities and show that low level-detection probabilities significantly affect the epidemic; increasing the detection probability of asymptomatic infections can effectively reduce the the scale of the epidemic. This study suggests that early detection is important for the control of the COVID-19 epidemic.

**Keywords:** COVID-19; nucleic acid testing; detection probability; epidemic scale

## 1. Introduction

Coronavirus disease 2019 (COVID-19), a pandemic disease caused by the novel severe acute respiratory syndrome (SARS)-like coronavirus (SARS-CoV-2), has spread globally since it was first reported in December 2019 in Wuhan, China. As of May 25, 2021, there were more than 166 million confirmed cases including more than 3 million deaths [1]. Controlling the rapid spread of COVID-19

has been an emergency global public health issue. Many countries have implemented different types of nonpharmaceutical interventions (NPIs) to control the COVID-19 epidemic, such as restricting travel, stopping parties, closing cities, closing schools, and self-protection [2–4]. Furthermore, vaccines have been widely used in many countries. However, the epidemic situation is far from under control, and the second wave of COVID-19 erupted in India in April 2021, resulting in more than 300,000 new cases reported every day. Importantly, the number of reported cases may be much lower than the exact number due to the low coverage rate of nucleic acid testing (NAT). Many countries/regions are suffering from an NAT shortage due to either resource limitations or control measures. Insufficient detection of COVID-19 may seriously affect the clinical intervention of infected patients and forecasting of epidemic tends. Nevertheless, the impact of insufficient detection of COVID-19 has not been clearly quantified.

Many mathematical models have been proposed to investigate the effects of various measures on controlling the epidemic and forecast the dynamics of the spread of COVID-19. Most models are formulated as differential equations that originate from the classical compartmental dynamics of SIR (susceptible-infectious-recovered) or SEIR (susceptible-exposed-infectious-recovered) models [5–9]. Many studies have attempted to forecast of the COVID-19 epidemic based on the established model formulation and parameters estimated from reported data to forecast the COVID-19 epidemic [10–15]. Alternatively, data-driven studies try to forecast epidemic dynamics directly from statistical analyses of reported data [16–20]. Other studies have attempted to improve the estimation of COVID-19 mortality by combining historical and current mortality data, statistical test models, and SIR epidemic models [21, 22]. However, some counties do not perform NAT for light or moderate symptomatic infections, potentially leading to missing data and serious prediction problems. More importantly, some dead cases are not detected and hence are missing from reported data. Hence, the estimation of real infected cases from reported dead case numbers can be misleading and modeling and forecasting the spread of COVID-19 remains a challenge [23].

This study was intended to investigate the impact of insufficient detection on the prediction and control of COIVD-19 outbreaks. We propose a mathematical model with random (Poisson process) detections and varying detection probabilities for infection. We use epidemic data from Guam, Texas, the Virgin Islands, and Wyoming in the United States as examples to estimate the model parameters and study the potential effects of varying the detection probabilities. Based on the proposed model, we discuss possible long-term scenarios for COVID-19 by analyzing the role of detection probabilities in reducing the final scale and duration of COVID-19 outbreaks.

## 2. Model formulation

We extended the classical SEIR model to the situation of insufficient detections of cases of infection and death. The model is illustrated in Figure 1. Clinically, COVID-19 infections can be separated into two subpopulations: asymptomatic infections ($I_1$) and symptomatic infections ($I_2$). The infection rate of a susceptible person is $\beta$ (day$^{-1}$), the transition rate from latent ($E$) to the asymptomatic infections compartment ($I_1$) is $\gamma_1$ (day$^{-1}$), and the transition rate from $I_1$ to the symptomatic infections compartment ($I_2$) is $\gamma_2$ (day$^{-1}$). To compare the model simulation with reported recovered and dead cases, we distinguished the compartments of recovered ($R$) and dead and further separated the death compartment into unreported ($D_1$) and reported ($D_2$) dead cases. The unreported dead cases mainly

come from undetected infections, with a rate $\mu_1$ (day$^{-1}$). The reported dead cases may come from the hospitalized compartment with a rate $\mu_2$ (day$^{-1}$) or the infectious compartment with a rate $\mu_3$ (day$^{-1}$). Here, we omitted death from asymptomatic infections. We assumed that asymptomatic ($I_1$) and symptomatic infected ($I_2$) patients are detected with daily detection probabilities $p_1$ and $p_2$, respectively, and the confirmed cases are moved to the hospitalized compartment ($H$). Moreover, we assumed perfect isolation so confirmed infectious cases are hospitalized or isolated immediately so that they no longer contribute to infections. We further assumed that patients in the compartments of $E$, $I_1$, and $I_2$ recover automatically, with rates $\alpha_E$, $\alpha_1$, and $\alpha_2$ (day$^{-1}$), respectively, and that hospitalized patients recover with a rate $\alpha_H$ (day$^{-1}$). The above model assumptions lead to the following differential equations for

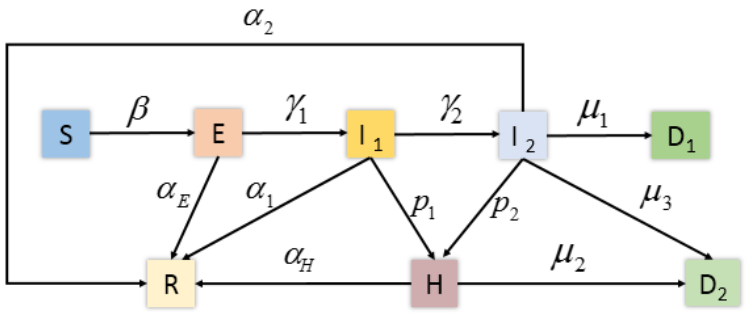


**Figure 1.** Illustration of the model of the COVID-19 epidemic with insufficient detection. Each individual transitions among states defined as susceptible ($S$), latent ($E$), asymptomatic infections ($I_1$), symptomatic infections ($I_2$), confirmed ($H$), recovered ($R$), unreported death ($D_1$), and reported death ($D_2$), following the direction of the arrows. The transition rates can vary with time.

the dynamics of different compartmental populations

$$\begin{cases} \frac{dS}{dt} &= -\beta S \frac{kI_1 + I_2}{N} \\ \frac{dE}{dt} &= \beta S \frac{kI_1 + I_2}{N} - \gamma_1 E - \alpha_E E \\ \frac{dI_1}{dt} &= \gamma_1 E - \gamma_2 I_1 - \alpha_1 I_1 - \Psi_1(I_1, p_1) \\ \frac{dI_2}{dt} &= \gamma_2 I_1 - \alpha_2 I_2 - \mu_1 I_2 - \Psi_2(I_2, p_2) - \mu_3 I_2 \\ \frac{dH}{dt} &= \Psi_1(I_1, p_1) + \Psi_2(I_2, p_2) - \alpha_H H - \mu_2 H \\ \frac{dR}{dt} &= \alpha_E E + \alpha_1 I_1 + \alpha_2 I_2 + \alpha_H H \\ \frac{dD_1}{dt} &= \mu_1 I_2 \\ \frac{dD_2}{dt} &= \mu_2 H + \mu_3 I_2 \end{cases} \tag{2.1}$$

Here, $N = S + E + I_1 + I_2 + H + R + D_1 + D_2$ represents the total population number, and is assumed to be a constant. We introduced a factor $k$ to represent the relative infection rate of asymptomatic infections ($I_1$) to symptomatic infections ($I_2$). The terms $\Psi_1(I_1, p_1)$ and $\Psi_2(I_2, p_2)$ are nonhomogeneous Poisson processes with varying arrival rates $\lambda_1(t) = p_1 I_1(t)$ and $\lambda_2(t) = p_2 I_2(t)$, respectively, which represent the number of infections patients testing positive per unit time. Thus, our model is implicitly stochastic since increments of infected individuals are randomly subtracted from $I_1$ and $I_2$ and added to the confirmed compartment $H$. The detection probabilities $p_1$ and $p_2$ are explicitly included in the model and are dependent on the epidemic control policy and NAT resources. The parameters associated with the infection rate ($\beta$), detection probabilities ($p_1$ and $p_2$), and death rates ($\mu_2$ and $\mu_3$) may vary with time, especially during the early stages of the outbreak of a novel epidemic disease, and hence are piecewise functions of time.

Model parameters and the range of parameter values are listed in Table 1.

The implicit stochastic model (2.1) can be solved numerically through a modified Euler method. Consider a differential equation of the form

$$\frac{d\vec{x}}{dt} = \vec{F}(\vec{x}, t) + A\vec{\Psi},$$

where $\vec{x} \in \mathbb{R}^n$, $\vec{F} : \mathbb{R}^n \times \mathbb{R} \mapsto \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$, and $\vec{\Psi} = (\Psi_1, \cdots, \Psi_n)^T$ with $\Psi_i$ a nonhomogeneous Poisson process with arrival rate $\lambda_i(t)$. The numerical scheme of the modified Euler method is given by

$$\vec{x}(t + \Delta t) = \vec{x}(t) + \vec{F}(\vec{x}, t)\Delta t + A \begin{bmatrix} \mathcal{P}(\lambda_1(t)\Delta t) \\ \mathcal{P}(\lambda_2(t)\Delta t) \\ \vdots \\ \mathcal{P}(\lambda_n(t)\Delta t) \end{bmatrix},$$

where $\mathcal{P}(\lambda)$ represents a Poisson distribution random number with parameter $\lambda$.

## 3. Results

### *3.1. Data collection*

Insufficient detection of COVID-19 is a common issue in many countries/regions for various reasons, such as limited testing resources, a large number of asymptomatic or moderately symptomatic infections, or government control policies. Here, we used reported epidemic data from Guam, Texas, the Virgin Islands and Wyoming in the United States from April 12, 2020 to February 28, 2021 according to the COVID-19 Map from the John Hopkins Coronavirus Resource Center [24]. The retrieved data include cumulative numbers of confirmed cases, recovered cases, and dead cases, which are shown in Figure 2.

### *3.2. Parameter estimation*

To estimate model parameters, we referred to the reported data, and randomly sample the parameter values for each parameter over the ranges listed in Table 1, and choose a parameter set that minimizes the mean square error between simulation results and the time series of reported cumulative numbers of confirmed, recovered, and dead cases. In parameter estimations, we first compared the data for

**Table 1.** Model parameters.

| Parameter | Description | Unit | Range | Resource |
|---|---|---|---|---|
| $\beta$ | Infection rate | $day^{-1}$ | [0.03,1.5] | [26–31] |
| $\gamma_1$ | Transition rate from latent infections to asymptomatic infections | $day^{-1}$ | $[0.2, 0.3]$ | [28–33] |
| $k$ | Relative infection rate of asymptomatic infections | - | $[0.5, 2]$ | Estimated[(a)] |
| $\gamma_2$ | Transition rate from asymptomatic to symptomatic infections | $day^{-1}$ | $[0.1, 0.5]$ | [31–33]] |
| $\alpha_E$ | Recovery rate of latent infections | $day^{-1}$ | $[0, 0.15]$ | Estimated[(b)] |
| $\alpha_1$ | Recovery rate of asymptomatic infections | $day^{-1}$ | $[0, 0.15]$ | [28–32, 34] |
| $\alpha_2$ | Recovery rate of symptomatic infections | $day^{-1}$ | $[0, 0.283]$ | [28–32, 34] |
| $\alpha_H$ | Recovery rate of confirmed infections | $day^{-1}$ | $[0.008, 0.25]$ | [26, 27, 32] |
| $p_1$ | Detection probability of asymptomatic infections | - | $[0, 1]$ | Estimated |
| $p_2$ | Detection probability of symptomatic infections | - | $[0, 1]$ | Estimated |
| $\mu_1$ | Death rate of undiagnosed symptomatic cases | $day^{-1}$ | $[0.0043, 0.035]$ | [29] |
| $\mu_2$ | Death rate of confirmed cases | $day^{-1}$ | $[0.002, 0.04]$ | [26, 29, 31, 32, 35] |
| $\mu_3$ | The rate of confirmed death from symptomatic infections cases | $day^{-1}$ | $[0.0017, 0.04]$ | [29, 31, 32] |

[(a)] There are no reference data for $k$ from the literature. Clinical evidence shows that the infection rate of COVID-19 reaches a maximum value $1 - 2$ days before symptoms arise.

[(b)] There are no reference data for the recovery rate $\alpha_E$. Here, we use the range for the rate $\alpha_1$.

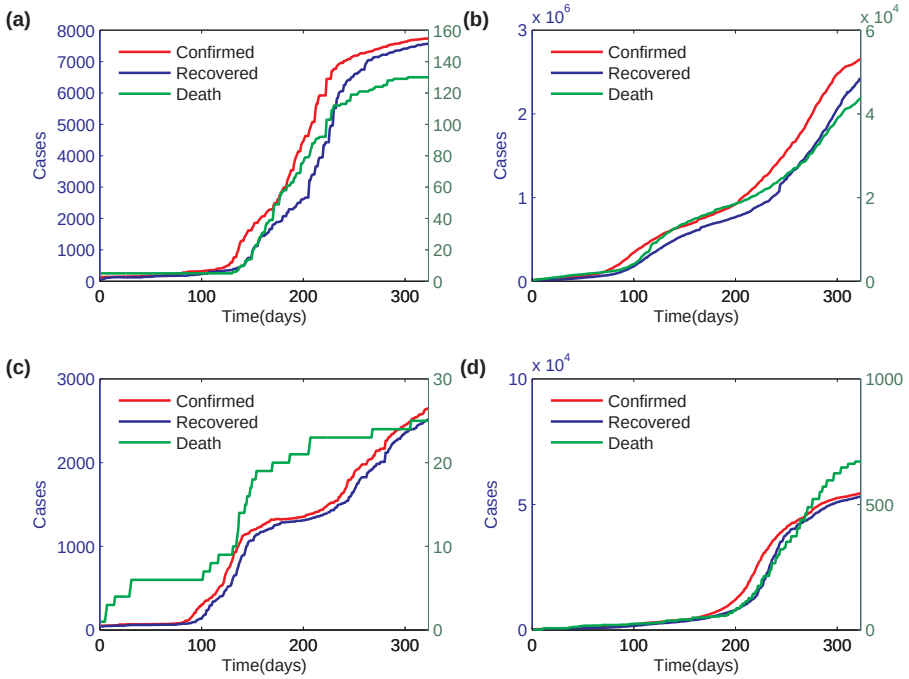


**Figure 2.** Reported COVID-19 epidemic data from **(a)** Guam, **(b)** Texas, **(c)** the Virgin Islands and **(d)** Wyoming in the United States from April 12, 2020 to February 28, 2021. Here, day 0 corresponds to April 12, 2020.

confirmed cases at different stages to obtain the estimated values of most epidemic parameters in the model; then, we compared the data for recovered cases and dead cases based on known results to obtain estimations of the other parameters. In sampling the parameters, we assumed that latent infections have a higher self-recovery rate ($\alpha_E$) than asymptomatic infections ($\alpha_1$) due to innate immune responses at the early stage after infection. Moreover, the detection probability of symptomatic infection ($p_2$) is usually higher than that of asymptomatic infection ($p_1$).

For comparisons with the reported data, we also need to estimate the initial values. The initial values of variables $H$, $R$, and $D_2$ were obtained from the reported data (on April 12, 2020). The initial value of susceptible persons ($S$) was retrieved from open sources [36]. The initial values of $E, I_1, I_2, D_1$ were estimated by minimizing the mean square error. Estimated initial values are shown in Table 2, and parameter values are shown in Table 3 and Figures 3 and 4. Here, we note that the infection rate $\beta$, detection probabilities $p_1$ and $p_2$, and death rate $\mu_2$ and $\mu_3$ are piecewise functions, since they may change with distancing policies and clinical conditions. Based on the parameter values in Table 3, the estimated initial values in Table 2 and equation (1), we can obtain the simulated value of Figure 2. Comparisons between simulations and epidemic data are shown in Figure 5.

According to the estimated parameters in Table 3, the infection rate $\beta$ obviously varies at different stages, which may reflect the distancing policies of the local government and people's attitudes towards the disease. The death rates $\mu_2$ and $\mu_3$ are lower in the later stage than in the early stage in clinical strategies in the later stage.

The estimated detection probabilities of the 4 states are shown in Figure 4 as piecewise functions, and suggest possible changes in NAT. According to our estimation, the detection probability for symp-

**Table 2.** Definitions and initial values of model variables. The initial values of model variables $H$, $R$ and $D_2$ for Guam, Texas, the Virgin Islands and Wyoming were obtained from real data, and other values were estimated based on model simulation and minimizing the mean square error between the simulation and reported data.

| Variables | Guam | Texas | Virgin Islands | Wyoming | Resource |
|---|---|---|---|---|---|
| $S$ | 149913 | 28984768 | 59976 | 578800 | [36] |
| $E$ | 6 | 5275 | 4 | 10 | Estimated |
| $I_1$ | 8 | 3860 | 6 | 15 | Estimated |
| $I_2$ | 10 | 755 | 8 | 17 | Estimated |
| $H$ | 0 | 450 | 0 | 3 | [24] |
| $R$ | 7 | 120 | 5 | 4 | [24] |
| $D_1$ | 2 | 35 | 0 | 0 | Estimated |
| $D_2$ | 5 | 283 | 1 | 0 | [24] |

**Table 3.** Parameter values for Guam, Texas, the Virgin Islands and Wyoming.

| Parameters | Guam | Texas | Virgin Islands | Wyoming |
|---|---|---|---|---|
| $\beta^{(a)}$ | $\beta_1(t)$ | $\beta_2(t)$ | $\beta_3(t)$ | $\beta_4(t)$ |
| $\gamma_1$ | 0.121 | 0.131 | 0.101 | 0.101 |
| $k$ | 1.18 | 1.18 | 1.18 | 1.18 |
| $\gamma_2$ | 0.06 | 0.082 | 0.082 | 0.082 |
| $\alpha_E$ | 0.04 | 0.04 | 0.04 | 0.04 |
| $\alpha_1$ | 0.02 | 0.02 | 0.02 | 0.02 |
| $\alpha_2$ | 0.02 | 0.02 | 0.02 | 0.005 |
| $\alpha_H$ | 0.05 | 0.074 | 0.074 | 0.074 |
| $p_1^{(b)}$ | $p_{11}(t)$ | $p_{12}(t)$ | $p_{13}(t)$ | $p_{14}(t)$ |
| $p_2^{(b)}$ | $p_{21}(t)$ | $p_{22}(t)$ | $p_{23}(t)$ | $p_{24}(t)$ |
| $\mu_1$ | 0.006 | 0.012 | 0.011 | 0.012 |
| $\mu_2^{(a)}$ | $\mu_{21}(t)$ | $\mu_{22}(t)$ | $\mu_{23}(t)$ | $\mu_{24}(t)$ |
| $\mu_3^{(a)}$ | $\mu_{31}(t)$ | $\mu_{32}(t)$ | $\mu_{33}(t)$ | $\mu_{34}(t)$ |

[(a)] These parameters are defined by the piecewise functions given in Figure 3.
[(b)] These parameters are defined by the piecewise functions given in Figure 4.

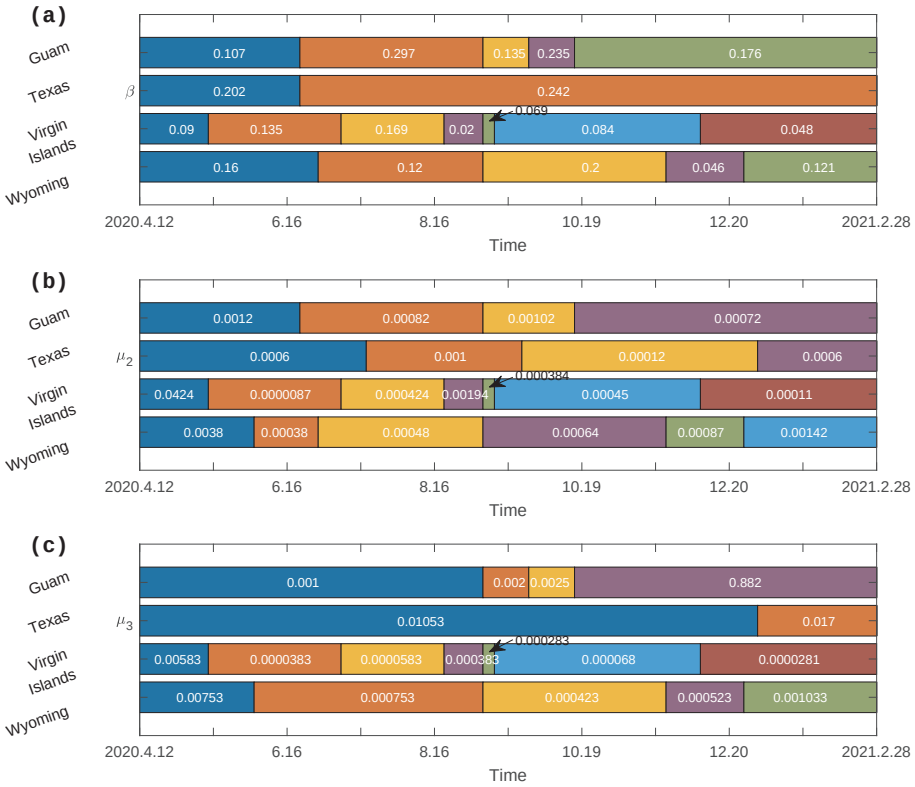


**Figure 3.** Estimated piecewise functions of the **(a)** infection rate $\beta$, death rates of **(b)** unconfirmed infections $\mu_2$ and **(c)** confirmed patients $\mu_3$ in the four states: Guam, Texas, the Virgin Islands, and Wyoming.

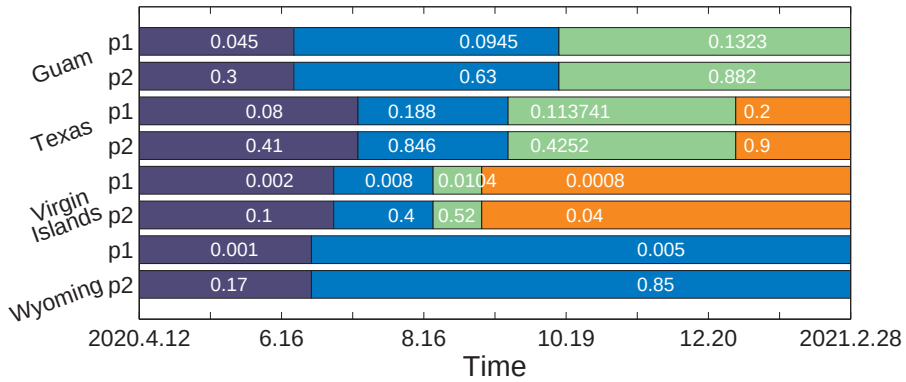


**Figure 4.** Estimated piecewise functions of the detection probabilities $p_1$ and $p_2$ in the four states: Guam, Texas, the Virgin Islands, and Wyoming.

tomatic infections ($p_2$) is higher in the later stage than in the early stage in all states, with a maximum detection probability larger than 0.8 in Guam, Texas, and Wyoming, and larger than 0.5 in the Virgin Islands. The detection probability for asymptomatic infections ($p_1$) also increases in Guam and Texas, but is much lower than that for symptomatic infections. Moreover, the detection probability for asymptomatic infections in Virgin Islands and Wyoming are extremely low.

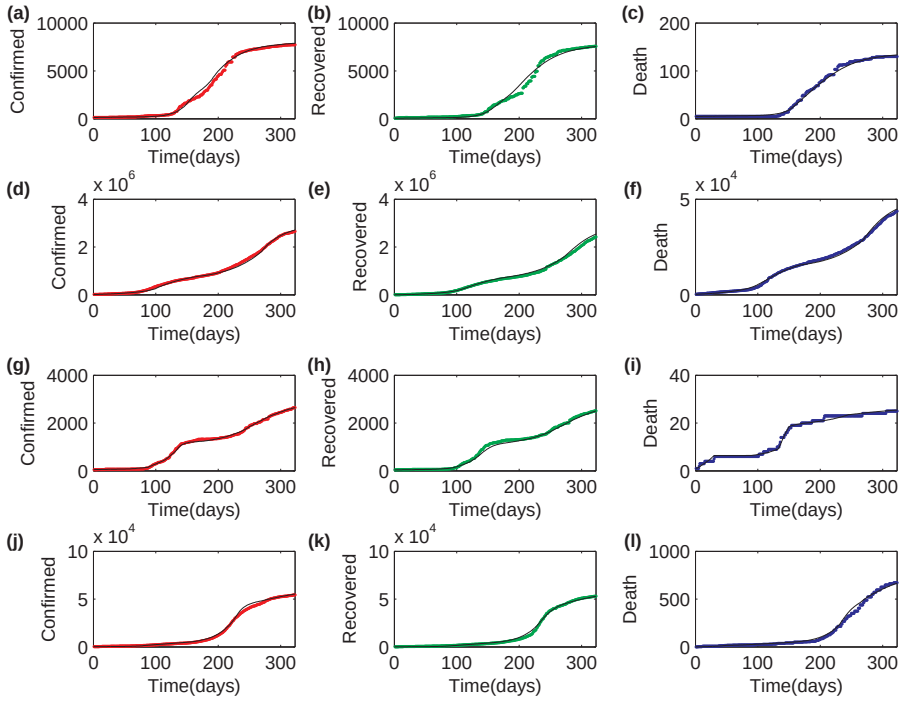


**Figure 5.** Model simulation and epidemic data. Data for **(a-c)** Guam, **(d-f)** Texas, **(g-i)** the Virgin Islands, and **(j-l)** Wyoming are shown from top to bottom. For each state, the data of cumulative confirmed cases, recovered cases, and dead cases are shown (from left to right). Epidemic data from April 12, 2020 to February 28, 2021 are shown with dots, and simulation results are shown by black solid lines.

### *3.3. Increasing the detection probability for asymptomatic infections can reduce the epidemic scale*

The above simulation shows that the proposed model is capable of reproducing epidemic dynamics. To further quantify the influence of the detection probability $p_1$ on the COVID-19 epidemic, we took the parameters for Guam as an example in the following study. Based on our parameter estimation, the probability $p_1$ in Guam increased from 0.045 at the early stage to 0.13 in the later stage. Here, we took $p_1 = 0.13$ as the default value. First, we varied the detection probability $p_1$ ($p_1 = 0.07, 0.1, 0.13, 0.16, 0.19$) for asymptomatic infections at constant values of the other parameters. Here, we set $p_1$ as a constant in the model simulations.

We note that $H(t)$ in the model equation represents the number of hospitalized patients, which varies with time due to newly confirmed cases, recovered patients and dead patients. Here, to quantify the

epidemic dynamics, we are interested in the daily confirmed cases defined as

$$H_{\text{new}}(t) = \Psi_1(I_1(t), p_1) + \Psi_2(I_2(t), p_2). \tag{3.1}$$

Moreover, we also examined the peak value of daily confirmed cases

$$H_{\text{new}}^{\max} = \max_{t\geq 0}\{H_{\text{new}}(t)\}, \tag{3.2}$$

and the cumulative new confirmed cases

$$H_{\text{C}}(t) = \int_0^t H_{\text{new}}(s)ds. \tag{3.3}$$

Similarly, we also consider the daily new infected cases

$$I_{\text{new}}(t) = \gamma_1 E(t), \tag{3.4}$$

the peak value

$$I_{\text{new}}^{\max} = \max_{t\geq 0}\{I_{\text{new}}(t)\}, \tag{3.5}$$

and the corresponding cumulative new infected cases

$$I_{\text{C}}(t) = \int_0^t I_{\text{new}}(s)ds. \tag{3.6}$$

Similar to the classical SIR or SEIR models, the daily confirmed case number increases to reach a peak value and then decreases to 0 as time $t$ approaches infinity. The cumulative confirmed case number saturates at a final value when $t$ is large enough. The daily new confirmed case number obviously increases if the detection probability $p_1$ is decreased (Figure 6a). If $p_1$ is reduced by half ($p_1 = 0.07$), the peak value of daily confirmed cases can be as high as 5000, and the number decreases to 98 if the detection probability increases to $p_1 = 0.16$, approximately 1% of that with $p_1 = 0.07$. We further examined the peak value of both daily new confirmed cases and new infections cases; both numbers exponentially decrease with the detection probability $p_1$, while the daily infection number is more sensitively dependent on changes in the detection probability (Figure 6b).

We further examined the dependence of cumulative confirmed cases on the detection probability $p_1$ (Figure 6c). The cumulative confirmed case number obviously increases with the reduction of $p_1$. Model simulations predict a final epidemic scale of $4 \times 10^4$ cases when $p_1 = 0.07$, and the number decreases to 2.7% (1160 cases) when $p_1$ increases to 0.16. These results suggest that increasing the detection probability $p_1$ can effectively reduce the epidemic scale.

To quantify the epidemic dynamics with various detection probabilities, we defined a relative increase index of daily new confirmed cases ($\Delta^{\ddagger}H$) as the ratio of changes in daily new confirmed cases to daily new confirmed cases, which is formulated as

$$\Delta^{\ddagger}H = \frac{\Delta\Delta H_{\text{new}}}{\Delta H_{\text{new}}}, \tag{3.7}$$

where $\Delta$ is a difference operator defined as $\Delta f(t) = f(t) - f(t-1)$ for any function $f(t)$. Based on our model simulations, the time evolution of $\Delta^{\ddagger}H$ is shown in Figure 6d. Despite the obvious dependence of epidemic dynamics on the detection probability $p_1$, the relative increase index $\Delta^{\ddagger}H$ is insensitive to $p_1$; however the underlying mechanism is not yet known. Hence, the proposed relative increase index may be used to quantify epidemic dynamics that are independent of the detection probability for asymptomatic infections.

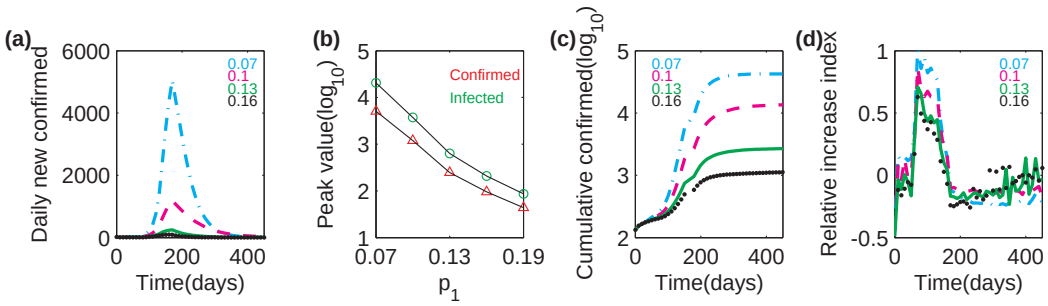


**Figure 6.** Influence of the detection probability $p_1$ on the COVID-19 epidemic. **(a)** Daily new confirmed cases ($H_{\text{new}}(t)$) under various probabilities $p_1$. **(b)** Dependence of the peak values (log scale) of the daily new confirmed cases ($H_{\text{new}}^{\max}$) and the infected cases ($I_{\text{new}}^{\max}$) on the probability $p_1$. **(c)** Cumulative confirmed cases ($H_{\text{C}}(t)$) under various probabilities $p_1$. **(d)** Time course of the relative increase index ($\Delta^{\ddagger}H$) for various probabilities $p_1$. Here, $p_1$ takes values $0.07, 0.1, 0.13, 0.16, 0.19$, respectively. In each case, the results were obtained from 50 independent runs.

### *3.4. Increasing the detection probability for symptomatic infections reduces the epidemic scale*

We further examined the influence of changing the detection probability for symptomatic infections ($p_2$). Based on our parameter estimation, the detection probability $p_2$ in Guam increased from 0.3 at the early stage to 0.882 in the later stage. Here, we took $p_2 = 0.7$ as the default value. We varied the detection probability $p_2$ ($p_2 = 0.5, 0.6, 0.7, 0.8, 0.9$) and fixed other parameters as their default values. Simulation results are shown in Figure 7. Similar to the results with changes in $p_1$, the daily new confirmed case number increases with $p_2$ (Figure 7a), and the peak value of daily new confirmed cases and the peak value of new infections both decrease with increasing detection probability (Figure 7b). Increasing the probability $p_2$ can decrease the number of cumulative confirmed cases (Figure 7c). We further examined the relative increase index $\Delta^{\ddagger}H$ and found that the index is independent of the detection probability $p_2$ (Figure 7d).

### *3.5. Prediction of epidemic scales with varying detection probabilities*

To examine the impact of the detection probabilities $p_1$ and $p_2$ on epidemic size in the four states of Guam, Texas, the Virgin Islands, and Wyoming in the US, we performed model simulations with varying detection probabilities $p_1 \in (0, 0.2)$ and $p_2 \in (0, 1)$ and fixed other parameters unchanged as their estimated values shown in Table 3 and Figure 3. The simulation results for the cumulative new infected cases $I_{\text{C}}(t)$ in Guam, Texas, the Virgin Islands and Wyoming are shown in Figure 8. The model simulation shows that cumulative infected cases obviously decrease if either the detection probability $p_1$ or $p_2$ is increased. Specifically, in Guam and Texas, the epidemic size obviously decreases when $p_1$ varies from 0 to 0.1. For the Virgin Islands, the epidemic size obviously decreases when $p_2 = 0$ and $p_1$ increases from 0 to 0.2, and a slight increase in $p_2$ from 0 may significantly reduce the epidemic size. Similar results are obtained for Wyoming; slight increases in the detection probabilities $p_1$ or $p_2$ from 0 would obviously reduce the epidemic size. These results suggest the importance of performing NAT detection and isolating of confirmed cases in controlling the COVID-19 epidemic.

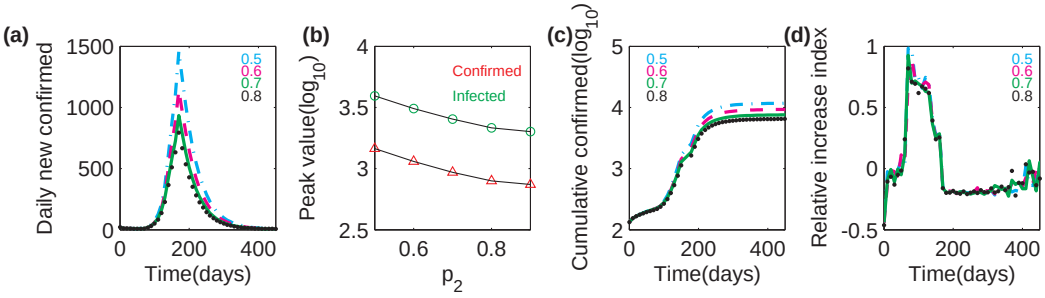


**Figure 7.** Influence of the detection probability $p_2$ on the COVID-19 epidemic. **(a)** Daily new confirmed cases ($H_{\text{new}}(t)$) under various probabilities $p_2$. **(b)** Dependence of the peak values (log scale) of the daily new confirmed cases ($H_{\text{new}}^{\max}$) and the infected cases ($I_{\text{new}}^{\max}$) on the probability $p_2$. **(c)** Cumulative confirmed cases ($H_{\text{C}}(t)$) under various probabilities $p_2$. **(d)** Time course of the relative increase index ($\Delta^{\ddagger}H$) for various probabilities $p_2$. Here, $p_2$ take values $0.5, 0.6, 0.7, 0.8, 0.9$, respectively. In each case, the results were obtained from 50 independent runs.

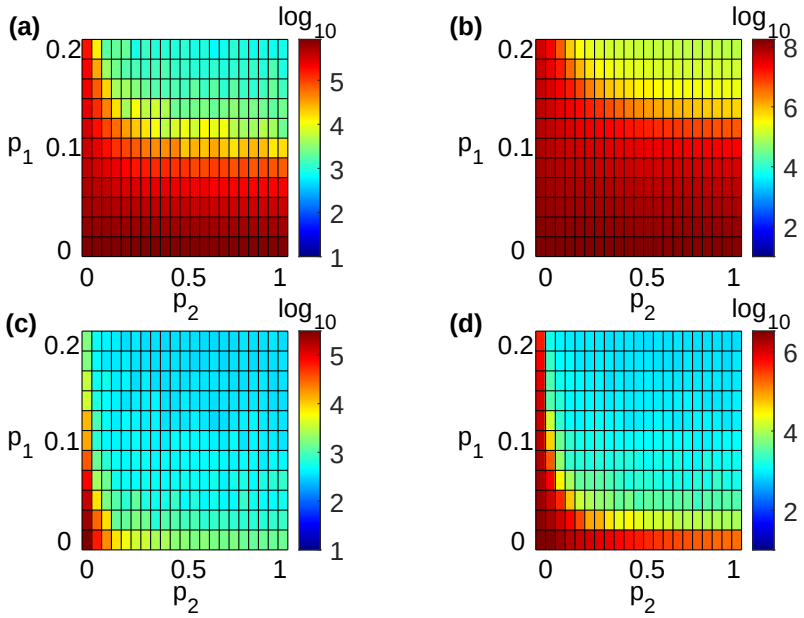


**Figure 8.** Impact of changing the detection probability $p_1$ and $p_2$ on the cumulative infected cases in **(a)** Guam, **(b)** Texas, **(c)** the Virgin Islands and **(d)** Wyoming. Color bars show the number ($\log_{10}$ scale) of cumulative infected cases $I_{\text{C}}(t)$ at the end of the model simulation. Here, the detection probability $p_1$ ranges from 0 to 0.2, and the detection probability $p_2$ ranges from 0 to 1.

## 4. Conclusions

Many counties suffer from insufficient detection of COVID-19 infections, which may result in underestimation of the epidemic size and, in turn, hamper appropriate epidemic control measures. Our study proposed a mathematical model to investigate how insufficient detection may affect the dynamics of the spread of COVID-19. The model explicitly considers various detection probabilities for asymptomatic and symptomatic infections. We took the reported data from four states in the US as an example in our study and tuned the model parameters. We found that detection probabilities may vary over time with different strategies of control measures.

Model simulations show that both infected and confirmed cases are sensitively dependent on the detection probability. Insufficient detection for either asymptomatic or symptomatic infections may worsen the situation of the COVID-19 epidemic, including increasing in the number of daily new confirmed cases, the peak value of daily new infections, and the cumulative number of confirmed cases.

We further investigated the influence of varying the detection probabilities for both asymptomatic and symptomatic infections on the epidemic scale of our model. Simulations show that increasing the detection probability can significantly reduce the epidemic size. The detection probability for asymptomatic infections is very important for reducing the size of the epidemic. Therefore, early detection and isolation of COVID-19 infections is important for the control of the epidemic. Nevertheless, asymptomatic infections often generate false-positive and false-negative results for asymptomatic infections, and there is a tradeoff between test sensitivity and test frequency when there are limitations in the testing budget. In this case, a multiscale modeling study has recommended that low-sensitivity tests be employed at high frequency [37].

## Acknowledgments

This work was supported by the National Natural Science Foundation of China under grant No.11831015.

## Conflict of interest

The authors declare there is no conflict of interest.

## References

1. World Health Organization, WHO Coronavirus Disease (COVID-19) Dashboard, available from: https://covid19.who.int/table.
2. S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, et al., Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe, *Nature*, **584** (2020), 257–261. doi:10.1038/s41586-020-2405-7.
3. N. Ferguson, D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, et al., Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand, 2020. doi:10.25561/77482.

4. S. Lai, N. W. Ruktanonchai, L. Zhou, O. Prosper, W. Luo, J. R. Floyd, et al., Effect of non-pharmaceutical interventions to contain COVID-19 in China, *Nature*, **585** (2020), 410–413. doi:10.1038/s41586-020-2293-x.
5. Y. Huang, L. Yang, H. Dai, F. Tian, K. Chen, Epidemic situation and forecasting of COVID-19 in and outside China, *Bull World Health Organization*, **4** (2020). doi:10.2471/BLT.20.255158.
6. J. Labadin, B. H. Hong, Transmission dynamics of 2019-nCOV in Malaysia proior to the movement Control Order, 2020. medRxiv 2020.02.07.20021188; doi: https://doi.org/10.1101/2020.02.07.20021188
7. B. Tang, X. Wang, Q. Li, N. L. Bragazzi, S. Y. Tang, Y. N. Xiao, et al., Estimation of the transmission risk of 2019-nCov and its implication for public health interventions, *J. Clin. Med.*, **9** (2020), 462.
8. Y. Chen, J. Cheng, Y. Jiang, K. Liu, A time delay dynamical model for outbreak of 2019-nCOV and the parameter identification, *J. Inv. Ill-Posed Problem*, **28** (2020), 243–250.
9. Y. Yan, Y. Chen, K. Liu, X. Luo, B. Xu, Y. Jiang, et al., Modeling and prediction for the trend of outbreak of NCP based on a time-delay dynamic system, *Sci. Sinica Math.*, **50** (2020), 385. doi:10.1360/SSM-2020-0026.
10. X. M. Rong, L. Yang, H. D. Chu, M. Fan, Effect of delay in diagnosis on transmission of COVID-19, *Math. Biosci. Eng.*, **17** (2020), 2725–2740.
11. V. Naveen, C. Aasish, M. Kavya, M. Vidhyalakshmi, Forecasting the number of infections of novel coronavirus with deep learning, *Int. J. Comput. Appl.*, **176** (2020), 21–24.
12. H. Abbasimehr, R. Paki, Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization, *Chaos Soliton Fract.*, **142** (2021), 110511.
13. L. Qin, Q. Sun, Y. D. Wang, K. F. Wu, M. Chen, B. C. Shia, et al., Prediction of number of cases of 2019 novel coronavirus (COVID-19) using social media search index, *Int. J. Environ. Res. Pub. Health*, **17** (2020), 2365.
14. A. R. Akhmetzhanov, K. Mizumoto, S. M. Jung, N. M. Linton, R. Omori, H. Nishiura, et al., Estimation of the actual incidence of coronavirus disease (COVID-19) in emergent hotspots: The example of Hokkaido Japan during February-March 2020, *Clin. Med.*, **10** (2021), 2392.
15. G. Pullano, L. D. Domenico, C. E. Sabbatini, E. Valdano, C. Turbelin, M. Debin, et al, Under detection of COVID-19 cases in France threatens epidemic control, *Nature*, **590** (2020), 134–139.
16. C. Xu, Y. Z. Pei, S. Q. Liu, J. Z. Lei, Effectiveness of non-pharmaceutical intervention against local transmission of COVID-19: An individual-based modelling study, *Infect. Dis. Model.*, **6** (2021), 848–858.
17. T. Alamo, D. G. Reina, P. M. Gata, V. M. Preciado, G. Giordano, Data-driven methods for present and future pandemics: monitoring, modeling and managing, Annu Rev Control, (2021), arXiv preprint arXiv:2102.13130. doi:10.1016/j.arcontrol.2021.05.003.
18. S. Venkatramanan ,B. Lewis, J. Chen, D. Higdon, A. Vullikanti, M. Marathe, Using data-driven agent-based models for forecasting emerging infectious diseases, *Epidemics*, **22** (2018), 43–49.

19. D. Bertsimas, L. Boussioux, R. Cory-Wright, A. Delarue, V. Digalakis, A. Jacquillat, et al., From predictions to prescriptions: A data-driven response to COVID-19, *Health Care Manag. Sci.*, **2** (2021), 1–20.

20. S. Y. Tang, B. Tang, N. L. Bragazzi, et al, Data mining of covid-19 epidemic and analysis of discrete random propagation dynamic model, *Sci. China*, **50** (2020), 1–16 (in Chinese).

21. L. Böttcher, M. R. D'Orsogna, T. Chou, Using excess deaths and testing statistics to determine COVID-19 mortalities, *Eur. J. Epidemiol.* **36** (2021), 545–558.

22. J. S. Faust, Z. Lin, C. Del Rio, Comparison of estimated excess deaths in New York City during the COVID-19 and 1918 influenza pandemics, *JAMA Netw. Open*, **3** (2020), e2017527.

23. A. L. Bertozzi, E. Franco, G. Mohler, M. B. Short, D. Sledge, The challenges of modeling and forecasting the spread of COVID-19, *Proc. Natl. Acad. Sci. USA*, **117** (2020), 16732–16738.

24. COVID-19 Map: Johns Hopkins Coronavirus Resource Center, https://coronavirus.jhu.edu/map.html.

25. C. Kuhbandner, S. Homburg, Commentary: Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe, *Front Med.*, **7** (2020), 580361.

26. C. Cakmakli, Y. Simsek, Bridging the COVID-19 data and the epidemiological model using time varying parameter SIRD model, (2020), arXiv preprint arXiv:2007.02726. doi:abs/2007.02726.

27. J. Rocklv, H. Sjdin, A. Wilder-Smith, COVID-19 outbreak on the Diamond Princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures, *J. Travel Med.*, **27** (2020), taaa030. doi:10.1093/jtm/taaa030.

28. S.Bentout, A. Chekroun, T. Kuniya, Parameter estimation and prediction for coronavirus disease outbreak 2019 (COVID-19) in Algeria, *AIMS Public Health*, **7** (2020), 306–318.

29. K. Chatterjee, K. Chatterjee, A. Kumar, S. Shankar, Healthcare impact of COVID-19 epidemic in India: A stochastic mathematical model, *Med. J. Armed. Forces India*, **76** (2020), 147–155.

30. T. Kuniya, Prediction of the epidemic peak of coronavirus disease in Japan, 2020, *Clin. Med.*, **9** (2020), 789.

31. Q. Li, Y. N. Xiao, J. H. Wu, COVID-19 epidemic time lag model construction and confirmed case-driven tracking and isolation measures analysis, *Acta Math. Appl. Sinica*, **43** (2020), 96–108(in Chinese).

32. M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Cararo, R. Casagrandi, et al., Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures, *Proc. Natl. Acad. Sci. USA*, **117** (2020), 202004978.

33. S. Y. Tang, B. Tang, N. L. Bragazzi, F. Xia, T. Li, S. He, et al., Analysis of COVID-19 epidemic traced data and stochastic discrete transmission dynamic model, *Sci. Sinica Math.*, **50** (2020), 1–16(in Chinese).

34. K. Prem, Y. liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, et al., 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*, **5** (2020), e261–e270.

35. B. Tang, F. Xia, S. Y. Tang, The effectiveness of quarantine and isolation determine the trend of the COVID-19 epidemic in the final phase of the current outbreak in China, *Int. J. Infect. Dis.*, **95** (2020), 288–293.
36. Population Ranking of American State Governments (2020), available from: http://blog.sina.com.cn/s/blog_5ce1af980102z91y.html.
37. J. E. Forde, S. M. Ciupe, Quantification of the tradeoff between test sensitivity and test frequency in a COVID-19 epidemic–A multi-scale modeling approach, *Viruses*, **13** (2021), 457.
38. M. T. Xia, L. Bőttcher, T. Chou, Controlling epidemics through optimal allocation of test kits and vaccine doses across networks, (2021), arXiv preprint arXiv:2107.13709. doi:abs/2107.13709.


AIMS Press