

# Statistical Inference of COVID-19 Outbreak: Delay Distribution Effect in EQIR Modeling of Epidemic

## Abstract

**Background:** The world is experiencing another pandemic called COVID-19. Several mathematical models have been proposed to examine the impact of health interventions in controlling pandemic growth. **Method:** In this study, we propose a fractional order distributed delay dynamic system, namely, EQIR model. In order to predict the outbreak, the proposed model incorporates changes in transmission rate, isolation rate, and identification of infected people through time varying deterministic and stochastic parameters. Furthermore, proposed stochastic model considers fluctuations in population behavior and simulates different scenarios of outbreak at the same time. Main novelty of this model is its ability to incorporate changes in transmission rate, latent periods, and rate of quarantine through time varying deterministic and stochastic assumptions. This model can exactly follow the disease trend from its beginning to current situation and predict outbreak future for various situations. **Results:** Parameters of this model were identified during fitting process to real data of Iran, USA, and South Korea. We calculated the reproduction number using a Laplace transform-based method. Results of numerical simulation verify the effectiveness and accuracy of proposed deterministic and stochastic models in current outbreak. **Conclusion:** Justifying of parameters of the model emphasizes that, although stricter deterrent interventions can prevent another peak and control the current outbreak, the consecutive screening schemes of COVID-19 plays more important role. This means that the more diagnostic tests performed on people, the faster the disease will be controlled.

**Keywords:** COVID-19, EQIR epidemic model, fractional differential equation, stochastic differential equation

Submitted: 18-Jun-2021

Revised: 02-Aug-2021

Accepted: 28-Oct-2021

Published: 12-May-2022

## Introduction

In December 2019, the epidemic was created by a member of coronaviruses, namely, 2019-nCoV. Till now, the virus has been spreading worldwide, and millions of people have been infected. Extensive mathematical studies in various fields have been done to find the policies to control the disease in its prevalence. Mathematical models can indicate how infectious diseases spread and what the trend of epidemic is. They can be divided into two general purpose models, forecasting models and mechanistic models. Forecasting models try to fit curves to data using different methods from extrapolations using simple models such as exponential growth<sup>[1]</sup> to machine learning methods<sup>[2]</sup> or regressions.<sup>[3]</sup> These models are simple and are only used in the early days of the outbreak. Mechanistic models can be divided into two groups as follows.

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow\_reprints@wolterskluwer.com

Self-exciting branching point process<sup>[4]</sup>: These models can easily fit to data and are used to estimate reproduction number and estimation rate. However, they can be used in initial phase of epidemic when the number of infected individuals are small compared to the population size.<sup>[5]</sup>

Dynamic epidemiological models<sup>[6]</sup>: In these models, the population is divided into several states or compartments, and different parameters are attributed to each state such as transmission rate and mortality rate. Then, considering specific assumptions for the population, differential equations for each state of the system are obtained and solved. In fact, these models are one of the best ways to explore long-term epidemiologic outcomes and to implement various assumptions about disease and controlling measures.<sup>[7,8]</sup> They have been widely used for the COVID-19

**How to cite this article:** Tajmirriahi M, Amini Z, Kafieh R, Rabbani H, Mirzazadeh A, Javanmard SH. Statistical inference of COVID-19 outbreak: Delay distribution effect in EQIR modeling of epidemic. *J Med Signals Sens* 2022;12:95-107.

**Mahnoosh Tajmirriahi<sup>1</sup>, Zahra Amini<sup>1</sup>, Rahele Kafieh<sup>1</sup>, Hossein Rabbani<sup>1</sup>, Ali Mirzazadeh<sup>2</sup>, Shaghayegh Haghjooy Javanmard<sup>3</sup>**

<sup>1</sup>Medical Image and Signal Processing Research Center, School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, <sup>3</sup>Applied Physiology Research Center, Isfahan Cardiovascular Research Institute, Isfahan University of Medical Sciences, Isfahan, Iran, <sup>2</sup>Department of Epidemiology and Biostatistics, Institute for Global Health Sciences, University of California, San Francisco, California, United States

## Address for correspondence:

Dr. Zahra Amini, Medical Image and Signal Processing Research Center, School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, Isfahan, Iran. E-mail: zahraamini64@yahoo.com.au  
Dr. Rahele Kafieh, Medical Image and Signal Processing Research Center, School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, Isfahan, Iran. E-mail: rkafieh@gmail.com

## Access this article online

Website: [www.jmssjournal.net](http://www.jmssjournal.net)

DOI: 10.4103/jmss.jmss\_134\_21

## Quick Response Code:



epidemic because of its simplicity, parametric approach, and good results. Some of these models have contributed improvements in susceptible-infected-recovered model with a variable total population<sup>[9]</sup> and time varying parameters.<sup>[10-12]</sup> Some others tried to use susceptible-exposed-infected-recovered (SEIR) models<sup>[13,14]</sup> including control strategies in the SEIR model,<sup>[15,16]</sup> and utilizing fractional order of differential equations in SEIR.<sup>[17,18]</sup> Meanwhile, some studies have been done to find new dynamic models with more states and relations to achieve more accurate results such as.<sup>[19-22]</sup>

However, time frame, randomness, basic assumptions of model, model generality, and amount of data matching are some important factors in evaluating a model, while a simple model may not satisfy all these conditions. Assumptions such as social knowledge, fluctuation in the behavior of populations, and randomness of transmission rate must be inserted in model through fractional stochastic differential equations (SDE).<sup>[23,24]</sup> SDE has been widely used in modeling of long range dependent events to indicate the memory of the systems.<sup>[25,26]</sup>

In this study inspired by Chen *et al.*,<sup>[27]</sup> we use a dynamic stochastic and deterministic model of fractional order to describe the regional outbreaks of COVID-19 and to predict its future trend. In the proposed model, in addition to considering specific states for different parts of the population, latent periods, isolation, and treatment durations are also considered by adding distributed delays to the model. Furthermore, we include memory effects in the model dynamics by means of fractional derivatives. The stochastic nature of population behavior is also considered in this model by assuming the transmission rate as a realization of Brownian motion stochastic process. All these assumptions lead to solve a fractional ordered distributed delay SDE.<sup>[28]</sup> We also obtain the basic reproduction number of the epidemic model. After finding parameters of deterministic model from available data through fitting process, results show a good performance of the model in fitting and prediction the trends of epidemic, besides its ability to justify the real behavior of different communities against disease. Stochastic model is used to explore short-term and long-term trajectories of COVID-19 and determines the range of outbreak size.

This paper is organized as follows. After this introduction, we explain deterministic and stochastic model and their parameters in Section 2. In this section, we also discuss about existence and persistence of the model by calculating reproduction number. The numerical results of the model are presented in Section 3. Discussion about the model and conclusions are presented in Sections 4 and 5.

## Materials and Methods

### Model description

The important point in spreading trend of COVID-19 is transmission from an exposed patient with no symptoms to

others which is challenging in SEIR models. We propose an EQIR model,<sup>[27]</sup> in which the population is divided into four states. The first state is exposed (E) individuals with no symptoms, due to government decisions some of the exposed individuals may be isolated in their latent period. Hence, the second state (Q) contains the quarantined individuals. Third state (I) includes individuals who are confirmed and have no ability of spreading disease. The last state contains removed (R) people (recovered or dead). Since environmental fluctuations can impact transmission rate  $\beta$ , it can be assumed a time-varying function such as exponentially decreasing function.<sup>[29,30]</sup> This decreasing trend may be caused because of government interventions such as isolations and public closures at specific time. Here, we assumed following function for transmission rate.

$$\beta(t) = \begin{cases} \beta_0 & 0 \leq t \leq T \\ \beta_0 \exp(-d(t-T)) & t > T \end{cases} \quad (1)$$

where  $T$  is the time of starting interventions and  $d$  is decaying coefficient. Another approach to display the time varying nature of transmission rate is considering a random generating process for it by adding a white Gaussian noise to transmission rate  $\beta$ . Hence, we expect extra random curves around the main curve indicating how far the outbreak will be from controllability in short terms. Further assumptions and parameters are adopted for both stochastic and deterministic models. Some important assumptions such as considering extremely large population of susceptible, independency, and homogeneity of infected individuals, and disease equilibrium condition are implicitly considered.

1. We assume that it takes averagely  $\tau_E$  days for an exposed individual to have symptoms and be identified. After this latent period, the exposed individual comes to infected state and in this state can no longer transmit the coronavirus to others
2. The second assumption is about those exposed individuals who are in the incubation period and are isolated from others. They cannot spread the disease and will be diagnosed during next  $\tau_Q$  days. After  $\tau_Q \leq \tau_E$  days, they will be identified as infected ones. We assume  $q$  as the isolation rate of currently exposed people
3. Infected individuals averagely experience  $\tau_R$  days before complete cure or death. We indicate mortality or recovery rate of this disease by parameter  $\delta$ . We ignore the natural mortality in our population
4. We assume that every exposed, isolated, and treated individual experiences different time period to change his/her states. This can be considered by means of distributed delay functions  $h_E(t)$ ,  $h_Q(t)$ , and  $h_R(t)$  respectively. Where  $h$  is a Gaussian probability distribution function. The mean value of this distribution is adopted to parameters  $\tau_E$ ,  $\tau_Q$ ,  $\tau_R$ . Different variances of each distribution can be determined by fitting the

model to real data. Zero variance value, converts  $h(t)$  to a Dirac delta function and means that all individuals experiences same latent. These parameters can determine how fast the exposed individuals are identified and isolated

5. Ordinary compartmental models implicitly follow a Markov epidemic process, in which the states of individuals are independent from previous steps at each time. This process is so-called memory less. However, when a disease spread within a human population, knowledge of individuals affects their response and it shows an implicit memory. However, the effect of memory of earlier times decreases over time. This decay of long-range memory can be controlled by the order of the fractional derivatives in the corresponding non-linear fractional differential equations of the system.<sup>[31]</sup> We assume such a fractional order model and use the Riemann-Liouville definition of fractional derivative of order  $\alpha$  as follows:<sup>[32]</sup>

$$D_t^\alpha f(t) = \frac{1}{\Gamma(n-\alpha)} \int_0^t \frac{f^n(s)}{(t-s)^{\alpha-n+1}} ds \quad (2)$$

where function is  $\Gamma()$  Gamma function and  $n$  is an integer. Considering above assumptions, the continuous time of deterministic EQIR model is constructed as:

$$\begin{cases} E'(t) = E(t) - Q(t) - I(t) \\ D_t^{\alpha_1} E = \beta E'(t) \\ D_t^{\alpha_2} I = \int_0^t h_E(t-\tau) \beta E'(\tau) d\tau \\ D_t^{\alpha_3} Q = q(E'(t)) - q \int_0^t h_Q(t-\tau) E'(\tau) d\tau \\ D_t^{\alpha_4} R = \delta \int_0^t h_R(t-\tau) \beta E'(\tau) d\tau \end{cases} \quad (3)$$

where  $\alpha_i, i = 1 \dots 4$  can be obtained by parameter identification. In order to simulate this model, we use discrete time implementation of the model in two cases deterministic and stochastic models, both of them are presented in Appendices A and B.

### Optimization problem

Each model has several parameters which must be carefully determined. Some of them must be extracted from epidemiological texts and others must be identified by fitting models to the available data. In second case, we encounter an optimization problem. For this purpose, Let  $\theta$  denotes vector of parameters of specific state which must be estimated, then our parameters identification problem has the following form:

$$\min_{\theta} \| \mathcal{L}(\theta, t) - \mathcal{C}_{obj} \|_2 \quad (4)$$

where  $\mathcal{C}_{obj}$  denotes the available data of that component and  $\| \cdot \|_2$  is related to the 2-norm. In order to solve this

problem, we use Levenberg-Marquadt (LM) method<sup>[37]</sup> for deterministic model. This method uses an iterative algorithm to solve nonlinear least square problems. In stochastic disease modeling,  $x_t$  is real number of infectious individuals which may be controlled by parameters  $\theta$  such as transmission rate,  $y_t$  is reported data, and the goal of modeling is determining the posteriori distribution of parameters that is  $p(\theta | y)$ .<sup>[38]</sup> An ordinary approach to estimate distribution of parameters is pseudomarginal Markov Chain Monte Carlo (PMCMC) method.<sup>[39]</sup> In this method, samples of desired probability density function  $p(\theta | y)$  are generated by using a proposal function which usually considered a normal distribution. A sampler generates some samples (particles) from proposed distribution. Then, the acceptance ratio of each sample is calculated and decision is made to keep or reject that sample. After lots of iterations, remaining particles represent  $p(\theta | y)$ .

### Reproduction number

By means of the same method used in Kiskinov, Zahariev and Özalp, Demirci's study,<sup>[40,41]</sup> one can show that there exists a unique, nonnegative solution for fractional differential equations depicted in Equation 2. For analysis the stability of the model, we adopt a Laplace transform approach to show the prevalence of individual in each state explained in.<sup>[42]</sup> For this purpose, we describe the basic reproduction number, typically denoted with  $R_0$ , of the model. For a deterministic epidemic model, if  $R_0 < 1$  the disease dies out and if  $R_0 > 1$  the disease persists and outbreak will happen, meaning that there exists an epidemic equilibria, in addition to the disease-free equilibria. For a stochastic model, this number indicates existence of unique stationary disease-free distribution or Epidemic distribution in states. There are several methods to estimate  $R_0$ .<sup>[43]</sup> However, in an ordinary SEIR model,  $R_0 = \beta \mu_1$  where  $\mu_1$  is average infectious latent period. In our proposed model, we generalize the SEIR model by first assuming that latent periods follow a Gaussian distribution and second by using fractional order derivatives. It is shown<sup>[41,44]</sup> that the fractional order of model does not affect  $R_0$ , however, distribution of latent periods affects it, and controlled reproduction number is defined as  $R_c = \beta \bar{\mu}_1$ .<sup>[42]</sup> Where  $\bar{\beta}$  indicates reduction in  $\beta$  due to system outflows occurred before exposed state. is defined as follows:

$$\bar{\mu}_1 = \int_0^\infty e^{-\omega x} \bar{F}_1(x) dx = L[\bar{F}_1](\omega) \quad (5)$$

where  $\omega$  denotes the outflow of state,  $\bar{F}_1$  is survival function and  $L[\cdot]$  denotes Laplace transform. For  $f_1 \sim N(\tau_1, \sigma_1)$  we have:

$$L[\bar{F}_1](s) = \frac{1}{s} \left[ 1 - \exp\left(\frac{1}{2} s^2 \sigma_1^2 - s \tau_1\right) \right] \quad (6)$$

Figure 1 shows the block diagram of the proposed model. It is worthy to mention that COVID-19 can spread during latent period and transmission rate is considered

for exposed state and  $\bar{\beta} = \beta$  What about infectious latent period? We assumed that individuals in state E are isolated with rate  $q$  and the out flow of this state is  $q$ . Isolated individuals are no longer infectious; they may recover or die after a latent period of distribution  $h_Q(t)$  or they may become symptomatic and escape to state I with outflow rate of  $q$ . Furthermore, Exposed individuals may become symptomatic and identified after a latent period of distribution  $h_E(t)$  and will recover or die after a removal latent period of distribution  $h_R(t)$ . Thus,  $R_c$  can be estimated as follows:

$$R_c = \theta\beta\bar{\mu}_Q + (1-\theta)(\beta\bar{\mu}_E + \beta\mu_R) \tag{7}$$

where  $\mu_R = \tau_R, \theta = pr(isolated|exposed)$  and  $\bar{\mu}_E, \bar{\mu}_Q$  defend as follows:

$$\bar{\mu}_Q = \frac{1}{q} \left( 1 - \exp\left(\frac{1}{2}q^2\sigma_Q^2 - q\tau_Q\right) \right) \tag{8}$$

$$\bar{\mu}_E = \frac{1}{q} \left( 1 - \exp\left(\frac{1}{2}q^2\sigma_E^2 - q\tau_E\right) \right) \tag{9}$$

Parameter  $R_c$  can be decreased by either increasing  $q$  or decreasing  $\tau_E$  and  $\tau_Q$  and this can be done by identifying exposed individuals via screening schemes. Parameter  $\theta$  can be considered equal 0.5 and means for every exposed individual the probability of being isolated or not is the same.

In stochastic case, we assumed that infectious rate  $\beta$  follows a normal distribution  $\beta \sim N(\beta_0, \sigma)$ , that is  $R_{cs}$  also has a normal distribution as follows:

$$R_{cs} \sim N(c\beta_0, c^2\sigma) \tag{10}$$

$$c = \theta\bar{\mu}_Q + (1-\theta)(\bar{\mu}_E + \mu_R) \tag{11}$$

It is clear that any change in transmission rate can directly be interpreted as change in the reproduction number.

### Results

In this section, we provide numerical solutions to illustrate main theoretical results on the proposed deterministic and stochastic models. For this purpose, the model is implemented in MATLAB 2017a. Starting points are considered to be  $E(0) = 2, I(0) = 0, Q(0) = 0, R(0) = 0$ . Parameters of these models and their corresponding descriptions are shown in Table 1.

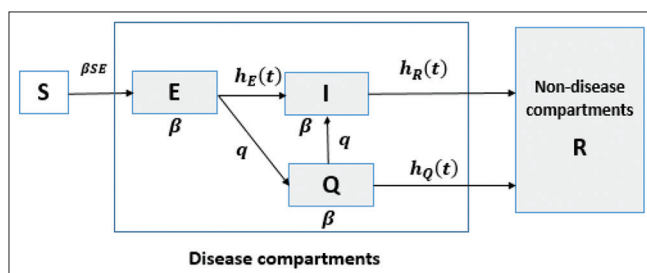


Figure 1: Block diagram of the proposed dynamic model

Model fit was done using cumulative confirmed, dead, and recovered regional data available on World Health Organization documents.<sup>[45]</sup> Based on the model, we used the LM and PMCMC methods to solve the optimization problem and identify parameters. To show that our model can adapt to changes in parameters over time, we fitted it with accumulated data of Iran, USA, and South Korea with two scenarios of short and long term of predicting data. Identified parameters of models for three selected countries are illustrated in Table 2.

In the left panel of Figure 2, we illustrated the sample distribution of transmission rate acquired from PMCMC method for three countries up to August [Table 2] for more clarity. According to this figure, posteriori distributions of transmission rate follow Gaussian distribution due to random walk assumption of transmission rate. In the right panel, we provided a plot of  $R_c$  with respect to parameters of (10), (11). As revealed in this figure, in the stochastic assumption,  $R_c$  has a normal distribution with mean value equals to the  $R_c$  of deterministic model and a variance depending on the variance of  $\beta$ .

### Results of deterministic model

#### Short term prediction

We started from the accumulated data of South Korea (from February 20 to May 5). This country was very successful in controlling the spread of the disease in early stages [Figure 3]. Besides low transmission rate, delay parameters ( $\tau_E, \tau_Q, \sigma_E$ ) were also very low which indicated that the exposed people have been identified and isolated very soon. The variance of this identification time, confirmed that approximately all exposed people were identified after at most 11 days. We then fitted the model to data of USA (from March 1 to May 5) and Iran (from February 20 to May 5) which their confirmed cases have been growing rapidly [Figure 4]. The spread rate of disease and delay parameters have higher values

Table 1: Parameters of model and their descriptions

Parameter	Value range	Source	Description
$\beta_0$	0-1	F	Transmission rate
$d$	0-0.1	F	Decaying coefficient
$T$	0-30	A	Starting point of interventions
$q$	0-1	F	Isolation rate
$\delta$	0-1	F	Mortality or recovery rate
$\alpha_1, \alpha_3$	1	A	Order of E, Q state
$\alpha_2, \alpha_4$	0-1	F	Fractional order of I, R states
$\tau_E$	2-11	F	Latent period
$\tau_Q$	4-10	F	Isolation period
$\tau_R$	7-12	A	Recovery period
$\sigma_E$	1-10	F	Latent standard deviation
$\sigma_Q$	1-10	F	Isolation standard deviation
$\sigma_R$	0-5	A	Treatment standard deviation
$\sigma$	0-0.1	F	Variance of $\beta$

F - Fitted with real data, A - Assumed



in the USA and Iran than South Korea's which indicate weaker performance to identify exposed cases. Indeed, approximately all exposed people were identified after at most 16 days [Table 2]. We assumed that, for both countries, government interventions have been applied 10 days after beginning of the epidemic ( $T = 10$ ). According to prediction results, if these interventions continued, the COVID19 pandemic could have been controlled in Iran by June and in the USA by August.

Figure 5 shows epidemic trajectories in the presence and absence of interventions in the USA and Iran. We plotted the effective reproduction number over time reflecting the impact of transmission rate reduction. It can be seen that although interventions effected the trend of epidemic and limited its spread, but does not decrease very much. We tried to find an appropriate decaying coefficient to predict if epidemic stops sooner. For  $d = 0.008$  reproductive number decreases more rapidly and equals to one late in June and the

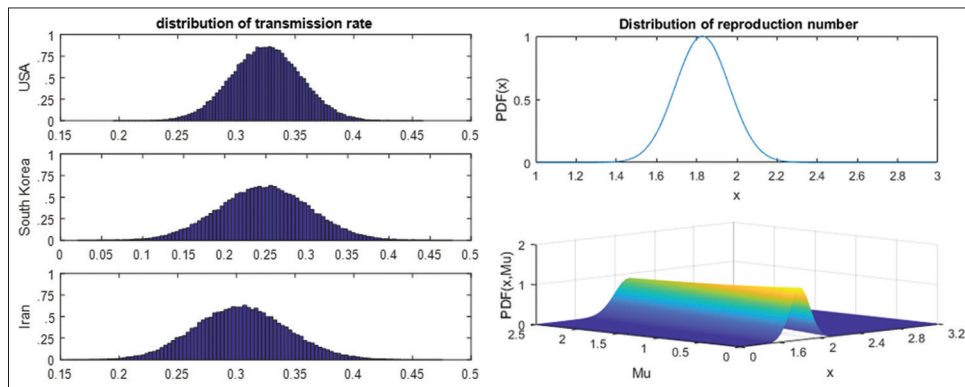


Figure 2: Left: Normal distributions of transmission rate in three countries up to August. Right: Variation of  $R_c$  with respect to parameters depicted in Equations (10), (11) for  $\sigma_q = 2, \tau_q = 5, \sigma_E = 2, \tau_E = 8, \mu_R = 11, \beta_0 = 0.27, q = 0.45, \sigma = 0.04$

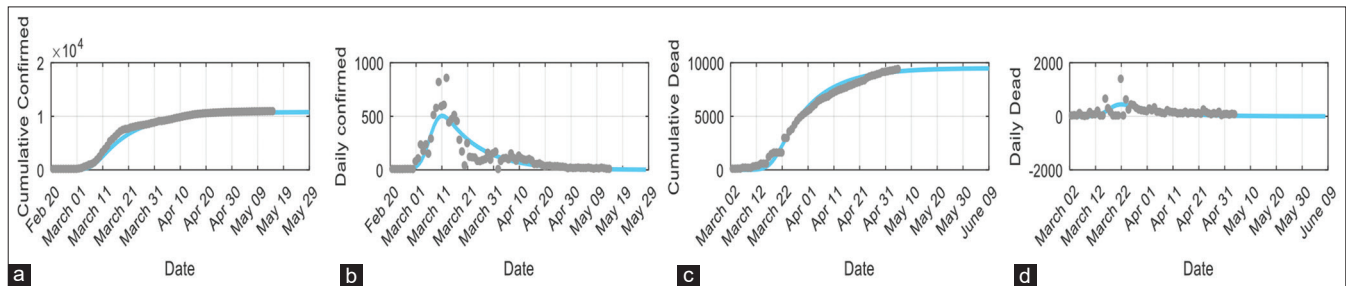


Figure 3: Numerical results of data fitting in deterministic model for South Korea. (a) Cumulative confirmed, (b) daily confirmed, (c) cumulative dead, (d) daily dead cases. Continuous line is prediction result and grey dots correspond to real data

Table 2: Fitted parameters of the countries

Parameter	S.K		USA			Iran		
	Up to August	After August	Up to June	June to September	After September	Up to June	June to September	After September
$\beta_0$	0.248	0.27	0.33	0.34	0.36	0.328	0.34	0.37
$d$	0	1e-4	2e-4	2e-4	1e-4	1.5e-3	1.2e-3	8e-4
$q$	0.5	0.45	0.65	0.5	0.47	0.64	0.64	0.62
$\delta$	0.02	0.032	0.1	0.07	0.07	0.06	0.056	0.07
$\alpha_2$	0.8	0.8	0.95	0.95	0.95	0.93	0.92	0.95
$\alpha_4$	1	1	0.7	0.7	0.7	1	0.98	0.98
$\sigma_E$	1.4	2	2	3.2	3.5	1.9	2.2	2.5
$\tau_E$	8	8	14	14	14	11	11	11
$\sigma_Q$	1.4	2	2	3.2	3.5	1.9	2.1	2.4
$\tau_Q$	5	5	11	11	11	8	8	8
$\sigma_R$	0	0	0	0	0	0	0	0
$\tau_R$	11	11	11	11	11	11	11	11
$\sigma$	0.05	0.04	0.03	0.02	0.02	0.08	0.06	0.06
$R_c$	1.7	1.82	2.16	2.25	2.5	2.14	2.16	2.35

epidemic completely goes to extinction in the USA and Iran. Comparing with current decreasing trend of transmission rate, either stricter rules must be enforced or patient screening schemes must be implemented continuously.

*Long-term prediction*

In early stage, South Korea was so successful in handling the COVID-19 pandemic, but this country experienced the second wave of infection from August 2020. We updated model's parameters and depicted the results in Figure 6. According to this figure, this pandemic will be controlled in South Korea by the end of 2020.

Furthermore, in Iran and the USA, social limitations decreased in June when public places, jobs, administrative

activities, etc., were reopened. Contact rate of individuals increased, and consequently, transmission rate was intensified, while isolation rate decreased. This caused spreading the disease into new populations or resurging in places that let down their guard too soon. We applied these points in our model by changing parameters of model from June to September (second peak) and after September (third peak) fitting to new data in Iran and the USA. Since readjustments of parameters have been done at a certain time (like a step function) and the derivative function of cumulative data has been used to indicate daily infected and death persons, discontinuities have been created in the daily curves (like Dirac delta functions) according to Figure 7c, f, k, and n. In order to precisely follow the outbreak trend, we

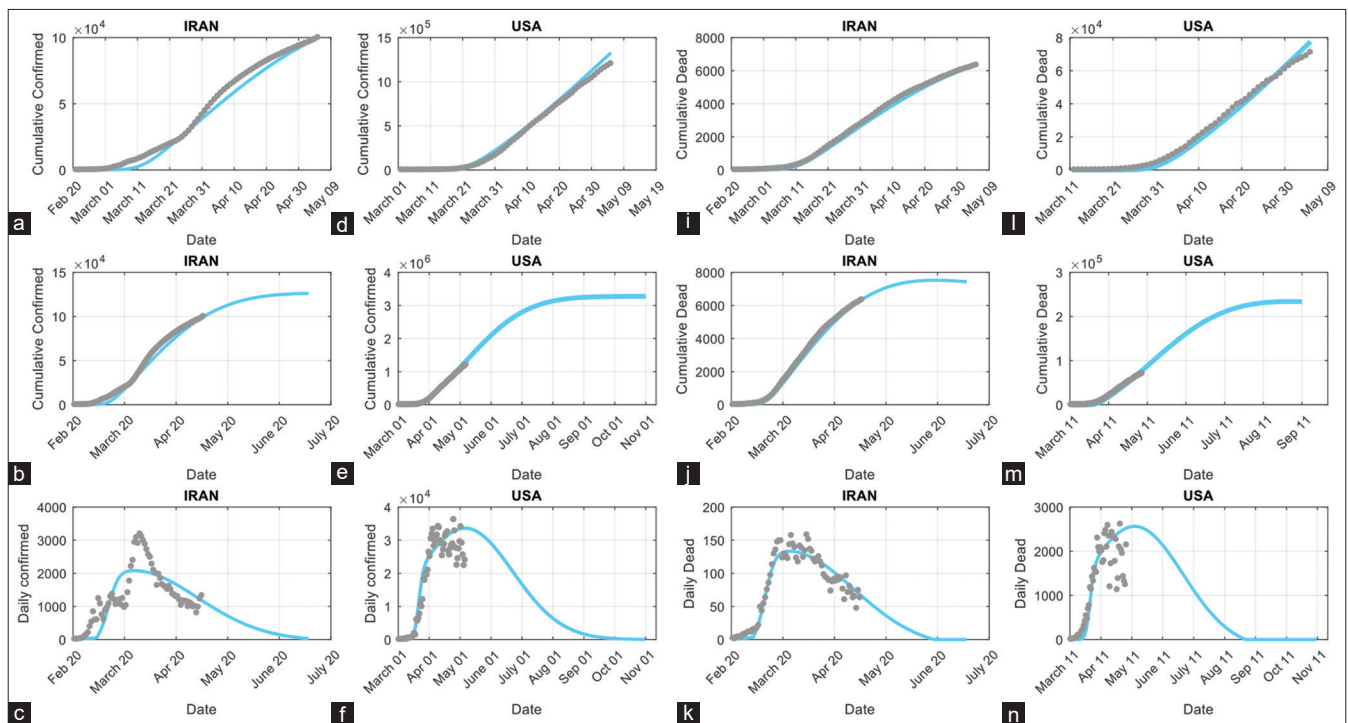


Figure 4: Numerical results of data fitting in deterministic model; (a), (b), (c), (i), (j), (k) for IRAN. (d), (e), (f), (l), (m), (n) for USA. Continuous line is prediction result and gray dots correspond to real data

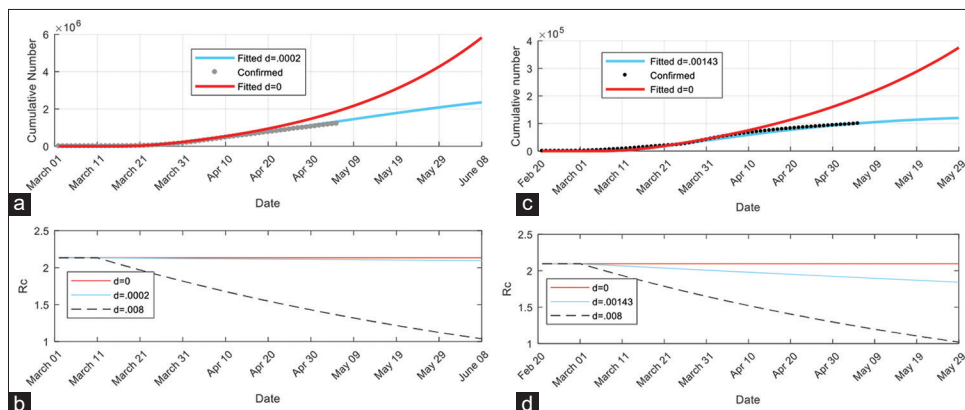
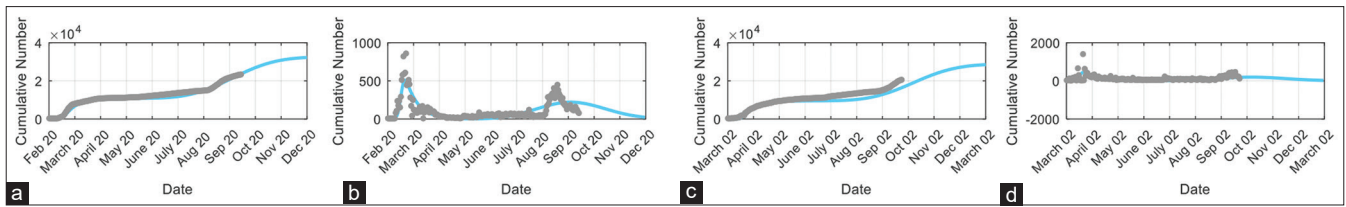
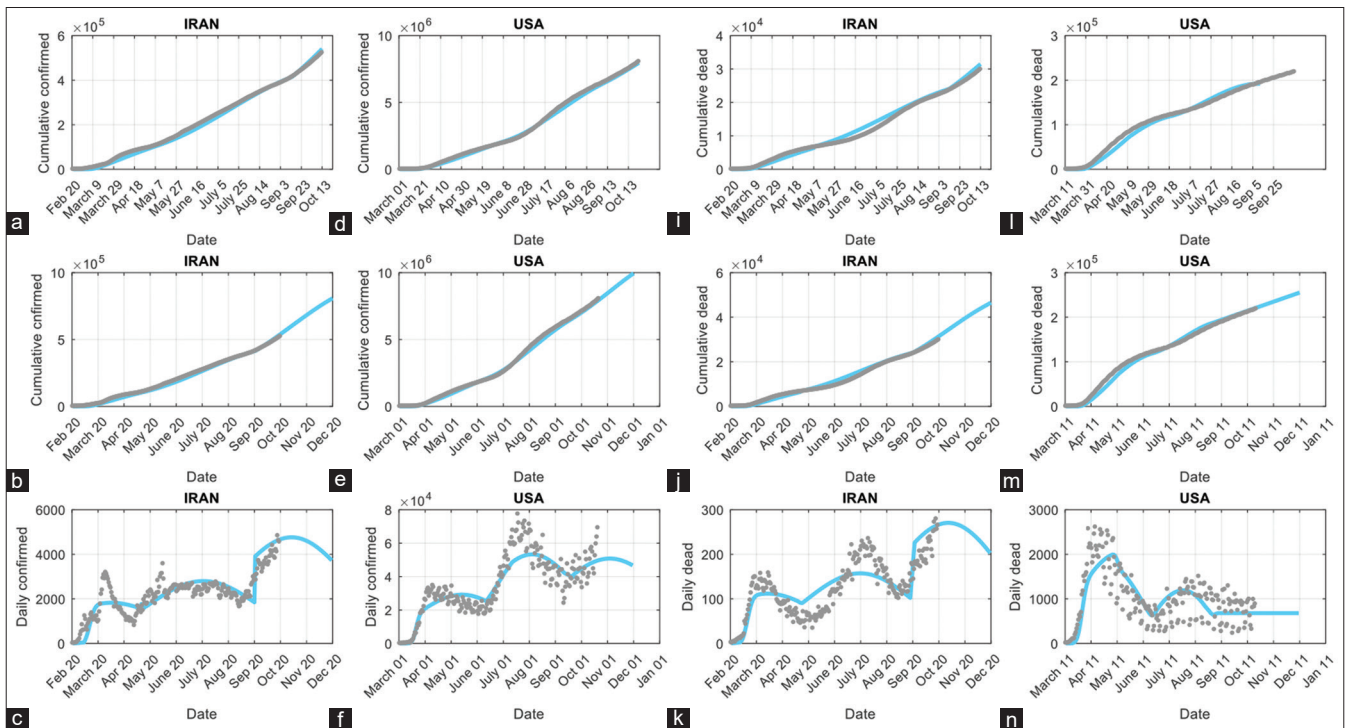


Figure 5: (a and c) Epidemic trajectories derived from the decaying and non-decaying transmission rates during the first 100 days of the epidemic for USA and Iran respectively. (b and d) Effective reproduction number over time reflecting the impact of transmission rate reduction for two decaying factors



**Figure 6:** Numerical results of data fitting in deterministic model for South Korea. (a) Cumulative confirmed, (b) daily confirmed, (c) cumulative dead, (d) Daily dead cases. Continuous line is prediction result and grey dots correspond to real data



**Figure 7:** Numerical results of data fitting in deterministic model; (a), (b), (c), (i), (j), (k) for Iran. (d), (e), (f), (l), (m), (n) for USA. Continuous line is prediction result and gray dots correspond to real data

have detected the amount of restrictions and cumulative data in mentioned countries to adjust the decaying coefficient,  $d$ , in (1). For interventions leading the reduction of  $\beta$ , we decreased decaying coefficient at the time point of starting interventions, and we increased decaying coefficient when the restrictive interventions decreased. We have done the precise adjustment by a new fitting of updated cumulative data to the model and have predicted outbreak trend.

We found that, as it was predicted, transmission rate and delay parameters have increased while isolation rate has decreased [Table 2 and Figure 7]. According to prediction results, if these conditions continued, COVID-19 pandemic will be continued in Iran and the USA by the end of 2020. Hence, unfortunately, there will be a deplorable situation. This indicates that stricter and more effective preventive laws and more diagnostic tests should be considered to avoid more disasters. We have considered a decreasing function to determine the decreasing nature of the transmission rate. Indeed, when countermeasures such as voluntary social distancing

by the population, short-term interventions such as global local lockdowns and general quarantine occur, its effect slowly decays the transmission rate for a period of time. Further abrupt changes in these interventions can be incorporated by adjusting the value of decaying coefficient at certain times. However, if restrictions were not sufficient from the beginning time ( $T$ ) or were not observed properly, their effect in reducing the transfer rate may disappear sooner, and amounts of underestimation may occur in the prediction when the model is used for long-term predictions. This may necessitate more increase in decaying coefficients.

### Results of stochastic model

In order to simulate the stochastic model, we used all fitted parameters of deterministic model except  $\beta$ . Parameter  $\beta$  identified in the deterministic model, was used as an initial guess of transmission rate for the PMCMC method to estimate variance of transmission rate [Figure 2 (left panel)]. Documentation and implementation of MATLAB example of this method are available on.<sup>[46]</sup>

Since transmission rate follows the random walk, general trend of epidemic is preserved during ensemble realization of the model. These fluctuations in transmission rate can model daily changes in the number of infected cases better than deterministic model

Results of stochastic model are shown in Figures 8 and 9. Corresponding to Figure 8 (short-term prediction), if health restrictions had been continued, probability of outbreak with greater size would have been very low and the outbreak would have been going to be slowly disappeared by the end of June in Iran and by the end of August in the USA.

However, due to removal of restrictions, we changed the parameters of model as we described in section 3.1.1 and observed that if current conditions remain, the pandemic will last at least by the end of 2020 in mentioned countries, corresponds to Figure 8 (long term prediction), and more and more spreading peaks will occur. In Figures 8 and 9 the corresponding distributions of outbreak size of cumulative [b, f, j] and daily [d, h, l] numbers are also represented for three countries. According to Figure 9, the distribution of outbreak size tends to have a large cumulative and daily frequencies of small sizes meanwhile a very heavier tail than its corresponding distribution in Figure 8. It indicates that final size of outbreak will dramatically increase compared to the scenario that could happen in short-term prediction. It confirms that in the absence of comprehensive controlling efforts, the transmission rate has increased rapidly, resulting in a sizable epidemic.

### Evaluation of the deterministic model

In order to evaluate the model in tracking the trend of the outbreak, we gathered new data which was not used in parameter estimation of the proposed EQIR model. We calculated the normalized mean square error (MSE) between predicted, and true number of daily confirmed infected and dead cases as follows:

$$MSE(x, \hat{x}) = \frac{\frac{1}{T} \sum_{t=1}^T (x(t) - \hat{x}(t))^2}{\bar{x} \times \bar{\hat{x}}} \quad (12)$$

where  $x(t), \hat{x}(t)$  denote the true and  $x, \hat{x}$  predicted time series respectively and  $\bar{\cdot}$  indicates the average values of corresponding time series. Results of MSE calculated for  $T = 30$  (from October 20,2020, to November 20,2020) are reported in Table 3 for more details. According to this table, normalized MSE of the proposed model is very low, and the proposed model has good performance in prediction of new cases. It must be mentioned that errors by day are probably greater than the cumulative errors. This can be visually observed through cumulative and daily plots as well [Figures 6 and 7].

**Table 3: Normalized mean square error of the countries**

	South Korea	USA	Iran
Infected MSE	0.00634	0.00876	0.13468
Dead MSE	0.00176	0.05307	0.09579

MSE – Mean square error

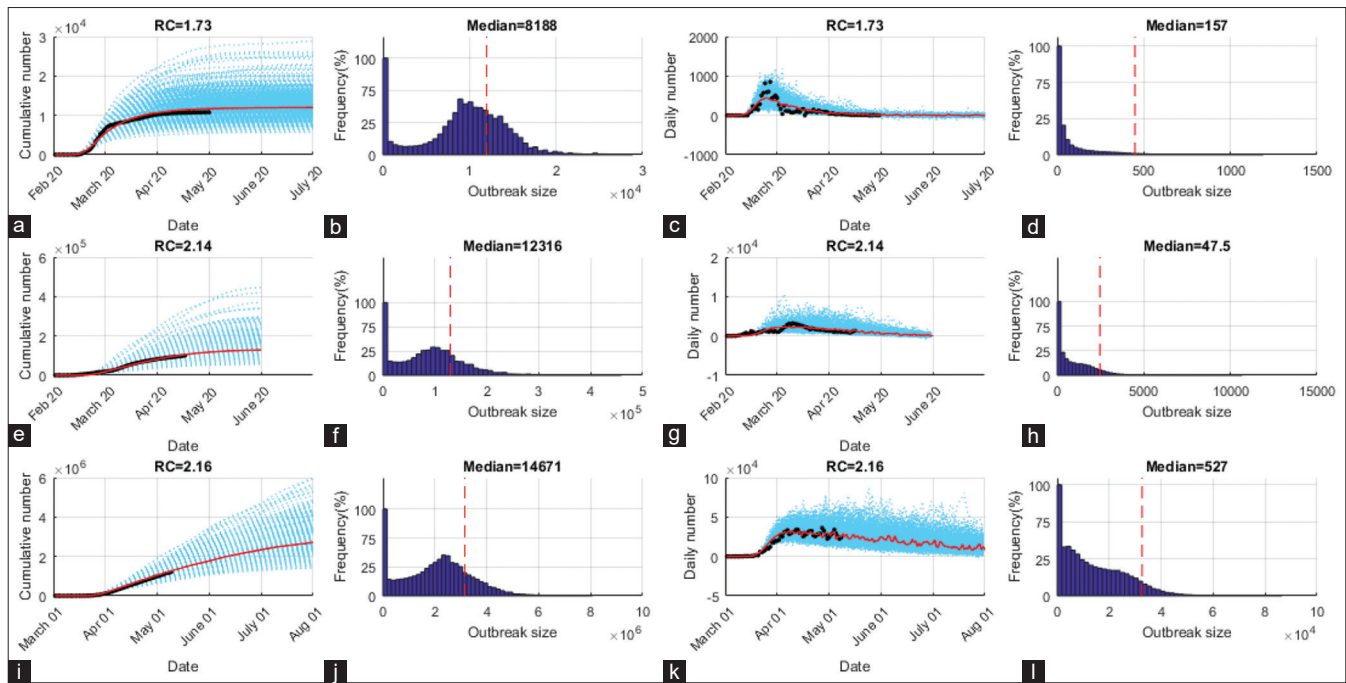


Figure 8: Short term stochastic epidemic realizations of the proposed model in South Korea (a and c), Iran (e and g), USA (I and k).The black dots correspond to the actual outbreak trajectory and the cyan blue lines correspond to 500 stochastic realizations of cumulative and daily number of infected cases respectively. Red line indicates median. The corresponding distribution of outbreak sizes are shown for South Korea in (b and d), Iran in (f and h), and USA in (j and l). The vertical dashed line indicates the predicted COVID-19outbreak size



### Forecasting the long-term trend of epidemic in Iran

We have adjusted parameters of the proposed dynamic model such as transmission rate, and isolation rate at different time points due to the changes occurred in the behavior of Iranian population resulting in the growth of infection rate in this country and caused several consecutive peaks in the outbreak trend. We provided the numerical results of data fitting of confirmed infected cases for the proposed EQIR deterministic and stochastic models from February 20,2020 to August 01,2021. Figure 10 shows epidemic trajectories in Iran since the beginning of the outbreak. We also provided prediction of number of

patients in Iran for the next 2 weeks in Table 4. According to the prediction results, if these conditions continued, the 5<sup>th</sup> peak of the COVID-19 pandemic will be continued in Iran till September 2021.

### Discussion

In this study, we proposed a dynamic EQIR model of COVID-19 spread. State-dependent structure of this model makes it possible to consider different descriptive parameters. We considered 14 parameters for this model among them up to 10 parameters were fitted to real data using optimization methods. Fitting process can

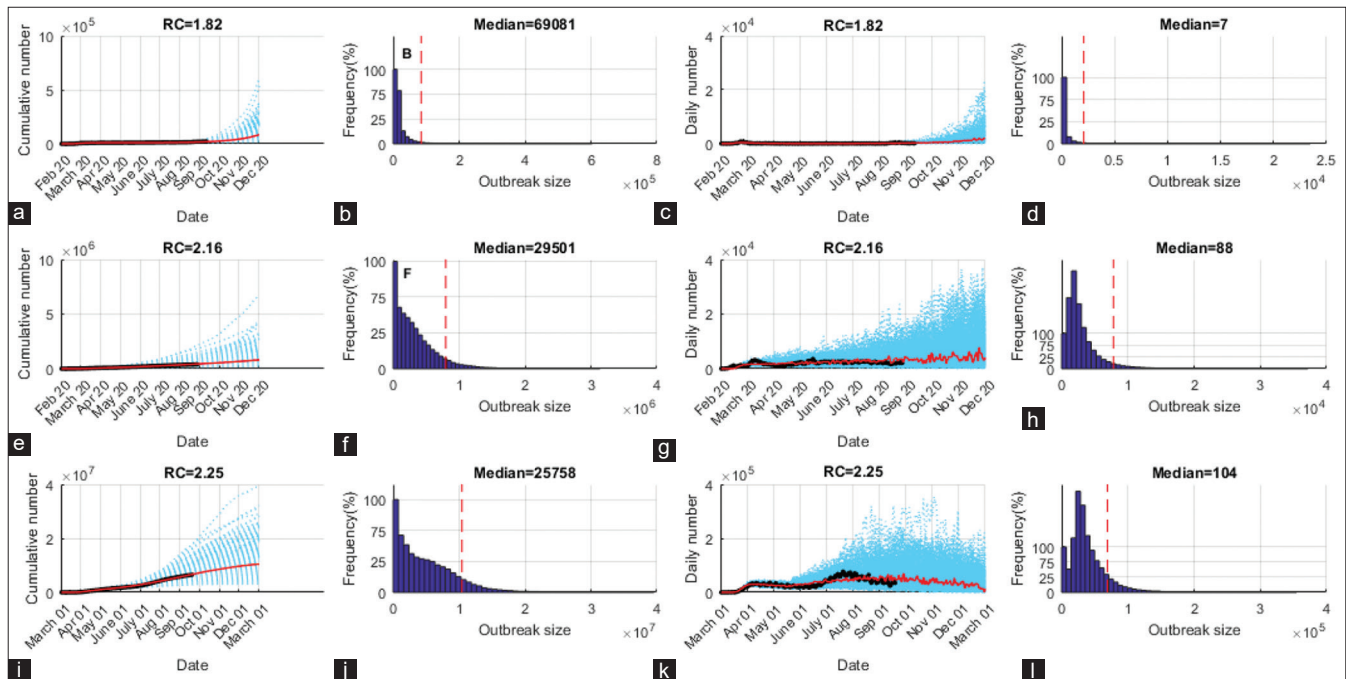


Figure 9: Long term stochastic epidemic realizations of the proposed model in South Korea (a and c), Iran (e and g), USA (i and k). The black dots correspond to the actual outbreak trajectory and the cyan blue lines correspond to 500 stochastic realizations of cumulative and daily number of infected cases respectively. Red line indicates median. The corresponding distribution of outbreak sizes are shown for South Korea in (b and d), Iran in (f and h), and USA in (j and l). The vertical dashed line indicates the predicted COVID-19 outbreak size

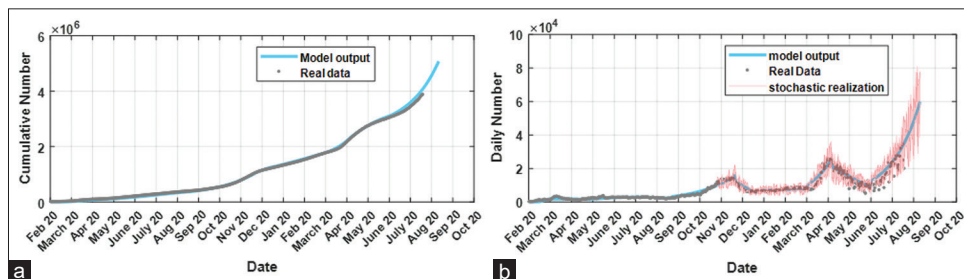


Figure 10: Numerical results of data fitting in deterministic and stochastic model in Iran; (a) cumulative infected cases, (b) daily infected cases

Table 4: Number of predicted patients and dead cases in Iran from 01 August 2021 to 14 August 2021

	August 1	August 2	August 3	August 4	August 5	August 6	August 7	August 8	August 9	August 10	August 11	August 12	August 13	August 14
Infected	33435	34467	35293	36989	37870	39701	40652	41628	42629	43565	44710	45792	46810	48040
Dead	307	322	324	331	365	391	410	428	442	457	469	478	500	533

be simply done because of the possibility of parameter description in real world. Furthermore, the model can be easily adapted with changing the interventions of human societies due to reality of parameters. For instance, effects of isolation and screening schemes of exposed people are expressed by isolation rate ( $q$ ) and different latent periods  $\tau_E$ ,  $\tau_Q$  respectively. The effect of contact rate decreasing activities such as home quarantine and closure of public places, which leads to lower transmission rates, is considered by an exponentially decreasing function of time for transmission rate. In addition, fractional order of differential equations determines the effect of society's memory in the face of disease. This degree of freedom allows the model to fit better to data as we observed for South Korea whose fractional order of infected state was 0.8 where for Iran and the USA this value was near one. We obtained the relation between parameters of model and reproduction number and clearly justified it through parameters of model. Results of model demonstrated that besides increasing isolation rate, decreasing  $\tau_E$ ,  $\tau_Q$  by means of identifying exposed people, help societies effectively control the outbreak.

We also implemented a stochastic model by adding a Gaussian noise to transmission rate of deterministic model. By considering such SDEs, we could insert model some of real-world fluctuations behavioral trend of disease. Results of this model can help us to estimate maximum intensive care units needed per day in the worst condition. In addition, the stochastic model shows that if the rate of spread has limited changes, outbreak size will be small.

We simulated short and long-term scenarios to predict outbreak size for both models and justified predictions due to changing in parameters in accordance to behavior of communities. Using different scenarios, this model can continuously monitor the trajectory of the disease and help control it.

## Conclusion

In this paper, we proposed a mathematical fractional order distributed delay dynamic model for spreading the trend of COVID-19 in two deterministic and stochastic versions. It is a flexible model that can observe and predict outbreak of diseases such as COVID-19, which can spread during their incubation period. In addition, multiple model parameters with physical descriptions allow us to apply these changes to the model at any time and forecast the future outbreak of COVID-19. However, some higher degrees of contacts such as transmission within health-care institutions and households are not considered in this model. Moreover, Due to unknown safety period obtained after infecting by COVID-19 and the possibility of renewed infectious after immunity, this model must be developed by adding new states and routes for accurate prediction of disease outbreak.

## Financial support and sponsorship

This work was supported in part by the Vice-Chancellery for Research and Technology, Isfahan University of Medical Sciences, under Grant 199067.

## Conflicts of interest

There are no conflicts of interest.

## References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020;382:1199-207.
2. Kafieh R, Arian R, Saeedizadeh N, Amini Z, Serej ND, Minaee S, *et al.* Haghjooy Javanmard S. COVID-19 in Iran: forecasting pandemic using deep learning. *Computational and mathematical methods in medicine*. 2021 Feb 25;2021.
3. Pirouz B, Shaffiee Haghshenas S, Shaffiee Haghshenas S, Piro P. Investigating a serious challenge in the sustainable development process: Analysis of confirmed cases of COVID-19 (new type of coronavirus) through a binary classification using artificial intelligence and regression analysis. *Sustainability* 2020;12:2427.
4. Schoenberg FP, Hoffmann M, Harrigan RJ. A recursive point process model for infectious diseases. *Ann Inst Stat Math* 2019;71:1271-87.
5. You C, Deng Y, Hu W, Sun J, Lin Q, Zhou F, *et al.* Estimation of the time-varying reproduction number of COVID-19 outbreak in China. *Int J Hyg Environ Health* 2020;228:113555.
6. Hethcote HW. Three basic epidemiological models. In: *Applied Mathematical Ecology*. Springer, Berlin, Heidelberg; 1989. p. 119-44.
7. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, *et al.* Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med* 2020;26:855-60.
8. López L, Rodo X. A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: simulating control scenarios and multi-scale epidemics. *Results in Physics*. 2021 Feb 1;21:103746.
9. Cooper I, Mondal A, Antonopoulos CG. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos Solitons Fractals* 2020;139:110057.
10. Law KB, Peariasamy KM, Gill BS, Singh S, Sundram BM, Rajendran K, *et al.* Tracking the early depleting transmission dynamics of COVID-19 with a time-varying SIR model. *Sci Rep* 2020;10:21721.
11. Chen YC, Lu PE, Chang CS, Liu TH. A time-dependent SIR model for COVID-19 with undetectable infected persons. *IEEE Trans Netw Sci Eng* 2020;7:3279-94.
12. Marinov TT, Marinova RS. Dynamics of COVID-19 using inverse problem for coefficient identification in SIR epidemic models. *Chaos Solitons Fractals X* 2020;5:100041.
13. Rădulescu A, Williams C, Cavanagh K. Management strategies in a SEIR-type model of COVID 19 community spread. *Sci Rep* 2020;10:21256.
14. Grimm V, Mengel F, Schmidt M. Extensions of the SEIR model for the analysis of tailored social distancing and tracing approaches to cope with COVID-19. *Sci Rep* 2021;11:4214.
15. He S, Peng Y, Sun K. SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dyn* 2020;101:1667-80.
16. Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, *et al.* Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac*

- Dis 2020;12:165-74.
17. Alrabaiah H, Arfan M, Shah K, Mahariq I, Ullah A. A comparative study of spreading of novel corona virus disease by using fractional order modified SEIR model. *Alexandria Eng J* 2021;60:573-85.
  18. Iwata K, Miyakoshi C. A simulation on potential secondary spread of novel coronavirus in an exported country using a stochastic epidemic SEIR model. *J Clin Med* 2020;9:E944.
  19. Liu Z, Magal P, Seydi O, Webb G. A COVID-19 epidemic model with latency period. *Infect Dis Model* 2020;5:323-37.
  20. Mahajan A, Sivadas NA, Solanki R. An epidemic model SIPHERD and its application for prediction of the spread of COVID-19 infection in India. *Chaos Solitons Fractals* 2020;140:110156.
  21. Korolev I. Identification and estimation of the SEIRD epidemic model for COVID-19. *J Econom* 2021;220:63-85.
  22. Rajendran S, Jayagopal P. Accessing COVID19 epidemic outbreak in Tamilnadu and the impact of lockdown through epidemiological models and dynamic systems. *Measurement (Lond)* 2021;169:108432.
  23. Lekone PE, Finkenstädt BF. Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics* 2006;62:1170-7.
  24. Britton T, Pardoux E, Ball F, Laredo C, Sirl D, Tran V.C. Stochastic epidemic models with inference. Berlin: Springer; 2019. p. 187.
  25. Tajmirriahi M, Amini Z. Modeling of seizure and seizure-free EEG signals based on stochastic differential equations. *Chaos Solitons Fractals* 2021;150:111104.
  26. Tajmirriahi M, Amini Z, Hamidi A, Zam A, Rabbani H. Modeling of retinal optical coherence tomography based on stochastic differential equations: Application to denoising. *IEEE Trans Med Imaging* 2021;40:2129-41.
  27. Chen Y, Cheng J, Jiang Y, Liu K. A time delay dynamical model for outbreak of 2019-nCoV and the parameter identification. *J Inverse Ill Posed Probl* 2020;28:243-50.
  28. Alofi A, Cao J, Elaiw A, Al-Mazrooei A. Delay-dependent stability criterion of Caputo fractional neural networks with distributed delay. *Discrete Dynamics in Nature and Society*. 2014 Jan 1;2014.
  29. 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. *Hal-Inria* 2018.
  30. Liu Z, Magal P, Seydi O, Webb G. Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data. *Mathematical Biosciences and Engineering: MBE*. 2020;17:3040-51.
  31. Saeedian M, Khalighi M, Azimi-Tafreshi N, Jafari GR, Ausloos M. Memory effects on epidemic evolution: The susceptible-infected-recovered epidemic model. *Phys Rev E* 2017;95:022409.
  32. Agarwal R, Belmekki M, Benchohra M. A survey on semilinear differential equations and inclusions involving Riemann-Liouville fractional derivative. *Adv Differ Equ* 2009;2009:1-47.
  33. Stetter HJ. Analysis of discretization methods for ordinary differential equations. Berlin-Heidelberg-New York: Springer; 1973 May.
  34. Liu J, Peng B, Zhang T. Effect of discretization on dynamical behavior of SEIR and SIR models with nonlinear incidence. *Appl Math Lett* 2015;39:60-6.
  35. Dimitrov Y. Numerical approximations for fractional differential equations. arXiv preprint arXiv:1311.3935. 2013 Nov 15.
  36. Durrett R, Durrett R. Brownian motion and martingales in analysis. In: Wadsworth Advanced Books & Software California. 1984.
  37. Moré JJ. The Levenberg-Marquardt algorithm: implementation and theory. In: *Numerical analysis*: Springer, Berlin, Heidelberg; 1978. pp. 105-16.
  38. Endo A, van Leeuwen E, Baguelin M. Introduction to particle Markov-chain Monte Carlo for disease dynamics modellers. *Epidemics* 2019;29:100363.
  39. Andrieu C, Roberts GO. The pseudo-marginal approach for efficient Monte Carlo computations. *Ann Stat* 2009;37:697-725.
  40. Kiskinov H, Zahariev A. On fractional systems with Riemann-Liouville derivatives and distributed delays-choice of initial conditions, existence and uniqueness of the solutions. *Eur Phys J Spec Top* 2017;226:3473-87.
  41. ÖZalp N, Demirci E. A fractional order SEIR model with vertical transmission. *Math Comput Model* 2011;54:1-6.
  42. Yan P, Chowell G. Quantitative methods for investigating infectious disease outbreaks: Cham, Switzerland: Springer, 2019.
  43. van den Driessche P. Reproduction numbers of infectious disease models. *Infect Dis Model* 2017;2:288-303.
  44. Farman M, Saleem MU, Ahmad A, Ahmad MO. Analysis and numerical solution of SEIR epidemic model of measles with non-integer time fractional derivatives by using Laplace Adomian Decomposition Method. *Ain Shams Eng J* 2018;9:3391-7.
  45. World Health Organization. Available from: <https://covid19.who.int/>.
  46. Pseudo-Marginal MCMC. Available from: <https://github.com/matt-graham/auxiliary-pm-mcmc>.

## Appendices

### Appendix A: Discrete deterministic model

For discretization of the model we use the backward Euler scheme.<sup>[33,34]</sup> Using this discretization method,  $N$  time intervals  $t_0 < t_1 < \dots < t_N = T$  with a uniform time step size  $u = \frac{T}{N}$  are defined. For approximating the fractional derivative at time  $t = t_n$ , the shifted Grunewald formula<sup>[35]</sup> is used as follows

$$D_t^\alpha f(t_n) \approx \frac{1}{u^\alpha} \sum_{k=0}^n w_k f_{n-k}$$

$$w_0 = 1, w_j = \frac{(-1)^k \Gamma(\alpha + 1)}{\Gamma(k + 1) \Gamma(\alpha - k + 1)} \quad k = 1, 2, \dots, n \quad (13)$$

Now, let  $\lambda = u^\alpha \beta_n$  where  $\beta_n$  is discrete version of time varying function introduced in Equation 1. Let  $\gamma = u^\alpha q$  where  $q$  is isolation rate. Let  $\nu = u^\alpha \delta$  where  $\delta$  is considered as mortality or recovery rate. Thus, discrete representation of model is illustrated as follows.

$$\begin{cases} (w_0 - \lambda)E_n + \lambda Q_n + \lambda I_n + \sum_{k=1}^{n-1} w_k E_{n-k} = 0 \\ \lambda h_E^n E_n + (w_0 + \lambda h_E^n) I_n + \lambda h_E^n Q_n = \sum_{k=1}^{n-1} (\lambda h_E^k E_k - (w_{n-k} + \lambda h_E^k) I_k - \lambda h_E^k Q_k) \\ (w_0 + \gamma + \gamma h_Q^n) Q_n - \gamma(1 + h_Q^n) E_n + \gamma(1 + h_Q^n) I_n = \sum_{k=1}^{n-1} (\gamma h_Q^k E_k - \gamma h_Q^k I_k - (\gamma h_Q^k - w_{n-k}) Q_k) \\ \nu \beta_n h_R^n (-E_n + I_n + Q_n) + w_0 R_n = \nu \beta_n \sum_{k=1}^{n-1} h_R^k (E_k - I_k - Q_k) - \sum_{k=1}^{n-1} w_{n-k} R_k \end{cases} \quad (14)$$

where variables  $E_n, I_n, Q_n$  are numerical solutions of  $E(t_n), I(t_n), Q(t_n)$  and  $h(\cdot)$  as explained in assumptions and defined as follows

$$h_{E,Q,R}^k = \frac{1}{\sigma_{E,Q,R} \sqrt{2\pi}} e^{-\frac{(n-\tau_{E,Q,R}-k)^2}{2\sigma_{E,Q,R}^2}} \quad (15)$$

### Appendix B: Discrete stochastic model

For stochastic model we consider that  $\beta$  is time varying function following the random walk. Modeling time varying nature of the transmission rate as a random walk, implicitly declare that the amount of transmission rate in one day is not much different from previous day. Thus, transmission rate is considered as stochastic differential equation as follows:

$$d\beta_t = \sigma dW_t \quad (16)$$

where  $W_t$  denotes the Wiener process<sup>[36]</sup> and is  $\sigma$  fluctuation of transmission rate. If we apply Ito's lemma to the process, we get:

$$\beta_t = \beta_0 + \sigma dW_t \quad (17)$$

Now, let  $A = u^\alpha \beta_0 + u^\alpha \sigma \sqrt{\Delta t} \xi_n$  where  $\Delta t = t_n - t_{n-1}$  and is assumed to be equal to  $u$ .  $\xi_n$  is Gaussian random variable  $\mathcal{N}(0,1)$ . Let  $A^k = u^\alpha \beta_0 + u^\alpha \sigma \sqrt{\Delta t} \xi_n^k$  where  $\xi_n^k$  denotes the  $k^{\text{th}}$  Gaussian random variable  $\mathcal{N}(0,1)$ . Then, Model's differential equations can be written as follows (backward Euler approximation)



$$\left\{ \begin{array}{l} (w_0 - A)E_n + AQ_n + AI_n + \sum_{k=1}^{n-1} w_k E_{n-k} = 0 \\ Ah_E^n E_n + (w_0 + Ah_E^n)I_n + Ah_E^n Q_n = \sum_{k=1}^{n-1} (A^k h_E^k E_k - (w_{n-k} + A^k h_E^k)I_k - A^k h_E^k Q_k) \\ (w_0 + \gamma + \gamma h_Q^n)Q_n - \gamma(1 + h_Q^n)E_n + \gamma(1 + h_Q^n)I_n = \sum_{k=1}^{n-1} (\gamma h_Q^k E_k - \gamma h_Q^k I_k - (\gamma h_Q^k - w_{n-k})Q_k) \\ \mu Ah_R^n (-E_n + I_n + Q_n) + w_0 R_n = \delta \sum_{k=1}^{n-1} A^k h_R^k (E_k - I_k - Q_k) - \sum_{k=1}^{n-1} w_{n-k} R_k \end{array} \right. \quad (18)$$

where variables  $E_n$ ,  $I_n$ ,  $Q_n$  are numerical solutions of  $E(t_n)$ ,  $I(t_n)$ ,  $Q(t_n)$  and  $h(\cdot)$  as explained in assumptions and defined as follows

$$h_{E,Q,R}^k = \frac{1}{\sigma_{E,Q,R} \sqrt{2\pi}} e^{-\frac{(n-\tau_{E,Q,R}-k)^2}{2\sigma_{E,Q,R}^2}} \quad (19)$$