

RESEARCH ARTICLE

Incorporating global dynamics to improve the accuracy of disease models: Example of a COVID-19 SIR model

Hadeel AlQadi^{1,2*}, Majid Bani-Yaghub¹

1 Department of Mathematics and Statistics, University of Missouri-Kansas City, Kansas City, Missouri, United States of America, **2** Department of Mathematics, Jazan University, Jazan, Saudi Arabia

* hadz6@mail.umkc.edu



Abstract

Mathematical models of infectious diseases exhibit robust dynamics, such as stable endemic, disease-free equilibriums or convergence of the solutions to periodic epidemic waves. The present work shows that the accuracy of such dynamics can be significantly improved by including global effects of host movements in disease models. To demonstrate improved accuracy, we extended a standard Susceptible-Infected-Recovered (SIR) model by incorporating the global dynamics of the COVID-19 pandemic. The extended SIR model assumes three possibilities for susceptible individuals traveling outside of their community:

- They can return to the community without any exposure to the infection.
- They can be exposed and develop symptoms after returning to the community.
- They can be tested positively during the trip and remain quarantined until fully recovered.

To examine the predictive accuracy of the extended SIR model, we studied the prevalence of the COVID-19 infection in six randomly selected cities and states in the United States: Kansas City, Saint Louis, San Francisco, Missouri, Illinois, and Arizona. The extended SIR model was parameterized using a two-step model-fitting algorithm. The extended SIR model significantly outperformed the standard SIR model and revealed oscillatory behaviors with an increasing trend of infected individuals. In conclusion, the analytics and predictive accuracy of disease models can be significantly improved by incorporating the global dynamics of the infection.

OPEN ACCESS

Citation: AlQadi H, Bani-Yaghub M (2022) Incorporating global dynamics to improve the accuracy of disease models: Example of a COVID-19 SIR model. PLoS ONE 17(4): e0265815. <https://doi.org/10.1371/journal.pone.0265815>

Editor: Yury E. Khudyakov, Centers for Disease Control and Prevention, UNITED STATES

Received: February 25, 2021

Accepted: March 8, 2022

Published: April 8, 2022

Copyright: © 2022 AlQadi, Bani-Yaghub. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper, [Supporting Information](#) files, and uploaded to <https://www.kaggle.com/hadeelalqadi/uscovid-19data>.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

1- Introduction

COVID-19 is the recent infectious disease caused by the severe acute respiratory syndrome novel coronavirus (SARS-CoV-2). Because the transmissibility of this virus is relatively high and the outbreaks remained undetected for several days, COVID-19 turned into a global pandemic. Almost all countries of the world have been exposed to this virus. Since January 2020, more 256 million individuals have become infected with the COVID-19. The infection has resulted more than 5 millions death as of December, 2021 [1]. Just in the US, the COVID-19 cases are over 48 millions and more than 786,000 deaths as of December, 2021 [2]. In order to

reduce the spread of COVID-19, businesses, communities, and governments have implemented different control measures such as mandatory lockdowns, social distancing, avoiding crowded events, using face masks in public, and vaccination [3]. Nevertheless, control of COVID-19 remains a major issue in several parts of the world [4].

COVID-19 is mainly transmitted from human-to-human via direct contact with contaminated surfaces and through the inhalation of respiratory droplets from infected individuals [5]. About 97% of the infected individuals will recover after period ranging between one to four weeks. Therefore, the use of mathematical modeling seems to be an appropriate approach to study COVID-19 transmission dynamics.

Mathematical modeling of infectious diseases has increasingly become an essential tool for prevention, prediction, and control of infectious diseases [6–8]. Since 1760, when Daniel Bernoulli developed the first disease model of smallpox, numerous mathematical models have been utilized to study disease transmission dynamics, and to predict, assess, and control infectious diseases [9–12]. The substance of mathematical modeling lies in formulating a set of mathematical equations that mimic reality [13]. Mathematical models have been evolved from small sets of ordinary differential equations to sophisticated compartmental models with several equations (see [14–16] for a review).

One of the simplest, yet powerful, disease models is the standard Susceptible-Infected-Recovered (SIR) model, which was first introduced by Kermack and McKendrick in a series of three papers [17–19]. In a standard SIR model, the host population is divided into susceptible, infected and recovered individuals, denoted by $S(t)$, $I(t)$ and $R(t)$, respectively. These quantities track the numbers of individuals in each compartment over different time periods [20, 21]. The standard SIR model without birth and death is represented by the set of ordinary differential equations [22]:

$$\begin{aligned}\frac{dS}{dt}(t) &= -\beta S(t)I(t). \\ \frac{dI}{dt}(t) &= \beta S(t)I(t) - \gamma I(t). \\ \frac{dR}{dt}(t) &= \gamma I(t).\end{aligned}\tag{1}$$

Where β is the average number of susceptible individuals infected by one infectious individual per contact per unit of time (the transmission rate), and γ is the average number of infected individuals recovered per unit of time (recovery rate).

For decades, the standard SIR model has been extended to various forms by adding different compartments to suit the biological, spatio-temporal and social aspects of the disease dynamics or to study the impact of intervention strategies on the disease transmission dynamics in different communities [23, 24]. For instance, it has been extended to SIR models with diffusion [25], contaminated environment [26, 27], delay terms [28], several strains of infection [29], and multiple routes of infection [30].

Recently, several researchers utilized mathematical modeling to analyze, and predict the transmission dynamics of COVID-19 pandemic [31–33]. Dynamics of COVID-19 epidemic has been simulated using different versions of SIR or SEIR (susceptible, exposed, infected and recovered) models [31, 32]. The main modification include adding asymptomatic and symptomatic infection compartments [33], hospitalization compartment [34], and quarantined and isolated compartments [31]. These models are presumably able to predict and simulate the number of infected cases by taking into consideration the asymptomatic and symptomatic

cases, deaths, needs of beds in hospitals, and effect of control measures and the interventions to decrease the number of cases.

The abovementioned extended SIR models contribute to the existing literature. However, they largely ignore the effects of global dynamics of infection on local communities. The presence of a global pandemic or a widespread infection can largely influence the dynamics of infection in a local community. Most communities are well-connected and the assumption that the disease exists only within the community is invalid [35].

There have been attempts to include the global dynamics in different SIR models [28, 36]. Nevertheless, such extended SIR models have several unknown parameters and poorly fit to data of host population. Due to lack flexibility and poor fitness to data, there is a need for develop SIR models that are more practical.

Moreover, regardless of the parameter values, most numerical simulations of SIR models are limited to three distinct dynamics. The first of these dynamics is the solution curve of infected individuals may exhibit an epidemic wave before converging to a disease-free equilibrium [37], secondly, the solution curve of infected converge to an endemic equilibrium [38], the third of these dynamics is the solution curve of infected converge to periodic epidemic waves [39]. For the standard SIR model (1), the dynamics are even more limited. Namely, the solution curves always represent the same qualitative dynamics: an epidemic wave of the infectious population, an inverted S shape for susceptible population, and S shape for the recovered population. Regardless of the set of parameter values and initial conditions, such qualitative behaviors will always remain the same (see panels Fig 1A–1C). A quick review of the number of individuals infected with COVID-19, at the country [1],

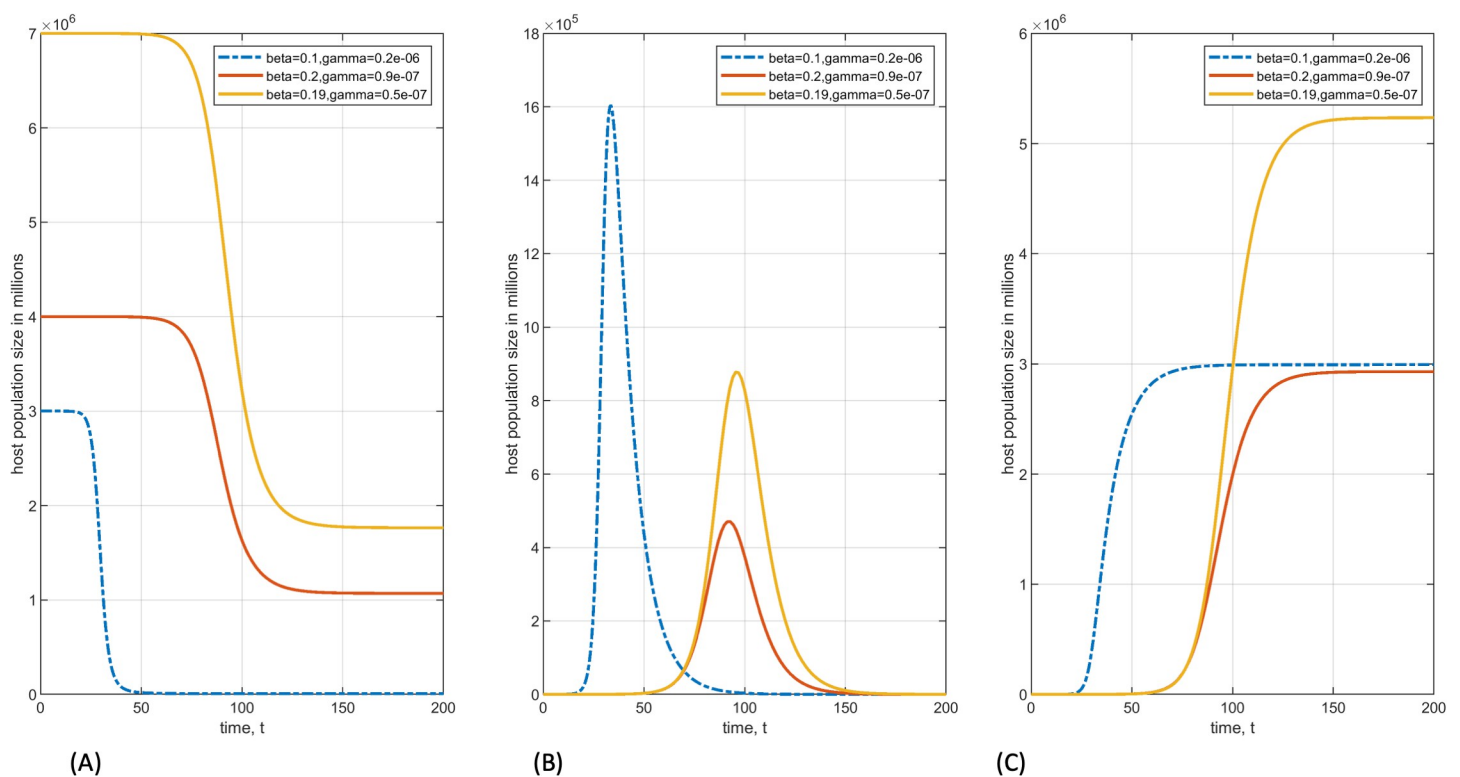


Fig 1. Qualitative behavior of the standard SIR model remains the same regardless of the parameter values (β and γ). (A) an inverted S-shape occurs for the susceptible population. (B) a bell-shaped epidemic wave of infected population. (C) a S-shaped curve of a recovered population.

<https://doi.org/10.1371/journal.pone.0265815.g001>

State [40], or city level [41], shows that the dynamics of COVID-19 is more complicated than a single epidemic wave. Therefore, it is essential to include the global dynamics of infection in a disease model.

The present work aims to address this issue. We extend the SIR model to a new model that includes the global impacts of the infection and is also capable of fitting well to infectious disease data. To incorporate the global effect and test the predictive accuracies of the extended model, we selected randomly six cities and states influenced by the COVID-19 global pandemic: Kansas City, Saint Louis, San Francisco, Missouri, Illinois, and Arizona.

We assumed three possibilities to consider in the SIR modeling of COVID-19: susceptible individuals from a local community can travel in and out of their community without any exposure to COVID-19, they can be exposed to COVID-19 while traveling and develop symptoms after they return to their community, or they can be diagnosed with COVID-19 during their traveling and return to their community after recovery. In the next section we include all these three possibilities in the extended SIR model.

The rest of the paper is organized as follows. In section. 2, we introduce the extended SIR model and its formulation. We explain the methodology that was used to fit the extended SIR model to COVID-19 data. In section. 3, we present the results of our analysis based on the six cities and states. In section. 4, we provide a discussion of the main results and additional factors to consider in the modeling process.

2- Materials and methods

2.1 Data

The COVID-19 data used in this study were obtained from the health department of Kansas City, Saint Louis, San Francisco, Missouri, Illinois, and Arizona [42–45]. The data were dated from March 10, 2020, to March 7, 2021 (a total of 363 days). We did not include data after March 7, 2021, because our model does not include the effects of vaccination and it would be inappropriate to include the data thereafter. Specifically, the data variables consisted of date, total number of cases, new cases, total deaths, new deaths, and total number of individuals tested for COVID-19.

We used abovementioned data to extract the daily number of recovered, susceptible, and infected individuals (see [S1 File](#) for the algorithms used for generating the data).

[Table 1](#) provides the basic descriptive statistics of Kansas City, Saint Louis, San Francisco, Missouri, Illinois, and Arizona of daily COVID-19 data of infected individuals.

Observe that the statistics of susceptible and recovered data are comparable (see [S1 File](#) for the basic descriptive statistics). However, the statistics of infected individuals are at a much

Table 1. Descriptive Statistics of Kansas City (KC), Saint Louis (SL), San Francisco (SF), Missouri (MO), Illinois (IL), and Arizona (AZ) daily COVID-19 data from March 10, 2020 to March 7, 2021.

	KC	SL	SF	MO	IL	AZ
Minimum	1	1	2	0	0	0
Maximum	4109	2201	4600	62456	168855	129240
Mean	1413.140	772.60	1305.50	18395.38	45387.31	31215.03
Median	1280	549	917	15793	27896	16242
Range	4108	2200	4598	62456	168855	129240
Standard Deviation	1137.49	565.87	1141.94	16777.20	42645.57	33702.66

<https://doi.org/10.1371/journal.pone.0265815.t001>

lower scale. Details of time series and spatial clusters of COVID-19 infection in Kansas City have been provided in [46, 47].

To estimate the number of susceptible individuals, we assumed an average incubation period of 5 days for COVID-19 [48]. We also considered one day for obtaining the COVID-19 test results. Hence, all of those who were tested positive were susceptible from the beginning until 6 days prior to obtaining the test results. Also, we added the individuals who take the test, but their results were negative. These individuals had presumably high risk of getting infected and therefore susceptible. The number of infected individuals were calculated by considering an average infection period 14 days [49]. Hence, we cumulatively added of new cases for 14 days until they recovered.

2.2 Model formulation

We divided our population of N individuals living in a local community into sub-populations (i.e., compartments) of susceptible compartment $S(t)$, infected compartment $I(t)$, and recovered compartment $R(t)$. As shown in Fig 2, the extended SIR model of COVID-19 transmission assumes three possibilities for susceptible individuals traveling outside of the community: They can return to the community without any exposure (the net rate is $f(t) = f_2(t) - f_1(t)$), they can be exposed COVID-19 and develop symptoms after returning to the community (the inflow rate of $g(t)$), or they can be tested positive during their trip and remain quarantined until fully recovered and thereafter return to the community (the inflow rate of $h(t)$). The extended SIR model is formulated by the following system of deterministic non-linear differential equations and Fig 2 gives the flow diagram of

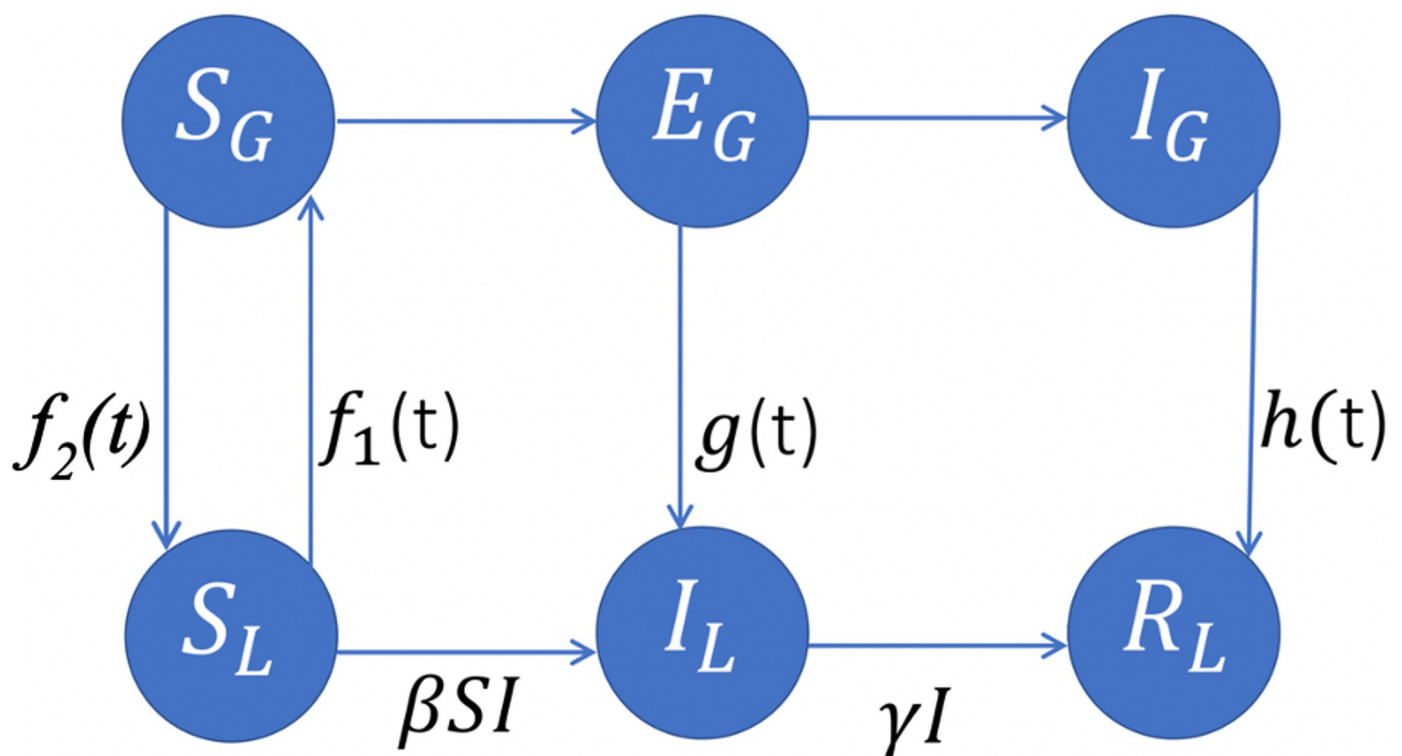


Fig 2. A schematic representation of the extended SIR model coupled with a global SEI model. L and G correspond to the number of individuals in the local and global communities, respectively.

<https://doi.org/10.1371/journal.pone.0265815.g002>

the model.

$$\begin{aligned}\frac{dS}{dt}(t) &= -\beta S(t)I(t) + f(t). \\ \frac{dI}{dt}(t) &= \beta S(t)I(t) - \gamma I(t) + g(t). \\ \frac{dR}{dt}(t) &= \gamma I(t) + h(t).\end{aligned}\tag{2}$$

where β and γ are the same parameters as in system (1). Functions $f(t)$, $g(t)$ and $h(t)$ are differentiable and bounded functions and take into account the global effects of the infection. To avoid overfitting, our goal is to estimate f , g and h using least complicated forms.

Although adding exposed population can provide interesting dynamics, we decided to exclude the exposed compartment from our modeling. This is due to lack of data associated with exposed population. Namely, there is no known method of accurately identify the time series of exposed population in a community. In addition, the more compartments are added the harder it becomes to accurately estimate the parameter values. In some cases, the confidence intervals of estimated parameter values become extremely large due to high number of parameters and insufficient amount of data. With this rationale in mind, we therefore employ a two-step method to estimate the parameters of the model.

Individuals can get infected both within and outside of the community. A standard SIR model has been considered for progression of infection within the community. The extended SIR model coupled with a global SEI model where S_G , E_G , and I_G are the global of susceptible, exposed, and infected compartments respectively, and S_L , I_L , and R_L are the susceptible, infected, and recovered compartments in the local community, respectively. The extended model assumes three possibilities for susceptible individuals traveling outside of the community: They can return to the community without any exposure (the net rate is $f(t) = f_2(t) - f_1(t)$), they can be exposed to the infection and develop symptoms after returning to the community (rate of $g(t)$), or they can be tested positive during their trip and remain quarantined until fully recovered and thereafter return to the community (rate of $h(t)$).

2.3 Model fitting

The single-step numerical methods such as linearization and discretization [50] to estimate the parameter values of model (2) fail to converge, due to high degrees of freedom and unknown intervals of parameter estimations. We therefore proposed a two-step process for parameters estimation of model (2). First, we estimated the parameter values of the standard SIR model (1) and then we determined the functions $f(t)$, $g(t)$ and $h(t)$ using the residual data of $S(t)$, $I(t)$ and $R(t)$ subpopulations. As mentioned before, we used the COVID-19 data of susceptible, infected, and recovered individuals in Kansas City, Saint Louis, San Francisco, Missouri, Illinois, and Arizona for an epidemic period starting from March 10, 2020, to March 7, 2021.

In the first step, we numerically solved the system (2) using the MATLAB ode45 solver which is based on the fourth order Runge-Kutta method. The stability of the method is well established in [51]. For data fitting, the optimization function “fmincon” was used along with the common technique of the least-squares method [52, 53]. This method minimizes the sum of the squared residuals, that is, the difference between model predictions and their corresponding data values. The sum of the squared residuals is calculated using the formula below

$$E = \frac{1}{M} \sum_{i=1}^M (y - y_i)^2\tag{3}$$

where M represents the total number of data points considered for fitting and y and y_i represent the values predicted by the model and those from the data, respectively.

We estimated the SIR parameter values by considering the following factors. Several studies indicated that the COVID-19 transmission rate of infection was 0.5 [54, 55]. Hence, we set $\beta = 0.5$. Also, some studies assumed the average recovery period (i.e $1/\gamma$) is about 7 days [54, 55], which results in the initial value of $\gamma = 0.13$. Also, to be consistent with the data, we set our initial conditions to the number of susceptible, infected and recovered at = 1. The estimated model parameters are provided in Table 2.

In the second step, we fitted the global effect of infection on the community by estimating functions $f(t)$, $g(t)$ and $h(t)$ in model (2).

Although model selection can be done using Akaike’s Information Criteria (AIC) and Bayesian Information Criteria (BIC) methods, we used MATLAB curve fitting toolbox to measure the goodness of fit (adjusted R^2 , sum of the squared residuals, etc.) to find the optimal forms of functions $f(t)$, $g(t)$ and $h(t)$. Specifically, the model fitting resulted in the following forms:

$$f(t) = \lambda_1 t + \lambda_2.$$

$$g(t) = a_1 b_1 \cos(b_1 T + c_1) + a_2 b_2 \cos(b_2 T + c_2) + a_3 b_3 \cos(b_3 T + c_3).$$

$$h(t) = p_1 t + p_2.$$

3- Results

Using the COVID-19 data of each city and states, we estimated the functions corresponding to the global effects $f(t)$, $g(t)$ and $h(t)$ in model (2) using the abovementioned two steps in Kansas City, Saint Louis, San Francisco, Missouri, Illinois, and Arizona. The estimated net rate for the susceptible individuals who can return to the community without any exposure is given by $f(t) = \lambda_1 t + \lambda_2$. The estimated net rate of the individuals who exposed to the infection and develop symptoms after returning to the community is given by $g(t) = a_1 b_1 \cos(b_1 T + c_1) + a_2 b_2 \cos(b_2 T + c_2) + a_3 b_3 \cos(b_3 T + c_3)$, and the estimated net rate of the individuals who tested positive during their trip and remain quarantined until fully recovered and thereafter return to the community is given by $h(t) = p_1 t + p_2$. All parameter estimations of step 1 and step 2 are summarized in Tables 2 and 3, respectively. From Table 2, we noticed that the transmission rate in the states; Missouri, Illinois, and Arizona are much higher than the cities; Kansas City, Saint Louis, and San Francisco, which is due to the larger population size of states and cities. Note that the estimated values for recovery rates are different than those of the standard SIR model because of the recovery rate is influenced by the global effects. In Table 3, there are three parameters, a , b and c where a represents the amplitude of each wave for each city or

Table 2. Estimated parameter values of model (2) based on data of Kansas City (KC), Saint Louis (SL), San Francisco (SF), Missouri (MO), Illinois (IL), and Arizona (AZ).

Parameter	Description	KC	SL	SF	MO	IL	AZ
λ_1	Linear recruitment rate	100	99.9997	71.42	-20	600	-20
λ_2	Constant recruitment rate	-0.2481	-0.3278	-0.0809	-1.32	8.62	-10
β	Transmission rate	2.9×10^{-16}	5.49×10^{-4}	4.9×10^{-4}	8.23	8.22	3.998
γ	Recovery rate*	7.3×10^{-15}	0.1275	1.97×10^{-9}	1.003	1.002	0.59
p1	Linear flow of recovered	0.633	0.288	0.553	10.2	23.92	14.92
p2	Constant flow of recovered	-8.01	2.876	-15.17	-477.6	-1190	-729.9

* The estimated values for the recovery rates are different than those of the standard SIR model.

<https://doi.org/10.1371/journal.pone.0265815.t002>

Table 3. Estimated parameter values of the global functions $f(t)$, $g(t)$ and $h(t)$ based on data of Kansas City (KC), Saint Louis (SL), San Francisco (SF), Missouri (MO), Illinois (IL), and Arizona (AZ).

Parameters	KC	SL	SF	MO	IL	AZ
a_1	2412	5763	1929	3.28×10^4	8.69×10^4	6.65×10^4
b_1	0.009	0.0149	0.005	0.009	0.0051	0.0142
c_1	2.548	1.75	2.94	2.43	2.794	2.008
$T1 = 2\pi/b1$	698	421.5	1256	698	1231.4	442.2
a_2	609.1	416.6	1.41×10^4	7048	3.85×10^4	3.007×10^5
b_2	0.046	0.05	0.037	0.048	0.031	0.027
c_2	-1.81	3.58	-4.875	3.88	-3.53	-2.812
$T2 = 2\pi/b2$	136.5	125.6	169.7	130.8	202.6	232.6
a_3	826.1	5072	1.42×10^4	1.3×10^4	2.03×10^4	2.7×10^5
b_3	0.025	0.017	0.037	0.025	0.048	0.029
c_3	3.64	-1.70	4.48	-2.4	-2.47	0.138
$T3 = 2\pi/b3$	251.2	369.4	169.7	251.2	130.8	216.5

Note: See the supplementary document for the goodness of fit (R^2).

<https://doi.org/10.1371/journal.pone.0265815.t003>

state, b represents the frequency of each wave, and c represents the phase shift of each wave. Also, the value of T represents the period of each wave. For instance, the state of Missouri and the city of Kansas City have almost the same of three periodic epidemic waves of COVID-19. One happens every two years, the other epidemic wave every four months and half, and the last wave every nine months. The periodic waves could be associated with several factors such

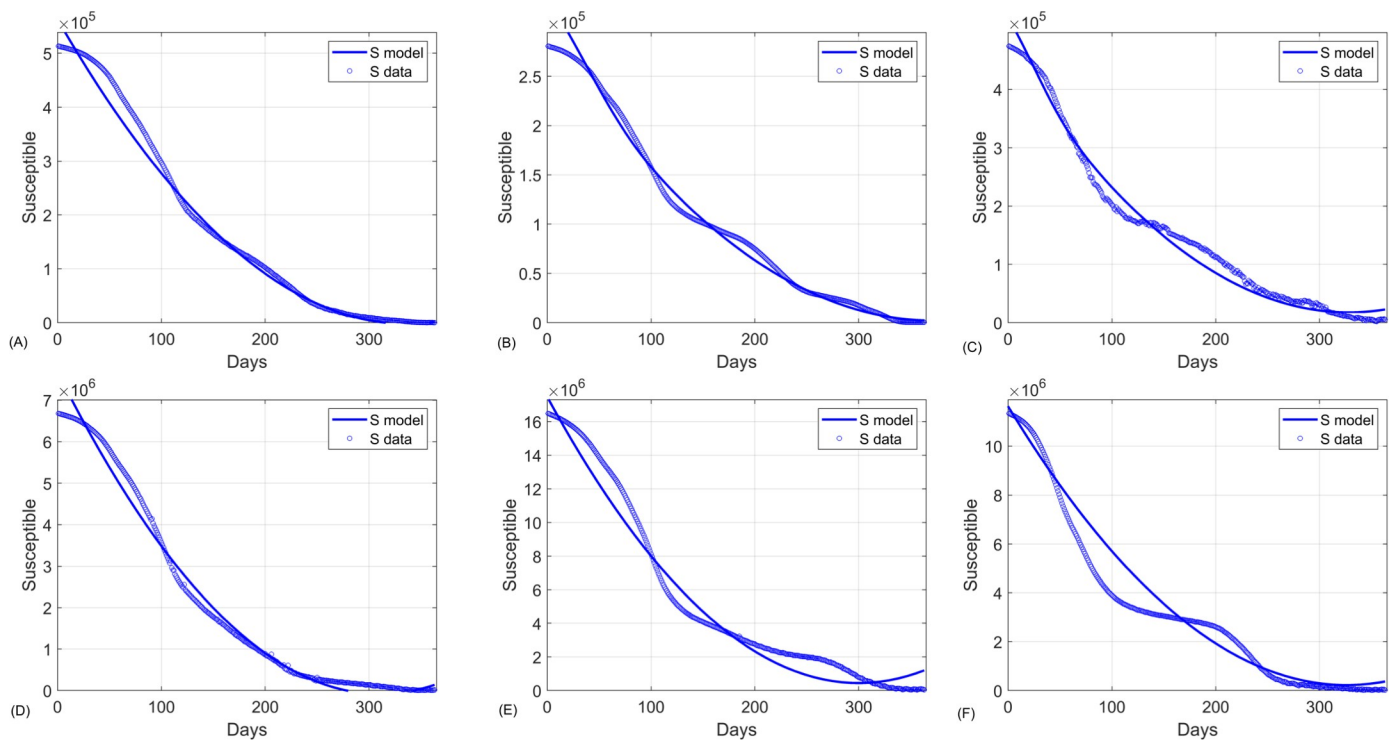


Fig 3. The extended SIR model fitted to the COVID-19 data of susceptible subpopulation in (A) Kansas City, (B) Saint Louis, (C) San Francisco, (D) Missouri, (E) Illinois, and (F) Arizona.

<https://doi.org/10.1371/journal.pone.0265815.g003>

as, traveling, major social events, infection prevention policies, changes to the coronavirus itself, and the increase of people who become susceptible because they have not developed some immunity. [56].

The panel of Fig 3A–3F shows that the extended SIR model of the susceptible solution curve fits well to the data of susceptible subpopulation in the cities and states. Likewise, S7 Table in S1 File shows the goodness of fit of susceptible subpopulation for Kansas City, Saint Louis, San Francisco, Missouri, Illinois, and Arizona data where $R^2 = 0.9909, 0.9911, 0.9802, 0.9878, 0.9761, 0.9583$, respectively. (See S7 Table in S1 File).

Similarly, the panel of Fig 4A–4F shows the extended SIR model in the subpopulations of recovered individuals fits well to the data and have a goodness of fit $R^2 = 0.9873, 0.9893, 0.9804, 0.9844, 0.9748, 0.9606$, respectively (See S7 Table in S1 File).

The panel of Fig 5A–5F shows that the dynamics of COVID-19 in the cities and states are more complicated than a single epidemic wave as common in the standard SIR model. The extended model is capable of revealing the underlying behaviors hidden in the data. Also, the extended SIR model in the subpopulations of infected individuals fits well to the data and have goodness of fit $R^2 = 0.9542, 0.9287, 0.962, 0.9438, 0.9015, 0.9492$ for Kansas City, Saint Louis, San Francisco, Missouri, Illinois, and Arizona data, respectively. (See S7 Table in S1 File). Therefore, the inclusion of global effects to the SIR model can substantially improve the predictive accuracy of the model.

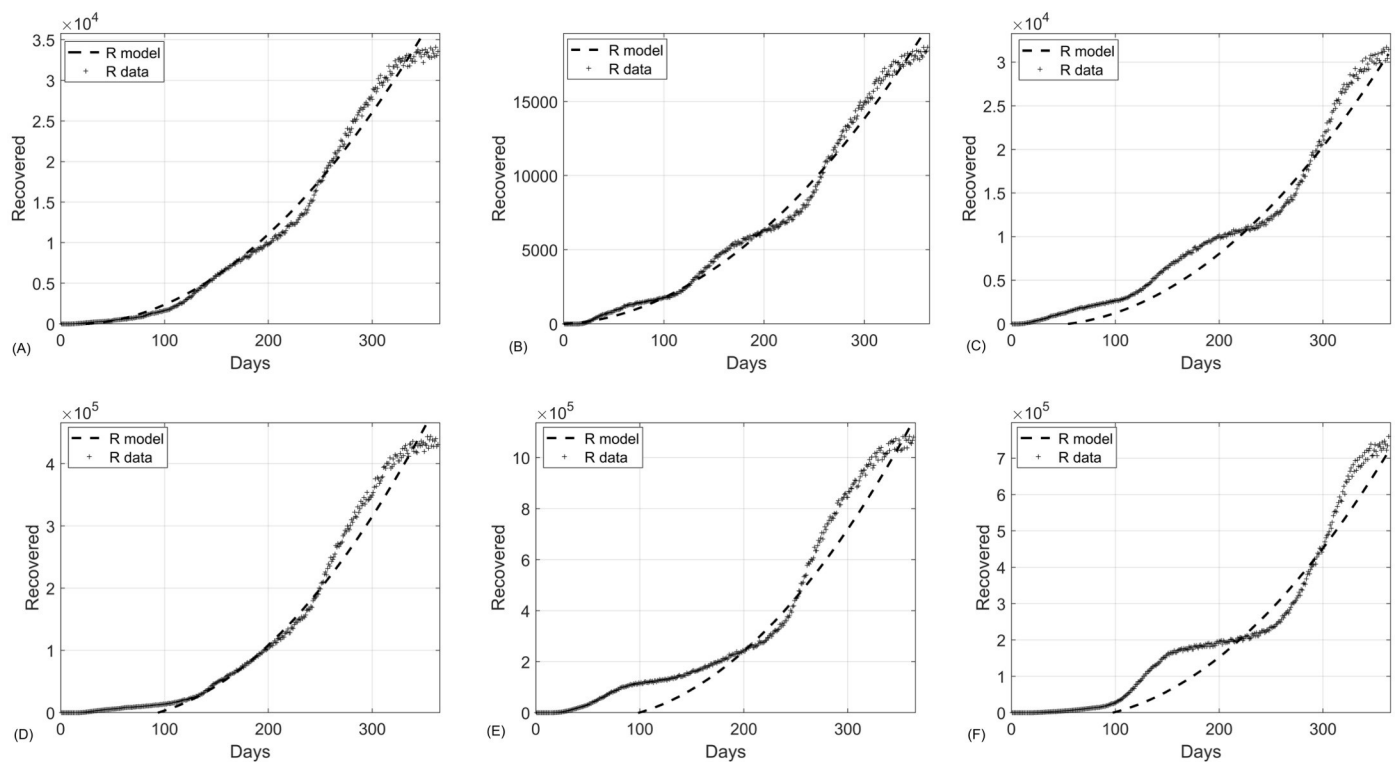


Fig 4. The extended SIR model fitted to the COVID-19 data of recovered subpopulation in (A) Kansas City, (B) Saint Louis, (C) San Francisco, (D) Missouri, (E) Illinois, and (F) Arizona.

<https://doi.org/10.1371/journal.pone.0265815.g004>

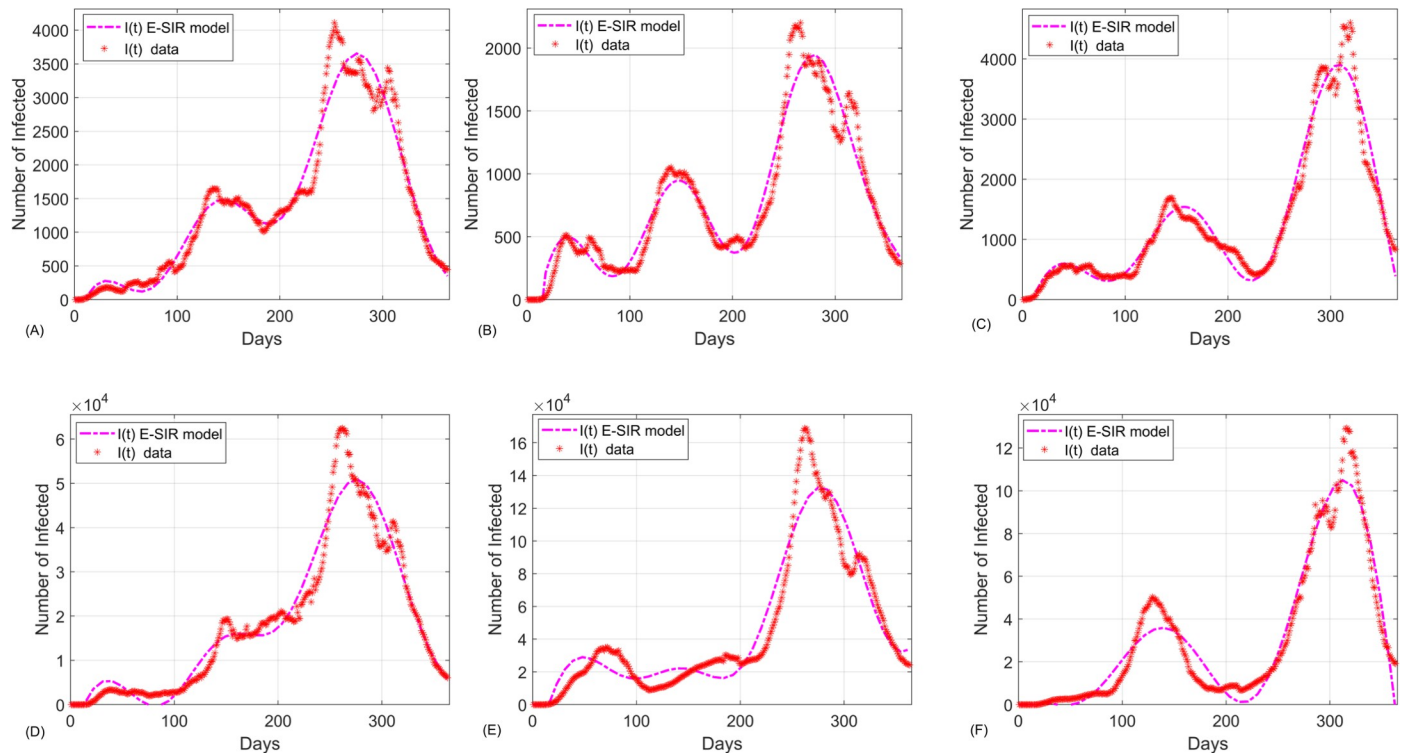


Fig 5. The extended SIR model fitted to the COVID-19 data of infected subpopulation in (A) Kansas City, (B) Saint Louis, (C) San Francisco, (D) Missouri, (E) Illinois, and (F) Arizona.

<https://doi.org/10.1371/journal.pone.0265815.g005>

4- Discussion

The present work highlights the importance of including global dynamics of infection in disease models to achieve higher predictive accuracies. We introduced a two-step algorithm for accurate estimation of infection parameters by considering both global and local effects of the infection spread in SIR models. The first step leads to estimation of local parameters (i.e., the transmission and recovery rates, β and, respectively) whereas the second step incorporates the global effects of the infection (i.e., estimation of functions $f(t)$, $g(t)$ and $h(t)$). To test the methodology, we applied the two-step model fitting algorithm to the extended SIR model (2) using Kansas City, Saint Louis, San Francisco, Missouri, Illinois, and Arizona data from March 10, 2020, to March 7, 2021. As shown in the panels of Figs 3–5, the two-step method resulted in solution curves that fit well to the COVID-19 data. The goodness of fit becomes more apparent when it is compared to that of the standard SIR model. Therefore, we compared the standard SIR model with the extended SIR model using the first 212 Kansas City COVID-19 data. As shown in Fig 6A and 6B, the solution curves of the standard SIR model poorly fit to the COVID-19. Moreover, Table 4 shows the comparisons of model fitness for the standard and extended SIR model. The extended SIR model of the susceptible solution curve has $R^2 = 0.9905$ while in standard model $R^2 = 0.1551$. Similarly the extended SIR model outperformed the standard SIR model in the subpopulations of recovered individuals ($R^2 = 0.9912$ versus $R^2 = 0.47$), and the subpopulation of infected individuals ($R^2 = 0.7083$ versus $R^2 = -258.65$). Note that the negative R^2 value is because the classical SIR model does not follow the trend of the data.

In addition to higher predictive accuracies of the extended SIR model (2), the solution curves revealed oscillatory behaviors with an increasing trend of infected individuals. This

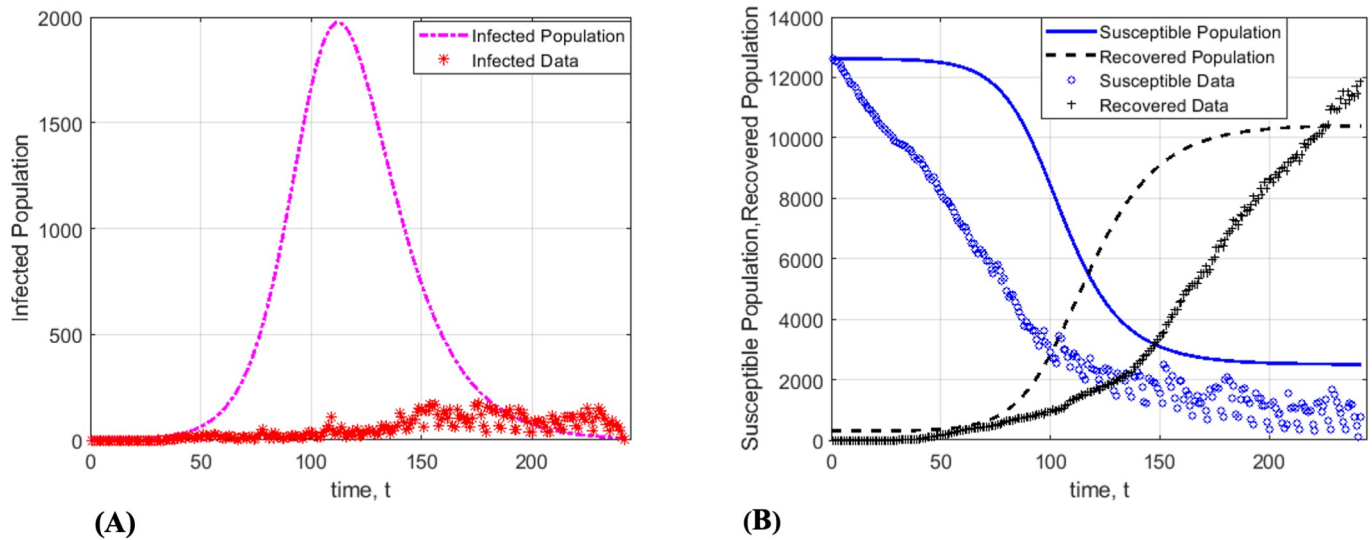


Fig 6. The standard SIR model fitted to fitted Kansas City COVID-19 data. (A) The standard SIR model poorly fits to the susceptible and recovered data. (B) The standard SIR model has poor fitness to data of infected population.

<https://doi.org/10.1371/journal.pone.0265815.g006>

contrasts with the standard SIR model, where regardless of chosen parameter values, the solution curves always exhibit the same qualitative behaviors (see Fig 1A–1C).

Although the standard SIR has been proven useful to study local dynamics of various infections, it fails to capture the global effects of a widespread disease. The failure of standard SIR model to forecast the COVID-19 pandemic can be described by a variety of factors. One of these factors is that the standard SIR model assumes the population is closed, isolated, and ignores the effects of the global dynamics of infection on neighboring communities which is not a valid assumption [35]. Hence, by including the global infection effects in the disease models, we can identify underlying mechanisms governing the dynamics of infectious diseases.

In spite of several studies indicate that include temperature factor in the SIR model could potentially improve the model outcomes [57], recent studies suggested that the temperature has no influence on the propagation of the COVID-19 virus [58, 59]. In fact, some strains of the virus alter depending on their environments. They may live and grow in a variety of geographical areas or temperatures. Outside of laboratory tests, there is no way to anticipate how the virus would react in heat and humidity or even cold and dry temperatures.

The inability of the standard SIR model to fit their COVID-19 data has been identified by other researchers [60]. Nonetheless, the presence of a breakpoint due to strong policy

Table 4. Comparisons of model goodness of fitness for the standard and extended SIR models.

Fitness	Extended	Standard
Corrected AIC	3.2799e+03	4.3130e+03
AIC	3.0351e+03	4.0689e+03
SSR (R^2) for S	3.22e+09 (0.9905)	5.0377e+08 (0.1551)
SSR (R^2) for I	4.1249e+05 (0.7083)	-1.5064e+08 (-258.65)*
SSR (R^2) for R	3.51e+09 (0.9912)	1.6640e+09 (0.4700)

* The standard SIR model has an extremely poor fitting with respect to the infected individuals.

<https://doi.org/10.1371/journal.pone.0265815.t004>

interventions, mentioned in [60], does not necessarily reduce the prevalence of infection in a community dealing with a pandemic.

In conclusion, including the global dynamics of the infection and applying the two-step model fitting algorithm can enable us to extract vital information (e.g., presence of epidemic waves) from the data.

Supporting information

S1 File. Supplementary document.
(DOCX)

Acknowledgments

We would like to thank the reviewers for their thoughtful comments and efforts towards improving our manuscript.

Author Contributions

Conceptualization: Hadeel AlQadi, Majid Bani-Yaghoub.

Data curation: Hadeel AlQadi.

Formal analysis: Hadeel AlQadi.

Investigation: Hadeel AlQadi.

Methodology: Hadeel AlQadi, Majid Bani-Yaghoub.

Project administration: Hadeel AlQadi.

Software: Hadeel AlQadi, Majid Bani-Yaghoub.

Supervision: Hadeel AlQadi, Majid Bani-Yaghoub.

Validation: Majid Bani-Yaghoub.

Visualization: Hadeel AlQadi, Majid Bani-Yaghoub.

Writing – original draft: Hadeel AlQadi.

Writing – review & editing: Hadeel AlQadi, Majid Bani-Yaghoub.

References

1. Ourworldindata. Coronavirus Data [Internet]. Available from: <https://ourworldindata.org/coronavirus>.
2. Centers for Disease Control and Prevention. COVID-19 data Tracker [Internet]. Available from: https://covid.cdc.gov/covid-data-tracker/#cases_casesper100klast7days
3. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet* [Internet]. 2020; 395(10242):1973–87. Available from: <http://www.sciencedirect.com/science/article/pii/S0140673620311429>.
4. Washingtonpost. Markets live updates coronavirus economy [Internet]. Available from: <https://www.washingtonpost.com/business/2020/03/12/markets-live-updates-coronavirus-economy/>
5. 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. *New England Journal of Medicine*. 2020; 382(13):1199–207. <https://doi.org/10.1056/NEJMoa2001316> PMID: 31995857
6. Nwankwo A, Okuonghae D. A Mathematical Model for the Population Dynamics of Malaria with a Temperature Dependent Control. *Differential Equations and Dynamical Systems* [Internet]. 2019; Available from: <https://doi.org/10.1007/s12591-019-00466-y>.

7. Keeling MJ, Danon L. Mathematical modelling of infectious diseases. *British Medical Bulletin* [Internet]. 2009 Dec 1; 92(1):33–42. Available from: <https://doi.org/10.1093/bmb/ldp038> PMID: 19855103
8. Huppert A, Katriel G. Mathematical modelling and prediction in infectious disease epidemiology. *Clinical Microbiology and Infection* [Internet]. 2013; 19(11):999–1005. Available from: <http://www.sciencedirect.com/science/article/pii/S1198743X14630019>. <https://doi.org/10.1111/1469-0691.12308> PMID: 24266045
9. Roberts MG, Heesterbeek J a P. Mathematical models in epidemiology. *Mathematical Models*. 2003; 49(0):6221.
10. Siettos CI, Russo L. Mathematical modeling of infectious disease dynamics. *Virulence* [Internet]. 2013 May 15; 4(4):295–306. Available from: <https://doi.org/10.4161/viru.24041> PMID: 23552814
11. Kretzschmar M, Wallinga J. *Mathematical Models in Infectious Disease Epidemiology*. Krämer A, Kretzschmar M, Krickeberg K, editors. *Modern Infectious Disease Epidemiology: Concepts, Methods, Mathematical Models, and Public Health* [Internet]. 2009 Jul 28;209–21. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7178885/>.
12. Brauer F. Mathematical epidemiology: Past, present, and future. *Infectious Disease Modelling* [Internet]. 2017; 2(2):113–27. Available from: <http://www.sciencedirect.com/science/article/pii/S2468042716300367>. <https://doi.org/10.1016/j.idm.2017.02.001> PMID: 29928732
13. Panovska-Griffiths J. Can mathematical modelling solve the current Covid-19 crisis? *BMC Public Health* [Internet]. 2020; 20(1):551. Available from: <https://doi.org/10.1186/s12889-020-08671-z> PMID: 32331516
14. Roddam AW. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*: O Diekmann and JAP Heesterbeek, 2000, Chichester: John Wiley pp. 303, £39.95. ISBN 0-471-49241-8. *International Journal of Epidemiology* [Internet]. 2001 Feb 1;30(1):186. Available from: <https://doi.org/10.1093/ije/30.1.186>.
15. Edoh K, MacCarthy E. Network and equation-based models in epidemiology. *International Journal of Biomathematics*. 2018 Mar 19;11.
16. Hejblum G, Setbon M, Temime L, Lesieur S, Valleron A-J. Modelers' Perception of Mathematical Modeling in Epidemiology: A Web-Based Survey. *PLOS ONE* [Internet]. 2011 Jan 31; 6(1):e16531. Available from: <https://doi.org/10.1371/journal.pone.0016531> PMID: 21304976
17. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics—I. *Bulletin of Mathematical Biology* [Internet]. 1991; 53(1):33–55. Available from: <https://doi.org/10.1007/BF02464423>.
18. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics—II. The problem of endemicity. *Bulletin of Mathematical Biology* [Internet]. 1991; 53(1):57–87. Available from: <https://doi.org/10.1007/BF02464424>
19. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics—III. Further studies of the problem of endemicity. *Bulletin of Mathematical Biology* [Internet]. 1991; 53(1):89–118. Available from: <https://doi.org/10.1007/BF02464425>.
20. Chen D. Modeling the Spread of Infectious Diseases: A Review. In 2014. p. 19–42.
21. Shrestha S, Lloyd-Smith J. Introduction to mathematical modeling of infectious diseases. 2010;0000:1–46.
22. Harko T, Lobo FSN, Mak MK. Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Applied Mathematics and Computation* [Internet]. 2014; 236:184–94. Available from: <http://www.sciencedirect.com/science/article/pii/S009630031400383X>.
23. Kim S, Lee J, Jung E. Mathematical model of transmission dynamics and optimal control strategies for 2009 A/H1N1 influenza in the Republic of Korea. *Journal of Theoretical Biology* [Internet]. 2017; 412:74–85. Available from: <http://www.sciencedirect.com/science/article/pii/S0022519316303228>. <https://doi.org/10.1016/j.jtbi.2016.09.025> PMID: 27769686
24. Del Valle S, Hethcote H, Hyman JM, Castillo-Chavez C. Effects of behavioral changes in a smallpox attack model. *Mathematical Biosciences* [Internet]. 2005; 195(2):228–51. Available from: <http://www.sciencedirect.com/science/article/pii/S0025556405000593>. <https://doi.org/10.1016/j.mbs.2005.03.006> PMID: 15913667
25. Gai C, Iron D, Kolokolnikov T. Localized outbreaks in an S-I-R model with diffusion. *Journal of Mathematical Biology* [Internet]. 2020; 80(5):1389–411. Available from: <https://doi.org/10.1007/s00285-020-01466-1> PMID: 31950259
26. GAUTAM R, LAHODNY G, BANI-YAGHOUB M, MORLEY PS, IVANEK R. Understanding the role of cleaning in the control of Salmonella Typhimurium in grower-finisher pigs: a modelling approach. *Epidemiology and Infection* [Internet]. 2013/08/07. 2014; 142(5):1034–49. Available from: [https://www.cambridge.org/core/article/understanding-the-role-of-cleaning-in-the-control-of-salmonella-](https://www.cambridge.org/core/article/understanding-the-role-of-cleaning-in-the-control-of-salmonella)

- typhimurium-in-growerfinisher-pigs-a-modelling-approach/0583A9978A127B4F47482EE6A8B048DF. <https://doi.org/10.1017/S0950268813001805> PMID: 23920341
27. Gautam R, Bani-Yaghoub M, Neill WH, Döpfer D, Kaspar C, Ivanek R. Modeling the effect of seasonal variation in ambient temperature on the transmission dynamics of a pathogen with a free-living stage: example of *Escherichia coli* O157:H7 in a dairy herd. *Preventive veterinary medicine* [Internet]. 2011; 102(1):10–21. Available from: <http://europepmc.org/abstract/MED/21764472>. <https://doi.org/10.1016/j.prevetmed.2011.06.008> PMID: 21764472
 28. Zhang S-P, Yang Y-R, Zhou Y-H. Traveling waves in a delayed SIR model with nonlocal dispersal and nonlinear incidence. *Journal of Mathematical Physics* [Internet]. 2018 Jan 1; 59:11513. Available from: <https://ui.adsabs.harvard.edu/abs/2018JMP....59a1513Z>.
 29. Bani-Yaghoub M, Gautam R, Shuai Z, van den Driessche P, Ivanek R. Reproduction numbers for infections with free-living pathogens growing in the environment. *Journal of Biological Dynamics* [Internet]. 2012 Mar 1; 6(2):923–40. Available from: <https://doi.org/10.1080/17513758.2012.693206> PMID: 22881277
 30. BANI-YAGHOUB M, GAUTAM R, DÖPFER D, KASPAR CW, IVANEK R. Effectiveness of environmental decontamination as an infection control measure. *Epidemiology and Infection* [Internet]. 2011/05/18. 2012; 140(3):542–53. Available from: <https://www.cambridge.org/core/article/effectiveness-of-environmental-decontamination-as-an-infection-control-measure/FA961694D9AAE4773379262975087191>. <https://doi.org/10.1017/S0950268811000604> PMID: 21676360
 31. Ivorra B, Ferrández MR, Vela-Pérez M, Ramos AM. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Communications in Nonlinear Science and Numerical Simulation* [Internet]. 2020; 88:105303. Available from: <http://www.sciencedirect.com/science/article/pii/S1007570420301350>. <https://doi.org/10.1016/j.cnsns.2020.105303> PMID: 32355435
 32. Kim S, Bin Seo Y, Jung E. Prediction of COVID-19 transmission dynamics using a mathematical model considering behavior changes in Korea. *Epidemiol Health* [Internet]. 2020 Apr 13; 42(0):e2020026–0. Available from: <https://doi.org/10.4178/epih.e2020026> PMID: 32375455
 33. Tomochi M, Kono M. A mathematical model for COVID-19 pandemic—SIIR model: Effects of asymptomatic individuals. *Journal of General and Family Medicine* [Internet]. 2020 Nov 1;n/a(n/a). Available from: <https://doi.org/10.1002/jgf2.382> PMID: 33457150
 34. Alshammari FS. A Mathematical Model to Investigate the Transmission of COVID-19 in the Kingdom of Saudi Arabia. Yetilmezsoy K, editor. *Computational and Mathematical Methods in Medicine* [Internet]. 2020; 2020:9136157. Available from: <https://doi.org/10.1155/2020/9136157> PMID: 33062043
 35. Moein S, Nickaeen N, Roointan A, Borhani N, Heidary Z, Javanmard SH, et al. Inefficiency of SIR models in forecasting COVID-19 epidemic: a case study of Isfahan. *Scientific Reports* [Internet]. 2021; 11(1):4725. Available from: <https://doi.org/10.1038/s41598-021-84055-6> PMID: 33633275
 36. Li W-T, Lin G, Ma C, Yang F-Y. Traveling wave solutions of a nonlocal delayed SIR model without outbreak threshold. *Discrete & Continuous Dynamical Systems—B* [Internet]. 19(2):467–84. Available from: <http://aimsciences.org/article/id/d99e6db9-456f-4a01-add9-e0b99b823bb4>.
 37. Köhler-Rieper F, Röhl CHF, De Micheli E. A novel deterministic forecast model for the Covid-19 epidemic based on a single ordinary integro-differential equation. *The European Physical Journal Plus* [Internet]. 2020; 135(7):599. Available from: <https://doi.org/10.1140/epjp/s13360-020-00608-0> PMID: 32834915
 38. Sirijampa A, Chinviriyasit S, Chinviriyasit W. Hopf bifurcation analysis of a delayed SEIR epidemic model with infectious force in latent and infected period. *Advances in Difference Equations* [Internet]. 2018; 2018(1):348. Available from: <https://doi.org/10.1186/s13662-018-1805-6> PMID: 32226452
 39. Oluyori DA. Backward and Hopf bifurcation analysis of an SEIRS COVID-19 epidemic model with saturated incidence and saturated treatment response. 2020;1–26. Available from: <https://www.medrxiv.org/content/10.1101/2020.08.28.20183723v1.full.pdf>.
 40. Wu SL, Mertens AN, Crider YS, Nguyen A, Pokpongkiat NN, Djajadi S, et al. Substantial underestimation of SARS-CoV-2 infection in the United States. *Nature Communications* [Internet]. 2020; 11(1):4507. Available from: <https://doi.org/10.1038/s41467-020-18272-4> PMID: 32908126
 41. MARC. Kansas City Region COVID-19 Data Hub [Internet]. Available from: <https://marc2.org/covidhub/>
 42. City of Kansas City Health Department Data. KCMO COVID-19 Case, Death and Trends Data [Internet]. Available from: <https://data.kcmo.org/Health/COVID-19-Case-Death-Trends-by-Date/nfta-sjx6>.
 43. Saint Louis Health Department. Saint Louis COVID-19 Data [Internet]. [cited 2021 Nov 1]. Available from: <https://www.stlouis-mo.gov/covid-19/data/index.cfm>.

44. San Francisco Department Health. San Francisco COVID-19 data [Internet]. [cited 2021 Nov 1]. Available from: <https://sf.gov/resource/2021/covid-19-data-and-reports>.
45. The Atlantic. The COVID Tracking Project [Internet]. [cited 2021 Nov 1]. Available from: <https://covidtracking.com/data>.
46. AlQadi H, Bani-Yaghoub M, Balakumar S, Wu S, Francisco A. Assessment of Retrospective COVID-19 Spatial Clusters with Respect to Demographic Factors: Case Study of Kansas City, Missouri, United States. Vol. 18, *International Journal of Environmental Research and Public Health*. 2021. <https://doi.org/10.3390/ijerph182111496> PMID: 34770012
47. AlQadi H, Bani Yaghoub M, Wu S, Balakumar S, Francisco A. Prospective Spatial–Temporal Clusters of COVID-19 in Local Communities: Case Study of Kansas City, Missouri, United States. *Epidemiology and Infection* [Internet]. 2022 Mar 9 [cited 2022 Mar 12];1–24. Available from: https://www.cambridge.org/core/product/identifier/S0950268822000462/type/journal_article. <https://doi.org/10.1017/S0950268822000462> PMID: 35260205
48. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of Autoimmunity* [Internet]. 2020; 109:102433. Available from: <http://www.sciencedirect.com/science/article/pii/S0896841120300469>. <https://doi.org/10.1016/j.jaut.2020.102433> PMID: 32113704
49. Qin J, You C, Lin Q, Hu T, Yu S, Zhou X-H. Estimation of incubation period distribution of COVID-19 using disease onset forward time: A novel cross-sectional and forward follow-up study. *Science Advances* [Internet]. 2020 Aug 1; 6(33):eabc1202. Available from: <http://advances.sciencemag.org/content/6/33/eabc1202.abstract>. <https://doi.org/10.1126/sciadv.abc1202> PMID: 32851189
50. 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)* [Internet]. 2007 Nov 1; 69(5):741–96. Available from: <https://doi.org/10.1111/j.1467-9868.2007.00610.x>.
51. May R, Noye J. The Numerical Solution of Ordinary Differential Equations: Initial Value Problems. In: Noye J, editor. *Computational Techniques for Differential Equations* [Internet]. North-Holland; 1984. p. 1–94. (North-Holland Mathematics Studies; vol. 83). Available from: <http://www.sciencedirect.com/science/article/pii/S0304020808712003>.
52. Byrd RH, Hribar ME, Nocedal J. An Interior Point Algorithm for Large-Scale Nonlinear Programming. *SIAM Journal on Optimization* [Internet]. 1999 Jan 1; 9(4):877–900. Available from: <https://doi.org/10.1137/S1052623497325107>.
53. Waltz RA, Morales JL, Nocedal J, Orban D. An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Mathematical Programming* [Internet]. 2006; 107(3):391–408. Available from: <https://doi.org/10.1007/s10107-004-0560-5>.
54. Cauchemez S, Fraser C, Van Kerkhove MD, Donnelly CA, Riley S, Rambaut A, et al. Middle East respiratory syndrome coronavirus: quantification of the extent of the epidemic, surveillance biases, and transmissibility. *The Lancet Infectious Diseases* [Internet]. 2014; 14(1):50–6. Available from: <http://www.sciencedirect.com/science/article/pii/S1473309913703049>. [https://doi.org/10.1016/S1473-3099\(13\)70304-9](https://doi.org/10.1016/S1473-3099(13)70304-9) PMID: 24239323
55. Chen T-M, Rui J, Wang Q-P, Zhao Z-Y, Cui J-A, Yin L. A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. *Infectious Diseases of Poverty* [Internet]. 2020; 9(1):24. Available from: <https://doi.org/10.1186/s40249-020-00640-3> PMID: 32111262
56. Maragakis L. Coronavirus Second Wave, Third Wave and Beyond: What Causes a COVID Surge. Johns Hopkins; 2021.
57. Huang D, Tao H, Wu Q, Huang S-Y, Xiao Y. Modeling of the Long-Term Epidemic Dynamics of COVID-19 in the United States. Vol. 18, *International Journal of Environmental Research and Public Health*. 2021. <https://doi.org/10.3390/ijerph18147594> PMID: 34300045
58. Jamil T, Alam I, Gojabori T, Duarte CM. No Evidence for Temperature-Dependence of the COVID-19 Epidemic [Internet]. Vol. 8, *Frontiers in Public Health*. 2020. p. 436. Available from: <https://www.frontiersin.org/article/10.3389/fpubh.2020.00436>. <https://doi.org/10.3389/fpubh.2020.00436> PMID: 32984240
59. Kassem AZE. Does Temperature Affect COVID-19 Transmission? [Internet]. Vol. 8, *Frontiers in Public Health*. 2020. p. 934. Available from: <https://www.frontiersin.org/article/10.3389/fpubh.2020.554964>. <https://doi.org/10.3389/fpubh.2020.554964> PMID: 33425828
60. Kim YJ, Seo MH, Yeom HE. Estimating a breakpoint in the pattern of spread of COVID-19 in South Korea. *International Journal of Infectious Diseases*. 2020 Aug 1; 97:360–4. <https://doi.org/10.1016/j.ijid.2020.06.055> PMID: 32569839