Challenges in Modeling of an Outbreak's Prediction, Forecasting and Decision Making for Policy Makers

Altaf H. Khan

Abstract In this work an attempt has been made to review the current state of arts in epidemiological modeling, assessment of predictive models as well as forecasting of new pathogen. The primary concern is the containment of the outbreak from wide spread of the disease among the whole population. This article also focuses for the development of management tools and techniques in decision making for policy makers that are based on scientific evidence. Moreover, the identification, detection and reporting for outbreak of an infectious disease particularly a new pathogen in timely manner is quite challenging and tedious. Apparently understanding and reporting of such events are commonly rely on statistical and mathematical tools and both these approaches commonly depend upon a priory estimates as well as some reliable data. For example: statistical models requires a sizable number of events to develop predictive models, which is impossible at the outset of an outbreak of the disease to collate enough number of samples. Whereas, the mathematical models are reliable as well as have better predictive behavior, but they also require better initial guess apart from some rigid constraints to fully satisfy the model's assumptions. Apart from these issues, the other important features to study in epidemiology of the disease is how fast and quickly the scientific community promptly can pinpoint and able to address any causal factor which may suffice to account for the magnitude and severity of the epidemics of new pathogen that may have been taken place to any geographic locations. Hence in this work, first of all the SIR model (susceptible: S, infected: I, and recovered: R) will be outlined, as it is the most commonly used model in epidemiology of infectious diseases. Moreover, the applicability and utilization of R_0 in public health domain especially adaptive policy with management tools will be developed for the healthcare workers as well as the higher management of healthcare facility.

Keywords Covid-19 · Compartmental model · Testing · Outbreak mitigation

A. H. Khan (✉)

Department of Biostatistics and Bioinformatics, King Abdullah International Medical Research Center, National Guard Health Affairs, Riyadh, Kingdom of Saudi Arabia

1   Introduction

The infectious diseases related acute respiratory infections (ARIs) are the primary source of morbidity and mortality around the world, since each and every year approximately 5 to 6 million people die due to ARIs, and 98% of these deaths are due to lower respiratory tract infections see: Fig. 1 (the percentage of fatality rate on log-scale for some influenza types are shown, Source: New York Times). Mortality rates are higher among infants, children, and the elderly, and it's commonly prevalent in low-income and middle-income countries [6–8]. Influenza may also cause of an increase risk of strokes and heart attacks apart from disease complications [9–11]. The emergence of the current outbreak of the new pathogen started from Wuhan City, Hubei Province, China. This new outbreak cases were reported in the *I*st week of December, 2019, and the World Health Organization (WHO) on December 31, 2019 had declared this new pathogen as a 'Global-Pandemic'. It was soon identified as a novel coronavirus and termed as Covid-19, and belongs to the family of viruses that include the common cold and viruses such as SARS and MERS. On January 20, 2020, this has been also confirmed that the coronavirus can be transmitted between humans, and has wider risk of spread globally. Hence, the identification, detection and reporting for outbreak of an infectious disease particularly a new pathogen in timely manner is vital for the safety of general public health. Apparently reporting of such events are quite challenging and difficult, since it requires a complete understanding of the new pathogen. And to comprehend and stop from spreading any new such vicious viruses (like Covid-19) commonly rely on statistical and mathematical tools. Both these approaches commonly depend upon a priory estimates and some reliable data. For example: statistical models requires a sizable number of events to develop predictive models, which is impossible at the outset of an outbreak of the disease to collate enough number of samples. Whereas, the mathematical models are reliable as well as have better predictive behavior, but they also require better initial guess apart from some rigid constraints to fully satisfy the model's assumptions. Apart from these issues, the other important features to study in epidemiology of the disease is how fast and quickly the scientific community can pinpoint any causal factor which may suffice to account for the magnitude and severity of the epidemics of new pathogen that may have been taken place to any geographic locations. Moreover, some commonly utilized modeling tools in infectious diseases will be outlined here, whereas taking into account the primary aim of infectious disease modeling is: (i) to understand the mechanisms of spread, (ii) to estimate the time period of the latent and infectious periods, and (iii) the size of the epidemic, and the main focus is to determine strategies for disease control. Some of the commonly utilized approaches to model epidemic diseases are briefly outline here:
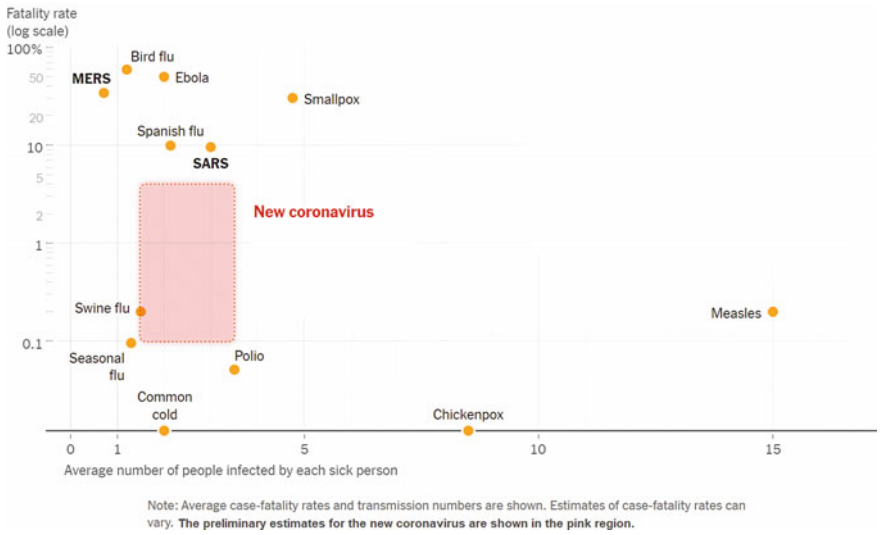
Fig. 1 The figure describes the percentage of different respiratory syndromes,the percentage of fatality rate is log scale: (*Source* New York Times)

1.1 Deterministic Compartmental Models (DCMs)

The DCMs based models are based on systems of differential equations which take into account the movement of the population through discrete states, including entry into and exit from the population, at specified rates. DCMs models are the most commonly used in the field of mathematical epidemiology, and can be solved analytically, or using numerical analysis. They can represent discrete forms of heterogeneity in the population. With DCMs, once the structure and parameters have been set explicitly, then there is no variation in model outcomes. In 1776, Daniel Bernoulli developed a model to analyze the life expectancy and death rates based on the inoculation or variolation in a public health enivironment, see: Dietz (2000) [37]. Some other scientists, for example namely: Philip-Charles Alexandre Louis, William Farr, Ronald Ross had made tremendous work in epidemiological sciences. Lately, epidemiologist now applying and utilizing new computational algorithms to analyze the infectious diseases based on modeling, and simulation of the dynamics of disease generation and propagation, see: Koopman (1996) [38].

1.2 Stochastic Individual Contact Models (ICMs)

The stochastic individual contact models (ICMs), also known as individual-based or agent based models, explicitly represent individual units in the population and the

Fig. 2 Visual display of a
two-dimensional
agent-based model. Each
square represents an
individually programmable,
mobile agent. Color-coding
allows easy visual tracking
of agents with different
properties. *Source* https://
www.ncbi.nlm.nih.gov/
books/NBK221490/figure/
mmm00027/?
report=objectonly
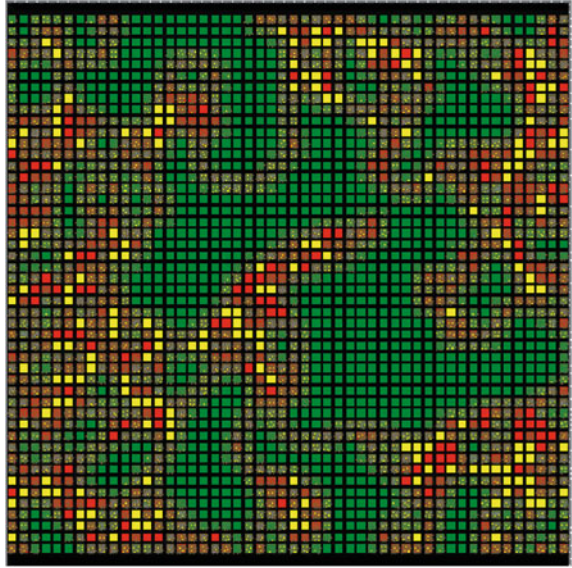
contacts between them are unique with discrete events. In contrast to DCMs, they
allow heterogeneity while specifying the contact process and other epidemiologi-
cally relevant events, and their stochasticity provides information on the range of
plausible outcomes resulting from a given set of parameters. The setbacks with these
models are they may require large amounts of input data that is needed for param-
eterization as well as the computational burden associated with running multiple
stochastic simulations. Agent-based modeling are used extensively in biology such
as: spread of epidemics, population dynamics, stochastic gene expression, plant-
animal interactions, vegetation ecology, as well as modeling 3D breast tissue forma-
tion/morphogenesis etc. They are agent-based computational models using computer
programs in which a population of individual entities is created, and each individual
is endowed with simple rules for interactions with the environment and with other
individuals, see: Holland (1995) [41]. They are used to model all manner of complex
scientific phenomena. Some important studies had used the agent-based modeling to
examine infectious diseases (e.g., influenza) and the immune response, see: Hofmeyr
and Forrest (2000), [40]. For example, the Swarm Development Group (ref: http://
www.swarm.org/wiki/Main_Page) had made the development of a wide variety of
infectious disease modeling using the agent-based modeling, see: Figure 2 display
of two-dimensional agent based model.

1.3 Network Models

Network models are also stochastic and represent individual units, but unlike ICMs, they providea flexible framework for representing repeated contacts with the same person or persons over time. These repeated contacts may give rise to persistent network configurations for example: pairs, triples, and larger connected components which in turn establishes the temporally ordered pathways for infectious disease transmission across a population. The R-package 'EpiModel' is a good tools for simulation of models for network analysis. It provides a generalized framework for both estimation and simulation of dynamically evolving networks. Network models provide the most accurate control over the contact process, but have greater computational burden than ICMs, both because they require statistical estimation of the network model parameters. Network models offer a versatile means of capturing heterogeneity in populations during an epidemic. In this approach, highly connected individuals tend to be infected at a higher rate early during an outbreak than those with fewer connections, see: Romanescu and Deardon [39]. Figure 3 shows the representation of evolving bit strings in a fitness landscape, and its clear that network models inspired by the Internet will productively inform the modeling of microbial pathogen networks, Albert et al. (2000) [42], Pastor-Satorras and Vesignani (2001) [43], and Lloyd and May (2001) [45]. As we know that the social networks had made a major role to determine the rate and pattern of epidemic spread of microbial diseases in human societies. Although it had been primarily focused in the role of population heterogeneity and sub-networks to study the spread of sexually transmitted diseases, especially HIV/AIDS. However, not too much attention had been paid in role of network topology to monitor the spread of other infectious diseases. Currently, computer scientists as well as physicists are more concerned about the spread of infectious agents, for example: computer viruses, worms, etc., through the Internet and the World Wide Web. This had made the development of new interest in network topology that have evolve as a revolution in network modeling, see Fig. 3, for references see the work of Barabasi (2002) [44], and Watts (1999).

1.4 Harmonic Decomposition Analysis

It should be noted that the biological scientists particularly in health-care's though were lagging behind to utilize sophisticated mathematical tools such as: Fourier transform as well as the wavelets theory, etc. but now these powerful analytical tools are also utilized by them for example: Fourier analysis has been used to decompose dengue and malaria data sets to reveal the weather-independence of interepidemic variability, Rogers et al. (2002) [47]; Hay et al. (2000) [48]. The power of wavelet analysis is evident as it was used to decompose measles epidemic harmonics to reveal recurrent spatial spreading patterns which were not evident in the undecomposed epidemic data, see Fig. 4 . With this successes the decompositional technique made

Fig. 3 Internet routing map (80,000 nodes). See http://www.cs.bell-labs.com/~ches/map/
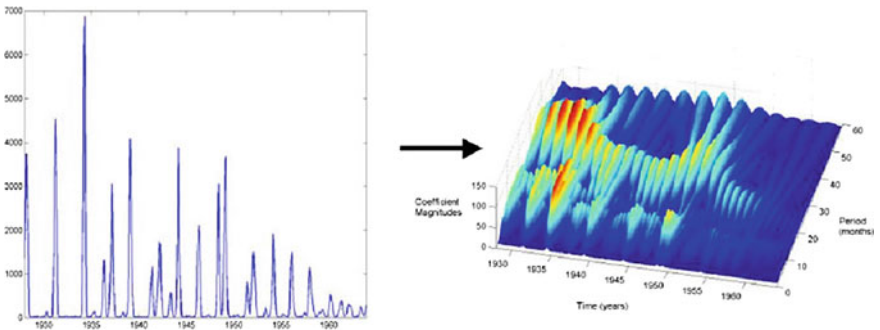

Fig. 4 Continuous wavelet transform decomposition of 1928–1964 Baltimore measles time series data showing that the incidence curve is decomposable into a shorter component with a periodicity of 12 months, and a longer component with a variable periodicity of 24–36 months. The longer component correlates closely with changes in birth rates. *Source* https://www.ncbi.nlm.nih.gov/books/NBK221490/figure/mmm00026/?report=objectonly

it possible to analyze and explain the dynamics of many infectious diseases, see: Grenfell et al. (2001) [49]; Strebel and Cochi (2001) [50].
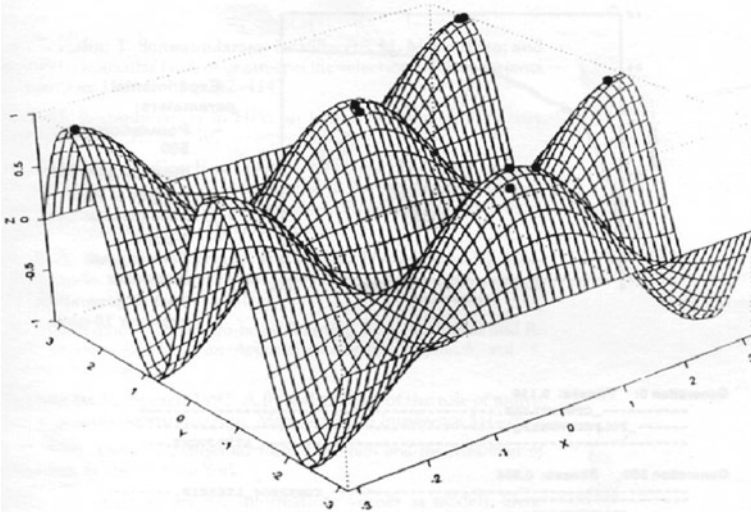
Fig. 5 Representation of evolving bit strings in a fitness landscape. In this example populations of strings are shown as dots colonizing local fitness optima in sequence space, *Source* https://www.ncbi.nlm.nih.gov/books/NBK221490/figure/mmm00029/?report=objectonly

1.5 Digital Microbes

The last decade had seen a see change after it's initial downfall in early nineties that evolutionary techniques would now be incorporated into machine learning, artificial intelligence, and computer programming. The genetic algorithms is the first and now a standard evolutionary computational technique—code strings are iteratively mutated, recombined, and selected for fitness, just as if they were nucleic acid strings evolving in nature, Burke et al. (1998) [51], see Fig. 5. This algorithms are widely employed to solve practical computationally intensive problems, such as protein folding, but only a few studies have appeared in which evolving code strings are used to simulate microbial evolution and adaptation. Preliminary studies suggest that the rules governing code string evolution may be independent of the stuff from which the evolving code strings are made, and that experiments on digital microbes with code string evolution and epidemiology "in silicon" may be a productive way to understand and solve problems that are difficult to study in nature, Ray (1995) [52]; Wilke et al., Adami et al. (2000) [53]; Radman et al. (1999) [54].[1]

In following sections: SIR model (susceptible: S, infected: I, and recovered: R) based on ordinary differential equations (ODEs) will be outlined, and it is the most

---

[1] Agent based, network based, harmonic decomposition analysis, and digital microbes notes were adopted from: Computational Modeling and Simulation of Epidemic Infectious Diseases Donald S. Burke (M.D.) Bloomberg School of Public Health, Johns Hopkins University (USA), and appendix from: Microbial Threats to Health: Emergence, Detection, and Response (2003), National Academic Press.

commonly used model in epidemiology of infectious diseases. The basic reproduction number known as: R_0 pronounced as R-naught will be estimated for the basic SIR and the extended SIR models and computed estimates of R_0 will be compared in different scenarios as well as computational algorithms will be outlined for solving SIR model. Next, the stochastic modeling will be briefly discussed. The applicability and utilization of R_0 in public health domain especially adaptive policy with management tools will be described for the healthcare workers as well as for higher management of healthcare facility.

2   Mathematical Modeling: The Basic SIR Model

The basic compartment models are: SIS, SIR, and SEIR, where the different compartments are symbolically denoted as letters S, E, I, and R. S: individual subjects in a population are susceptible to the disease, similarly E, I, and R mean that subjects are exposed (E), infected (I) from the disease and able to transmits to others, and R compartment signifies subjects have recovered from the disease, immune or have died. The disease parasite or virus dictates the choice or selection of the compartments and depends on the characteristics of the particular disease. It ought to be also noted that the inclusion of too many compartments into the model could be computationally intensive as well as tedious apart from risk of making unreliable prediction as well as may pose greater challenges in policy and decision making.

Historically, Daniel Bernoulli had formulated and solved a model for smallpox in 1760, and based on this he evaluated the efficacy by inoculating on healthy subjects the smallpox virus [12]. A discrete time model by Hamer in 1906 was formulated to understand the recurrence of the outbreak of measles [13]. Sir Ronald Ross was awarded the Nobel Prize, as he had developed a mathematical model for malaria as a host-vector disease in 1911 which was based on the differential equation differential [14]. Another interesting model was developed by Kermack and McKendrick [1] on epidemic models and introduced the threshold result that the density of susceptible must exceed a critical value in order for an epidemic outbreak to occur [8]. Moreover, recent development in mathematical modeling are numerous such as passive immunity, gradual loss of vaccine [3] and disease-acquired immunity, stages of infection, vertical transmission, disease vectors, vaccination, quarantine, social and sexual mixing groups and age structure [15–21]. The SIS, SIR, and SEIR models are graphically shown in Fig. 1. In SIR modeling the population is divided into three groups namely: (i) the group of individuals who are not infected and susceptible (S) of catching the disease, the group of individuals who are infected (I) by the concerned pathogen, and (iii) the group of recovered (R) individuals who have acquired a permanent immunity to the disease. Some of the basic ideas, assumptions, transmission, and recovery for an SIR model (adopted from [22]) are summarized in Table 1. Moreover, a system of differential equations for an SIR model for three compartments are modeled as follows:

Fig. 6 The compartmental model: S represents the group of the subjects not infected and susceptible to the risk of the disease, I is the group of the subjects are infected by the new pathogen: virus or bacteria and R are those group of people who recovered and acquired immunity from the new pathogen
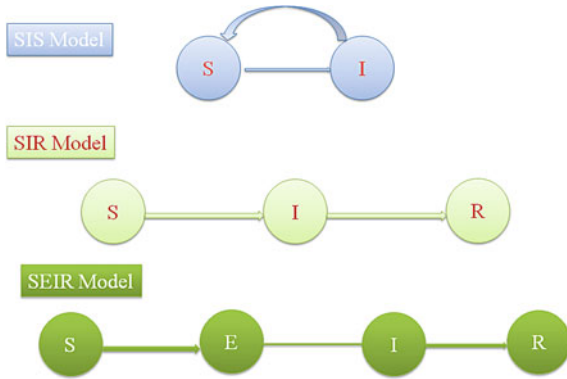
$$\frac{dS}{dt} = -\beta I S \tag{1}$$

$$\frac{dI}{dt} = -\beta I S - \gamma I \tag{2}$$

$$\frac{dR}{dt} = -\gamma I \tag{3}$$

where in the above equations: S denotes the number of susceptible, I the number of infected individuals and R the number of immune individual at time t, and the total population is given by: $N = S + I + R$ is constant by assumption as we have: $\frac{dN}{dt} = \frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$. In equations (1) and (2): the first term $\beta I S$ represents the disease transmission rate by contact between susceptible and infected individuals. This rate is assumed to be proportional to the sizes of both groups with a proportionality coefficient β and equations (2) and (3) the parameter γ is the specific rate at which infected individuals recover from the disease. For example: consider an epidemic outbreak in a population where, at the initial time, only a few individuals are infected, then the initial conditions for SIR model can be assumed as: $S(0) \approx N, \ I(0) = N - S(0) \approx 0, \ R(0) = 0$.

2.1 Phase Analysis

Sometimes, it's desired to have the dynamics in the phase-plane via deriving the isocline's and divide the plane into regions of increase and decrease of the various state variables. The phaseR package is a wrapper around ode that makes it easy to analyze 1D and 2D ode's. The R-state in the SIR model does not influence the dynamics, so we can rewrite the SIR model as a 2D system. So divide equation (1) by (2) gives the ODE:
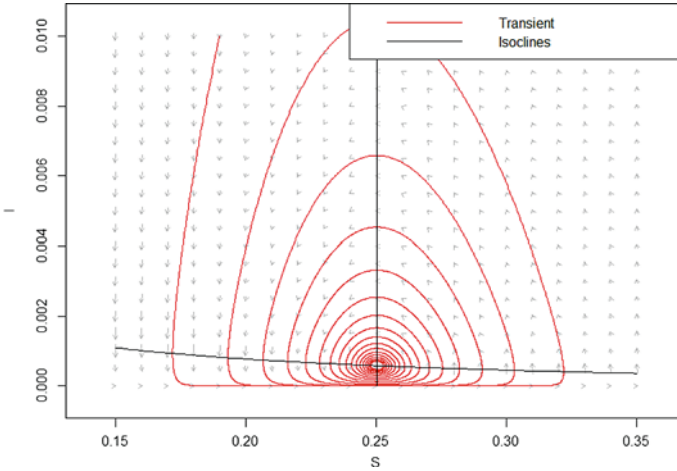
Fig. 7 The phase-plane diagram for the SI model

$$\frac{dS}{d}I = \frac{-\beta SI}{\beta SI - \gamma I} \tag{4}$$

The solution of the above ODE can be found analytically, using separation of variables the above equation can be rewritten as (for I > 0):

$$\int \frac{\beta S - \gamma}{\beta S} dS = -\int dI \tag{5}$$

After integrating the above equation and for every $t \geq 0$

$$I(t) + S(t) - \frac{\gamma}{\beta} \log S(t) = I(0) + S(0) - \frac{\gamma}{\beta} \log S(0) \tag{6}$$

The above expression gives the solutions (S(t), I(t)) in the S-I plane contains the level curves of the function (S(t); I(t)) viewed in the S–I plane (orbits) are contained in the level curves curves of this function $F(S; I) = I(t) + S(t) - \frac{\gamma}{\beta} \log S(t)$, and it's shown in Fig. 7.

2.2 Endemic of the Disease

The above SIR model describes the long term state of the epidemic. The question may naturally arise: *"How long the pandemic may last?"*. It should be noted that always there will be a some portion of the population of susceptible individuals can never get infected. Mathematically, this can be answered by the SIR model by dividing Eq.

Fig. 8 The stochastic model for the endemic of the disease based on SIR branching approximation
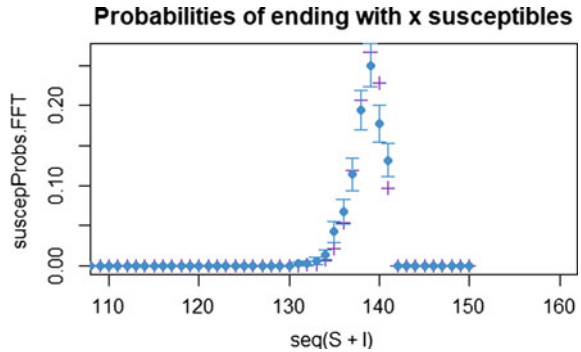
(1) by Eq. (3), and integrating with respect to R, we have

$$S(t) = S(0)e^{-R(t)R_0} \tag{7}$$

From the above expression $S(0)$ will be always positive, whereas the exponential term $R(t)R_0$ has negative sign, ultimately there will no susceptible individual get infected over the time, and pandemic will be ceased to an end. Using stochastic SIR branching approximation and the MultiBD an R-package https://cran.r-project.org/web/packages/MultiBD/.

2.3 Computational Methods for Solving SIR Model

The SIR model's equations can be solved numerically using: Explicit time, Backward Euler, and Crank-Nicolson discretization schemes. The Explicit time discretization are explicit ODE (ordinary differential equation) methods, for example Forward Euler scheme, Runge-Kutta methods, Adams-Bashforth methods, and all these schemes evaluate the function at time levels. The Backward Euler method is an implicit method, also used to solve ODEs. The Crank-Nicolson method is based on the finite difference scheme and commonly used to solve ODEs/PDEs, and is a 2nd order method in time. This method is implicit in time and can be written as an implicit Runge–Kutta method, and this method has numerically stability. Rewriting the SIR model as follows:

$$S' = -\beta SI \tag{8}$$
$$I' = \beta SI - \gamma I \tag{9}$$
$$R' = \gamma I \tag{10}$$

where S(t) , I(t), and R(t) are susceptible, infected and recovered respectively, whereas the constants, $\beta > 0$ and $\gamma > 0$ should be given as the initial conditions: S(0), I(0),

Table 1 The assumptions for SIR model (*Source* lecture notes by V. A. Bokil, "Mathematical Modeling and Analysis of Infectious Disease Dynamics")

SIR model's basic assumptions transmission, and recovery		
Basic ideas and assumptions	**Transmission assumptions**	**Recovery assumptions**
1. Populations under study are divided into compartments.	1. β is the average number of adequate contacts (i.e., contacts sufficient for transmission) of a person per unit time.	1. A fraction α of infectives leave the infective class in unit time.
2. Rates of transfer between compartments are expressed mathematically as derivatives with respect to time of the sizes of the compartments: systems of ordinary differential equations	2. $\frac{\beta I}{N}$ is the average number of contacts with infectives per unit time of one susceptible.	2. There is no entry or departure from the population except possibly through death from the disease.
3. The community size is constant over the duration of the epidemic and is a large number, N	3. $(\frac{\beta I}{N})S$ is the number of new cases per unit time due to the S susceptibles. (Horizontal Incidence)L	
4. The infection is transmitted primarily by person-to person contacts (e.g., measles)		
5. Individuals are homogeneous and mix uniformly.		
Ignore demography, i.e., births and deaths		

and R(0). Now, applying the Implicit time discretization for the Crank-Nicolson scheme will make a 33 system of non-linear algebraic equations in the unknowns as: S^{n+1}, I^{n+1}, and R^{n+1}, and they are written below

$$\frac{S^{n+1} - S^n}{\Delta t} = \beta[SI]^{n+0.5} \approx 0.5\beta(S^n I^n + S^{n+1} I^{n+1}) \tag{11}$$

$$\frac{I^{n+1} - I^n}{\Delta t} = \beta[SI]^{n+0.5} - \gamma I^{n+0.5} \approx 0.5\beta(S^n I^n + S^{n+1} I^{n+1}) - 0.5\gamma(I^n + I^{n+1}) \tag{12}$$

$$\frac{R^{n+1} - R^n}{\Delta t} = \gamma I^{n+0.5} \approx 0.5\gamma(I^n + I^{n+1}) \tag{13}$$

Denoting S for S^{n+1}, $S^{(1)}$ for S^n, I for I^{n+1}, $I^{(1)}$ for I^n, and R for R^{n+1}, $R^{(1)}$ for R^n, now writing the system of equations as

$$F_S(S, I, R) = S - S^{(1)} + 0.5\Delta t\beta(S^{(1)}I^{(1)} + SI) = 0 \tag{14}$$

$$I_S(S, I, R) = I - I^{(1)} - 0.5\Delta t\beta(S^{(1)}I^{(1)} + SI) + 0.5\Delta t\gamma(I^{(1)} + I) = 0 \tag{15}$$

$$R_S(S, I, R) = R - R^{(1)} - 0.5\Delta t\gamma(I^{(1)} + I) = 0 \tag{16}$$

Applying Picard's iterative approximation method and assume that $\hat{S}$, $\hat{I}$ and $\hat{R}$, for S, I, and R as to linearize the non-linear terms., and solving the above equations with respect to the unknowns: S, I, and R.

$$S = \frac{S^{(1)} - 0.5\Delta t\beta S^{(1)}I^{(1)}}{1 + 0.5\Delta t\beta\hat{I}} \tag{17}$$

$$I = \frac{I^{(1)} + 0.5\Delta t\beta S^{(1)}I^{(1)} - 0.5\Delta t\gamma I^{(1)}}{1 - 0.5\Delta t\beta\hat{S} + 0.5\Delta t\gamma} \tag{18}$$

$$R = R^{(1)} - 0.5\Delta t\gamma(I^{(1)} + \hat{I}) \tag{19}$$

The non-linear system of equation (10), (11), and (12) can be written as G(u) = 0, where G = G_S, G_I, G_R), so the Jacobian can computed as

$$J = \begin{bmatrix} \frac{\partial G_S}{\partial S} & \frac{\partial G_S}{\partial I} & \frac{\partial G_S}{\partial R} \\ \frac{\partial G_I}{\partial S} & \frac{\partial G_I}{\partial I} & \frac{\partial G_I}{\partial R} \\ \frac{\partial G_R}{\partial S} & \frac{\partial G_R}{\partial I} & \frac{\partial G_R}{\partial R} \end{bmatrix} = \begin{bmatrix} 1 + 0.5\Delta t\beta I & 0.5\Delta t\beta S & 0 \\ 0.5\Delta t\beta I & 1 - 0.5\Delta t\beta S + 0.5\Delta t\gamma & 0 \\ 0 & 0.5\Delta t\gamma & 0 \end{bmatrix} \tag{20}$$

Updating after each iteration with new updates, using Newton method to solve the Jacobian as an algebraic equation, the solution could be obtained for unknowns S, R, and R. Moreover, for the above SIR model, an explicit time integration approaches work well, the 4th order Runge-Kutta method is a suitable choice since it is efficient, accurate and is based on the simple algorithms. Moreover, in order to fit the model with the observed or real data, broadly, two things need to be taken into account, first a solver for the system of differential equations and an optimizer. For solving DEs the function 'ode' from the 'deSolve' an R package, and to optimize 'optim' function from base R; both these functions are also available in other software MATLAB, MAPLE, R and Mathematica (computational packages). So to minimize the sum of the squared differences between the number of infected I at time t and the corresponding number of predicted cases by our model $I(\hat{t})$.

$$RSS(\beta, \gamma) = \sum_t (I(t) - \hat{I}(t))^2 \tag{21}$$

Using these tools from COVID19.analytic and R-package, the estimated cases for susceptible, infected, and recovered are shown below in Fig. 9, the plots on the right side are on semi-log-scale, the plots show that the model fit with the observed data quite well.
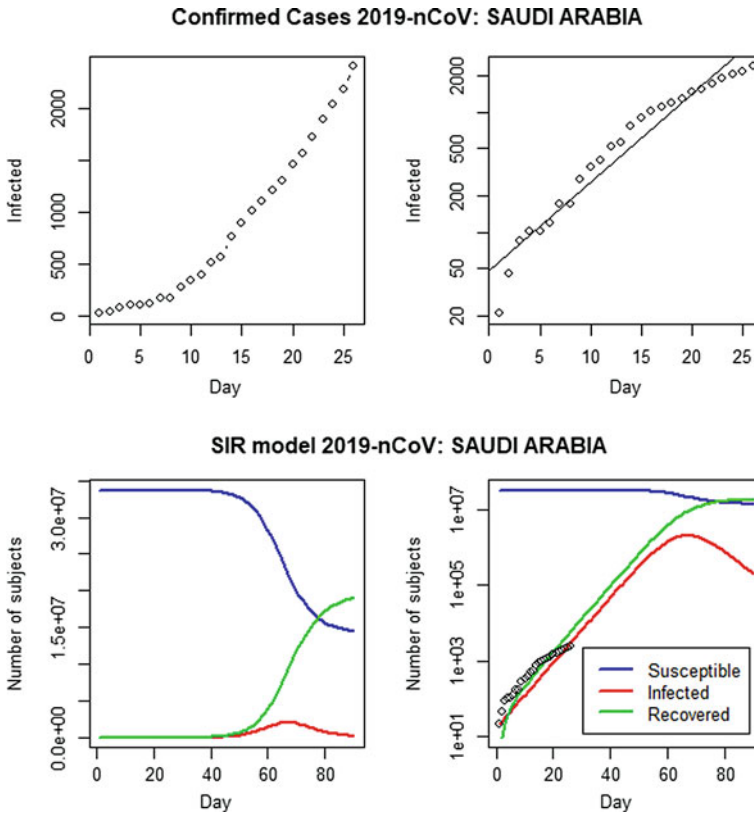
Confirmed Cases 2019-nCoV: SAUDI ARABIA

SIR model 2019-nCoV: SAUDI ARABIA

Fig. 9 Based on the SIR model the susceptible, infected, and recovered cases for Saudi Arabia is shown right hand side graphs on log-scale

Moreover, if the graphical plots do not fit well due to algorithm does not converge to the optimal solution. The reason could be the'optim' stops too early before it could not find an appropriate solution. Now, further explore the'optim' algorithm, the optim function uses the gradient algorithms such as:" "BFGS (by Broyden, Fletcher, Goldfarb and Shanno)", "CG (Fletcher and Reeves developed the conjugate gradients method" and "L-BFGS-B (method is by Byrd et al.)" methods. as well as a finite-difference approximation algorithms. These gradient based algorithms may try to find an optimum estimate via repeatedly improving the current estimate and finding a new solution with a lower residual sum of squares (RSS) each time. Gradient methods do this by computing for a small change of the parameters in which direction the RSS will change the fastest and is based on the linear search approach.

2.4   Estimating the Reproducing Number R_0

The basic reproduction number was introduced in 1886 by (the Director of the Statistical Office of Berlin) Richard Bockh, see: [23] and [24]. The basic reproduction number also commonly known as R_0 pronounced as 'R-nought' could be defined as the expected number of secondary cases produced by a single individual infected subject in a completely susceptible population [25]. It is a dimensionless number and not a rate. We can use the fact that R0 is a dimensionless number to help us in calculating it

$$R_0 \propto \left(\frac{infection}{contact} \right) \times \left(\frac{contact}{time} \right) \times \left(\frac{time}{infection} \right) \tag{22}$$

R_0 can be estimated from the above SIR model's equations (1) to (3), since it depends on the transmissibility, contact rates and expected duration of infection. Based on the model's assumption the population N is closed have N number of subjects, whereas number of susceptible S and infected I , and R subjects are removed, So rewriting the SIR model in terms of proportion, we have

$$\frac{ds}{dt} = -\beta i s \tag{23}$$

$$\frac{di}{dt} = -\beta i s - \gamma i \tag{24}$$

$$\frac{dr}{dt} = -\gamma i \tag{25}$$

where $s = \frac{S}{N}$, $i = \frac{I}{N}$, and $r = \frac{R}{N}$. The trajectory of the system solution in the $I - S$ plane is presented in Fig. 2; from this the existence of a 'threshold effect' can be observed. The maximum value of the curve occurs at $S = \frac{\gamma}{\beta}$. It implies that an epidemic will start and amplify only if $S(0) \approx N$ is larger than $\frac{\gamma}{\beta}$, or equivalently if

$$R_0 = \frac{N\beta}{\gamma} > 1$$

So under this condition, the number of infectious people will increase until the number of susceptible is reduced to $\frac{\gamma}{\beta}$ and will decrease thereafter. Thus the number R_0 represents a threshold for an epidemic to happen, and this number is also commonly known as 'basic reproduction ratio', since it represents the average number of susceptible which are contaminated by one infectious person. now divide Eq. 2 by 1, we get:

$$\frac{dI}{dS} = (\frac{\gamma}{\beta S} - 1)$$

Integrating this equation, we get:

$$I = \frac{\gamma}{\beta} \log S - S + C \quad with \quad C \approx N - \frac{\gamma}{\beta} \log N$$

From the above equation, the instantaneous maximum number of infectious subjects can computed as:

$$I_{max} = N(1 - \frac{1 + \log R_0}{R_0})$$

The trajectory terminates on the S-axis at a positive value as shown in Fig. 7, which can be seen from Eq. 4 that I must vanish at some positive value of S. So the epidemic terminates before all susceptible have become infected and some individual subjects escape the new pathogen completely. Further, we can estimate how many susceptible subjects remain or equivalently the final value $R(\infty)$ of immune population size. Divide Eq. 2 by Eq. 3, so we have:

$$\frac{dS}{dR} = -\frac{\beta}{\gamma} S \implies S(R) = S(0) e^{-\frac{\beta}{\gamma} R} \approx N e^{-\frac{\beta}{\gamma} R}$$

So

$$\frac{dS}{dt} = \gamma I = \gamma (N - S - R) = \gamma (N - N e^{-\frac{\beta}{\gamma} R})$$

Therefore

$$t \to \infty \implies I \to 0 \implies \frac{dR}{dt} = 0 \implies N[1 - e^{-\frac{\beta}{\gamma} R(\infty)}] = R(\infty) \quad (26)$$

Eq. 5 has unique solution $R(\infty)$ between 0 and N as long as $R_0 > 1$. Denote $x = \frac{R(\infty)}{N}$ the fraction the population that has contracted the disease before the epidemic collapses. Solving Eq. 5, we have: $R_0 = \frac{\log(1-x)}{x}$, and R_0 estimates for different pandemics are shown in Fig. 10, see: https://www.the-scientist.com/features/why-r0-is-problematic-for-predicting-covid-19spread-67690ga=2.205136600.9300868 60.1594988120-662736184.1594988120.

2.4.1   Challenges and Issues in Estimating R_0

As above an estimate for R_0 is described for an SIR model, and has been called "arguably the most important quantity in the study of epidemics". Since it's playing a vital role as well as desiderata especially for public health professionals in their decision and policy making. Hence, it's very crucial and important to produce accurate and reliable estimates of this quantity R_0. This quantity precisely presents the whole outbreak of a disease, and it assess the magnitude and severity as well as helps to quantify the percentage of the population needed to be vaccinated to avoid the
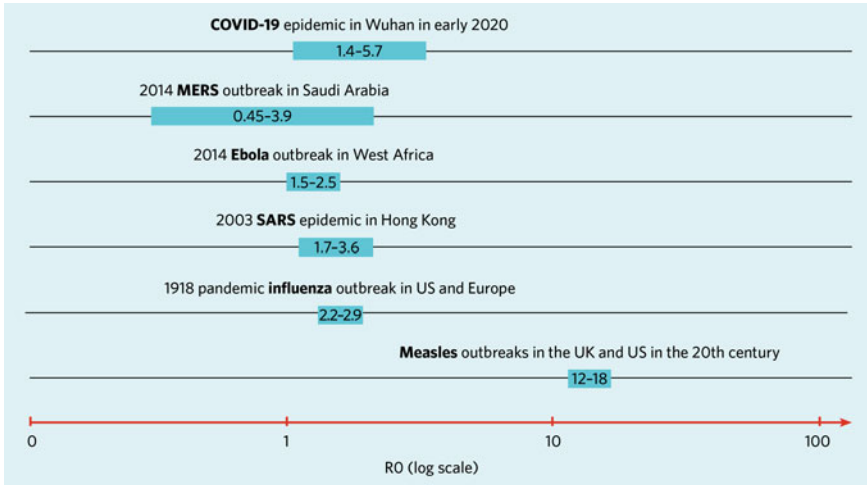
Fig. 10 R_0 estimates for different pandemic is shown adopted from The Scientist an article by Katarina Zimmer, July13, 2020

epidemic roughly as $1 - \frac{1}{R_0}$ is utilized to estimate final size of the total number of infected individuals; and is related to the probability of observing an outbreak under the same conditions (Anderson and May, 1992; Britton, 2010). Although there is an explicit definition of R_0, it is still difficult for an epidemiologists to standardize an estimator for R_0 (Hethcote 2000). An obvious issue in quantifying an estimate for R_0 is that it's solely depend on the property of the disease model apart from the commonly encountered noises inhibit in statistical models as well assumptions made by the researchers about the disease which had been transmitted in a population (Brown, Oleson, & Porter, 2016; Diekmann, Heesterbeek, & Roberts, 2009). To develop a good estimate for R_0 numerous research works had been done and the difficulties and nuances that's involved in estimating R0 can be found in Diekmann et al. (2009) [26], Heathcote [21], and Van den Driessche (2017), [27]. Based on eight different approaches Gallagher, et al. [25] had discussed the nuances pertaining in estimating R_0 for the 2009 pandemic influenza. The authors utilized the basic SIR model by adding the random into the model, and adding the noise the new compartment model as an stochastic model, and the expression with "hats" and without "hats" are distinguishes as stochastic and deterministic in equations (as observations were generated from the ODEs), are noises are given as below:

$$\hat{S(t)} = S(t) + \epsilon_{S,t} \tag{27}$$

$$\hat{I(t)} = N - \hat{S(t)} - \hat{R(t)} \tag{28}$$

$$\hat{R(t)} = \hat{R(t)} + \epsilon_{R,t} \tag{29}$$

where and $\epsilon_{S,t}$ and $\epsilon_{R,t}$ are the random noises in the model. Moreover, Gallagher, et al. [25] has utilized the above Eqs. 28-30 to estimate the R_0 by the data generated from these equations as follows:

$$Data = \left\{ \left(\hat{S(t)} = s(t), \hat{I(t)} = i(t), \hat{R(t)} = r(t) \right) : t = t_0, t_1, ..., t_T \right\} \quad (30)$$

$$\hat{R_0} = m(Data) \quad (31)$$

where m is a function of the data, and the eight models by these authors have been briefly outlined and their comparison of estimates for R_0 for the pandemic influenza is shown in Table 2.

- **Exponential Growth(EG)**
  The effective reproduction number R0 and hence the initial reproduction number R_0, was derived by Wallinga and Lipsitch (2007) [31] on the hypothesis that "counts increase exponentially in the initial phase of an epidemic." So to estimate r, the per capita change in the number of new cases per unit of time and ω the serial interval, the distribution of time between a primary and secondary infection, then, $R_0 = e^{r\omega}$. This equation is based on the Lotka-Euler survival model, commonly utilized in demography, ecology as well as evolutionary biology. Expanding this equation $R_0 = e^{r\omega}$ by using Taylor series expansion up to first order to estimate R_0, whereas in Nishiura, Chowell, Safan, and Castillo-Chavez (2010) [32] had derived it's variant. This approach assumes an exponential growth during early phase as well as the occurrence of initial phase growth. The model has an advantage since it relies on estimates of the number of susceptible, how and when such a method should be used because of the initial growth assumption, Nishiura et al. (2010) [32] had given some guidelines. Moreover, there are several adjustments that could be done to this approach, for example: Wallinga and Lipsitch (2007) [31] describes to estimate R_0 by assuming ω a random variable, whereas Obadia, Haneef, & Boëlle, (2012) [33] assumed that r has it's own distribution.
- **Ratio Estimator (RE)**
  The second approach the Gallagher et al. applied to the SIR model is to minimize the joint mean square error for the data collected at each time point:

$$(\hat{\beta}, \hat{\gamma}) = argmin_{\beta,\gamma} \sum_t \left[(s(t) - S(t; \beta, \gamma))^2 + (i(t) - I(t; \beta, \gamma))^2 + (r(t) - R(t; \beta, \gamma))^2 \right] \quad (32)$$

  So the equation for the ratio estimator (RE) for R_0 will be as:

$$\hat{R_0} = \frac{\hat{\beta}}{\hat{\gamma}} \quad (33)$$

  The above estimate could be found for the β and γ either with optimization algorithm or grid search methods.
- **Re-parameterized Ratio Estimator(rRE)**
  As the approach used to estimate β and γ to compute R_0 can also be estimated

by simply reparametrization of ODEs directly with R_0 via using the relationship $R_0 = \frac{\beta}{\gamma}$, and we have:

$$(\hat{R}_0, \hat{\gamma}) = argmin_{0,\gamma} \sum_t \left[(s(t) - S(t; R_0, \gamma))^2 + (i(t) - I(t; R_0, \gamma))^2 + (r(t) - R(t; R_0, \gamma))^2 \right] \tag{34}$$

The above estimate can again be found either grid search algorithm or optimization tools.

- **Log Linear (LL)**

  In Log-Linear model, the SIR model was reduced to ODEs by Harko, Lobo, and Mak(2014) [34] in two ODEs with one constraint for each equations as follows:

$$log\left(\frac{S(t)}{S(0)}\right) = -R_0 \frac{R(t)}{N} \tag{35}$$

and estimated the R_0 as below:

$$\hat{R}_0 = -\frac{\sum_{t_0}^T log\left(\frac{S(t)}{S(0)}\right)}{\sum_{t_0}^T \frac{R(t)}{N}} \tag{36}$$

- **Markov Chain (MC)**

  Based on Reed-Frost model Abbey(1952) [35] using the Reed-Frost Chain Binomial which as specific form of I(t), the number of infected individuals at time point t. So we have

$$\hat{S(t)} = \hat{S}(t-1) - \hat{I(t)} \tag{37}$$

$$\left(\hat{I(t)}|\hat{S}(t-1), \hat{I}(t-1)\right) \sim Binomial\left(\hat{S}(t-1), 1 - (1-\alpha)^{\hat{I}(t-1)}\right) \tag{38}$$

$$\hat{R(t)} = \hat{R}(t-1) + \hat{I(t)} \tag{39}$$

Using the likelihood method and the optimization tools, $\hat{R}_0$ can be obtained as:

$$\hat{R}_0 = log\left(\frac{1}{1-\alpha}\right) \tag{40}$$

- **Likelihood-Based Estimation (LBE)**

  Using the likelihood based estimate the R_0 is given by see for detail Gallagher et al. [25]

$$R_0 = \frac{\hat{\beta}}{\hat{\gamma}} \tag{41}$$

- **Incidence to Prevalence (IPR)**

  The incidence to prevalence method was described by Nishiura and Chowell (2009) [32], the estimate for $\hat{R}_0$ is:

Table 2 Comparison of R_0 estimates for 2009 influenza, *Source* Exploring the nuances of R_0: eight estimates and application to 2009 pandemic influenza, by Shannon Gallagher, Andersen Chang, and William F. Eddy (a preprint), March 25, 2020

Estimation of R_0 using eight models: a comparison

Methods	SEIR R_0 (SE)	SIR R_0 (SE)
EG	1.001 (<1e-04)	1.002 (2e-04)
RE	1.977 (0.0161)	1.767 (0.137)
rRE	1.977 (0.0162)	1.767 (0.0795)
LL	1.989 (0.021)	1.789 (0.0042)
MC	1.31 (0)	1.619 (0)
LBE	1.893 (0.0135)	1.742 (5.5501)
IPR	4.254 (0.8687)	1.177 (0.9889)
LMA	-0.645 (2.5491)	2.565 (0.175)

$$\hat{R}_0 = \frac{1}{\gamma} \cdot IPR(t^*) \tag{42}$$

- **Linear model approximation (LMA)**
  Chen and Li, 2009; and Hu, Teng, andLong (2014) [36] described method to estimate $\hat{R}_0$ by applying the linear approximation of the Kermack and McKendrick SIR model, this emthod is further extended by Gallagher et al. [25], so the estimate for R_0 is given as:

$$\hat{R}_0 = \frac{S'\hat{(0)}}{R'\hat{(0)}} \cdot \frac{N}{S\hat{(0)}} \tag{43}$$

2.4.2 Next Generation Method: R_0

As R_0 is the number of secondary infections in which a single individual subject is infected in a population. The issue is: how we deal if there were multiple types of infected subjects; for example malaria which is vector-borne disease or sexually transmitted disease (HIV). Such type pf problems can be handled using the structured epidemic models, the basic idea is simply average the expected number of new infections over all possible infected types. Now assuming that a system have multiple discrete type of infected individual subjects. Now introducing a 'next generation matrix' as square matrix G, and the elements of this matrix are denoted as g_{ij}, where i and j are the rows and columns of the matrix G. The g_{ij} gives the expected number of secondary infection type i, and it is caused by a single individual infected individual subjects of type j, whereas the population of type i completely susceptible. Hence all the elements in the matrix G are the reproduction number. The spectral radius of the matrix G gives the reproduction number, which is also known is the dominant eigenvalue. For example, consider G as a 2 by 2 matrix defined as:

$$G = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

The eigenvalue of matrix G can be given as

$$\lambda_i = \frac{T}{2} \pm \sqrt{\left(\frac{T}{2}\right)^2 - D} \tag{44}$$

where T = a + d is the trace and D = ad - bc is the determinant of the matrix G.

The next generation matrix has a number of desirable properties such as it is a non-negative matrix and, it guarantees that there will be a single, unique eigenvalue which is positive, real, and strictly greater than all the others.

The estimation of reproduction number R_0 is discussed, and the next generation method is outlined in this subsection, a detail work can be found in [26–28].

2.5 Stochastic Modeling

Since with any modeling tools, there are always limitation, and it also exist with compartmental models. For example, the model may not able to describe the real or observed data. Due to the assumptions were not fully met namely: homogeneity assumption, not a close system, imbalance equations, where as in general with real world data compartments models fail to describe the system. These limitation could be avoided by extending the deterministic compartmental models into a stochastic model which incorporates the probabilistic theory. So it can be done by keeping time discrete and utilizes the stochastic processes. Another approach may be using continuous framework and time to infection as stochastic.

2.5.1 Reed-Frost Model

The Reed-Frost model is based on the *chain binomial model* since the infection spread dynamically through direct contact and assume that it's independent and have constant probability, and this model has the following characteristics [29, 30]:

- It is similar to compartmental model where each individuals are either susceptible, infectious or recovered.
- The population of the study population is closed and constant and have initial values such as: s_0 and $i_0 \in N$, where $s_0 = S_0$, and $i_0 = I_0$ are susceptible and infected individual subjects.
- The infection dynamic can be explained via discrete time Markov chain.

$$\begin{cases} I_{t+1} | S_t = s_t, \ I_t = i_t \sim Bin\left(s_t, 1 - (1-\omega)^{i_t}\right), \\ S_{t+1} = S_t - I_{t+1} \end{cases} \tag{45}$$

where $t = 1, 2, \ldots$ are time steps and ω is the probability of an infectious individual who is a susceptible subjects in span of one time step.

- The epidemic final size is: $Z = \sum_{i=0}^{\infty} I_i$.

Now consider if ω is the probability of infection, then the probability of not infected subject will be $1 - \omega$, so the probability to escape from infection from contact will be $(1 - \omega)^i i$, hence the probability of infection will be $1 - (1 - \omega)^i$. The Reed-Frost model can be interpreted as an SIR model in which the incubation period and the recovery time is one unit time, whereas the basic reproduction number can be given as ωS_0. Moreover, the likelihood of the Reed-Frost model can be given as

$$L(\omega : \{i_0, i_1, , \ldots, i_T, s_0\}) = \prod_{i=0}^{T-1} \theta_t^{i_{t+1}} (1 - \theta_t)^{s_i - s_{i+1}}, \quad \theta_t = 1 - (1 - \omega)_t^i \quad (46)$$

where T denotes the number of time steps.

2.5.2 Gillespie's Direct Method

The Gillespie's direct method asks two questions if the system is in a given state:

- When does the next event occur? The time to the next event (τ) is exponentially distributed and the rate equal to the sum of the rates over all possible events. The probability density function is given by

$$f(\tau) = \left(\sum_i a_i\right) e^{\left(-\tau \sum_i a_i\right)} \quad (47)$$

- Which event occurs next? We convert event rates into probabilities, and randomly select one of these events according

$$P(Event = v) = \frac{a_v}{\sum_i a_i} \quad (48)$$

where a_i are event rates.

Assuming the above distributions the algorithm is as follows:

1. Set initial population numbers $t \mapsto 0$.
2. Calculate the a_i for all i.
3. Choose τ from an exponential distribution with parameter $\sum_i a_i$ as in Eq. (30).
4. Choose the event v according to the distribution in Eq. (31).
5. Change the number of individuals to reflect the event, v. Set $t \mapsto t + \tau$.
6. Go to step 2.

The above algorithm simulates stochastic realizations of the exact process described by what's known as the master equation. Assume that $p_{SIR}(t)$ is the probability and it is in state (S,I,R) at a given time t, and N = S + I + R, then the master equation is given as follows describes as how this probability distribution evolves over the time:

$$
\begin{aligned}
\frac{dp_{SIR}(t)}{dt} =& p_{S-1,I,R}\left[\mu(N-1)\right] + p_{S+1,I,R}\left[\mu(S+1)\right] + \\
& p_{S+1,I-1,R}\left[\beta\frac{(I-1)}{N}(S+1)\right] + p_{S,I+1,R-1}\left[\gamma(I+1)\right] \\
& + p_{S,I+1,R}\left[\mu(I+1)\right] + p_{S,I,R+1}\left[\mu(R+1)\right] \\
& - p_{S,I,R}\left[\mu N + \mu S + \beta\frac{I}{N}S + \gamma I + \mu I + \mu R\right]
\end{aligned}
\tag{49}
$$

2.5.3 Example Based on SIR Model Using SimInf R-Package

The following example is based on the 'SimInf' an R-package https://cran.r-project.org/web/packages/SimInf/vignettes/SimInf.pdf .

Specification of the SIR model without scheduled events

This example is based on the predefined three compartments (SIR) model (susceptible: S, infected: R, and recovered: R). The mode of transmission of infection is to susceptible individuals is through direct contact between susceptible and infected individuals, this model has two transitions at each node i as follows:

$$
S_i \xrightarrow{\beta S_i I_i / (S_+ I_i + R_i)} I_i
$$

$$
I_i \xrightarrow{\gamma I_i} R_i
$$
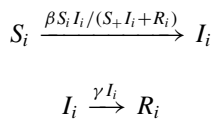
where β, and γ are the transmission and recovery rates respectively. In order to create and SIR model object, define u0, a data.frame with the initial number of individuals in each compartment when the simulation starts, assume that a node has 999 susceptible with 1 infected and there are no recovered individuals. As the assumptions made in this example is that there are no interaction between nodes, so the stochastic model does not disturb any nodes in the model. So the R-code is as follows:

```
> install.packages('SimInf')
> library(SimInf)
> n <- 1000
> u0 <- data.frame(S = rep(999, n), I = rep(1, n), R = rep(0, n))
> tspan <- seq(from = 1, to = 180, by = 7)
> model <- SIR(u0 = u0, tspan = tspan, beta = 0.16, gamma = 0.077)
> model
> model
Model: SIR
Number of nodes: 1000
Number of transitions: 2
Number of scheduled events: 0

Local data
----------
 Parameter Value
 beta       0.160
 gamma      0.077

Compartments
------------
 - Empty, please run the model first
> set.seed(123)
> set_num_threads(1)
> result <- run(model = model)
> result
Model: SIR
Number of nodes: 1000
Number of transitions: 2
Number of scheduled events: 0

Local data
----------
 Parameter Value
 beta       0.160
 gamma      0.077

Compartments
------------
    Min. 1st Qu. Median  Mean 3rd Qu.  Max.
 S 108.0  368.0  993.0 755.4  999.0 999.0
 I   0.0    0.0    1.0  30.4   38.0 235.0
 R   0.0    1.0    5.0 214.2  484.0 891.0
> plot(result)
> plot(result, node = 1:10, range = FALSE)
> plot(result, node = 1:5, range = FALSE)
%%%%%
> head(trajectory(model = result, node = 1))
  node time   S  I  R
1    1    1 999  1  0
2    1    8 998  1  1
3    1   15 991  8  1
4    1   22 973 21  6
5    1   29 935 42 23
6    1   36 886 61 53
```

Specification of scheduled events in the SIR model Further continuing with predefined SIR model, and taking into account the demographic data. So specify each event as one column in the select matrix E using the select attribute of the event. The non-zero entries in the selected column in E specify the compartment involved, define E as
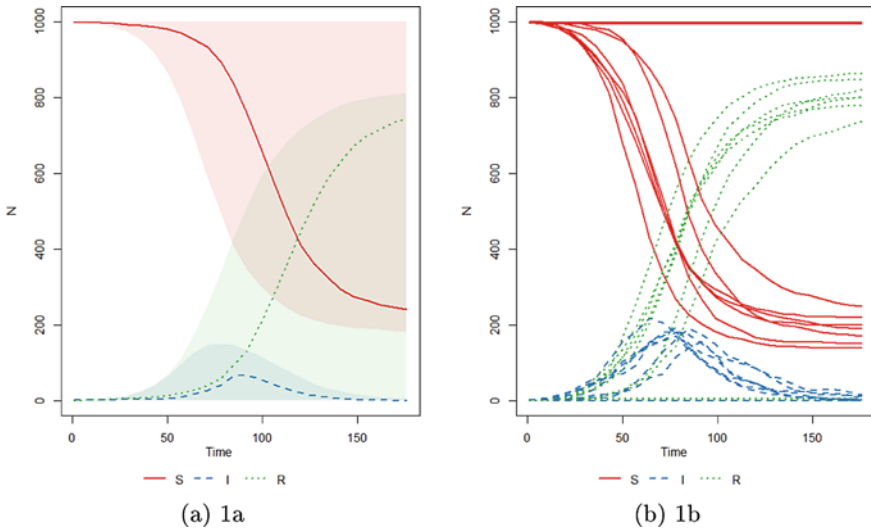
(a) 1a   (b) 1b

Fig. 11 The output result from a stochastic SIR model in 1000 nodes starting with 999 susceptible, 1 infected and 0 recovered individuals in each node ($\beta = 0.16$, $\gamma = 0.077$). There are no between-node interactions. Left (1a: The default plot shows the median and inter-quartile range of the count in each compartment through time across all nodes. Right 1b: Realizations from a subset of 10 nodes

$$
\begin{array}{c}
1\ 2\ 3\ 4 \\
\begin{pmatrix} 1\ 0\ 0\ 1 \\ 0\ 1\ 0\ 1 \\ 0\ 0\ 1\ 1 \end{pmatrix}
\begin{array}{l} S \\ I \\ R \end{array}
\end{array}
$$

In order to operate on a single compartment (S, I or R) as well as an event that involves all three compartment, we need to specify a scheduled event. When several compartments are involved in an event, the individuals affected by the event will be sampled without replacement from the specified compartments. The numerical solver performs an extensive error checking of the event before it is processed. And an error will be raised if the event is invalid, for example, if the event tries to move more individuals than exists in the specified compartments. Consider we have 4 scheduled events to include in a simulation. Below is a data.frame, that contains the events.

```
> u0 <- data.frame(S = rep(0, 5), I = rep(0, 5), R = rep(0, 5))
> add <- data.frame(event = "enter", time = rep(1:10, each = 5),
+                   node = 1:5, dest = 0, n = 1:5, proportion = 0, select = 1, shift = 0)
> infect <- data.frame(event = "enter", time = 25, node = 5,
+                   + dest = 0, n = 1, proportion = 0, select = 2, shift = 0)
> infect <- data.frame(event = "enter", time = 25, node = 5,
+                   dest = 0, n = 1, proportion = 0, select = 2, shift = 0)
> move <- data.frame(event = "extTrans", time = 35:45, node = c(5, 5, 5,
+                   5, 4, 4, 4, 3, 3, 2, 1), dest = c(4, 3, 3, 1, 3, 2, 1, 2, 1, 1, 2),
```

```
+                              n = 5, proportion = 0, select = 4, shift = 0)
> remove <- data.frame(event = "exit", time = c(70, 110),
+                          node = rep(1:5, each = 2), dest = 0, n = 0, proportion = 0.2,
+                        select = 4, shift = 0)
> model <- SIR(u0 = u0, tspan = 1:180, events = events, beta = 0.16,
+                  + gamma = 0.077)
> set.seed(3)
> set_num_threads(1)
> result <- run(model)
> plot(result, node = 1:5, range = FALSE)
> result
Model: SIR
Number of nodes: 1000
Number of transitions: 2
Number of scheduled events: 0

Local data
----------
 Parameter Value
 beta      0.160
 gamma     0.077

Compartments
------------
    Min. 1st Qu. Median  Mean 3rd Qu.  Max.
 S 103.0   337.0  992.0 744.5   999.0 999.0
 I   0.0     0.0    2.0  31.4    41.0 239.0
 R   0.0     1.0    5.0 224.1   528.0 896.0
```

3 Role of Reproduction Number and Growth Curve in Decision and Policy Making

The basic reproduction number plays a vital role in epidemiological sciences as well as in public health management, since it has been used to explain the dynamic of epidemics in population. For example Covid-19 (an infectious disease), R_0 is estimated between 2 and 2.5, whereas for measles it lies between 12 to 18. R_0 spread is order of exponential, for example if $R_0 = 2$, then a single person can generate new infections exponentially as 2^n, whereas if this number if less than 1 then it decays fast as well exponentially:

$$I \; generation = 2 \; new \; infections$$
$$II \; generation = 4 \; new \; infections$$
$$III \; generation = 8 \; new \; infections$$
$$IV \; generation = 16 \; new \; infections$$
$$V \; generation = 32 \; new \; infections$$

Moreover in practice or real life *the effective reproduction number* is utilized and denoted as R and defined as: the average number an infected person goes on to infect in a population where some people are immune (or some other interventions are in place). It related with R_0 as: $R = s R_0$, where is the proportion of susceptible
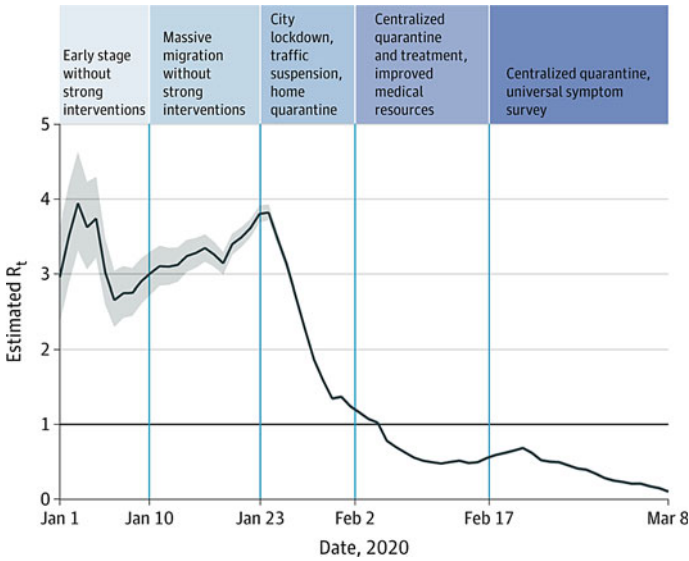
Fig. 12 The effective reproduction number based on laboratory-confirmed Corona virus (Covid-19) Cases in Wuhan, China (*Source* https://jamanetwork.com/journals/jama/fullarticle/2765665)

subjects for a population. As R is not a rate, and it cannot explain how fast the epidemic is growing in the population, and this can by quantified by using the *growth curve*, so the growth curve can be defined as exponential curve:

$$N(t) = constant \ e^{\lambda t} \tag{50}$$

where N is the number of cases and depends on time t in days and λ is the growth rate of of the disease per day. If the growth rate is positive this implies the rise in the epidemic cases whereas if sign of growth rate is negative means there are decrease in number of epidemic cases, and for growth rate zero gives number of cases constant. Now the question is: R or growth rate λ is better? The pros and cons as given in Table 3 below:

3.1 Challenges and Issues in Modeling Infectious Diseases

- Provide a systematic framework for when we should try to eradicate
- Develop quantitative models of the economics of control versus eradication
- Identify the most effective approaches to achieve eradication
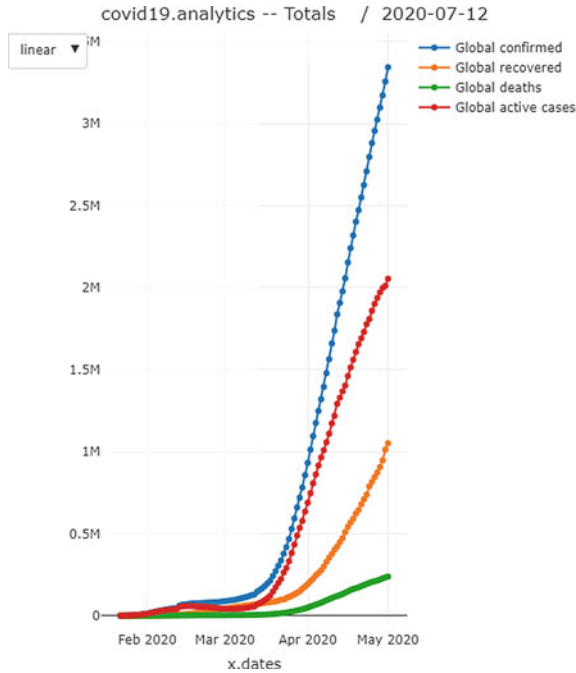- Quantify the landscape of susceptibility

Fig. 13 Global growth rate of Corona virus: confirmed, recovered, deaths and active cases by using Covid-19analytics (R-package)
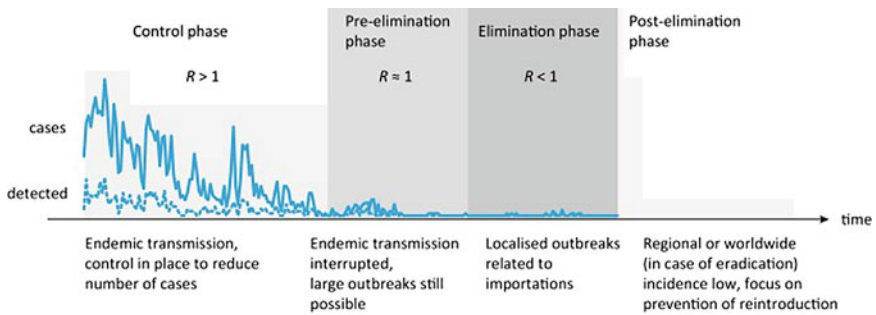

Fig. 14 Stages towards and after elimination in a given location and milestones on the path to elimination. Adapted from (Townsend et al., 2013b, World Health Organization, 2007). Shading illustrates control intensity (darker grey for heightened efforts), also see: https://www.sciencedirect.com/science/article/pii/S175543651400070X

Table 3 Comparing R and the growth rate, *Source* https://plus.maths.org/content/epidemic-growth-rate)

Comparison between reproduction number and the growth rate	
Reproduction ratio: R	**Growth rate per day:**

1. In planning to deal with the spread of epidemic using any intervention, then the reproduction number is most suitable in planning of the control measures such as:
 	
   • if R=1.5 means then we need to plan for one third of reduction in transmission.
   • if R=2 means need to vaccinate half the susceptible people in the population;
 	
An intuitive way of predicting the strength of future interventions needed to stop an epidemic then consideration of R is a better choice.
2. if R>1 then exponential growth in epidemic cases if R=1 then the epidemic cases are constant or flat if R<1 then exponential decay in epidemic cases.
3. R a ratio of cases by infection generation. It is not a rate: there is no timescale involved.
4. R is not at all easy to measure in practice, but can be fitted using models if the timescales of infection are known. In principle it could be estimated by detailed epidemiological data on exactly who got infection from whom, but this is not usually feasible in typical settings.
	1. Growth rate is more natural for thinking about how cases are oscillating over time. For example
	
	   • $\lambda = 0.01$ per day means cases will increase by about 1% a day.
	   • $\lambda = -0.02$ per day means cases will decrease by about 2% a day.
	
	The growth rate is a good description of what's happening now: if we have a certain number of cases today then we can work out how many to expect tomorrow, the day after, and so on.
	2. if $\lambda > 0$ then exponential growth if $\lambda = 0$ then the growth is flat or constant if $\lambda < 0$ then growth curve is exponential decay
	3. The growth rate λ is a rate, usually given in days for Covid-19.
	4. The growth rate λ is relatively easy to estimate from time series data of cases or deaths (but see below about small numbers). A simple approach is just to find the gradient of the logged cases. More advanced approaches, which can take into account a time-varying growth rate, or heterogeneous population, again involve fitting epidemic models.

- Improve monitoring during and after the endgame
- Identify post-eradication opportunities and threats

4 Summary

The basic SIR, and stochastic models were outlined along with some notes on computational tools for these models. Briefly, the reproduction number R_0 was discussed based on different approaches. A famous quote by great mathematician Daniel Bernoulli:

> I simply wish that, in a matter which so closely concerns the well being of the human race, no decision shall be made without all the knowledge which a little analysis and calculation can provide." (Daniel Bernoulli,1760)

References

1. Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. Proc. Roy. Soc. Lond. A **115**, 700–721, 192

2. Brauer, F.: The Kermack-McKendrick epidemic model revisited. Math Biosci. **198**(2), 119–31 (2005 Dec). Epub 2005 Aug 30

3. Jenness, S.M., Goodreau, S.M., Morris, M.: EpiModel: An R Package for mathematical modeling of infectious disease over networks. J. Stat. Softw. **84**(8) (2018)

4. Biomarkers Definition Working Group: Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin. Pharmacol. Therapeut. **69**, 89–95 (2001)

5. Simon, R.: Development and evaluation of therapeutically relevant predictive classifiers using gene expression profiling. J. Natl. Cancer Inst. **98**, 1169–1171 (2006)

6. World Health Organization (WHO). The World Health Report: Changing History [electronic resource], p. 2004. WHO, Geneva (2004)

7. WHO, Infuenza (seasonal). Fact sheet No. 211, March 2014., [online], http://www.who.int/mediacentre/factsheets/fs211/en/

8. Nair, H., Brooks, W.A., Katz, M., et al.: Global burden of respiratory infections due to seasonal influenza in young children: a systematic review and meta-analysis. The Lancet **378**(9807), 1917–1930. http://www.sciencedirect.com/science/article/pii/S0140673611610519

9. CDC, Influenza signs and symptoms and the role of laboratory diagnostics. http://www.cdc.gov/flu/professionals/diagnosis/labrolesprocedures.htm

10. CDC, People with heart disease and those who have had a stroke are at high risk of developing complications from influenza (the flu), http://www.cdc.gov/flu/heartdisease/

11. Berman S. Epidemiology of acute respiratory infections in children of developing countries. Rev. Infect. Dis. **13** Suppl 6, S454–462 (1991). http://www.ncbi.nlm.nih.gov/entrez/query.fcgicmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1862276

12. Bernoulli, D.: Essai d'une nouvelle analyse de la mortalite causee par la petite verole. Mem Math Phy Acad Roy Sci Paris 1766. (English translation entitled 'An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it' In Smallpox Inoculation: An Eighteenth Century Mathematical Controversy, Bradley L. Adult Education Department: Nottingham, 1971, 21)

13. Hamer, W.H.: Epidemic disease in England. Lancet **1**, 733–739 (1906)

14. Ross, R.: The Prevention of Malaria, 2nd edn. John Murray, London (1911)

15. Anderson, R.M. (ed.): Population Dynamics of Infectious Diseases. Chapman and Hall, London (1982)

16. Anderson, R.M., May, R.M. (eds.): Population Biology of Infectious Diseases. Springer Verlag, Berlin, Heidelberg, New York (1982)

17. Anderson, R.M., May, R.M. (eds.): Vaccination against rubella and measles: Quantitative investigations of different policies, J. Hyg. Camb. **90**, 259–325 (1983)

18. Anderson, R.M., May, R.M. (eds.): Infectious Diseases of Humans: Dynamics and Control, Oxford University Press, Oxford, UK, (1991)

19. Bailey, N.T.J.: The Mathematical Theory of Infectious Diseases, 2nd edn. Hafner, New York (1975)

20. Castillo-Chavez, C., Feng, Z., Huang, W.: On the computation of R0 and its role on global stability. In: Castillo-Chavez, C., Blower, S., van den Driessche, P., Kirschner, D., Yakubu, A.A. (eds.) Mathematical Approach for Emerging and Reemerging Infectious Diseases: An Introduction, Springer, (2002), p. 229

21. Hethcote, H.W.: The mathematics of infectious diseases. SIAM Soc. Ind. Appl. Math. **42**, 599–653 (2000)

22. Bokil, V.A.: Mathematical Modeling and Analysis of Infectious Disease Dynamics, Department of Mathematics Oregon State University, Corvallis, OR (USA)

23. Heesterbeek, J.A.P., Dietz, K.: The concept of R0 in epidemic theory. Statistica Neerlandica **50**(1), 89–110 (1996)

24. Heesterbeek, J.: A brief history of R0 and a recipe for its calculation. Acta Biotheoretica **50**, 189–204 (2002)
25. Notes On R_0 by James Holland Jones, Department of Anthropological Sciences, Stanford University
26. Diekmann, O., Heesterbeek, J.A.P., Metz, J.A.J.: On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. J. Math. Biol. **28**(4), 365–382 (1990). https://doi.org/10.1007/BF00178324. hdl:1874/8051. PMID 2117040
27. Van den Driessche, P., Watmough, J.: Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. **180** (1–2), 29–48. https://doi.org/10.1016/S0025-5564(02)00108-6. PMID 12387915
28. Hefferenan, J.M., Smith, R.J., Wahl, L.M.: Prospective on the basic reproductive ratio. J. R. Soc. Interface **2**(4), 281–93. https://doi.org/10.1098/rsif.2005.0042. PMC 1578275. PMID 16849186
29. Furrer, R.: Stochastic Modeling: An Excursion. Applied Statistics Group, July 10, 2018
30. Wearing, H.J.: Lecture Notes. Simple Stochastic Models for Epidemics, July 23, 2014
31. Wallinga, J., Lipsitch, M.: How generation intervals shape the relationship between growth rates and reproductive numbers. Proc. R. Soc. Lond. B: Biol. Sci. **274**(1609), 599–604 (2007)
32. Nishiura, H., Chowell, G., Safan, M., Castillo-Chavez, C.: Pros and cons of estimating the reproduction number from early epidemic growth rate of influenza a (h1n1) 2009. Theoret. Biol. Med. Modell. **7**(1), 1 (2010)
33. Obadia, T., Haneef, R., Boëlle, P.-Y.: The r0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. BMC Med. Inf. Decis. Making **12**(1), 147 (2012)
34. Harko, T., Lobo, F.S., Mak, M.: Exact analytical solutions of the susceptible-infected-recovered (sir) epidemic model and of the sir model with equal death and birth rates. Appl. Math. Comput. **236**, 184–194 (2014)
35. Abbey, H.: An examination of the Reed-Frost theory of epidemics. Hum. Biol. **24**(3), 201 (1952)
36. Chen, G., Li, T.: Stability of stochastic delayed sir model. Stochast. Dyn. **9**(2), 231–252 (2009)
37. Dietz, K., Heesterbeek, J.A.P.: Bernoulli was ahead of modern epidemiology. Nature **408**, 513–514 (2000)
38. Koopman, J.S.: Emerging objectives and methods in epidemiology. Am. J. Public Health **86**, 630–632 (1996)
39. Razvan, G.: Romanescu Rob Deardon. Fast Inference for Network Models of Infectious Disease Spread, Scandinavian J. Statist. Theory Appl. **44**(3), 666–683 (2017)
40. Hofmeyr, S.A., Forrest, S.: Architecture for an artificial immune system. Evol. Comput. **8**, 443–473 (2000). [PubMed]
41. Holland, J.H.: Hidden Order: How Adaptation Builds Complexity. Addison Wesley (1995)
42. Albert, R., Jeong, H., Barabasi, A.-L.: Error and attack tolerance of complex networks. Nature **406**, 378–382 (2000)
43. Pastor-Satorras, R., Vespignani, A.: Epidemic spreading in scale-free networks. Phys. Rev. Lett. **86**, 3200–3203 (2001)
44. Barabasi, A-L.: Linked: The new science of networks. (2002)
45. Lloyd, A.L., May, R.M.: Epidemiology: how viruses spread among computers and people. Science **292**, 1316–1317 (2001)
46. Watts, D.: Small worlds. The dynamics of networks between order and randomness (1999)
47. Rogers, D., Randoph, S., Snow, R.W., Hay, S.I.: Satellite imagery in the study and forecast of malaria. Nature **415**, 710–715 (2002)
48. Hay, S.I., Myers, M.F., Burke, D.S., et al.: Etiology of interepidemic periods of mosquito-borne disease. Proc. Natl. Acad. Sci. **97**, 9335–9339 (2000)
49. Grenfell, B.T., Bjornstad, O.N., Kappey, J.: Travelling waves and spatial hierarchies in measles epidemics. Nature **414**, 716–723 (2001)
50. Strebel, P.M., Cochi, S.L.: Waving goodbye to measles. Nature **414**, 695–696 (2001)

51. Burke, D.S., De Jong, K.A., Grefenstette, J.J., Ramsey, C.L., Wu, A.S.: Putting more genetics into genetic algorithms. Evol. Comput. **6**, 387–410 (1998)
52. Ray, T.S.: Evolution, ecology, and optimization of digital organisms. http://www.isd.atr.co.jp/ray/pubs/tierra (1995)
53. Wilke, C.O., Wang, J.L., Ofria, C., Lenski, R.E., Adami, C.: Evolution of digital organisms at high mutation rates leads to survival of the flattest
54. Radman, R., Matic, I., Taddei, F.: Evolution of evolvability. Ann. N Y Acad. Sci. **870**, 146–155 (1999)