On SIR-type epidemiological models and population heterogeneity effects

Physica A 624 (2023) 128928

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

On SIR-type epidemiological models and population heterogeneity effects

Silke Klemm [a,b], Lucrezia Ravera [c,d,*]

[a] Dipartimento di Fisica, Università di Milano, Via Celoria 16, 20133 Milano, Italy
[b] INFN, Sezione di Milano, Via Celoria 16, 20133 Milano, Italy
[c] DISAT, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
[d] INFN, Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy

ARTICLE INFO

Article history:
Received 9 March 2023
Received in revised form 27 April 2023
Available online 28 May 2023

Keywords:
Statistical mechanics
Epidemiology
SIR models
Statistical methods for epidemiology
Mathematical methods for physics
Epidemiology and analogue gravity

ABSTRACT

In this paper we elaborate on homogeneous and heterogeneous SIR-type epidemiological models. We find an unexpected correspondence between the epidemic trajectory of a transmissible disease in a homogeneous SIR-type model and radial null geodesics in the Schwarzschild spacetime. We also discuss modeling of population heterogeneity effects by considering both a one- and two-parameter gamma-distributed function for the initial susceptibility distribution, and deriving the associated herd immunity threshold. We furthermore describe how mitigation measures can be taken into account by model fitting.

1. Introduction

The SARS-CoV-2 pandemic led, in many countries, to lockdown measures aiming to control and limit the spreading of the virus. A key role when facing global events of this type is played by mathematical modeling of infectious diseases, which allows direct validation with real data. This consequently permits to evaluate the effectiveness of control and prevention strategies, giving support to public health.

In this context, Susceptible–Infected–Removed (SIR) models of epidemics (see e.g. [1–8]) capture key features of a spreading epidemic as a mean field theory based on pair-wise interactions between infected and susceptible individuals, without aiming to describe specific details. In particular, in the presence of I infected individuals in a population of N individuals, the infection can be transmitted to susceptible individuals S. They stay infectious during an average time γ^{-1}, after which they no longer contribute to infections. The fraction of immune individuals in the population beyond which the epidemic can no longer grow defines the herd immunity threshold (HIT).

Simple SIR models commonly assume the population to be homogeneous; each individual has the same probability of being infected by the disease. However, in order to take into account that the infection probability actually depends on age, sex, connections with other individuals, etc., SIR models for heterogeneous populations have been considered [9–15]. In these models, a parameter, usually denoted by α, is commonly introduced to describe population heterogeneity and, hence, variation in susceptibility of individuals.

Studying the transmission of the virus SARS-CoV-2, in [9] it was shown that the percentage of a homogeneous population to be immune given some value for R_0 (which is the basic reproduction number, namely the average number

* Corresponding author at: DISAT, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy.
E-mail addresses: silke.klemm@mi.infn.it (S. Klemm), lucrezia.ravera@polito.it (L. Ravera).

https://doi.org/10.1016/j.physa.2023.128928

of new infected generated by an infected individual at the early epidemic stage) noticeably drops if the population is considered to be highly heterogeneous. More specifically, while herd immunity is expected to require 60–75 percent of a homogeneous population to be immune given an R_0 (that is the basic reproduction number) between 2.5 and 4, these percentages drop to the 10–20 percent range for the coefficients of variation in susceptibility considered in [9] between 2 and 4. In particular, it was shown that individual variation in susceptibility or exposure (connectivity) accelerates the acquisition of immunity in populations due to selection by the force of infection. More susceptible and more connected individuals have a higher propensity to be infected and thus are likely to become immune earlier. Due to this selective immunization, heterogeneous populations require less infections to cross their HITs than homogeneous (or not sufficiently heterogeneous) models would suggest. In [9] the initial susceptibility was considered to be gamma-distributed, with a one-parameter gamma distribution. Besides, the case of a lognormal distribution was treated numerically. The gamma distribution was also considered in [16] to model the first-wave COVID-19 daily cases, and it was proven, in this context, to provide better results than the Gaussian, Weibull (and Gumbel) distributions.

Taking into account heterogeneity effects has proven to be relevant also in the spread of smallpox (cf. [12]), where homogeneous models are not capable to explain the data, as well as for tuberculosis and malaria (see, e.g., [9] and references therein).

In this work we discuss modeling of population heterogeneity effects by considering both a one-parameter gamma-distributed function and a two-parameter one for the initial susceptibility distribution, deriving the associated HIT. The latter is computed analytically in both cases. We also describe a possible way to take into account mitigation measures when performing model fitting in the case of the one-parameter initial gamma distribution, while the two-parameter initial gamma distribution appears to automatically accommodate this external action on diseases spread. On the other hand, regarding homogeneous SIR models, we present an intriguing feature of a simple model of this type, which paves the way to future analytically tractable studies of epidemiological models.

The remainder of this paper is structured as follows: In Section 2, we review homogeneous SIR-type models of epidemics and, in Section 2.1, we present a correspondence between the epidemic trajectory in a homogeneous SIR model and radial null geodesics in the Schwarzschild spacetime. Subsequently, in Section 3, we discuss modeling of population heterogeneity effects to capture the fact that the probability of being infected is not the same for all individuals. Section 4 is devoted to final remarks and possible future developments of our analysis.

2. Homogeneous SIR-type models

In an SIR-type model [1], the population is divided into susceptible, infected and recovered individuals, whose numbers are denoted respectively by S, I, and R. Their dynamics is governed by the equations

$$\dot{S} = -f(I,S)\,, \qquad \dot{I} = f(I,S) - g(I)\,, \qquad \dot{R} = g(I)\,. \tag{1}$$

Here $f(I,S)$ denotes the infection force, i.e., the rate at which susceptible persons acquire the infectious disease, while $g(I)$ is some function to be specified below. The upper dot symbol denotes the time derivative. From (1) one obtains the conservation law

$$\dot{S} + \dot{I} + \dot{R} = 0 \quad \Rightarrow \quad S + I + R = \text{const.} = N\,, \tag{2}$$

with N the total number of individuals in the population. A common choice is $f(I,S) = \beta IS$, $g(I) = \gamma I$, where β is the transmission (or infection) rate (per capita),[1] and γ denotes the rate of recovery. It is related to the average recovery time D by $D = 1/\gamma$. We have thus[2]

$$\dot{S} = -\beta IS\,, \qquad \dot{I} = \beta IS - \gamma I\,, \qquad \dot{R} = \gamma I\,. \tag{3}$$

This implies

$$\frac{dI}{dS} = -1 + \frac{\gamma}{\beta S}\,, \tag{4}$$

which can be integrated to give the epidemic trajectory

$$I - I_0 = S_0 - S + \frac{\gamma}{\beta}\ln\frac{S}{S_0}\,, \tag{5}$$

with $I_0 = I(t=0)$ and $S_0 = S(t=0)$. In order to obtain the early growth of the epidemic, one linearizes (3) around $S = S_0 \approx N$ and $I \approx 0$, i.e., sets

$$S = N - \delta S\,, \qquad I = \delta I\,, \qquad\qquad \delta S, \delta I \ll N\,. \tag{6}$$

This leads to the exponential law

$$\delta I = I_0\, e^{\gamma(R_0-1)t}\,, \tag{7}$$

[1] The infection rate β can in general depend on time t; this time dependence could correspond to seasonal changes or mitigation measures [17–19].

[2] In this work, we multiply the quantity β with the constant factor N with respect to the one defined, e.g., in [12], that is $\beta \to N\beta$.

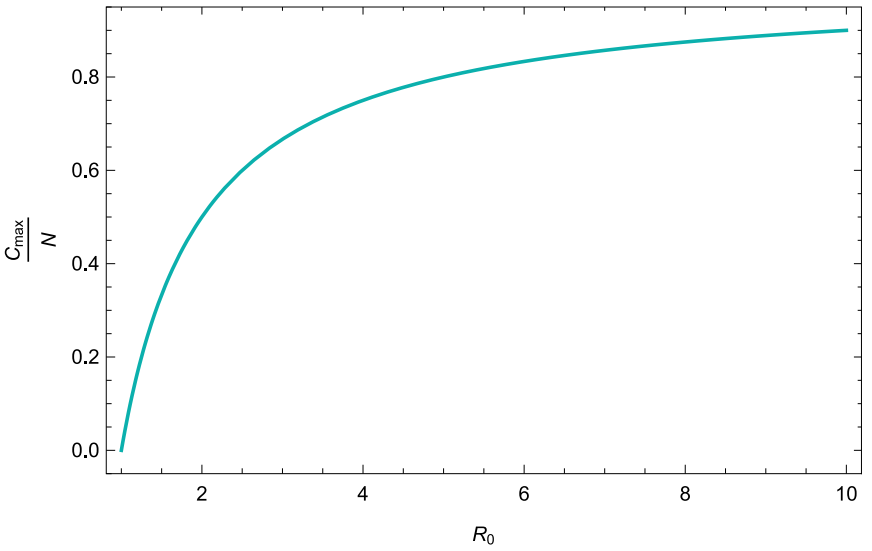

Fig. 1. Herd immunity threshold $\frac{C_{max}}{N}$ of a homogeneous population as a function of R_0 (more specifically, here we consider $1 \leq R_0 \leq 10$, which is a typical range for the basic reproduction number of common diseases). The larger R_0, the more the HIT rises.

where

$$R_0 = \frac{N\beta}{\gamma} \tag{8}$$

is the basic reproduction number. It denotes the average number of new infections generated by an infected individual (at the early epidemic stage).

The function $I(S)$ has a maximum at $S = \gamma/\beta$. At this peak, a fraction $S/N = 1/R_0$ of individuals remains susceptible. The cumulative number of infections $C = N - S$ at the maximum of I thus obeys

$$\frac{C_{max}}{N} = 1 - \frac{1}{R_0}\,. \tag{9}$$

This is the well-known formula for the herd immunity level (or herd immunity threshold, HIT), i.e., the fraction of immune individuals in the population beyond which the epidemic can no longer grow. Here we are not considering mitigation measures, nor reinfections. Hence, in particular, the threshold to reach herd immunity is estimated by considering natural infections without restrictions (lockdown, social distancing, etc.) and without taking into account possible vaccinations. The plot in Fig. 1 displays the herd immunity level (9) as a function of R_0.

When the epidemic stops we have $I = 0$. Using (5), it is easy to show that the number of susceptible individuals left over at the end of an epidemic is given by

$$S_f = -\frac{S_0}{R_e} W_0\left[-R_e \exp\left(-R_e\left(1 + \frac{I_0}{S_0}\right)\right)\right], \tag{10}$$

where

$$R_e = S_0\beta/\gamma = S_0 R_0/N \tag{11}$$

is the effective reproduction number, while W_0 represents a particular branch of the Lambert W-function, which is defined as the inverse of the function $f : x \mapsto xe^x$.

Let us also mention that a different choice for the functions f and g in (1) was made by Mickens [4], namely $f = \beta\sqrt{IS}$ and $g = \gamma\sqrt{I}$, and thus

$$\dot{S} = -\beta\sqrt{IS}\,, \qquad \dot{I} = \beta\sqrt{IS} - \gamma\sqrt{I}\,, \qquad \dot{R} = \gamma\sqrt{I}\,. \tag{12}$$

The square root leads to a power-law (instead of exponential) early growth of the epidemic. Indeed, linearizing (12) according to (6), one gets

$$\delta I = \left((\beta\sqrt{N} - \gamma)\frac{t}{2} + I_0\right)^2 . \tag{13}$$

As a consequence of this, the model (12) is particularly adapted to describe effects of mitigation measures imposed by governments, like social distancing and so on. From (12) we obtain

$$\frac{dI}{dS} = -1 + \frac{\gamma}{\beta\sqrt{S}}\,, \tag{14}$$

which leads to the epidemic trajectory

$$I = I_0 - (S - S_0) + \frac{2\gamma}{\beta}(\sqrt{S} - \sqrt{S_0})\,. \tag{15}$$

This has a maximum at $S = (\gamma/\beta)^2$. In this case, the formula for the herd immunity level becomes

$$\frac{C_{\max}}{N} = 1 - \frac{\gamma^2}{\beta^2 N}\,. \tag{16}$$

Observe that the latter depends, in particular, on the squared of the ratio γ/β, and therefore may be also rewritten as $\frac{C_{\max}}{N} = 1 - \frac{N}{R_0^2}$.

Note in this context, as it was shown in [20], that certain heterogeneous models (that we will consider in the next section) can be reduced to homogeneous models with a nonlinear transmission function $f = \beta S^p I^q$.

2.1. Epidemic trajectory as radial null geodesics in Schwarzschild spacetime

Remarkably, the epidemic trajectory (4) coincides with radial null geodesics in the Schwarzschild spacetime,[3] which are given by

$$0 = -\left(1 - \frac{2m}{r}\right)d\tau^2 + \frac{dr^2}{1 - \frac{2m}{r}} \quad \Rightarrow \quad d\tau = \pm\frac{dr}{1 - \frac{2m}{r}}\,, \tag{17}$$

where m is the black hole mass, r the radial coordinate, and τ the time coordinate (time measured by a stationary clock at infinity).[4] If we identify

$$I = \mp a\tau\,, \qquad S = a(r - 2\,m)\,, \tag{18}$$

with a an arbitrary scale factor, (4) becomes precisely (17), provided that $-2ma = \gamma/\beta$. One can also map the evolution Eqs. (3) into

$$\frac{d\tau}{d\lambda} = \frac{E}{1 - \frac{2m}{r}}\,, \qquad \frac{dr}{d\lambda} = \pm E\,, \tag{19}$$

that follow respectively from the conservation law $g_{\mu\nu}u^\mu\xi^\nu = -E$ and $g_{\mu\nu}u^\mu u^\nu = 0$. Here λ denotes an affine parameter, $g_{\mu\nu}$ is the spacetime metric, E is the conserved energy, $u = d/d\lambda$ is tangent to the geodesic, while $\xi = \partial_\tau$ is the timelike Killing vector. Using (18), the Eqs. (19) reduce to (3) if $E = \mp 1/a$ and

$$\frac{d\lambda}{dt} = \beta IS\,. \tag{20}$$

The scale factor a is thus related to the energy E of the geodesic, and t is not an affine parameter. Since βIS is the infection force in this specific model, we have from (20) that $\lambda = f(I, S)$, so that λ can be interpreted as infection momentum. Moreover, from the first of (3) one gets $\dot{S} + \dot{\lambda} = 0$, and therefore $S + \lambda$ is constant. If we choose this constant to be equal to N, then

$$\lambda = N - S\,, \tag{21}$$

which is the cumulative number of infections C.

One may now map the epidemic trajectory into a part of a lightlike (null) geodesic in (the Kruskal extension of) the Schwarzschild geometry. It is not difficult to see that the descending part of the epidemic trajectory (i.e., the part that starts from the point where the herd immunity is reached until the point where the epidemic has completely died out) maps into a light ray trajectory that comes out of the white hole singularity (the latter corresponding to the point where the herd immunity is reached), and travels towards the white hole horizon until the point $r = \frac{1}{a}\left(S_{\rm f} - \frac{\gamma}{\beta}\right)$, where $S_{\rm f}$ was defined in (10). Notice that the relation $-2ma = \gamma/\beta$ implies $a < 0$ if $m > 0$. This unveiled correspondence between the epidemic trajectory (4) and radial null geodesics in the Schwarzschild geometry, in practice, may be useful to obtain new analytically

[3] The Schwarzschild metric describes a static, non-rotating black hole solution to the Einstein's field equations. It was discovered by Karl Schwarzschild within a year of Einstein's publication of the theory of general relativity. A null geodesic is the path that a massless particle, such as a photon, follows. It is called null since its interval (its "distance" in four-dimensional spacetime) is equal to zero and it does not have a proper time associated with it.

[4] We adopt geometrized units, that is $c = G = 1$, where c is the speed of light in vacuum and G the gravitational constant.

tractable epidemiological models. In this regard, it would be interesting to consider, for instance, timelike geodesics, nonradial geodesics, or null geodesics in the Reissner–Nordström metric case, analyzing the predictions of epidemiological models in this context. Also, the epidemic trajectory of other models with different number of compartments (e.g., SEIR or SI) might correspond to geodesics of some black holes as well.

3. Modeling population heterogeneity effects

In order to capture the fact that the probability of being infected is not the same for all individuals, we use and review the model of [12], i.e., we introduce a distribution s of susceptibilities x, and denote by $s(x,t)dx$ the number of individuals with susceptibility between x and $x+dx$ at time t. The total number of susceptible individuals reads

$$S(t) = \int_0^\infty s(x,t)dx\,, \tag{22}$$

and s obeys

$$\frac{\partial s}{\partial t} = -\beta x I s\,, \tag{23}$$

which generalizes the first of (3). Integrating (23) over x leads to

$$\dot{S}(t) = -\beta \bar{x}(t) I S\,, \tag{24}$$

where the average susceptibility $\bar{x}(t)$ is given by[5]

$$\bar{x}(t) = \frac{1}{S(t)} \int_0^\infty s(x,t)x dx\,. \tag{25}$$

In the special case $\bar{x}=1$, (24) boils down to the corresponding Eq. in (3). The dynamical equations are completed by

$$\dot{I}(t) = \beta \bar{x}(t) I S - \gamma I\,. \tag{26}$$

The time course of an epidemic is often provided as the number of new cases per day. This corresponds to the rate of new infections per unit time,

$$J = \beta \bar{x} I S\,, \tag{27}$$

with $J = -\dot{S}$.

Notice that the authors of [9–11] considered two cases, namely variable susceptibility or variable connectivity (individuals that have many contacts are both more likely to get infected and to infect others). They describe these situations with a susceptible–exposed–infected–recovered (SEIR) model,

$$\begin{aligned} \partial_t s(x,t) &= -x\rho(t)s(x,t)\,, \qquad \partial_t e(x,t) = x\rho(t)s(x,t) - \delta e(x,t)\,,\\ \partial_t i(x,t) &= \delta e(x,t) - \gamma i(x,t)\,, \qquad \partial_t r(x,t) = \gamma i(x,t)\,, \end{aligned} \tag{28}$$

where

$$\rho(t) = \begin{cases} \beta \int i(x,t)dx & \text{variable susceptibility,}\\ \beta \int i(x,t)x dx & \text{variable connectivity,} \end{cases} \tag{29}$$

while γ and δ are constants, the latter denoting the rate of progression from exposed to infectious. For variable susceptibility, the first of (28) is identical to (23), if we set $I(t) = \int i(x,t)dx$. We see that in (28), also $E(t)$, $I(t)$ and $R(t)$ are divided into infinitely many compartments $e(x,t)$, $i(x,t)$ and $r(x,t)$, which is different form the model used in [12], where only $S(t)$ is split. In this paper, we shall limit ourselves to the case where only effects of heterogeneity in the degree of susceptibility to infection are taken into account, as it was done in [12].

In what follows, it will prove useful to introduce a new time variable τ, defined by [12]

$$\dot{\tau} = \beta I\,. \tag{30}$$

Let us stress that $\tau = \tau(t)$. Eq. (23) can then easily be integrated, and (22) gives

$$S(\tau) = \int_0^\infty s_0(x) e^{-\tau x} dx\,. \tag{31}$$

In other words, $S(\tau)$ is the Laplace transform of the initial distribution $s_0(x)$.

[5] Notice that the average infection susceptibility $\bar{x}$, which is introduced to capture effects of population heterogeneity, also allows to modulate the infection rate β.

3.1. HIT for the case of a one-parameter gamma distribution

As in [9–12], we shall now assume that $s_0(x)$ is gamma-distributed with shape parameter $\alpha > 0$ (here we start by considering, following [9–12], the case of a one-parameter, that is α, gamma distribution with rate parameter $\eta := \frac{1}{\theta} = \alpha$, being θ the scale parameter),[6]

$$s_0(x) = S_0 \frac{\alpha^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\alpha x} , \tag{32}$$

where S_0 is the initial value of susceptible individuals and $\Gamma(\alpha)$ denotes Euler's gamma function. This gives for the Laplace transform (31)

$$S(\tau) = \frac{S_0}{\left(1+\frac{\tau}{\alpha}\right)^\alpha} = S_0 \frac{\alpha^\alpha}{(\alpha+\tau)^\alpha} . \tag{33}$$

Since $s(x,\tau) = s_0(x)\exp(-x\tau)$, one obtains for the average susceptibility (25)

$$\bar{x}(\tau) = \frac{1}{1+\frac{\tau}{\alpha}} = \frac{\alpha}{\alpha+\tau} , \tag{34}$$

which starts from $\bar{x} = 1$ at the initial time $\tau = 0$ and then decays to zero for increasing τ, slowing down the epidemic. Note that $s(x,\tau)$ obeys the scaling law [12]

$$s(x,\tau) = \bar{x}^{\alpha-1} s_0(x/\bar{x}) . \tag{35}$$

Thus, $s(x,\tau)$ is shape invariant, i.e., the gamma distribution is kept during the whole time evolution, instead of being just an initial condition.

Using (33) and (34), Eq. (26) can be rewritten as

$$\frac{dI}{d\tau} = S_0 \left(1+\frac{\tau}{\alpha}\right)^{-1-\alpha} - \frac{\gamma}{\beta} , \tag{36}$$

which can be integrated to give

$$I(\tau) = I_0 + S_0 \left[1 - \left(1+\frac{\tau}{\alpha}\right)^{-\alpha}\right] - \frac{\gamma}{\beta}\tau . \tag{37}$$

This has a maximum for

$$\left(1+\frac{\tau}{\alpha}\right)^{1+\alpha} = R_e , \tag{38}$$

with R_e the effective reproduction number defined in Eq. (11). The maximum value of I is given by

$$\frac{I_{\max}}{S_0} = \frac{I_0}{S_0} + 1 - \frac{1}{R_e} - \frac{1}{R_e}(1+\alpha)\left(R_e^{\frac{1}{1+\alpha}} - 1\right) . \tag{39}$$

In the special case $I_0 \ll S_0 \approx N$, this boils down to Eq. (18) of [12]. Then, the cumulative number of infections reads

$$C(\tau) = N - S(\tau) = N\left[1 - \left(1+\frac{\tau}{\alpha}\right)^{-\alpha}\right] , \tag{40}$$

which, evaluated at the maximum of I, becomes

$$\frac{C_{\max}}{N} = 1 - R_0^{-\frac{\alpha}{1+\alpha}} . \tag{41}$$

This is the generalization of the herd immunity level (9) to a heterogeneous population. In Fig. 2 we give a three-dimensional plot of the HIT as a function of R_0 and α.

On the other hand, the plot in Fig. 3 displays the herd immunity level (41) as a function of α for fixed values of R_0 (we consider $R_0 = 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5$ as sampling values), while the plot in Fig. 4 shows the HIT (41) as a function of R_0 for different values of α (we consider $\alpha = 0.5, 1, 2, 3, 4, 10, 100$ as sampling values).

We observe that, for finite α, (41) is smaller than (9), to which it reduces in the homogeneous limit $\alpha \to \infty$. In particular, for an R_0 between 2.5 and 4, herd immunity is expected to require 60–75 percent of the population in the homogeneous limit $\alpha \to \infty$, while these percentages drop to the range 26–50 percent for α between 0.5 and 1. In other words, for fixed values of R_0, the smaller α, the more heterogeneous the population and the lower the threshold for achieving herd immunity.

One easily verifies that for constant α and large R_0 the logarithm of $\frac{C_{\max}}{N}$ goes like an inverse power law of R_0, $\ln \frac{C_{\max}}{N} \approx -R_0^{-\frac{\alpha}{1+\alpha}}$, whereas for $R_0 \to 1$ (which is the minimal allowed value for R_0 in order for the epidemic to start) we

[6] This subsection is mainly devoted to a brief review of the one-parameter gamma distribution considered in [9–12].

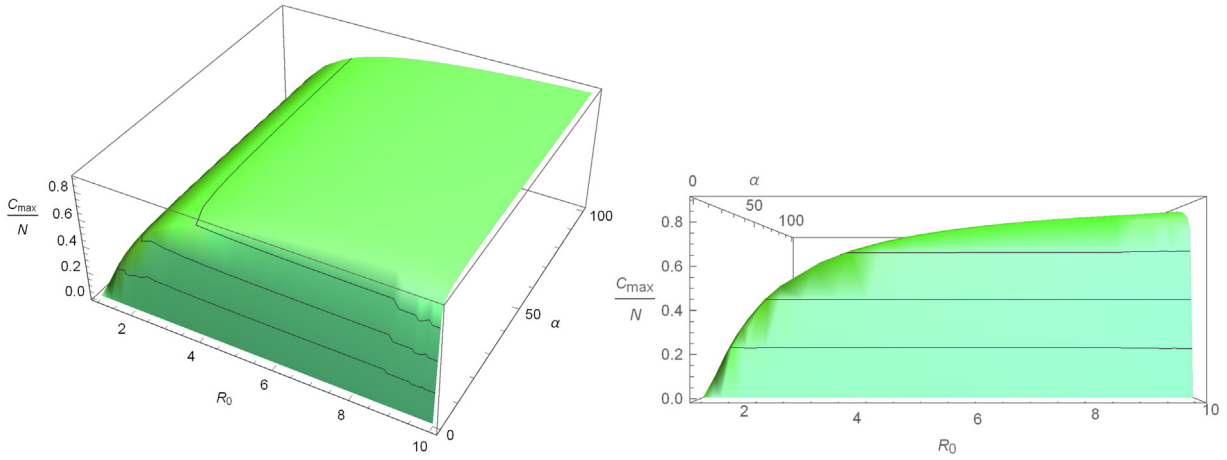

Fig. 2. Three-dimensional plot of the herd immunity threshold $\frac{C_{max}}{N}$ of a heterogeneous population as a function of R_0 and α in the case of a one-parameter gamma distribution.

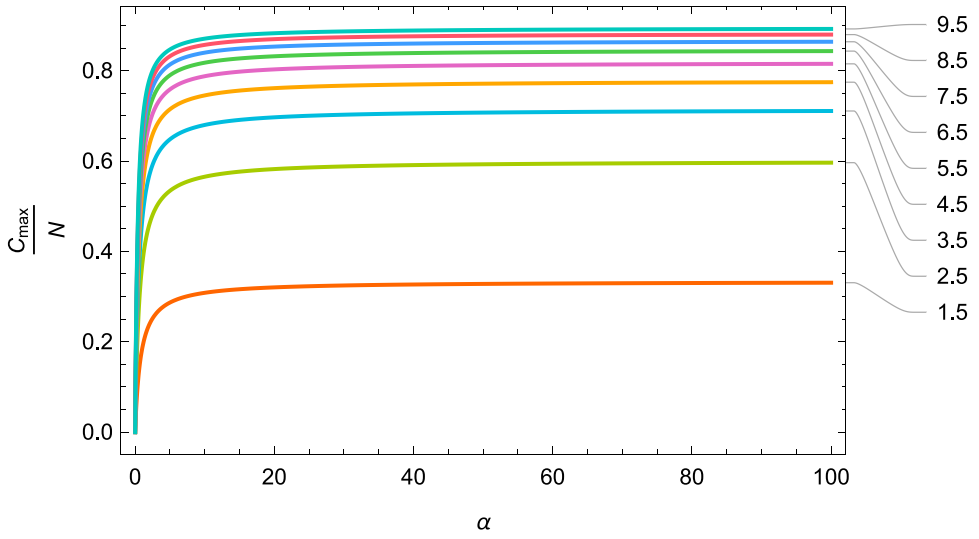

Fig. 3. HIT $\frac{C_{max}}{N}$ of a heterogeneous population as a function of α, for different values of R_0 (indicated on the right side of the plot), in the case of a one-parameter gamma distribution.

have the linear behavior $\frac{C_{max}}{N} \approx \frac{\alpha}{1+\alpha}(R_0 - 1)$, with coefficient < 1, which lowers the HIT with respect to the case of a homogeneous population. On the other hand, for constant R_0 and large α (nearly homogeneous population), we have an inverse power law correction to the usual formula of the homogeneous case, namely $\frac{C_{max}}{N} \approx 1 - \frac{1}{R_0} - \frac{\ln R_0}{\alpha R_0}$, again lowering the HIT. Finally, for constant R_0 and $\alpha \to 0$ (strongly inhomogeneous population), one gets the linear law $\frac{C_{max}}{N} \approx \alpha \ln R_0$.

3.1.1. A possible way to take into account mitigation measures in model fitting

Let us comment on a possible new way to take into account mitigation measures in model fitting.

When performing model fitting of J (defined in Eq. (27)), one has to deal with data, let us call them J_{exp}, which are affected by mitigation measures (lockdown, etc.). Here we propose a possible way to take this into account, which consists in fitting J from the model (with the one-parameter initial gamma distribution), namely

$$J(\tau) = \beta I S_0 \left(1 + \frac{\tau}{\alpha}\right)^{-\alpha-1}, \tag{42}$$

with J_{exp} from data and solve a system of differential equations *at the same time*, considering $\beta = \beta(t)$ (cf. [12]). In particular, the time dependence of the infection rate β can be seen as taking into account averaged seasonal changes or mitigation measures.

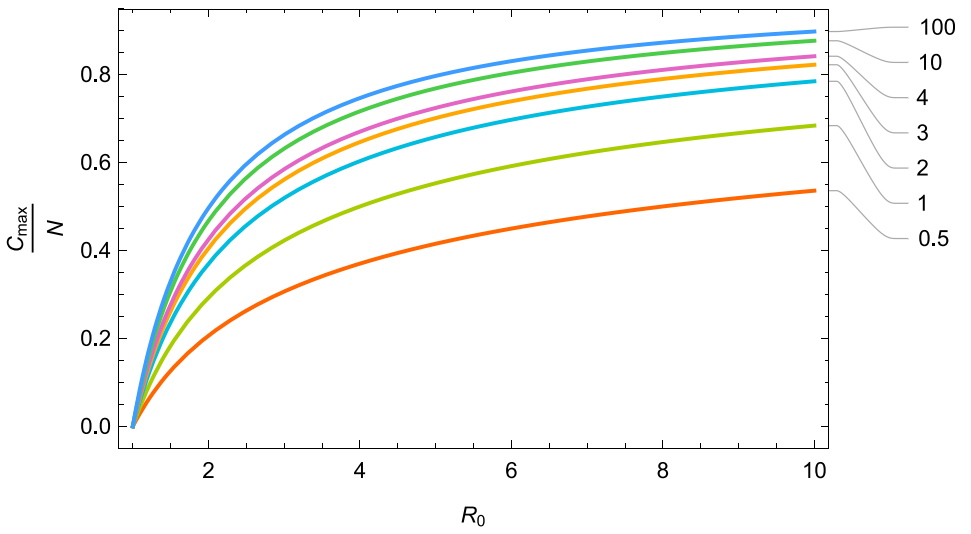

Fig. 4. HIT $\frac{C_{max}}{N}$ of a heterogeneous population as a function of R_0, for different values of α (indicated on the right side of the plot), in the case of a one-parameter gamma distribution.

We start from the differential equation for β, obtained by taking the derivative of J. Indeed, this provides a differential equation for β (more precisely, for $\ln\beta$) if $\ln J(t)$ is given, which does not require knowledge of the amplitude of J. As $J(t) = J_{\exp}$, we get

$$\dot{\beta} = \beta \left[\frac{d}{dt} \ln J_{\exp} - \frac{\dot{I}}{I} + \left(\frac{\alpha+1}{\alpha} \right) \left(1 + \frac{\tau}{\alpha} \right)^{-1} \beta I \right], \tag{43}$$

with

$$\frac{\dot{I}}{I} = \frac{dI}{d\tau} \frac{d\tau}{dt} \frac{1}{I} = \left[S_0 \left(1 + \frac{\tau}{\alpha} \right)^{-1-\alpha} - \frac{\gamma}{\beta} \right] \beta, \tag{44}$$

where we have used (30) and (36). Moreover, recall that, integrating the latter, we find (37). Plugging all of this back into (43), we get the following differential equations:

$$\begin{aligned} \dot{\beta} &= \beta \left\{ \frac{d}{dt} \ln J_{\exp} - \left[S_0 \left(1 + \frac{\tau}{\alpha} \right)^{-1-\alpha} - \frac{\gamma}{\beta} \right] \beta \right. \\ &\left. + \left(\frac{\alpha+1}{\alpha} \right) \left(1 + \frac{\tau}{\alpha} \right)^{-1} \beta \left[I_0 + S_0 \left(1 - \left(1 + \frac{\tau}{\alpha} \right)^{-\alpha} \right) - \frac{\gamma}{\beta} \tau \right] \right\}, \\ \dot{\tau} &= \beta \left[I_0 + S_0 \left(1 - \left(1 + \frac{\tau}{\alpha} \right)^{-\alpha} \right) - \frac{\gamma}{\beta} \tau \right]. \end{aligned} \tag{45}$$

These have to be solved while performing the fitting of J from the model, that is

$$J(\tau) = \beta S_0 \left[I_0 + S_0 \left(1 - \left(1 + \frac{\tau}{\alpha} \right)^{-\alpha} \right) - \frac{\gamma}{\beta} \tau \right] \left(1 + \frac{\tau}{\alpha} \right)^{-\alpha-1}, \tag{46}$$

with the data $J(t) = J_{\exp}$.

3.2. HIT for the case of a two-parameter gamma distribution

We will now present and analyze a novel generalization of the above discussion to the case of the usual two-parameter gamma distribution. Therefore we assume that $s_0(x)$ is gamma-distributed with shape parameter $\alpha > 0$ and rate parameter $\eta > 0$,

$$s_0(x) = S_0 \frac{\eta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\eta x}. \tag{47}$$

Note that the case previously discussed is obtained from the above setting $\eta = \alpha$. For the Laplace transform (31) we get

$$S(\tau) = \frac{S_0}{\left(1 + \frac{\tau}{\eta} \right)^\alpha} = S_0 \frac{\eta^\alpha}{(\eta + \tau)^\alpha}. \tag{48}$$

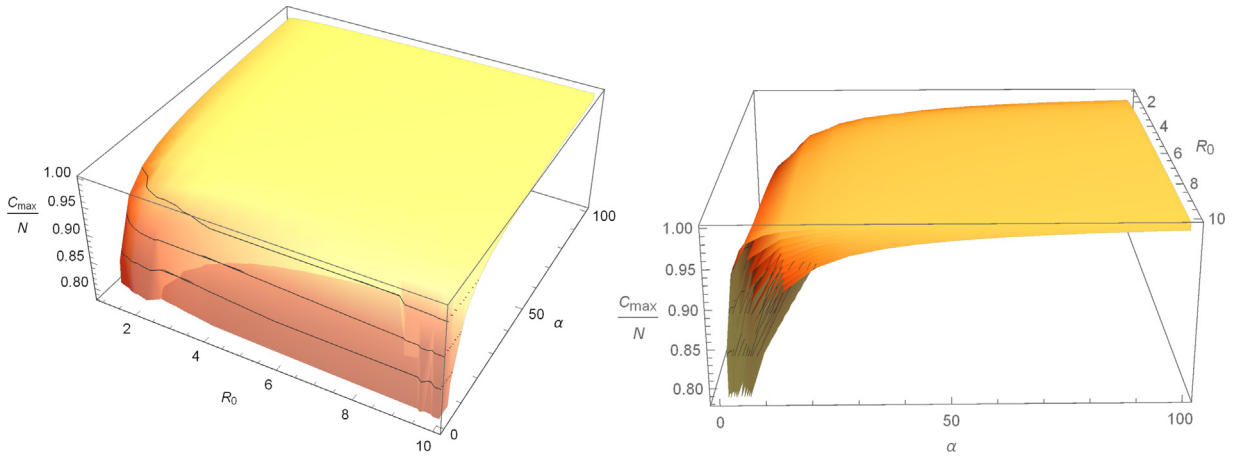

Fig. 5. Three-dimensional plot of the HIT $\frac{C_{max}}{N}$ of a heterogeneous population as a function of R_0 and α in the case of a two-parameter gamma distribution, for $1 \leq \eta \leq 10$.

Since $s(x, \tau) = s_0(x)\exp(-x\tau)$, now for the average susceptibility (25) we obtain

$$\bar{x}(\tau) = \frac{1}{\frac{\eta}{\alpha} + \frac{\tau}{\alpha}} = \frac{\alpha}{\eta + \tau}\,. \tag{49}$$

Note that for $\tau = 0$ we have $\bar{x}(0) = \alpha/\eta$. Then, using (48) and (49), Eq. (26) can be rewritten as

$$\frac{dI}{d\tau} = S_0 \alpha \eta^{\alpha} (\eta + \tau)^{-1-\alpha} - \frac{\gamma}{\beta}\,, \tag{50}$$

which, integrated, yields

$$I(\tau) = I_0 + S_0 \left[-\frac{\eta^{\alpha}}{(\eta + \tau)^{\alpha}} \right] - \frac{\gamma}{\beta}\tau\,. \tag{51}$$

The latter has a maximum for

$$\eta^{-\alpha} \frac{(\eta + \tau)^{1+\alpha}}{\alpha} = R_e\,. \tag{52}$$

Consequently, considering $S_0 \approx N$, the cumulative number of infections reads

$$C(\tau) = N - S(\tau) = N \left[1 - \frac{\eta^{\alpha}}{(\eta + \tau)^{\alpha}} \right], \tag{53}$$

which, evaluated at the maximum of I, becomes

$$\frac{C_{max}}{N} = 1 - \eta^{\alpha} \left(\alpha \eta^{\alpha} R_0\right)^{\frac{-\alpha}{1+\alpha}} = 1 - \left(\frac{\alpha}{\eta} R_0\right)^{\frac{-\alpha}{1+\alpha}}\,. \tag{54}$$

This is the generalization of the herd immunity threshold (9) to the case of a heterogeneous population with an initial susceptibility given by a two-parameter gamma distributions. In the limit $\alpha/\eta \to 1$ we recover the HIT of the particular case of the initial one-parameter gamma distribution (41). The HIT homogeneous for a homogeneous population (9) is recovered in the limit $\alpha/\eta \to 1$, $\alpha \to \infty$. Besides, we can observe that, in case of the initial two-parameter gamma distribution case, a redefinition of R_0 emerges. Indeed, (54) may be rewritten as

$$\frac{C_{max}}{N} = 1 - \hat{R}_0^{\frac{-\alpha}{1+\alpha}}\,, \tag{55}$$

where we have introduced

$$\hat{R}_0 = \hat{R}_0(R_0, \alpha, \eta) := \frac{\alpha}{\eta} R_0\,. \tag{56}$$

In Fig. 5 we give a three-dimensional plot of the HIT as a function of R_0 and α as η runs from 1 to ten. We can see that this outlines different surfaces depending on the value of η. The larger η, the higher the threshold is raised to achieve herd immunity, with the same R_0 and α. This is better highlighted in Fig. 6, where we plot the herd immunity level $\frac{C_{max}}{N}$ of a heterogeneous population, with $R_0 = 2.5$, as a function of α for two different values of η (we take as sample values $\eta = 2$ and $\eta = 9$).

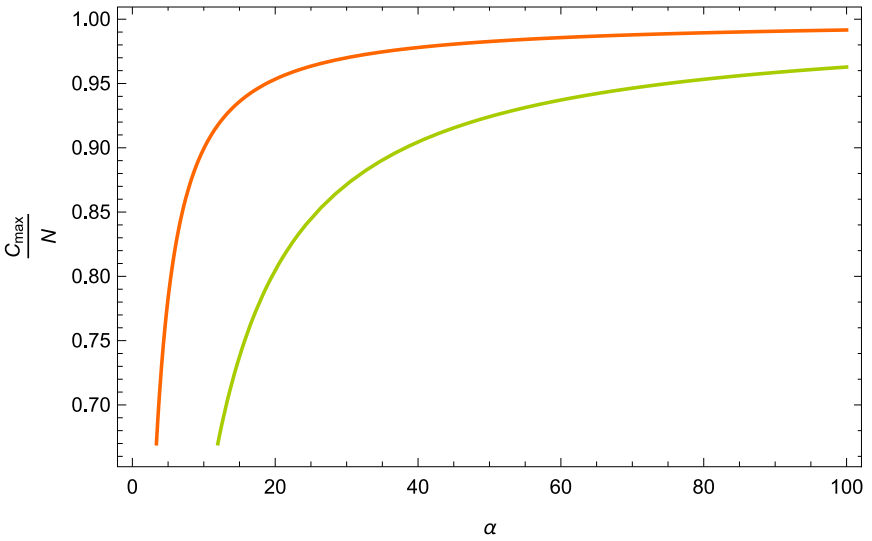

Fig. 6. HIT $\frac{C_{max}}{N}$ of a heterogeneous population, for $R_0 = 2.5$, as a function of α in the case of a two-parameter gamma distribution, for two different values of η. The orange curve originates from taking $\eta = 2$, while the green one from $\eta = 9$.

Of course, as it is commonly understood also for R_0, for the $\hat{R}_0 = \hat{R}_0(R_0, \alpha, \eta)$ defined in (56) we have that if $\hat{R}_0 < 1$, each existing infection causes less than one new infection. In this case, the disease will decline and eventually die out. If $\hat{R}_0 = 1$, each existing infection causes one new infection. The disease will stay alive and stable, but there will not be an outbreak or an epidemic. On the other hand, if $\hat{R}_0 > 1$, each existing infection causes more than one new infection. The disease will be transmitted between people, and there may be an outbreak or epidemic.

For the sake of completeness, let us also say that the maximum value of I in this case is

$$\frac{I}{S_0} = \frac{I_0}{S_0} + \frac{\eta}{R_e} - \frac{1}{R_e}\left[(\alpha\eta^\alpha R_e)^{\frac{1}{1+\alpha}} - \eta^\alpha R_e\,(\alpha\eta^\alpha R_e)^{\frac{-\alpha}{1+\alpha}}\right]. \tag{57}$$

All of this suggests that the parameter η may take into account mitigation measures, which in fact affects both the average susceptibility (in particular, the initial one) and the HIT. Such mitigation measures can be even natural, that is driven by the epidemic history of the human being. For fixed R_0 and α, decreasing η, which can also be understood as increasing the scale parameter θ of the initial gamma distribution, can be translated into contrasting heterogeneity effects, rising the HIT, and vice versa. For fixed R_0, if η is bigger enough with respect to α, its presence lowers the herd immunity threshold. For instance, in the two-parameter initial gamma distribution case, for an R_0 between 2.5 and 4 and values of α between 0.5 and 1, herd immunity is expected to require 7–29 percent of the population for η between 1 and 2, with respect to the range 60–75 percent of a homogeneous population, with the same R_0, and the range 26–50 percent for a heterogeneous population with $\eta = \alpha$ between 0.5 and 1.

4. Final remarks

In this paper we have elaborated on SIR-type models of epidemics, especially with respect to to population heterogeneity effects, which capture the fact that the probability of being infected with a transmissible disease is not the same for all individuals of a population.

We have first reviewed homogeneous SIR-models, presenting an intriguing correspondence between the epidemic trajectory in a homogeneous SIR-type model and radial null geodesics in the Schwarzschild spacetime. The correspondence between the SIR model and radial null geodesics in the Schwarzschild geometry may be useful to obtain new analytically tractable epidemiological models, e.g. by considering timelike and/or nonradial geodesics, or null geodesics in the Reissner–Nordström metric case, analyzing the predictions of epidemiological models in this context. Moreover, one could construct an action principle that gives rise to the Eqs. (3).

Subsequently, we have analyzed modeling of population heterogeneity effects. We have first reviewed and discussed the case in which the initial susceptibility is given in terms of a one-parameter gamma distribution, deriving the associated HIT. The study has been done without considering mitigation measures and reinfections, which means, in particular, that the threshold to reach herd immunity has been estimated by considering natural infections without restrictions on individuals (lockdown, social distancing, etc.) and without taking into account possible vaccinations. Consequently, in the same setup we have also described a possible way to take into account mitigation measures when performing model fitting. Our proposal consists in considering a time-dependent infection rate $\beta = \beta(t)$ to derive a differential equation for

it (actually, for $\ln\beta$) to be solved while implementing the fitting of J from the model with J_{exp} from data (J is the number of new cases per day).

Finally, we have derived the HIT in the case of an initial two-parameter gamma distribution. We find that the additional parameter (η) induces a shift of R_0, which can be interpreted as a possible way of taking into account mitigation measures in this setup. One could also say that η allows to take into account the effects of whatever makes a population more or less heterogeneous.

It would be interesting to extract expressions for S, I or τ in the limit of small times. Such a solution, which shows sub-exponential growth, would be important to model the initial stages of a epidemic before mitigation measures are taken. Elements of related analysis can be found in [21,22].

To conclude, although the gamma distribution has been shown to be better than other ones in developing epidemiology models, it would be interesting to try to compute analytically the HIT for distributions other than gamma, in the context of heterogeneous SIR-type models. Consequently, a fundamental study would be comparison with data. Besides, a critically important question is: how variable are humans in their susceptibility and exposure to transmissible diseases (such as SARS-CoV-2)? Hitherto, there is no definite answer to this question. Such issue, which, in particular, can be translated, in the context of this paper, to determine the value of α from model fitting, is left to a future investigation on heterogeneous models of epidemic.

CRediT authorship contribution statement

Silke Klemm: Conceptualization, Methodology, Computations and analysis, Writing – original draft. **Lucrezia Ravera:** Conceptualization, Methodology, Computations and analysis, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

We thank F. Lingua for stimulating discussions. This work was supported partly by INFN, Italy and by MIUR-PRIN contract 2017CC72MK003. L. R. would like to thank the Department of Applied Science and Technology of Politecnico di Torino, Italy and the INFN, Italy for financial support.

References

[1] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. Roy. Soc. A 115 (1927) 700.
[2] W. Hethcote Herbert, The mathematics of infectious diseases, SIAM Rev. 42 (4) (2000) 599–653.
[3] J.D. Murray, Mathematical biology, in: Interdisciplinary Applied Mathematics, third ed., Springer, New York, 2002.
[4] R.E. Mickens, An exactly solvable model for the spread of disease, College Math. J. 43 (2012) 114.
[5] R. Beckley, C. Weatherspoon, M. Alexander, M. Chandler, A. Johnson, G.S. Bhatt, Modeling Epidemics with Differential Equations, Tennessee State University Internal Report, 2013.
[6] T. Harko, F.S. Lobo, M.K. Mak, 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 (2014) 184–194, arXiv:1403.2160.
[7] M. Kröger, R. Schlickeiser, Analytical solution of the SIR-model for the temporal evolution of epidemics, Part A: Time-Independent Reprod. Factor J. Phys. A 53 (50) (2020) 505601.
[8] R. Schlickeiser, M. Kröger, Analytical solution of the SIR-model for the temporal evolution of epidemics, Part B: Semi-Time Case, J. Phys. A 54 (17) (2021) 175601.
[9] M.G.M. Gomes, R.M. Corder, J.G. King, K.E. Langwig, C. Souto-Maior, J. Carneiro, G. Gonçalves, C. Penha-Gonçalves, M.U. Ferreira, R. Aguas, Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold, J. Theoret. Biol. 540 (2022) 111063, http://dx.doi.org/10.1101/2020.04.27.20081893, Epub 2022 Feb 18. PMID: 35189135; PMCID: PMC8855661 (medRxiv).
[10] R. Aguas, G. Gonçalves, M.U. Ferreira, M.G.M. Gomes, Herd immunity thresholds for SARS-CoV-2 estimated from unfolding epidemics, medRxiv. http://dx.doi.org/10.1101/2020.07.23.20160762.
[11] A. Montalbán, R.M. Corder, M.G.M. Gomes, Herd immunity under individual variation and reinfection, J. Math. Biol. 85 (1) (2022) 1–22, [q-bio] arXiv:2008.00098.
[12] J. Neipel, J. Bauermann, S. Bo, T. Harmon, F. Jülicher, Power-law population heterogeneity governs epidemic waves, PLoS One 15 (10) e0239678, [q-bio] arXiv:2008.00471.
[13] J. Ma, D.J.D. Earn David, Generality of the final size formula for an epidemic of a newly invading infectious disease, Bull. Math. Biol. 68 (3) (2006) 679–702.
[14] M.G. Roberts, Epidemic models with uncertainty in the reproduction number, J. Math. Biol. 66 (2013) 1463–1474.
[15] R.I. Hickson, M.G. Roberts, How population heterogeneity in susceptibility and infectivity influences epidemic dynamics, J. Theoret. Biol. 350 (2014) 70–80.

[16] J. Duchesne, O.A. Coubard, First-wave COVID-19 daily cases obey gamma law, Infect. Dis. Model. 7 (2) (2022) 64–74, http://dx.doi.org/10.1016/j.idm.2022.02.004, Epub 2022 Mar 11. PMID: 35291224; PMCID: PMC8912979.
[17] J. Dehning, J. Zierenberg, F.P. Spitzner, M. Wibral, J.P. Neto, M. Wilczek, V. Priesemann, Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions, Science 369 (6500) (2020) eabb9789, [q-bio.PE] arXiv:2004.01105.
[18] S. Flaxman, S. Mishra, A. Gandy, H.J.T. Unwin, T.A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J.W. Eaton, M. Monod, Imperial College COVID-19 Response Team, A.C. Ghani, C.A. Donnelly, S. Riley, M.A.C. Vollmer, N.M. Ferguson, L.C. Okell, S. Bhatt, Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe, Nature 584 (7820) (2020) 257–261, [stat.AP] arXiv:2004.11342. Epub 2020 Jun 8. PMID: 32512579.
[19] S.M. Kissler, C. Tedijanto, E. Goldstein, Y.H. Grad, M. Lipsitch, Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period, Science 368 (6493) (2020) 860–868, May 22; 368(6493)860-868. Epub 2020 Apr 14. PMID: 32291278, PMCID: PMC7164482.
[20] A.S. Novozhilov, On the spread of epidemics in a closed heterogeneous population, Math. Biosci. 215 (2) (2008) 177–185.
[21] G. Dwyer, J.S. Elkinton, J.P. Buonaccorsi, Host heterogeneity in susceptibility and disease dynamics: tests of a mathematical model, Am. Nat. 150 (6) (1997) 685–707, PMID: 18811331.
[22] R.M. May, R.M. Anderson, M.E. Irwin, The transmission dynamics of human immunodeficiency virus (HIV) [and discussion], Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 321 (1207) (1988) 565–607.