medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

A probabilistic approach for the study of epidemiological dynamics of infectious diseases: basic model and properties

José Giral Barajas[1*], Carlos Ignacio Herrera-Nolasco[1,2*], Marco Arieli Herrera-Valdez[1,2✉], Sergio I. López[1✉]

[1] Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México.

[2] Laboratorio de Dinámica, Biofísica, y Fisiología de Sistemas, Facultad de Ciencias, Universidad Nacional Autónoma de México.

[*] Authors with equal contributions.

✉ silo@ciencias.unam.mx, marcoh@ciencias.unam.mx

Contents

1 Introduction **2**

2 Modelling framework for stochastic epidemiological dynamics **5**

2.1 Working definitions and important conceptual distinctions 5

2.2 Modelling rationale. 7

2.2.1 Epidemiological sampling and state switching 8

2.2.2 Time-dependent switch in epidemiological state 9

2.2.3 The start of the chain of transmission: switching from susceptible to infected 9

2.3 Infections and possibly fatal outcomes 10

2.3.1 Probabilities for recovery and death 12

2.3.2 The routing probability and the mean infection time 13

2.4 Continuous time Markov chain approximation 15

3 Applications **19**

3.1 The classical SIR model from the basic sED model 19

3.1.1 Is it possible to write a simple and well posed extension for the continuous SIR to account for deaths? 20

3.2 Limiting the exposure of the susceptible vs. limiting the exposure of the infected 21

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

3.3 Case fatality ratios and the necessity of considering infectious with different severity stages . . . 23
3.3.1 Extension to consider infections with multiple states . . . 23
3.3.2 Probability distribution of the final stage reached by an individual 26

4 Discussion **27**
4.1 Why do we need a different model than the SIR model? . . . 28
4.2 The exposure of susceptible and infected have different effects on the epidemic dynamics. . . . 29
4.3 Infection vs infectivity . . . 30
4.3.1 Multiple stages for infection and illness progression. . . . 31
4.4 On numerical simulations and approximations . . . 32
4.5 Concluding remarks . . . 32

Abstract

The dynamics of epidemiological phenomena associated to infectious diseases have long been modelled with different approaches. However, recent pandemic events exposed many areas of opportunity to improve the existing models. We develop a model based on the idea that infection and also, transitions between different stages during an infection and possibly the disease that it may cause, are alike sampling processes. The resulting modelling scheme proves to be easy to implement but very robust, and easy to extend to answer questions that emerge from close examination of data trends in the COVID-19 pandemic, and other infectious diseases.

1 Introduction

The epidemiological dynamics of infectious diseases are random, nonlinear, and result from a combination of phenomena occurring in different levels of biological organisation. A variety of theoretical approaches, of which the most notable is the family of SIR-type models (Kermack and McKendrick, 1927) has proven useful to study general aspects of epidemics. However, the existing deterministic and stochastic classes of epidemiological models based on equations, at large, fail to offer reasonable links between the mathematical expressions that make up the models and the underlying biochemical or biophysical mechanisms they are supposed to capture from a macroscopic perspective.

For instance, there are stochastic models capable of describing different aspects of the dynamics in epidemics. The work of (Greenwood, 1931) analyses the pertinence

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

of different probabilistic models as explanatory mechanisms of a disease spreading at some given households. The number of cases of the disease registered in a period of time at each household are used as input for the models, hence having the possibility of measure the fit of each model to real data. More recently, (Greenwood and Gordillo, 2009) review compartmental stochastic models, focused on the distribution of the final size of the epidemics, the stochastic periodicity of the number of infectious individuals and the random period that an individual remains infectious after been infected, questions which arise naturally in the random setting. The model by (Tuckwell and Williams, 2007) assumes discrete time with Markovian infection dynamics that occur as binomial processes within a constant population. Some surveys about similar stochastic models can be found in the work of (Allen, 2017) and (Britton, 2010).

Significant improvements to study epidemiological dynamics can be made by replacing one or more of the phenomenological construction postulates of the SIR family (e.g. homogeneous mixing), by making assumptions and deriving expressions explicitly based on physical, biological, or biochemical principles, however macroscopic. For instance, the infection rate is typically assumed to be proportional to the prevalence of infection (density of infected). Such assumption may sound reasonable at first and it may even be possible to fit a multiple of prevalence to case data after mild corrections for testing and other factors. However, the proportionality constant implicit in the assumption depends on factors that include the source of inoculation (e.g. a water reservoir), the collective behavioural patterns of people susceptible to infection and those infected, and the particular characteristics of the individuals in those populations (e.g. obesity, age, gender).

Of note, a more explicit description of the dependence of the infection probability on the availability of inoculum would be useful to assess different questions related to mitigation strategies. The inoculum available at any point in time is particularly difficult to assess, as it depends on factors like the number of infected people and their shedding levels, the pathogen's infectiousness and transmission mechanisms (e.g. aerosol, droplets, direct contact).

Another issue of importance that is often ignored is the difference between an infection process, typically detected by laboratory tests, and the natural and progressive evolution of the disease possibly caused by that infection, which may be detected as symptoms develop. However, the clinical manifestations possibly caused by an infection are not necessarily detected, as it is the case with diseases with asymptomatic phases. This is a problem from an epidemiological stand point because modelling of these two phenomena is usually lumped

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

together by construction (see (Allen, 2008) for example), limiting the scope of models and the resulting dynamics (Shayak and Sharma, 2021). By extension, decision-making by health authorities and government officials is hindered by the lack of connection between models and the reality driven by biological (Holmdahl and Buckee, 2020) and sociological factors (Eker, 2020; Roozenbeek et al., 2020), to name a few.

Infections can be regarded as pathogen invasions at the cellular and possibly tissue levels (e.g. start of viral replication in susceptible cells). Hence, infections by any disease-causing pathogen do not necessarily result in confirmed cases, but may be the main underlying cause of large epidemic outbreaks (Gerba, 2009). Examples of communicable diseases of current importance due to their large burden on economical, social, and health systems, that are asymptomatically transmitted, and therefore difficult to keep under control include malaria (Bousema et al., 2014), dengue (Grange et al., 2014), influenza (Cohen et al., 2021), COVID-19 (Gandhi et al., 2020), aids (Hollingsworth et al., 2008), among many well documented examples.

Models that explicitly make distinctions between infectious and illness states, or between parameters like the exposure of susceptibles versus inoculum shedding by infected. For instance, it should be possible to use models constructed from a macroscopic perspective to explain seemingly paradoxical situations in which infections may cause severe illness or even death in older people and young athletes, but not in children, as it has been observed during the COVID-19 pandemic. Of particular interest, there is a lack of computationally simple, mathematically robust models based on by pathophysiological considerations that distinguish the clinical outcome from infections taking into account different susceptibility profiles (Butler and Barrientos, 2020; Fakhroo et al., 2020; Fricke-Galindo and Falfán-Valencia, 2021), or explaining different severity in the clinical outcomes of reinfections depending on the time of the last infection (Katzelnick et al., 2016; Montoya et al., 2013).

As a starting step, in this paper we present a modelling framework based on probabilistic considerations and a simple concept: individuals in a given epidemiological stage are sampled according to physical or biological principles, possibly switching between stages that are common to all individuals, at least in principle. For instance, all people become infected before becoming ill because of the infection, if they do so at all. Similarly, before recovery, pathologies typically develop to a point in which the immune system takes over and clears the pathogen; and so on. We start by giving basic descriptions for macroscopic, necessary factors involved in the transmission of infectious diseases, and make punctual distinctions between them. We then proceed to present a simple model for epidemiological

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

dynamics, hereby called sED, using a paradigm motivated by the SIR class of models. Once the construction of the basic model is explained, we demonstrate a few of its mathematical properties and illustrate some of its advantages over classical epidemiological models, including the possibility of modelling infection spread in small populations with integer state variables. We also show that the sED recovers different types of deterministic SIR models under particular sets of assumptions, and illustrate the application of the sED model with specific examples. We finish the article showing how the modelling framework we propose can be extended.

2 Modelling framework for stochastic epidemiological dynamics

2.1 Working definitions and important conceptual distinctions

Susceptible individuals are, by definition, those that can be infected. *Exposed* individuals are those in contact with the pathogen by means of direct or indirect interaction. Exposure may change over time depending on different factors like physical distancing policy, personal mobility, the time of the day, etcetera. Naturally, only susceptible individuals that are *exposed* can become infected, and an individual is considered *infected* if the pathogen finds a susceptible tissue and starts replicating. After some time in which the pathogen replicates, infected individuals become able to transmit the pathogen and start contributing to the chain of transmission. The time relative to the initial inoculation at which this happens is a random variable (Fenner et al., 1987). The *incubation* time can be regarded as the time interval between the start of replication and the start of pathological processes associated to pathogen invasion (Li, 2010, p.25). However, since symptoms are usually the way in which a pathology is detected, the incubation time is usually referred in medical epidemiology as the time lapse between the suspected time of infectious contact and the start of symptoms (Dicker et al., 2006; He et al., 2020). This is typically a problem since infected people may not be symptomatic (they do not show obvious symptoms), as their infection progresses. In the case that disease ensues, the severity may progress from mild to worse until recovery or death.

The time individuals spend being infected is a random variable that often is modelled as a gamma-distributed random variable (Anderson and Watson, 1980; Feng et al., 2007; Lloyd, 2001; Vazquez, 2021). For practical purposes, the infection time can be thought of

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

as a sum of an initial replication period of the pathogen causing the infection, and a second period during which an individual may have detectable symptoms. The individual may be infectious shortly after the first pathogen invasion, during the incubation period, and before symptoms appear (e.g. COVID-19). This can be done more generally by dividing the infectious period in several different stages. After the infection period, individuals may develop other phases of disease that are not necessarily infectious. For instance, in both symptomatic and asymptomatic COVID-19, the initial viral phase might be followed by an inflammatory phase (García, 2020; Manjili et al., 2020) where the probability of infectious virus shedding decreases dramatically (van Kampen et al., 2021). A longitud study of people infected with SARS-CoV-2 (Boucau et al., 2022) suggest the possibility of shedding culturable virus for more than 5 days -possibly till 15 days- after symptom onset or an initial PCR positive test. In the best case scenario, the immune system develops a protection profile against the pathogen and the individual recovers. In a worse case scenario, the immune system may over-react, or other opportunistic diseases may also appear, causing severe illness and potentially death.

Individuals can also be born immune or become immune for reasons that include having beneficial mutations in their genome (Allers and Schneider, 2015), because of acquired immunity upon exposure to other pathogens (Doolan et al., 2009), or after vaccination (Clem, 2011).

The immune system of an individual can be determinant of her physiological responses to particular disease. Subsequently, the physiological state of the individual during the initial phase can also affect the evolution of the infection and the disease that may follow it, possibly influencing different subsequent phases of the infection or disease. Hence, the individual's immune system profile is essential to evaluate her possible particular disease evolution, regardless of whether key characteristics of the immune profile are available, measurable, or understood at all.

In consideration of the observations made above, we developed a hierarchy of models that can be reduced or extended depending on the level of detail sought and the questions of interest. Importantly, this hierarchy of models can be used to model disease spread in small and large populations that may be composed of individuals with different levels of susceptibility (see companion paper). One case of special interest to us that these models address the study of epidemiological dynamics including reinfection and the possibility of becoming more or less susceptible to reinfection and severe disease with subsequent infections.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2.2 Modelling rationale.

We regard any epidemiological classification of individuals from a population as a dynamical partition into subsets that change in time. In turn, the subsets that make up the partition can be represented as nodes in a directed graph. For visualization purposes, nodes and edges can be respectively weighed according to the size of the population they represent, and the probability that the epidemiological stage that an individual changes to the one pointed to by that edge (Fig. 1).

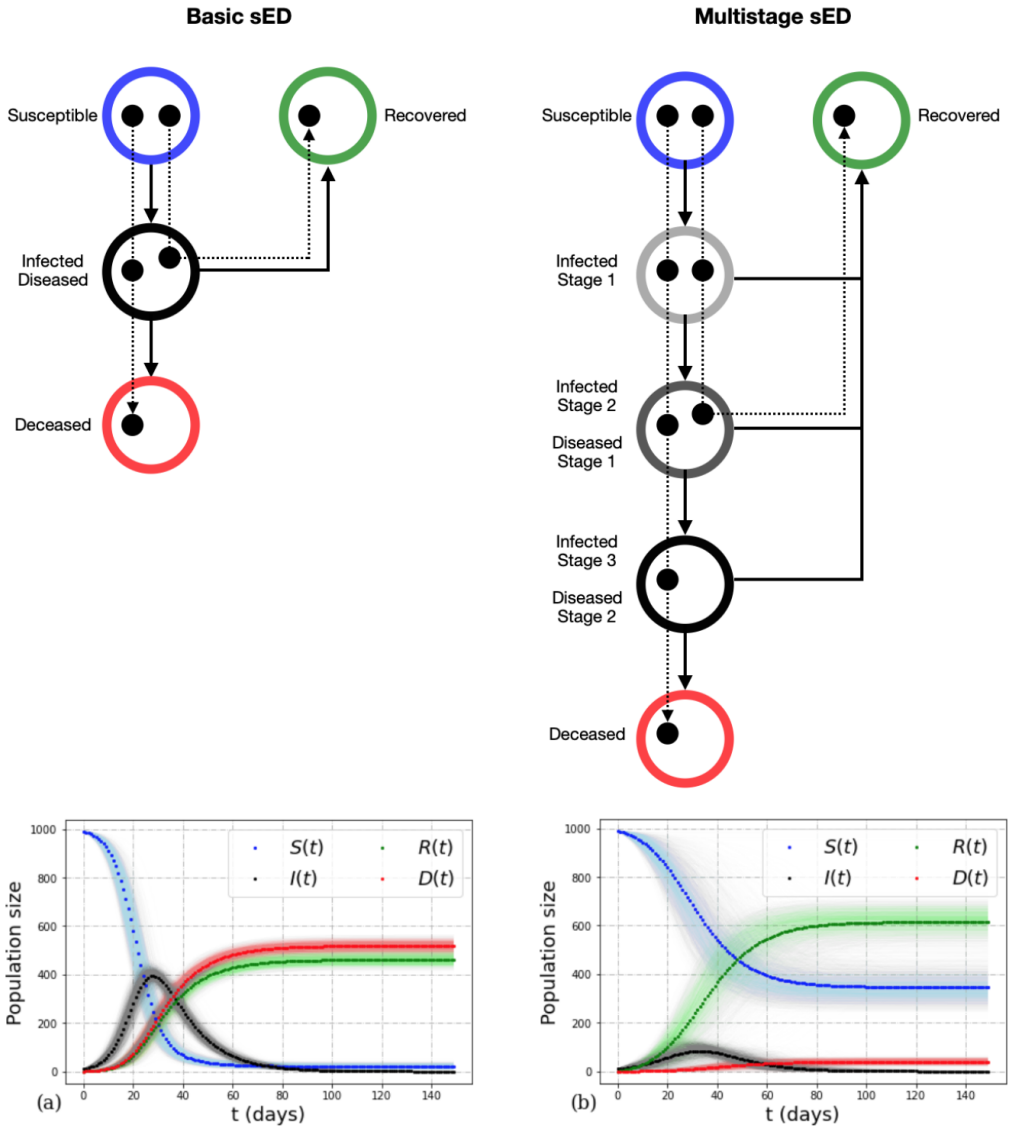

Figure 1: Different dynamical scenarios contemplated by the sED modelling scheme. The light shades represent the confidence intervals generated with 2,000 simulations and the solid dots are taken as the average of those simulations. The two models show different epidemiological dynamics with the same average recovery and death time, but different time courses for infection and disease before recovery or death. See Tables 1 and 2 for further information.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The epidemiological dynamics at the population-level can then be based on two simple but powerful concepts.

- First, each infection can be thought of as a *one-way, path* starting at the node representing the susceptible sub-population, and ending at one of the terminal nodes in the graph, which represent, by construction, the state for death due to infection, or the recovered state.

- Second, the epidemiological state $s(t)$ of any individual may change during the time interval $[t, t+h]$ with a certain probability $p_s^h(t)$, and remain the same with probability $1 - p_s^h(t)$; such changes should occur independently, where s is a label indicating the type of switch (e.g. from susceptible to infected).

2.2.1 Epidemiological sampling and state switching

The epidemiological state s of any individual at time $t+h$ can be thought of as the result of *routing* individuals via *Bernoulli* trials occurring between times t and $t+h$, with probability $p_s^h(t)$. The number of people that change their epidemiological state between times t and $t+h$ is thus the number of successful trials in the *Bernoulli* sample drawn at time t. If circumstances allow assuming independence between switching events for different individuals, then the number of people who switch can be regarded as a *Binomial random variable* in which the number of trials is the number of individuals initially considered. If there is more than one possible new epidemiological state to switch to, $p_s^h(t)$ can regarded as a sum of the probabilities of switching to the new different states, and the individuals that switch to the different states can be thought of as a *Multinomial* sample with a number of trials equal to the number of people in state s at time t. The epidemiological dynamics of a population can then be modelled by keeping track the individual changes in epidemiological state, and updating the number of people in the different stages after each time step.

Reasonable approximations for *populations* can be derived by assuming homogeneity among individuals (e.g. similar co-morbidities). For instance, in some cases it would be reasonable to assume that the average probability of switch describes the population dynamics, and therefore, sampling for those individuals switching epidemiological states can be done by assuming that success occurs with a probability equal to the average probability of switching between epidemiological states.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2.2.2 Time-dependent switch in epidemiological state

Let τ_s be a random variable representing the waiting time for a switch in state s after infection has begun. The probability of switching within h time units can then be written as $p_s^h(t) = P_s(h)Q_s(t,h)$ where $P_s(h)$ represents the probability that the switch occurs within h time units, and $Q_s(t,h)$ represents the probability that additional, necessary conditions for the switch occur at time t. Assuming that the time τ_s for a switch from the epidemiological state s is an exponential random variable with mean waiting time μ_s,

$$P_s(h) = P(\tau_s < h) = 1 - \exp\left(-h/\mu_s\right). \tag{1}$$

Numerical simulations can then be accelerated by noticing that the probability $P_s(h) \approx h/\mu_s$ for small enough h (Fig. 2, red curves) but the approximation diverges quickly for $\mu_s < 1$ (Fig. 2, blue curves).

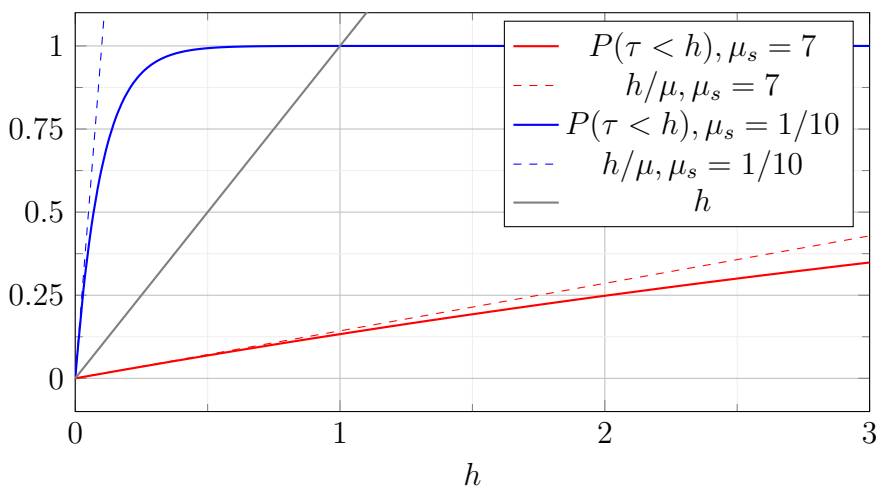

Figure 2: Geometric approximation of exponential waiting times with mean switching time $\mu_s \in \left\{\frac{1}{5}, 7\right\}$.

We now start deriving a hierarchy of models capable of capturing different aspects of epidemiological dynamics, showing some of their properties and advantages over classical models. We do so taking into account the characteristics and data reported for COVID-19.

2.2.3 The start of the chain of transmission: switching from susceptible to infected

Suppose that every member of a population is either susceptible to infection, non-susceptible, or infected by a certain pathogen (e.g. SARS-CoV-2), at any point in time. Let non-negative integers $T(t)$, $S(t)$, and $I(t)$ represent the sizes of the whole population, those susceptible

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

to infection, and those infected at time t (in days). Prior to any infection, $T(t) - S(t)$ is the size of the non-susceptible, *i.e.* those that do not participate in the chain of infections.

Assume independence between individual infections are *independent*, with an average probability of infection $p^h(t, I)$ that depends on factors that include the time of exposure h and the availability inoculum, which can be thought of as a monotonically increasing function of I. In addition, suppose that each of the exposed susceptibles at time t can be infected within h time units of exposure with probability $p^h(t, I)$, or remain susceptible with probability $1 - p^h(t, I)$. The number $X_I^h(t)$ of newly infected people (absolute incidence) can then be thought of as a *Binomial* random variable sampled from $S(t)$ individuals. One possibility is to define

$$p^h(t, I) = P_I(h)Q_h(t, I) = P_I(h)\beta\frac{I(t)}{T(t)} \tag{2}$$

where β is the probability that a susceptible person is infected after having an infectious contact (contact with the pathogen), $I(t)/T(t)$ represents the proportion of infected individuals within the population at time t, and $P_I(h)$ represents the probability that the infection occurs within a time window of length h, while being *exposed* to the pathogen causing the disease of interest. The number of newly infected individuals between times t and $t + h$ can be thought of as a random variable $X_I^h(t)$ with $Binomial\left(S(t), p^h(t, I)\right)$ distribution, and the expected number of newly infected individuals would then be $p^h(t, I)S(t)$. We will say that a new epidemic starts when the expected number of infections becomes an increasing function of time.

The number of susceptible people at a time $t + h$ can then be calculated as,

$$S(t + h) = S(t) - X_I^h(t), \tag{3}$$

which means that the expected change between $S(t)$ and $S(t + h)$ is a decreasing function of t, for all $h > 0$.

We now proceed to write a first model that takes into account the possibility of death due to disease,

2.3 Infections and possibly fatal outcomes

Assume that an initial number of infections $I(0) = I_0$ are caused by a primal source of pathogens, and that those infections are transmissible to the rest of the population. Assume

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

also that at any point in time, infected individuals shed an average inoculum per unit time. For now, all susceptible and infected individuals are assumed to participate in the chain of infection without restrictions. In a later section, we will explore the effects of extending the model to explicitly include exposure factors for susceptible and infected, separately.

Immunity, and death. For this first model, assume that some of the infected individuals die after becoming infected, and the rest recover and become immune. Notice that the assumptions above do not take into consideration any clinical aspects associated to the infection, except the possibility of eventual death as a consequence of infection. Let $Y_R^h(t)$ and $Y_D^h(t)$ respectively represent the numbers of people that that recover or die between t and $t+h$. Those that remain infected after that time interval are then

$$I(t+h) = I(t) - Y_D^h(t) - Y_R^h(t). \tag{4}$$

Note that the probabilities that a person is removed from the infected group (i.e. recovers or dies), or that the person remains infected within a small time interval of length h, do not depend on the state of the epidemics at time t but only on the physiological state of the individual facing the infection. Therefore, assuming independence between infections for different individuals, $Y_R^h(t)$ and $Y_D^h(t)$ can be thought of as outcomes from a *Multinomial* sample from $I(t)$, with probabilities p_R^h and p_D^h, for $t \in [t, t+h]$. As a consequence, the triplet

$$\left(Y_R^h(t), Y_D^h(t), I(t) - Y_R^h(t) - Y_D^h(t)\right) \sim \text{Mult}\left(I(t), p_R^h, p_D^h, 1 - p_R^h - p_D^h\right),$$

and the dynamics of the infected, recovered, and dead at time $t+h$ can be written as

$$\begin{aligned} I(t+h) &= I(t) + X_I^h(t) - (Y_R^h(t) + Y_D^h(t)), \\ R(t+h) &= R(t) + Y_R^h(t), \\ D(t+h) &= D(t) + Y_D^h(t). \end{aligned} \tag{5}$$

where $R(t)$ and $D(t)$ are non-negative integers respectively representing the number of recovered or deceased individuals at time t.

The model with evolution given by equations (3)-(5) yields a stochastic dynamical system, referred to as a Stochastic Epidemiological Dynamics model, or sED model from here on.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2.3.1 Probabilities for recovery and death

Let τ_R and τ_D be positive-valued random variables respectively representing the waiting time for an infected person before recovery or death. Bearing in mind that the distributions of τ_R and τ_D may vary substantially depending on the pathogen (Baron, 1964), assume that infected individuals can either clear a pathogen within an average time μ_R, or alternatively, the individual may die from disease caused after an average time μ_D. The probabilities that a person is removed from the infected group or that remains infected within a small time interval of length h can be estimated by assuming *Geometrically* distributed times for recovery and death and writing

$$p_s^h \approx h\,\mu_s^{-1}, s \in \{R, D\}\,. \tag{6}$$

As a consequence, $1 - (p_R^h + p_D^h)$ is the probability of remaining infected within h time units. Large enough values of $I(t)$ yield a *Poisson* approximation of the form

$$p_s^h \approx 1 - e^{-h/\mu_s}, \quad s \in \{R, D\}\,. \tag{7}$$

Take for instance the epidemiological dynamics of COVID-19 (Fig. 3). Data from the first symptomatic cases of COVID-19 during 2020 in China suggested that deaths in hospitalised individuals occurred between 2 and 3 weeks after start of symptoms (Zhou et al., 2020). In contrast, the hospitalization time for survivors was reported to be between 3 and 4 weeks (Zhou et al., 2020). Also, the average incubation time was reported to be between 4 and 6 days (Quesada et al., 2020), and during this interval of time time those infected could shed inoculum while remaining asymptomatic. Taking the above observations into account, we assume values for μ_R and μ_D to be 21 and 28 days, and 28 and 35 days, respectively.

The joint dynamics of the decreasing size of the susceptible population and the probability of infection result in a sequence of probability mass functions for the newly infected that have many remarkable features (Fig. 4). First, the peak probability has a "U" shape as a function of time, taking the highest values at the beginning and end of the epidemic, and the lowest at the peak of infections. The number of newly infected that corresponds to that peak probability has the opposite behaviour, starting and ending at very low values, and reaching a maximum at the peak of the epidemic. This is in line with the idea that unless the value of β is very large, the probability that a large number of new infections occur is low, at any point in time, regardless of the population size.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

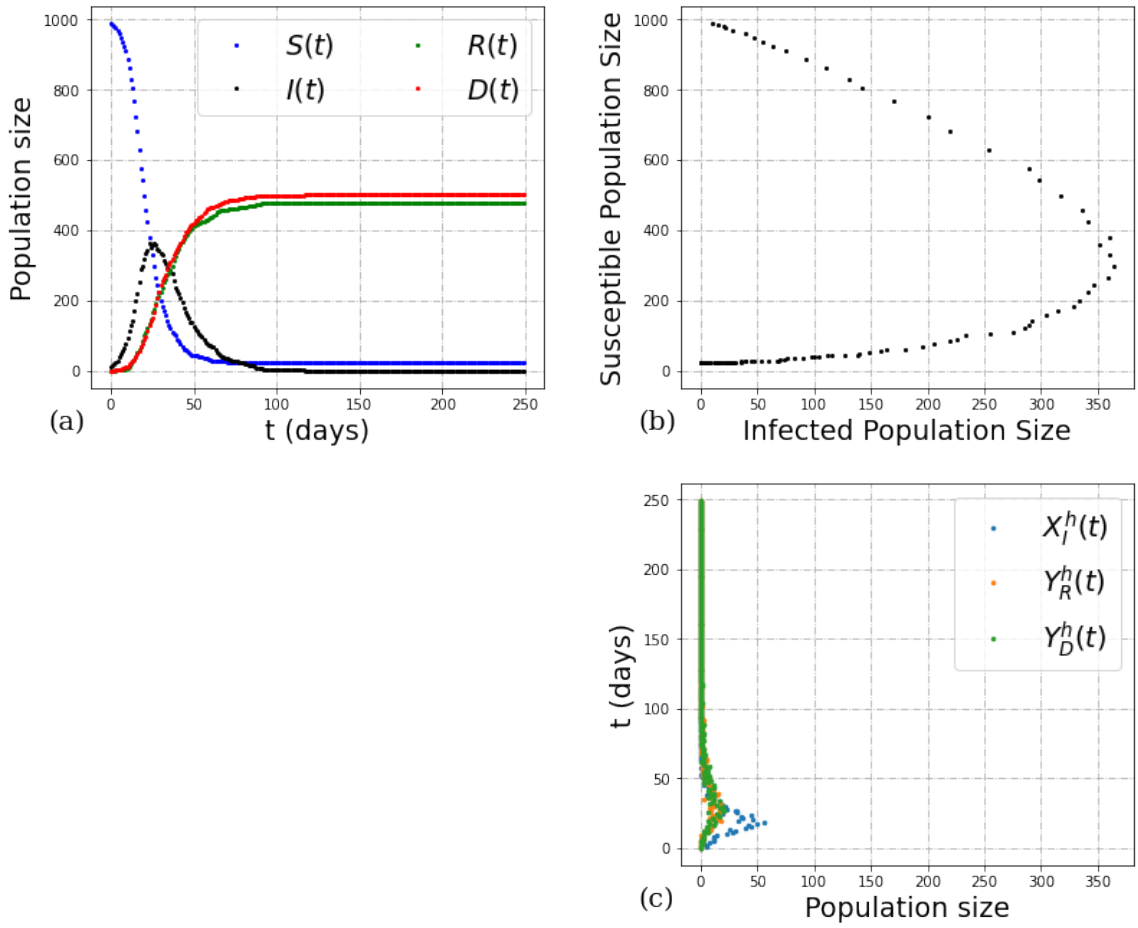

Figure 3: Dynamics and geometry of the sED model. The parameters used in this simulation were $\mu_R = 27$, and $\mu_D = 24$. Other parameters can be found in Table 1.

Before presenting applications and further extensions of this basic model, we demonstrate some mathematical properties derived directly from equations (3)-(6).

2.3.2 The routing probability and the mean infection time

We first answer a fundamental question: What is the probability that a given infected person recovers?

The multinomial trial with probabilities p_R^h, p_D^h, and $1 - (p_R^h + p_D^h)$ for any time t is independently repeated for time windows of size h. The event in which an individual *stops being infected* can be thought of as a sequence of independent trials, each lasting h time units, in which the individual remains infected and in the last trial the person either recovers or dies. The *routing probability* to the recovered state, that is, the probability that

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

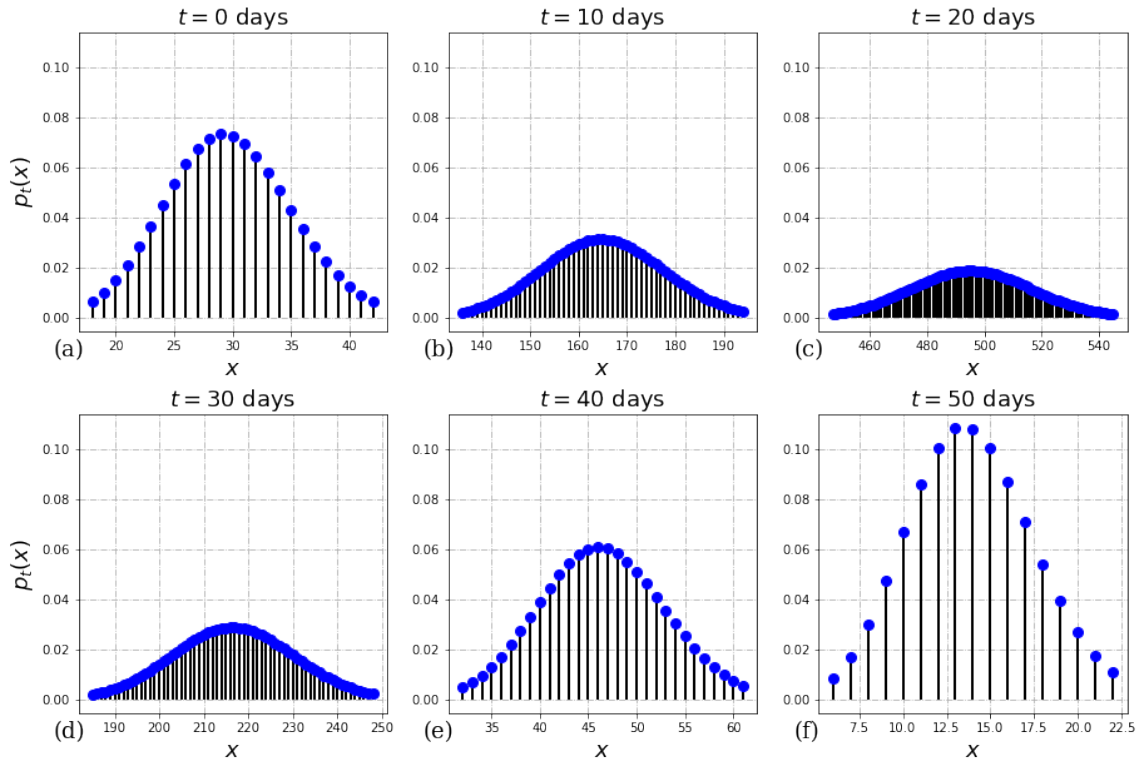

Figure 4: Example of the evolution of the probability mass function $p_t(x)$ of new infections x in the basic sED model as a function of time t. The parameters used in the simulation were $T_0 = 10,000$, $\beta = 0.3$, $\mu_R = 27$, $\mu_D = 24$ and $h = 1$.

an individual eventually recovers, is then given by

$$r_R = \sum_{j=1}^{\infty} \left(1 - p_R^h - p_D^h\right)^{j-1} p_R^h = \frac{p_R^h}{p_R^h + p_D^h} = \frac{\mu_D}{\mu_R + \mu_D}. \tag{8}$$

Similarly, the probability that an infected individual eventually dies is

$$r_D = \frac{\mu_R}{\mu_R + \mu_D}. \tag{9}$$

Note that these two probabilities do *not* depend on the time step h and are complementary.

The dynamics of the multinomial trials for each infected person can also be thought of as Bernoulli trials with probability of success $p_R + p_D$, and the elapsed time τ_I can then be thought of as a random variable with a mean $\bar{\tau}_I$ given by the inverse of the probability of leaving the infectious state,

$$\bar{\tau}_I = \frac{1}{h\left(\mu_R^{-1} + \mu_D^{-1}\right)} \tag{10}$$

Importantly, the model can use the routing probabilities and the mean time spent in the

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

infection stage, $(r_R, r_D, \bar{\tau}_I)$, as input parameters, instead of the mean times that a person passes in the infection phase for each given possible outcomes (μ_R, μ_D). Using equations (8)-(10) we have that

$$p_R^h = \frac{h}{\mu_R} = \frac{r_R}{\bar{\tau}_I}. \tag{11}$$

The inverse of the mean time spent in the infected phase is thus proportional to the probability that within h time units an individual stops being infected, regardless of the outcome.

2.4 Continuous time Markov chain approximation

Now consider an approximation for the simple sED model, in the case when the step size h (the discrete units of time for the simple sED model) decreases to zero for integer population sizes. The approximation gives insight about the theoretical properties in the limit for small step sizes, and the limit process is relatively easy to simulate since it is a continuous time Markov chain process (Fig. 5).

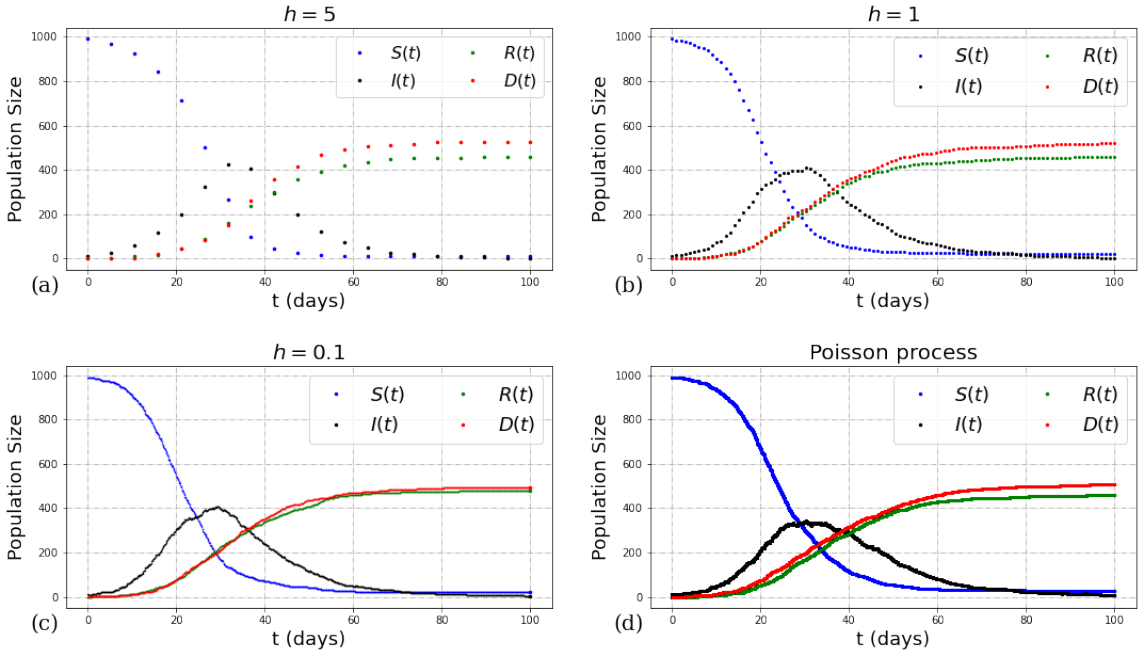

Figure 5: Realisations of the sED model dynamics for a decreasing step sizes (h) and the approximation of the model using the Poisson process. For this simulation $\mu_R = 27$ and $\mu_D = 24$, and the other parameters can be found in Table 1.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Recall that a non-homogeneous Poisson process $\{N(t) : t \geq 0\}$ with intensity $\lambda(t)$ and starting at the origin, can be assumed to have independent increments such that,

$$\begin{aligned} \mathbb{P}(N(t+h) - N(h) = 1) &= h\lambda(t) + o(h), \\ \mathbb{P}(N(t+h) - N(h) \geq 2) &= o(h), \end{aligned} \tag{12}$$

for any $t \geq 0$ (see for instance Basu, 2003, p. 142). Consider the sED state vector $\{(S(t), I(t), R(t), D(t)) : t \geq 0\}$ with dynamics given by equations (3)-(5), starting from some initial condition $(S(0), I(0), R(0), D(0))$. Define the cumulative sampling processes after k steps as

$$\hat{X}_I^h(kh) = \sum_{i=1}^{k} X_I^h(ih), \qquad \hat{Y}_D^h(kh) = \sum_{i=0}^{k} Y_D^h(ih), \qquad \hat{Y}_R^h(kh) = \sum_{i=0}^{k} Y_R^h(ih),$$

and assume that $\hat{X}_I^h(0) = \hat{Y}_R^h(0) = \hat{X}_D^h(0) = 0$ with the super index h emphasising dependence on the size of the time step. These processes are defined for a discrete set of times $\{0, h, 2h, ...\}$, but they can be extended for any positive time as left-continuous step functions on continuous time, taking a discrete set of values. For example, by defining $\hat{X}_I^h(t) := \hat{X}_I^h(kh)$ for any $kh \leq t < (k+1)h$, with $k \in \mathbb{N}$. As a consequence, the limit processes

$$N_I(t) := \lim_{h \to 0} \hat{X}_I^h(t), \qquad N_R(t) := \lim_{h \to 0} \hat{Y}_R^h(t), \qquad N_D(t) := \lim_{h \to 0} \hat{Y}_D^h(t),$$

are non-homogeneous Poisson processes. We show that in the following lines.

By construction N_I, N_R and N_D start at the origin, and the random variables $\left\{X_I^h(kh)\right\}_{k \in \mathbb{N}}$ are independent, which means that the process $\hat{X}_I^h$ has independent increments on disjoint intervals of the form $[n_1 h, n_2 h]$ where $n_1 < n_2$ are non negative integers. Therefore, the limit processes N_I, and by a similar argument, N_R and N_D, have independent increments. Further, for $t \in [kh, (k+1)h)$, $S(t) = S(kh)$, $I(t) = I(kh)$, and

$$\mathbb{P}(\hat{X}_I^h(t+h) - \hat{X}_I^h(t) = 1) = \mathbb{P}(X_I^h(kh) = 1).$$

Since $X_I^h(kh)$ has a Binomial distribution with parameters $S(kh)$ and $h\beta I(kh)/T$, respec-

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

tively, it follows that

$$\begin{aligned}\mathbb{P}(\hat{X}_I^h(t+h)-\hat{X}_I^h(t)=1) &= S(kh)\left(h\beta\frac{I(kh)}{T(kh)}\right)^1\left(1-h\beta\frac{I(kh)}{T(kh)}\right)^{S(kh)-1}\\ &\approx S(t)h\beta\frac{I(t)}{T(t)}+o(h),\end{aligned}$$

for $h>0$. Also the probability

$$\mathbb{P}(\hat{X}_I^h(t+h)-\hat{X}_I^h(t)\geq 2)=\sum_{i=2}^{S(kh)}\binom{S(kh)}{i}\left(h\beta\frac{I(kh)}{T(kh)}\right)^i\left(1-h\beta\frac{I(kh)}{T(kh)}\right)^{S(kh)-i},$$

has order $o(h)$, for $h>0$. Taking the limit as h goes to zero, it is verified that the process N_I satisfies equation (12) and we conclude that $\hat{X}_I$ is a non-homogeneous Poisson process with intensity $S(t)\beta I(t)/T(t)$. Since, $(Y_R^h(kh), Y_D^h(kh), I(kh)-Y_R^h(kh)-Y_D^h(kh))$ is Multinomial distributed for kh, we have that $Y_R^h(kh)$ is Binomial distributed (with parameters $I(kh)$ and p_R^h) and $Y_D^h(kh)$ is Binomial distributed (with parameters $I(kh)$ and p_D^h). The processes N_R and N_D satisfy (12), which can be proven by following the same reasoning as for N_I.

Taking into account the arguments just presented, as the step size h decreases to zero, the sED system $\{(S(t), I(t), R(t), D(t)) : t\geq 0\}$ based on Binomial samples converges in distribution with order $o(h)$ to the system $\{(\hat{S}(t), \hat{I}(t), \hat{R}(t), \hat{D}(t)) : t\geq 0\}$ based on Poisson sampling, starting from the same initial conditions and dynamics

$$\begin{aligned}\hat{S}(t) &= S(0)-N_I(t),\\ \hat{I}(t) &= I(0)+N_I(t)-(N_R(t)+N_D(t)),\\ \hat{R}(t) &= R(0)+N_R(t),\\ \hat{D}(t) &= D(0)+N_D(t),\end{aligned}$$

where N_I, N_R, N_D are non-homogeneous Poisson processes with corresponding intensity rates given by $\gamma_S=\hat{S}(t)\beta I(t)/T(t)$, $\gamma_R=\hat{I}(t)\mu_R^{-1}$, and $\gamma_D=\hat{I}(t)\mu_D^{-1}$.

As a direct consequence of last result, if the total size of the population $T(t)$ is constant, the limit process $(\hat{S}(t), \hat{I}(t), \hat{R}(t), \hat{D}(t))$ is a continuous time Markov chain with state

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

space $\{0, 1, ..., T\}^4$ and its infinitesimal generator matrix is given by its transitions rates:

$$\begin{aligned} q((s,i,r,d),(s-1,i+1,r,d)) &= s\beta\frac{i}{T}\,1_{\{1,2,...,T\}}(s), \\ q((s,i,r,d),(s,i-1,r+1,d) &= \mu_R^{-1}\,i\,1_{\{1,2,...,T\}}(i), \\ q((s,i,r,d),(s,i-1,r,d+1) &= \mu_D^{-1}\,i\,1_{\{1,2,...,T\}}(i), \end{aligned}$$

for $s, i, r, d \in \{0, 1, ..., T\}$.

The most basic models in the hierarchy have limitations. In the next section we expose some of the limitations and show how to overcome them by extending the basic model. Importantly, we do so by following the same rationale used for the basic construction. More specifically, we will modify the basic sED assumptions to capture finer dynamics. For instance, we will include different stages for the severity of infection to analyse the impact they have in the case-fatality ratios, a quantity often ignored in modelling studies.

Table 1: Notation and values used for simulations with the basic sED model.

Symbol	Values	Units	Description
T_0	1000	persons	Initial population size
S_0	10	persons	Initial size of susceptible population
I_0	990	persons	Initial size of infected population
R_0	0	persons	Initial size of recovered population
D_0	0	persons	Initial size of population dead due to disease
X^h	$[0, S]$	persons	Size of newly infected sample (depends on time step)
Y_R^h	$[0, I]$	persons	Size of newly recovered sample (depends on time step)
Y_D^h	$[0, I]$	persons	Size of n sample (depends on time step)
μ_R	[22–32]	days	Expected time between initial infection and death due to disease Li et al. (2020); Zhou et al. (2020)
μ_D	[19–29]	days	Expected time between initial infection and death due to disease Li et al. (2020); Zhou et al. (2020)
β	0.3	—-	Probability of infection given an contact with an infected person Phucharoen et al. (2020)
ε	1	1	Proportion of exposed susceptibles
κ	1	1	Proportion of exposed infected
h	1	days	Time step

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

3 Applications

3.1 The classical SIR model from the basic sED model

The classical SIR model describes the dynamics of an epidemic in a large population of constant size, in terms of three non-overlapping subpopulations representing susceptible, infected, and those that cannot participate in the chain of infections (either formerly infected or immune from the start). One key assumption of the classical SIR model is that individuals that cease being infected do not become susceptible again. In its continuous time and continuous state version, the SIR model can be written as a system of differential equations constructed under the assumption that the population is heterogeneously mixed and of fixed size (no deceases due to infection), with transmission occurring after contacts between susceptible and infected individuals.

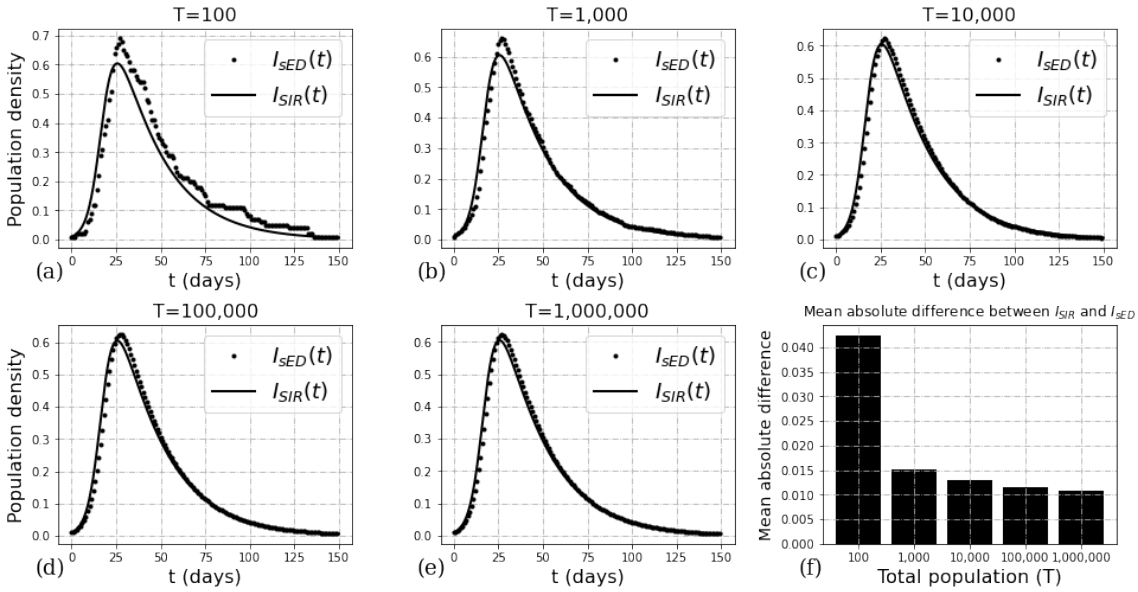

Figure 6: Density of infections for different population sizes of SIR vs sED dynamics obtained by replacing the sampling processes by their mean at each time step. The bars indicate the size of the difference between the total infected over time between the two models as a function of the population size. For these realisations, $\mu_R = (27 + 24)/2$ for the sED model without deaths, and $\beta = 0.3$ and $\gamma = 2/(27 + 24)$ for the SIR model.

An analogue of the classical SIR model can be obtained as a particular case of the basic sED model (equations (3)-(5)) by assuming that (0) there are no deaths due to infections, which makes the population size T a constant; (1) the samples $X_I^h(t)$ and $Y_R^h(t)$ are replaced by their expected values, which yields a deterministic system of difference equations; (2) the population size is very large; (3) the (integer) sizes of the epidemiological classes

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

are replaced by subpopulation densities, thus enabling the possibility of thinking about state variables as continuous; and (4) the time step can be arbitrarily small to replace the difference equations with differential equations where the densities are continuous variables changing with respect (continuous) time.

Briefly and explicitly, the expected number of new infected between times t and $t+h$ is $h\ \beta\ S(t)\ I(t)/T(t)$, which means that, on average, the absolute incidence (new cases) between t and $t+h$ can be written as

$$S(t+h) - S(t) = -h\beta S(t)I(t)/T(t). \tag{13}$$

Similarly, the number of people expected to stop being infected between t and $t+h$ is $p_R^h I(t)$, and the average change in a window of h time units is

$$I(t+h) - I(t) = h\left[\beta\frac{S(t)}{T} - \frac{1}{\mu_R}\right] I(t). \tag{14}$$

Then, if the (initial) population size T is large enough, we can let $x(t) = S(t)/T$, $y(t) = I(t)/T$, and $z(t) = 1 - (x(t) + y(t))$, and then write differential equations from equations (13)-(14) describing the dynamics of (x, y, z) by taking the limit as h tends to 0. If we let ∂_t denote the instantaneous rate of change with respect to time, then the dynamics for x and y can be described by differential equations of the form

$$\partial_t x = -\beta x y, \tag{15}$$

$$\partial_t y = (\beta x - \gamma)\, y, \tag{16}$$

with $\partial_t z = -(\partial_t x + \partial_t y) = \gamma y$ representing the time-dependent change in the density of the population that is immune to the disease (*i.e.* does not participate in the chain of infections), and $\gamma = 1/\mu_R$.

3.1.1 Is it possible to write a simple and well posed extension for the continuous SIR to account for deaths?

Can we change equation (16) to include the possibility of death due to disease? One way to do that would be to split γ as a sum of two rates, one representing recovery, the other death, respectively. As a consequence, $z(t)$ would be now a sum of population densities $d(t) + r(t)$ where $d(t) = D(t)/T(t)$ represents the proportion of people dying at time t, and $r(t) = R(t)/T(t)$ represents the proportion of people that do not participate in the

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

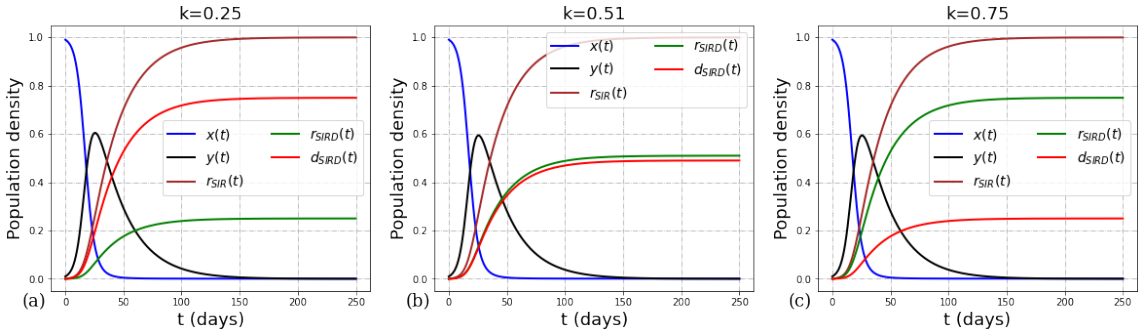

Figure 7: Systematic comparison between the dynamics of the SIR and SIRD models. The comparison was made asumning that $\gamma = \gamma_d + \gamma_r$ with $\gamma_d = k\gamma$, and $\gamma_r = (1-k)\gamma$ and $k \in \{0.25, 0.51, 0.75\}$, with $\gamma = 2/(27+24)$.

chain of infection at time t (either because they recovered or because they were immune to infection from the start). A typical setup for this would be to extend the system (15)-(16) to include

$$\partial_t r(t) = \frac{y(t)}{\mu_R}, \quad \partial_t d(t) = \frac{y(t)}{\mu_D}, \tag{17}$$

which could, in principle, be solved numerically to obtain approximations for the dynamics of the state vector (x, y, r, d), for each time t (Fig. 7). However, the total population size at time t should then be $T(t) = S(t) + I(t) + R(t)$, so that $1 = x(t) + y(t) + r(t)$ for all $t \geq 0$, and

$$\partial_t r \quad = \quad -\partial_t(x+y) = \gamma y.$$

But $\gamma y = \partial_t z = \partial_t r + \partial_t d$, which means that $\partial_t d \approx 0$ for all $t \geq 0$.

3.2 Limiting the exposure of the susceptible vs. limiting the exposure of the infected

Assume that a certain proportion ε of the susceptible are exposed to inoculation, and also, that only a proportion κ of those infected actively sheds inoculum.

Suppose that at any time t (days), non-negative integers $T(t)$ and $S(t)$ represent the size of the whole population and those susceptible to infection at time t. Let $\varepsilon(t)$ and $\kappa(t)$ respectively represent the proportion of the susceptible population that are exposed to the pathogen at any time t, and the proportion of infected people that shed inoculum. The values of $\varepsilon(t)$ and $\kappa(t)$ depends on factors that include the mobility of the population and other behavioural patterns.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

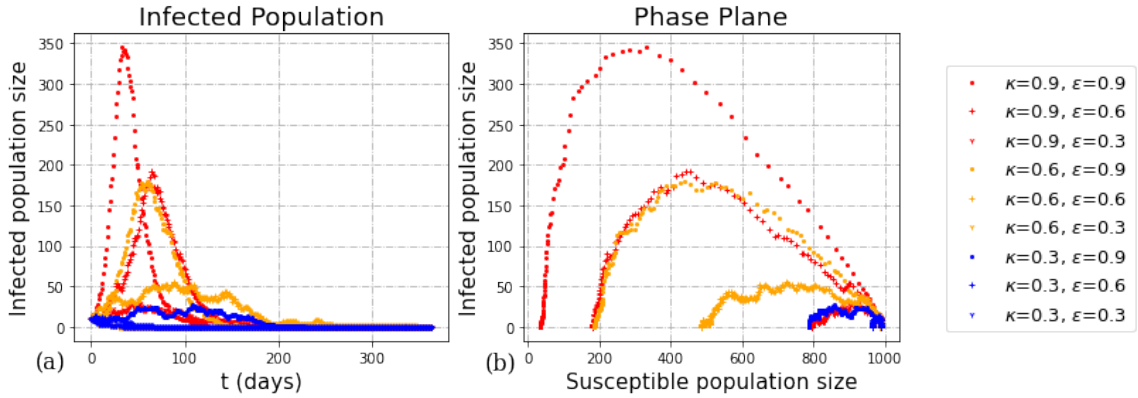

Figure 8: Infection processes varying the exposure for susceptible and infected (ε and κ). $S - X^h$ phase plane for different values of p. Realisations were calculated using $\mu_R = 27$, and $\mu_D = 24$.

As before, assume that infections are *independent* between individuals, with an average probability of infection $p^h(t, I)$ that depends on factors that include the time of exposure h and the availability inoculum, which can be thought of as a monotonically increasing function of I. In addition, suppose that each of the exposed susceptibles at time t can be infected within h time units of exposure with probability $p^h(t, I)$, or remain susceptible with probability $1 - p^h(t, I)$. The number $X_I^h(t)$ of newly infected people (absolute incidence) can then be thought of as a *Binomial* random variable sampled from $\lfloor \varepsilon(t)S(t) \rfloor$ individuals. Assuming homogeneous mixing between the exposed susceptibles and the exposed infected, define

$$p^h(t, I) = P_I(h)Q_h(t, I) = P_I(h)\beta\frac{\kappa(t)I(t)}{T(t)} \tag{18}$$

where β, $\kappa(t)$ and $I(t)/T(t)$, respectively represent the probability that a susceptible person is infected after having an infectious contact (contact with the pathogen), the proportion of infected individuals shedding inoculum, and the proportion of infected individuals within the population at time t. $P_I(h)$ represents the probability that the infection occurs within a time window of length h, while being *exposed* to the pathogen causing the disease of interest. As a consequence, the number of newly infected individuals between times t and $t + h$ can be thought of as a random variable $X_I^h(t) \sim Bin\left(\lfloor \varepsilon(t)S(t) \rfloor, p^h(t, I)\right)$, and the expected number of newly infected individuals would then be $p^h(t, I)\lfloor \varepsilon(t)S(t) \rfloor$. We will say that a new epidemic starts when the expected number of infections becomes an increasing function of time.

In principle, it is possible to estimate the value of $p^h(t, I)\varepsilon(t)$, if data of new infections is available. For infectious diseases like COVID-19, or ebola, if time is in minutes, or larger, and h is only a few seconds ($h << 1$), then $p_I(h) \approx 1$, and the probability of infection

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

given a contact with an infected individual is approximately $\beta\,\kappa(t)I(t)/T(t)$, as it has been proposed for models like the classical SIR.

3.3 Case fatality ratios and the necessity of considering infectious with different severity stages

The probability that an individual recovers eventually (equation (8)) depends on the mean time spent by an individual in the infectious state μ_R, and the mean time spent infected before death, μ_D; two quantities mainly statistical in nature.

However, equation (8) does not capture the complexity of the relation between epidemiological waiting times like μ_R and μ_D, and the probability of having a good or a bad outcome. For example, for some individuals the time between the initial inoculation and the emergence of symptoms may be very small, as they quickly develop an increasingly severe disease, significantly increasing the probability of death in comparison to other cases in which the disease is not as severe (Salinas-Escudero et al., 2020). Also, for diseases like COVID-19, the probability of recovery decreases as individuals spend more days in a hospital (Faes et al., 2020; Salinas-Escudero et al., 2020). In fact, during the first months of the COVID-19 pandemic, there were several studies reporting shorter hospitalization times with fatal outcomes in comparison to those for people who recovered (Alimohamadi et al., 2021; Porcheddu et al., 2020). Naturally, assuming $\mu_R > \mu_D$ for the basic sED model (and the SIR model) yields case-fatality ratios $D/(D+R) > 1/2$ (Figure 9 (a), (c), (e)), which does not necessarily happen for all infectious diseases, and certainly not for COVID-19, in which $\mu_R < \mu_D$ and the CFR has been less than 0.1 for most variants and most populations (Cao et al., 2020; Luo et al., 2021). As a consequence, the most simple form of the sED model cannot yield small case-fatality ratios and simultaneously, capture the macroscopic trends in the curve of cases vs time. To address this issue, we propose an adaptation of the model that takes a more realistic description of the disease history of individuals into account, therefore allowing the possibility of reproducing the macroscopic infection dynamics and also, the more detailed evolution of the recovered and deceased after infection.

3.3.1 Extension to consider infections with multiple states

To get a better idea of the progression through the infection and possibly the subsequent clinical stages, imagine that an individual starts in the susceptible state. At some time she

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

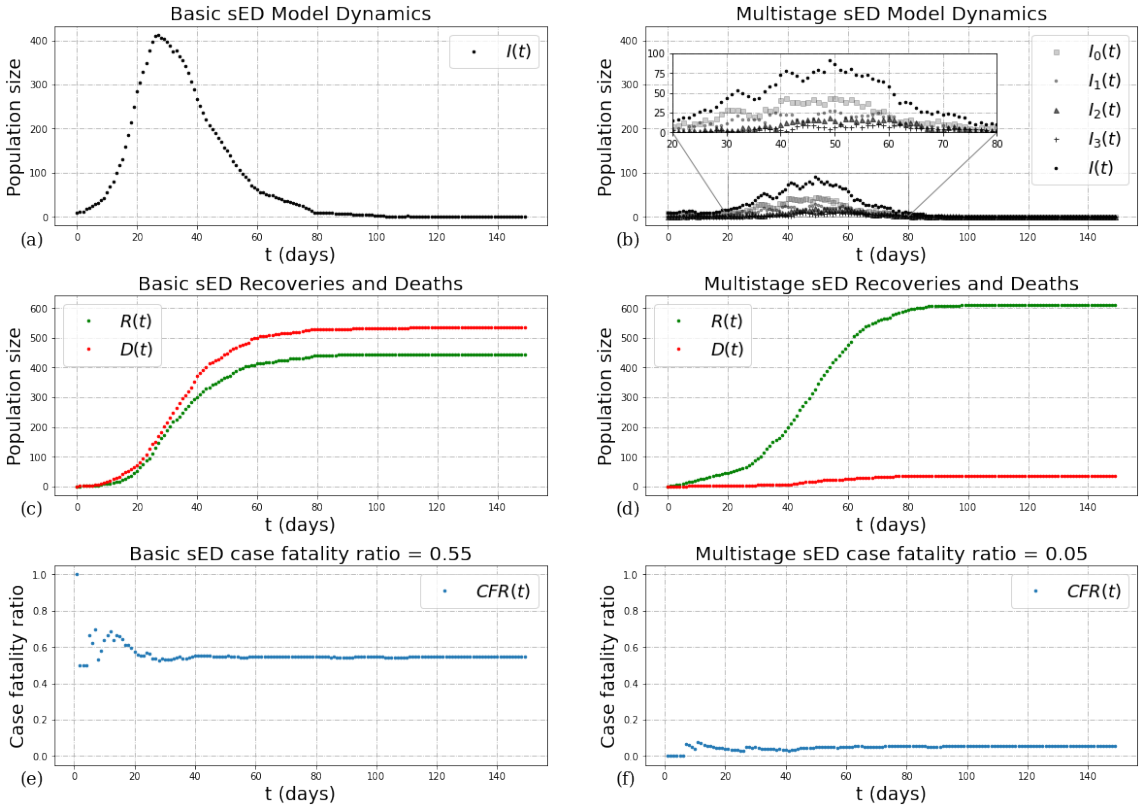

Figure 9: Case fatality ratios can be adjusted by assuming multiple infection-related epidemiological stages. Epidemiological dynamics of for multiple infection-related stages (a,c,e) vs a single infectious stage (b,d,f). Realisations in both cases were constructed assuming the same total expected infection time with $\mu_R = 27$ and $\mu_D = 24$ for the basic sED and $\mu_R^0 = 5$, $\mu_R^1 = 6$, $\mu_R^2 = 12$, $\mu_R^3 = 4$, and $\mu_D^0 = 5$, $\mu_D^1 = 5$, $\mu_D^2 = 5$, $\mu_D^3 = 9$ for the multistage sED.

might become infected, becoming part of those individuals in the initial state of infection. From there, she could recover or the disease may increase in severity. In that case, we model the increase in the severity of the disease as a switch to a more advanced infection state, often characterised by the clinical traits. Importantly, the switching is assumed to sequentially progress toward more severe illness, until possibly reaching the death state in the absence of recovery. In other words, the individual either recovers, or advances to the following illness state, and if the individual reaches the last infection stage, the only possibilities become to either recover or die.

Before developing an extension for the basic sED model, it is important to remark that the time period in which an individual is capable of transmission of the disease does not necessarily coincide with the manifestation of clinical symptoms. For instance, shedding of inoculum may occur during an incubation period for many infectious diseases. Alternatively, it is also possible that an infected person stops transmitting the disease before recovering

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(Li, 2010). These events depend strongly on the disease (Baron, 1964). Of interest, it could also be the case that individuals in a very advanced stage of the disease do not participate in the chain of transmission. This may happen because of becoming spatially isolated (*e.g.* due to hospitalisation or self isolation), or because of losing the ability to move. To start considering these cases, we propose the aforementioned multistage sED model.

Assume that the subpopulation sizes of infected or ill people are represented by k infectious stages $\{I_0,..., I_{k-1}\}$, with increasing indices depending on severity of the disease. The states do not necessarily represent clinical states. The total infected population is therefore $I = I_0 + ... + I_{k-1}$. Of note, it is assumed here that, an individual may recover from any infection state, and cannot become part of the subpopulation in infection state j without having been through all the infection states of less severity. By extension, an individual cannot die without having been through all the infection states $I_0, ..., I_{k-1}$. For instance, for COVID-19 it is reasonable to assume that there are at least four infectious states labelled as $0, 1, 2$, and 3, with population sizes I_0, I_1, I_2, I_3, respectively representing an incubation period, an initial symptomatic state I_1, a second (mild severity) symptomatic state, and those in a severe disease state after which they could die (Figure 9 (b)).

Suppose that in a time window of size h, given that a susceptible individual has had contact with an individual in the stage I_i, the probability of having an effective transmission is given by $h\beta_i\kappa_i(t)$. Then, by using the law of total probability counting over the partition on the different possible stages of the infected individual, the probability that a susceptible individual becomes infected in a time window of size h is given by

$$p^h(I_0, ..., I_{k-1}, t) = h\sum_{i=0}^{k-1} \beta_i\kappa_i(t)\frac{I_i(t)}{T(t)}. \tag{19}$$

which generalises (18).

For the subsequent stages, in analogy with the Subsection 2.3.2, we can derive geometrically distributed times for advancing or recovering to the next stage for each individual. The probabilities that a person in stage I_i advances to the next stage I_{i+1} or recovers, within a small time interval of length h are estimated by

$$p_I^{(i)} \approx \frac{h}{\mu_I^{(i)}}, \tag{20}$$

$$p_R^{(i)} \approx \frac{h}{\mu_R^{(i)}}, \tag{21}$$

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

where $\mu_I^{(i)}$ is the mean elapsed time that a patient spends in stage I_i before moving forward to the stage I_{i+1}, and $\mu_R^{(i)}$ is the mean elapsed time that a patient who recovers after stage I_i spends in that stage. Accordingly, an individual remains infected in the ith stage with probability $1-(p_I^{(i)}+p_R^{(i)})$. For computational purposes, we can think the death stage as the stage I_{k+1}, to be included in last equation. Therefore, a Multinomial approach can be derived in each stage, to compute the number of individual that advances or recover in each state to give a complete description of the dynamics for this extension. A system that extends equations (3)-(5) taking into account the different infectious stages is thus given by

$$S(t+h) = S(t) - X_0^h(t) \tag{22}$$

$$I_j(t+h) = I_j(t) + X_{j-1}^h(t) - (Y_j^h(t) + X_j^h(t)) \qquad \forall\, j = 0, ..., k, \tag{23}$$

$$R(t+h) = R(t) + \sum_{j=1}^{k} Y_j^h(t), \tag{24}$$

$$D(t+h) = D(t) + X_k^h(t), \tag{25}$$

where $S(t)$, $I_0(t)$, ..., $I_k(t)$, $R(t)$, and $D(t)$ represent the numbers of individuals in each stage at time t and

$$X_0^h(t) \sim Bin\left(\lfloor \varepsilon(t)S(t) \rfloor, p^h(t)\right),$$

$$(X_j^h(t), Y_j^h(t), I_j(t) - X_j^h(t) - Y_j^h(t)) \sim Mult\Big(I_j(t), p_I^{(j)}, p_R^{(j)}, 1 - p_I^{(j)} - p_R^{(j)}\Big),$$

for all $j = 0, 1, ..., k$, assuming that all random elements are independent. Note that the $X's$ represent the quantity of people passing from some stage to the next (the last one could be death), while the $Y's$ represent the number of people recovering from the disease, at some given time.

The dynamics that can be obtained from this extension of the basic sED can be such that the total expected time before recovery is larger than the total expected time before death due to disease (Fig. 9), but such that the case fatality ratio is much smaller than 1/2 depending on the parameters of the model.

3.3.2 Probability distribution of the final stage reached by an individual

In this section we illustrate some mathematical properties of the multistage extension of the sED model just presented. Of interest, we focus on the maximal stage of infection or disease that a given individual reaches before recovering, once that she has been infected.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The reason is that the last stage and the final clinical outcome is of interest for clinical and epidemiological purposes (Roger, 2011). This is developed in the context of the model with multiple infection and disease stages sED (equations (22)).

Consider the sED model with multiple infection (and disease) stages, given by $\{S, I_0, ..., I_k, R, D\}$ (equations (22)). Assume that one individual has been initially infected at some time t. By construction, once the individual has been infected, her or his recovery time and the maximal stage reached before recovering will not depend on the evolution of the epidemic, but only in its (random) individual evolution through the infection and disease stages.

First, using the same rationale as in the basic sED model, if an individual is in stage i of disease, the probability that the individual goes to the recovery state without passing to the next stage (without getting worse) is given by the routing probability to recovery $\frac{\mu_I^{(i)}}{\mu_R^{(i)}+\mu_I^{(i)}}$ (see Subsection 2.3.2).

By considering last equation for each stage, and assuming independence between the sampling processes in different stages, the probability that the infected individual reaches the stage I_l ($l \leq k$) as maximal stage of infection or disease before recovering is given by

$$\left(\prod_{i=1}^{l-1} \frac{\mu_R^{(i)}}{\mu_R^{(i)} + \mu_I^{(i)}}\right)\left(\frac{\mu_I^{(l)}}{\mu_R^{(l)} + \mu_I^{(l)}}\right). \tag{26}$$

Using the same reasoning, the probability of death of such individual is

$$\prod_{i=1}^{k} \frac{\mu_R^{(i)}}{\mu_R^{(i)} + \mu_I^{(i)}}. \tag{27}$$

As aforementioned, last probabilities are expressed in only in terms of the expected elapsed times in each stage of infection or disease.

4 Discussion

We have presented a derivation to model epidemiological dynamics for infectious diseases based on a simple key assumptions: at any given time, individuals in any epidemiological stage are the subjects of a random sampling process in which they either remain in their epidemiological stage, or move to a different stage. Then we have distinguished two different types of sampling: individuals from one population may get sampled depending on an interaction with another subpopulation, as is the case in the infection process; or sampling to switch epidemiological states may depend on waiting times given by individual,

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2: Notation and values used in the multistage extension of the sED model.

Symbol	Value	Units	Description
μ_R^0	[4.1-7.0]	days	Expected incubation time (Li et al., 2020)
μ_R^1	[4.3-7.5]	days	Expected time spent in first symptomatic stage (Li et al., 2020)
μ_R^2	[9.0–15.0]	days	Expected time spent in the second (mild severity) symptomatic stage (Zhou et al., 2020)
μ_R^3	[2.0–9.0]	days	Expected time spent in the third (severe) symptomatic stage (Zhou et al., 2020)
μ_D^0	[4.1-7.0]	days	Expected incubation time (Li et al., 2020)
μ_D^1	[4.3-7.5]	days	Expected time spent in first symptomatic stage (Li et al., 2020)
μ_D^2	[5.0–11.0]	days	Expected time spent in the second (mild severity) symptomatic stage (Zhou et al., 2020)
μ_D^3	[4.0–12.0]	days	Expected time spent in the third (severe) symptomatic stage (Zhou et al., 2020)

pathophysiological, or economical factors, to name a few. As a result, we have obtained a simple, easy to implement model derived from common sense rules, that takes into account different sources of randomness in the evolution of the dynamics in infectious diseases. The model works for small and large populations as well, and also, allows extensions to study the effects of specific factors that may become important determinants of the epidemics (e.g. exposure of infected *vs.* exposure of susceptible, single *vs.* multiple disease stages). *En passage*, we explain how to derive a continuous time, continuously valued SIR model from the sED model, and in what circumstances do the classical SIR models work, and expose some of the problems that arise in extensions of the SIR models intended take into account deaths due to disease. We also show a few mathematical properties related to the expected times spent by an individual in different epidemiological stages. Of particular interest, we show how case fatality ratios in models with only one infectious stage would be larger than or equal to 1/2 unless the expected time for recovery is shorter than the expected time for death due to disease. We also show that lower values for the CFR smaller than 0.1 can be only be obtained in models with more than one infectious (or clinical) stage that allows percolation of the population through different stages and recovery instead of dying.

4.1 Why do we need a different model than the SIR model?

Let us point out some similarities and differences between sED and SIR models.

One issue of interest is the theoretical impossibility of using the SIR model for small

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

populations: population densities are discrete by construction, but they can be thought of as continuous variables for large population sizes. However, there are many epidemics in which the populations of interest are small (e.g. small villages in Liberia during Ebola outbreaks). Nevertheless, for large enough populations the dynamics of the deterministic SIR model in equations (15)-(16) can be compared to simulations of the dynamics obtained with the sED model after dividing the sED state variables by the population size (Fig. 6). This is illustrated by analyzing the distance between the normalized 3-dimensional vector $(S(t), I(t), R(t))/T(t)$ and its deterministic counterpart $(x(t), y(t), z(t))$, which decreases robustly as T increases (Fig. 6).

There are many possible extensions that could be implemented from the sED model. For instance, models taking into account different susceptibility. For instance, the need for such extensions has been repeatedly suggested by surveillance data during the COVID-19 pandemic, from genetic profiling data (Velavan et al., 2021), from CT-value readings in RT-PCR tests (Waudby-West et al., 2021), and from serological testing (Toulis, 2020). These are examples of implementations that have been suggested, then regarded as unnecessary and too complicated, and subsequently discarded. However, making such considerations adds a layer of knowledge and understanding that renders realistic decision making impossible to do without them.

The dynamics in the of SIR "with deaths" and sED are not supposed to be similar. The results obtained in Subsection 3.1.1 show that simply adding another variable to the regular SIR deterministic dynamics to account for deceases would not be coherent mathematically, and also, with what would be expected from the sED dynamics, so it would be profitless.

4.2 The exposure of susceptible and infected have different effects on the epidemic dynamics.

Infection processes occurring between pairs of individuals depend at least in three aspects: the infectivity potential of the infected individual, the vulnerability of being infected of the susceptible individual, and the physical exposure between them, typically ocurring through physical contact. These aspects are very difficult to measure precisely. However, different indirect statistical measures have been used in each of them. Examples of these include viral load as an indicator of infectivity potential (Jones et al., 2021; Marc et al., 2021); existence of comorbidities to measure vulnerability (MK et al., 2021); presence of

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

antibodies in an individual as indicators of infection severity (Legros et al., 2021); mobility in cities as a proxy of physical exposure (Lu and Gan, 2022), especially in consideration of lock downs and social distance and public health measures (Wu et al., 2021). The sED model allows us to explicitly separate exposition of the susceptible population and the exposure of people with the potential to infect others. This allows us to observe explicitly the different dynamics resulting from having drastic reductions in one factor but not in the other one, opening the possibility of establishing or discarding the efficacy of public health measures directed to the reduction of such factors. As it should have been obvious, the most effective strategy to mitigate the epidemics is to isolate the infected. Unfortunately, this is impossible for epidemics in which the asymptomatic population is present, as is the case for COVID-19.

4.3 Infection vs infectivity

The sED model can describe some aspects concerning the infectivity and lethality of a disease that cannot be directly explained by the force of infection and the basic reproduction number. For a fact, but also for epidemiological purposes, the infection time may not necessarily be be equal to the time interval during which a person is infectious. For instance, an infected person during the incubation period, or already in the process of clearing the virus, may have a sufficiently low viral load such that the person is effectively not infectious. Similarly, the infection time may include a period of symptoms that may be severe enough to reduce the exposure of the infected person (e.g. hospitalisation or self isolation), potentially decreasing the availability of inoculum shedding for other people to get infected. The sED model can be easily extended to study the case in which infectivity periods last less than the infection time and possibly include decreased levels of exposure by the infected; we study those cases in a companion paper dedicated to study COVID-19.

Of note, hypothesis in the sED model that infection depends on the exposure of those infected in environments where susceptibles may be inoculated applies also to infections that depend on non-human vectors like dengue. This is justified by recent data from dengue outbreaks and endemicity in densely populated urban areas (Falcón-Lezama et al., 2017; Villabona-Arenas et al., 2016). Consider the case in which there is one female mosquito per building in an apartment complex. The radius of interaction of the mosquito can easily be 25 meters, without taking into account vertical interactions. This means that one case of dengue in one building could result in many more cases, without the need of an abundant population of mosquitoes. The sED model uses simple parameters to

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

represent such complex process of infection. Those parameters can be calibrated using the statistical information available. Then the model can be used to analyse qualitatively different scenarios depending on those parameters. In particular, the multistage sED model allows to divide the process of infection using different layers related to different possible mechanisms of infection and transmission of the diseased studied.

4.3.1 Multiple stages for infection and illness progression.

As pointed out earlier, an adequate scheme to model the evolution of many diseases, should be one in which classification of infection states is constructed according to the physiological changes of the individual, whether or not they depend on clinical assessment. The clinical stages associated with infections and disease due to such infections can then also be considered. Note, however, that only taking into consideration clinical considerations may result in very inaccurate assessments with possibly bad consequences. For instance, infections with SARS-CoV-2 cause a large percentage of asymptomatic cases (non-clinical), but not taking them into consideration has proven harmful for the sake of public health (Zhao et al., 2020). This is at large our motivation for thinking about the pathophysiological states of an individual during an infection as the base for our modelling scheme.

An important remark is that, constructing a model by taking the physiology related to the infection as a basis to classify different subpopulations may result in an inability to estimate parameters in an statistically meaningful way. However, the sED model is simple enough and easy to expand so that anyone could use it to explore qualitatively different behaviours of an epidemics to identify mechanisms underlying the epidemic guided by fitting real data (e.g. prevalence and incidence curves). That was the case leading to the extension of the basic sED model into the multistage model introduced in Section 3.3, where we show how CFR with very low values can be obtained by simply introducing multiple, subsequent stages and the condition that individuals who die from the infection must go through all disease stages before dying. By means of exploration, the multistage sED model thus provided some insight about why it would not be possible to obtain low CFR with shorter recovery intervals in comparison to hospitalization times, as documented during the first months of the COVID-19 pandemic. *En passage*, that analysis also shows that it would not be possible to model COVID-19 with classical SIR dynamics unless the model is modified by stratifying the disease stages before death.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

4.4 On numerical simulations and approximations

Of note, the fact that $P_s(h) \approx 1$ for a small enough ratio μ_s/h suggest that sampling events at time steps that are close to 1 may not be approximated as accurately as those with mean waiting time larger than 1. As a note of caution, the sampling in the sED model would not be well defined for mean waiting times shorter than the step size.

4.5 Concluding remarks

In this article we presented a new mathematical model to describe epidemiological dynamics. Its purpose is to describe some crucial aspects observed in the evolution of pandemics often not taken into account in the existing modelling literature, while simultaneously being intuitive, based on common sense arguments, and robust enough to be easily adapted to different particular features of diseases.

Disclosures. Some of the work contained in this article is part of the thesis written by CINH (Herrera-Nolasco, 2020) to obtain a Bachelor's degree in Mathematics at Facultad de Ciencias, UNAM. MAHV's contribution was supported by DGAPA-PAPIIT-UNAM grant IN-228820.

References

Yousef Alimohamadi, Habteyes Hailu Tola, Abbas Abbasi-Ghahramanloo, Majid Janani, and Mojtaba Sepandi. Case fatality rate of covid-19: a systematic review and meta-analysis. *Journal of preventive medicine and hygiene*, 62(2):E311, 2021.

Linda JS Allen. An introduction to stochastic epidemic models. In *Mathematical epidemiology*, pages 81–130. Springer, 2008.

Linda J.S. Allen. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2(2):128–142, 2017.

Kristina Allers and Thomas Schneider. Ccr5δ32 mutation and HIV infection: basis for curative HIV therapy. *Current Opinion in Virology*, 14:24–29, 2015.

Dorothy Anderson and Ray Watson. On the spread of a disease with gamma distributed latent and infectious periods. *Biometrika*, 67(1):191–198, 1980.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Samuel Baron. Mechanism of Recovery from Viral Infection. volume 10 of *Advances in Virus Research*, pages 39–64. Academic Press, 1964.

Adhir K Basu. *Introduction to Stochastic Process*. Alpha Science Int'l Ltd., 2003.

Julie Boucau, Caitlin Marino, James Regan, Rockib Uddin, Manish C. Choudhary, James P. Flynn, Geoffrey Chen, Ashley M. Stuckwisch, Josh Mathews, May Y. Liew, Arshdeep Singh, Taryn Lipiner, Autumn Kittilson, Meghan Melberg, Yijia Li, Rebecca F. Gilbert, Zahra Reynolds, Surabhi L. Iyer, Grace C. Chamberlin, Tammy D. Vyas, Marcia B. Goldberg, Jatin M. Vyas, Jonathan Z. Li, Jacob E. Lemieux, Mark J. Siedner, and Amy K. Barczak. Duration of Shedding of Culturable Virus in SARS-CoV-2 Omicron (BA.1) Infection. *New England Journal of Medicine*, 387(3):275–277, 2022. doi: 10.1056/NEJMc2202092. URL https://doi.org/10.1056/NEJMc2202092.

Teun Bousema, Lucy Okell, Ingrid Felger, and Chris Drakeley. Asymptomatic malaria infections: detectability, transmissibility and public health relevance. *Nature Reviews Microbiology*, 12(12):833–840, 2014.

Tom Britton. Stochastic epidemic models: A survey. *Mathematical Biosciences*, 225(1): 24–35, 2010.

Michael J Butler and Ruth M Barrientos. The impact of nutrition on COVID-19 susceptibility and long-term consequences. *Brain, behavior, and immunity*, 87:53–54, 2020.

Y. Cao, A. Hiyoshi, and S.Montgomery. Covid-19 case-fatality rate and demographic and socioeconomic influencers: worldwide spatial regression analysis based on country-level data. *BMJ Open*, 3(10):(11):e043560, 2020.

Angela S Clem. Fundamentals of Vaccine Immunology. *Journal of global infectious diseases*, 3(1):73, 2011.

Cheryl Cohen, Jackie Kleynhans, Jocelyn Moyes, Meredith L McMorrow, Florette K Treurnicht, Orienka Hellferscee, Azwifarwi Mathunjwa, Anne von Gottberg, Nicole Wolter, Neil A Martinson, et al. Asymptomatic transmission and high community burden of seasonal influenza in an urban and a rural community in South Africa, 2017–18 (PHIRST): a population cohort study. *The Lancet Global Health*, 9(6):e863–e874, 2021.

Richard C Dicker, Fatima Coronado, Denise Koo, and R Gibson Parrish. Principles of Epidemiology in Public Health Practice; An Introduction to Applied Epidemiology and Biostatistics. 2006.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Denise L Doolan, Carlota Dobaño, and J Kevin Baird. Acquired immunity to malaria. *Clinical microbiology reviews*, 22(1):13–36, 2009.

S. Eker. Validity and usefulness of COVID-19 models. *Humanit Soc Sci Commun*, 7(54), 2020.

C. Faes, S. Abrams, D. Van Beckhoven, G. Meyfroidt, E. Vlieghe, and N. Hens. Time between Symptom Onset, Hospitalisation and Recovery or Death: Statistical Analysis of Belgian COVID-19 Patients. *International Journal of Environmental Research and Public Health*, 17(20):7560, 2020.

Aisha D Fakhroo, Asmaa A Al Thani, and Hadi M Yassine. Markers associated with COVID-19 susceptibility, resistance, and severity. *Viruses*, 13(1):45, 2020.

Jorge Abelardo Falcón-Lezama, René Santos-Luna, Susana Román-Pérez, Ruth Aralí Martínez-Vega, Marco Arieli Herrera-Valdez, Ángel Fernando Kuri-Morales, Ben Adams, Pablo Antonio Kuri-Morales, Malaquías López-Cervantes, and José Ramos-Castañeda. Analysis of spatial mobility in subjects from a dengue endemic urban locality in Morelos State, Mexico. *PloS one*, 12(2):e0172313, 2017.

Z. Feng, D. Xu, and H. Zhao. Epidemiological Models with Non-Exponentially Distributed Disease Stages and Applications to Disease Control. *Bull. Math. Biol.*, 69:1511–1536, 2007.

Frank Fenner, Peter A Bachmann, E Paul J Gibbs, Frederick A Murphy, Michael J Studdert, and David O White. Viral replication. *Veterinary Virology*, page 55, 1987.

Ingrid Fricke-Galindo and Ramcés Falfán-Valencia. Genetics insight for COVID-19 susceptibility and severity: a review. *Frontiers in immunology*, 12:622176, 2021.

Monica Gandhi, Deborah S Yokoe, and Diane V Havlir. Asymptomatic transmission, the Achilles' heel of current strategies to control Covid-19, 2020.

Luis F. García. Immune Response, Inflammation, and the Clinical Spectrum of COVID-19. *Frontiers in Immunology*, 11:1441, 2020.

Charles P Gerba. Environmentally transmitted pathogens. In *Environmental microbiology*, pages 445–484. Elsevier, 2009.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Laura Grange, Etienne Simon-Loriere, Anavaj Sakuntabhai, Lionel Gresh, Richard Paul, and Eva Harris. Epidemiological risk factors associated with high global frequency of inapparent dengue virus infections. *Frontiers in Immunology*, 5:280, 2014. ISSN 1664-3224. doi: 10.3389/fimmu.2014.00280. URL https://www.frontiersin.org/article/10.3389/fimmu.2014.00280.

Major Greenwood. On the statistical measure of infectiousness. *Epidemiology & Infection*, 31(3):336–351, 1931.

Priscilla E Greenwood and Luis F Gordillo. Stochastic epidemic modeling. In *Mathematical and statistical estimation approaches in epidemiology*, pages 31–52. Springer, 2009.

Xi He, Eric HY Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y Wong, Yujuan Guan, Xinghua Tan, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature medicine*, 26(5):672–675, 2020.

Carlos Ignacio Herrera-Nolasco. Modelo epidemiológico probabilista con extensión a grupos metapoblacionales. Master's thesis, Facultad de Ciencias, UNAM, Av. Universidad 3000, Circuito Exterior SN, 7 2020.

T Déirdre Hollingsworth, Roy M Anderson, and Christophe Fraser. Hiv-1 transmission, by stage of infection. *The Journal of infectious diseases*, 198(5):687–693, 2008.

Inga Holmdahl and Caroline Buckee. Wrong but Useful — What Covid-19 Epidemiologic Models Can and Cannot Tell Us. *New England Journal of Medicine*, 383(4):303–305, 2020. doi: 10.1056/NEJMp2016822. URL https://doi.org/10.1056/NEJMp2016822.

Terry C. Jones, Guido Biele, Barbara Mühlemann, Talitha Veith, Julia Schneider, Jörn Beheim-Schwarzbach, Tobias Bleicker, Julia Tesch, Marie Luisa Schmidt, Leif Erik Sander, Florian Kurth, Peter Menzel, Rolf Schwarzer, Marta Zuchowski, Jörg Hofmann, Andi Krumbholz, Angela Stein, Anke Edelmann, Victor Max Corman, and Christian Drosten. Estimating infectiousness throughout SARS-CoV-2 infection course. *Science*, 373(6551):eabi5273, 2021.

Leah C Katzelnick, Magelda Montoya, Lionel Gresh, Angel Balmaseda, and Eva Harris. Neutralizing antibody titers against dengue virus correlate with protection from symptomatic infection in a longitudinal cohort. *Proceedings of the National Academy of Sciences*, 113(3):728–733, 2016.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

WO Kermack and AG McKendrick. Contributions to the mathematical theory of epidemics. *Proc Roy Soc Lond*, 115:700–721, 1927.

V. Legros, S. Denolly, and M. et al. Vogrig. A longitudinal study of SARS-CoV-2-infected patients reveals a high correlation between neutralizing antibodies and COVID-19 severity. *Cell Mol Immunol*, 18:318–327, 2021.

Michael Y. Li. *An Introduction to Mathematical Modeling of Infectious Diseases*, volume 2. Mathematics of Planet Earth, Springer, 2010.

Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy SM Leung, Eric HY Lau, Jessica Y Wong, et al. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England journal of medicine*, 2020.

Alun L. Lloyd. Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics. *Theoretical Population Biology*, 60(1):59–71, 2001.

Huan Lu and Hongcheng Gan. Evaluation and prevention and control measures of urban public transport exposure risk under the influence of COVID-19—Taking Wuhan as an example. *PLOS ONE*, 17(6):1–15, 06 2022.

G. Luo, X. Zhang, H. Zheng, and D. He. Infection fatality ratio and case fatality ratio of COVID-19. *Int J Infect Dis*, 113:43–46, 2021.

Rose H. Manjili, Melika Zarei, Mehran Habibi, and Masoud H. Manjili. Covid-19 as an acute inflammatory disease. *The Journal of Immunology*, 205(1):12–19, 2020.

Aurélien Marc, Marion Kerioui, François Blanquart, Julie Bertrand, Oriol Mitjà, Marc Corbacho-Monné, Michael Marks, and Jeremie Guedj. Quantifying the relationship between SARS-CoV-2 viral load and infectiousness. *eLife*, 10:e69302, 2021.

Singh MK, Mobeen A, Chandra A, Joshi S, and Ramachandran S. A meta-analysis of comorbidities in COVID-19: Which diseases increase the susceptibility of SARS-CoV-2 infection? *Comput Biol Med*, 130:104219, 2021.

Magelda Montoya, Lionel Gresh, Juan Carlos Mercado, Katherine L Williams, Maria José Vargas, Gamaliel Gutierrez, Guillermina Kuan, Aubree Gordon, Angel Balmaseda, and

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Eva Harris. Symptomatic versus inapparent outcome in repeat dengue virus infections is influenced by the time interval between infections and study year. *PLoS Negl Trop Dis*, 7 (8):e2357, 2013.

Chayanon Phucharoen, Nichapat Sangkaew, and Kristina Stosic. The characteristics of covid-19 transmission from case to high-risk contact, a statistical analysis from contact tracing data. *EClinicalMedicine*, 27:100543, 2020.

Rossella Porcheddu, Caterina Serra, David Kelvin, Nikki Kelvin, and Salvatore Rubino. Similarity in case fatality rates (CFR) of COVID-19/SARS-CoV-2 in Italy and China. *The Journal of Infection in Developing Countries*, 14(02):125–128, 2020.

JA Quesada, A López-Pineda, VF Gil-Guillén, JM Arriero-Marín, F Gutiérrez, and C Carratala-Munuera. Incubation period of COVID-19: A systematic review and meta-analysis. *Revista Clínica Española (English Edition)*, 2020.

VL Roger. Outcomes research and epidemiology: the synergy between public health and clinical practice. *Circ Cardiovasc Qual Outcomes.*, 4(3):257–9, 2011.

Jon Roozenbeek, Claudia R Schneider, Sarah Dryhurst, John Kerr, Alexandra LJ Freeman, Gabriel Recchia, Anne Marthe Van Der Bles, and Sander Van Der Linden. Susceptibility to misinformation about COVID-19 around the world. *Royal Society open science*, 7 (10):201199, 2020.

G. Salinas-Escudero, M.F. Carrillo-Vega, V. Granados-García, and et al. A survival analysis of COVID-19 in the Mexican population. *BMC Public Health*, 20:1616, 2020.

B. Shayak and M. Sharma. A new approach to the dynamic modeling of an infectious disease. *Math. Model. Nat. Phenom.*, 16:33, 2021. ISSN 1359-5938.

P. Toulis. Estimation of Covid-19 prevalence from serology tests: A partial identification approach. *J Econom.*, 220(1):193–213, 2020.

Henry C Tuckwell and Ruth J Williams. Some properties of a simple stochastic epidemic model of SIR type. *Mathematical biosciences*, 208(1):76–97, 2007.

J.J.A. van Kampen, D.A.M.C. van de Vijver, and P.L.A. et. al. Fraaij. Covid-19: Does the infectious inoculum dose-response relationship contribute to understanding heterogeneity in disease severity and transmission dynamics? *Nature Communications*, 12:267, 2021.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.16.22278844; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Alexei Vazquez. Exact solution of infection dynamics with gamma distribution of generation intervals. *Phys. Rev. E*, 103:042306, Apr 2021.

TP Velavan, SR Pallerla, J Rüter, Y Augustin, PG Kremsner, S Krishna, and Meyer CG. Host genetic factors determining COVID-19 susceptibility and severity. *EBioMedicine*, (72):103629, 2021.

Christian Julián Villabona-Arenas, Jessica Luana de Oliveira, Carla de Sousa-Capra, Karime Balarini, Celso Ricardo Theoto Pereira da Fonseca, and Paolo Marinho de Andrade Zanotto. Epidemiological dynamics of an urban Dengue 4 outbreak in São Paulo, Brazil. *PeerJ*, 4:e1892, 2016.

Rupert Waudby-West, Benjamin J. Parcell, Colin N.A. Palmer, Samira Bell, James D. Chalmers, and Moneeza K. Siddiqui. The association between SARS-CoV-2 RT-PCR cycle threshold and mortality in a community cohort. *European Respiratory Journal*, 58 (1), 2021.

Shishi Wu, Rachel Neill, Chuan De Foo, Alvin Qijia Chua, Anne-Sophie Jung, Victoria Haldane, Salma M Abdalla, Wei-jie Guan, Sudhvir Singh, Anders Nordström, and Helena Legido-Quigley. Aggressive containment, suppression, and mitigation of COVID-19: lessons learnt from eight countries. *BMJ*, 375:e067508, 2021.

Hongjun Zhao, Xiaoxiao Lu, Yibin Deng, Yujin Tang, and Jiachun Lu. Covid-19: asymptomatic carrier transmission is an underestimated problem. *Epidemiology & Infection*, 148, 2020.

Fei Zhou, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu, Zhibo Liu, Jie Xiang, Yeming Wang, Bin Song, Xiaoying Gu, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 395(10229):1054–1062, 2020.